

Mathematics and Its Applications

K. Vajravelu (Ed.)

**Differential Equations
and Nonlinear Mechanics**



Kluwer Academic Publishers

Differential Equations and Nonlinear Mechanics

Mathematics and Its Applications

Managing Editor:

M. HAZEWINKEL

Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

Differential Equations and Nonlinear Mechanics

Edited by

K. Vajravelu

University of Central Florida



KLUWER ACADEMIC PUBLISHERS

DORDRECHT / BOSTON / LONDON

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-13:978-1-4613-7974-4 e-ISBN-13:978-1-4613-0277-3
DOI: 10.1007/978-1-4613-0277-3

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

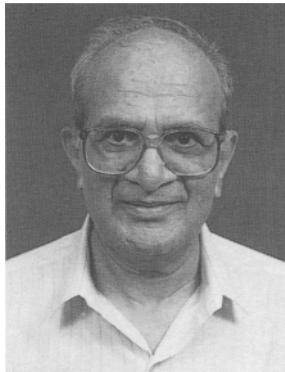
Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

All Rights Reserved
© 2001 Kluwer Academic Publishers
Softcover reprint of the hardcover 1st edition 2001

No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any information storage and
retrieval system, without written permission from the copyright owner.



Professor V. Lakshmikantham has had an illustrious career as a mathematician, administrator, teacher, and a promoter of mathematics. As a mathematician, Professor Lakshmikantham has published hundreds of articles and dozens of books and monographs in the areas of differential equations, dynamical systems, and nonlinear analysis. His work has had an international impact on research in these areas. Currently many mathematicians and followers worldwide are expanding and developing his ideas.

As an administrator, he is well known and held in high regard for his leadership and motivational skills. He has served as department head in several universities. He has been able to construct, and convert obscure departments into prominent Ph.D. granting research departments, which have received national renown. Professor Lakshmikantham has achieved this with his extraordinary leadership and motivational skills, and without a large influx of funds. He has encouraged inactive faculty to start doing research and enjoy their productive careers.

As a teacher, Professor Lakshmikantham is a master in motivating and inspiring his students to perform at their highest capabilities. To him, each student is a member of his family with none treated better than another. He follows his students' careers and lives wherever they may go. As a result, most of his students have become productive citizens of the mathematical community and several of them are serving as chief editors of mathematical journals.

Professor Lakshmikantham has devoted his whole life to mathematics. He had envisioned the importance of unifying and solidifying the area of nonlinear analysis. The Journal of Nonlinear Analysis and other journals that he founded have been useful and timely.

Furthermore, Professor Lakshmikantham has touched many lives academically, intellectually, and professionally. He has played an important role in the development and nurturing of careers of mathematicians all over the world. His nurturing nature and fine people skills have had an effect on everyone who have been in his presence. Professor Lakshmikantham will live forever, in the heart, spirit, and minds of his family, friends, and students.

TABLE OF CONTENTS

PREFACE.....	xi
1. Properties of the Radii of Stability and Instability	
Patricia Anderson and S.R. Bernfeld.....	1
2. Extremal Solutions of Hemivariational Inequalities with D.C.-Superpotentials	
S. Carl	11
3. Degenerate Quasilinear Parabolic Problems With Slow Diffusions	
C.Y. Chan and W.Y. Chan.....	27
4. Superasymptotic Perturbation Analysis of the Kelvin-Helmholtz Instability of Supersonic Shear Layers	
S. Roy Choudhury	31
5. Solitary Waves with Galilean Invariance in Dispersive Shallow-Water Flows	
C.I. Christov	49
6. Discrete Dynamical Systems Described by Neutral Equations	
C. Corduneanu.....	69
7. Oscillation of Third Order Differential Equations With and Without Delay	
R.S. Dahiya.....	75
8. Numerical Techniques for Solving a Biharmonic Equation in a Sectorial Region	
Elias Deeba, Suheil A. Khuri, and Shishen Xie.....	89
9. The Seamount on a Sloping Seabed Problem	
R.P. Gilbert, Miao Ou, and Yongzhi S. Xu.....	101

10. Discrete Simulation in Nonlinear Dynamics With Applications	
Donald Greenspan	113
11. Ergodic Type Solutions of Some Differential Equations	
Jialin Hong and Rafael Obaya	135
12. Synchronous Solutions of Delayed Neural Networks	
Ying Sue Huang	153
13. Coherent Structures and Statistical Equilibrium States in a Model of Dispersive Wave Turbulence	
Richard Jordan and Christophe Josserand.....	163
14. Fuzzy Sets and Fuzzy Differential Equations	
V. Lakshmikantham and R.N. Mohapatra.....	183
15. Numerical Solutions of Coupled Parabolic Systems With Time Delays	
Xin Lu.....	201
16. Global Analysis for the Fluids of a Power-Law Type	
Josef Málek.....	213
17. Nonlinear Hyperbolic Partial Differential and Volterra Integral Equations: Analytical and Numerical Approaches	
Roger G. Marshall and Sudhakar G. Pandit.....	235
18. Kronecker Product Operations of Tensors	
David W. Nicholson.....	249
19. Global Behavior of Solutions of a Certain Nth Order Differential Equation in the Vicinity of an Irregular Singular Point	
T.K. Puttaswamy	265
20. On the Modelling of Dissipative Processes	
K.R. Rajagopal	285
21. Pathwise Average Cost Per Unit Time Problem for Stochastic Differential Games With a Small Parameter	
K.M. Ramachandran and A.N.V. Rao.....	293

22. Interaction of Surface Radiation With Natural Convection	
N. Ramesh, C. Balaji, and S.P. Venkateshan	309
23. Mathematical Results and Numerical Methods for Steady Incompressible Viscoelastic Fluid Flows	
Adélia Sequeira and Juha H. Videman	339
24. Full Conversion in Gas-Solid Reactions	
Ivar Stakgold	367
25. New Analysis Procedure in Predicting Rotor Vibration	
R. Subbiah	373
26. Equivalent Conditions for Disconjugacy in Self-Adjoint Systems	
Betty Travis and Ramón Navarro	385
27. Some Results on Reaction Diffusion Equations With Initial Time Difference	
A.S. Vatsala	391
28. Dynamics of Neural Networks With Delay: Attractors and Content-Addressable Memory	
Jianhong Wu	401
29. Invariant Sets and Global Attractor of a Class of Partial Differential Equations	
Daoyi Xu and Qingyi Guo	419
Index	431

PREFACE

The International Conference on Differential Equations and Nonlinear Mechanics was hosted by the University of Central Florida in Orlando from March 17-19, 1999. One of the conference days was dedicated to Professor V. Lakshmikantham in honor of his 75th birthday. 50 well established professionals (in differential equations, nonlinear analysis, numerical analysis, and nonlinear mechanics) attended the conference from 13 countries. Twelve of the attendees delivered hour long invited talks and remaining thirty-eight presented invited forty-five minute talks. In each of these talks, the focus was on the recent developments in differential equations and nonlinear mechanics and their applications. This book consists of 29 papers based on the invited lectures, and I believe that it provides a good selection of advanced topics of current interest in differential equations and nonlinear mechanics.

I am indebted to the Department of Mathematics, College of Arts and Sciences, Department of Mechanical, Materials and Aerospace Engineering, and the Office of International Studies (of the University of Central Florida) for the financial support of the conference. Also, to the Mathematics Department of the University of Central Florida for providing secretarial and administrative assistance. I would like to thank the members of the local organizing committee, Jeanne Blank, Jackie Callahan, John Cannon, Holly Carley, Brad Pyle, Pete Rautenstrauch, and June Wingler for their assistance. Thanks are also due to the conference organizing committee, F.H. Busse, J.R. Cannon, V. Girault, R.H.J. Grimshaw, P.N. Kaloni, V. Lakshmikantham, R.N. Mohapatra, D. Nicholson, K.R. Rajagopal, and A. Sequeira. The invited speakers of the conference, especially Shair Ahmad who delivered the banquet talk, and everyone who attended the conference deserve a special mention for making this a success. My special thanks are due to Jackie Callahan for typing the manuscript carefully. Also, I wish to thank J.R. Martindale, Editor, and the staff of Kluwer Academic Publishers.

Finally, I thank my wife, Rani, for her ideas, devotion, and for having a vision for me; and my older son, Ravy, for his computer assistance, and to my younger son, Gopi, for his support and understanding, throughout the stages of the conference.

K. Vajravelu

Orlando, Florida

June 2000

1 PROPERTIES OF THE RADII OF STABILITY AND INSTABILITY

Patricia Anderson
Union College
Lincoln, NE 68506-4316
and
Stephen R. Bernfeld
University of Texas at Arlington
Arlington, TX 76019-0408

1. INTRODUCTION

The radius of stability and the radius of instability of the zero solution of the differential equation $x' = f(t, x)$ were introduced by Salvadori and Visentini [9], [10]. These radii in some sense provide a measure of the “region” of stability or instability of the zero solution. This knowledge has been used in the study of small solutions $x_p(t)$ of perturbations of the differential equation $x' = f(t, x)$ given by $x'_p = f(t, x_p) + h(t, x_p)$. In particular a relationship between the radius of stability of the zero solution of $x' = f(t, x)$ and its total stability was also introduced in [9] and [10]. Having been motivated by mechanical systems subject to conservative perturbations these authors analyzed the total stability of $x' = f(t, x)$ using the perturbed differential equation $x' = g(t, x, \lambda)$ where $g(t, x, 0) = f(t, x)$ and λ is a parameter in some Banach Space β . In this paper we often will assume β is the real line.

In this paper we wish to study the continuity properties of the radii of stability and instability in terms of the total stability of the zero solution of $x' = f(t, x)$. We generally only consider the scalar case (although some extension to higher dimensions are discussed). We provide results in the case where the perturbed differential equation has a bifurcation phenomenon and study the properties of the radii of stability and instability in this case.

2. PRELIMINARIES

Consider the unperturbed differential equation given by

$$x' = f(t, x) \quad (2.1)$$

where $f : [R \times D, R]$ and D is a neighborhood of the origin.

Let the perturbations of (2.1) be given by

$$x'_p = g(t, x_p, \lambda) \quad (2.2)$$

where $g(t, x, \lambda)$ is scalar in x and $g \in C[R \times D \times \Lambda, \mathbb{R}]$ for some set $\Lambda \subseteq \mathbb{R}$ such that the origin is an accumulation point of Λ . We also assume that

$$(*) \quad \begin{cases} a(|\lambda|) \leq \|g(t, x, \lambda) - f(t, x)\| \leq b(|\lambda|), & \text{where} \\ b(\cdot), a(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}^+ & \text{with} \\ a(0) = b(0) = 0 & \end{cases}$$

in which both $a(\cdot)$ and $b(\cdot)$ are strictly increasing and continuous.

We shall often restrict our attention to the case in which the unperturbed system is the autonomous differential equation

$$x' = f(x) \quad (2.3)$$

where $f \in C[D, \mathbb{R}]$. Assume also that $f(0) = 0$.

Let the perturbations of (2.3) be given by

$$x'_p = g(x_p, \lambda), \quad (2.4)$$

where $g(x, \lambda)$ satisfies (*).

Denote the solutions of (2.3) through (t_0, x_0) by $x(t, t_0, x_0)$ and denote the solutions of (2.4) through $(t_0, x_{p,0})$ by $x_p(t, t_0, x_{p,0})$.

The following definitions of the radius of stability and the radius of instability of the zero solution were given by Salvadori and Visentin [10]. (See also [11].) For completeness we shall consider (2.1) and provide the definition in the case in which $x \in \mathbb{R}^n$.

Definition 1. [11] The *radius of stability* is defined to be $r(t_0, \varepsilon)$ where

$$r(t_0, \varepsilon) = \sup \left\{ \delta \geq 0 : \|x_0\| < \delta : \text{implies that } \|x(t, t_0, x_0)\| < \varepsilon \text{ for all } t \geq t_0 \right\}.$$

Definition 2. [11] The *radius of uniform stability* is defined to be $r(\varepsilon)$ where

$$r(\varepsilon) = \inf \left\{ r(t_0, \varepsilon) : t_0 \in I \right\}.$$

Definition 3. [11] The *radius of instability* is defined to be $R(t_0)$ where

$$R(t_0) = \left\{ \sup \eta > 0 : \text{there exists two sequences } \{x_i\} \text{ and } \{t_i\} \text{ with } x_i \in D \text{ and } t_i \in I \text{ for all } i \in \mathbb{R} \text{ such that } \|x(t_i, t_0, x_i)\| \leq \eta \text{ for all } i \in \mathbb{N} \right\}.$$

Definition 4. [11] The *radius of non-uniform stability* is defined to be R where

$$R = \left\{ \sup \eta \geq 0 : \text{there exists three sequences } \{x_i\}, \{t_i\} \text{ and } \{t_{0,i}\} \text{ such that } \|x(t_i, t_{0,i}, x_i)\| \geq \eta \text{ for all } i \in \mathbb{N}, \text{ where } \lim_{i \rightarrow \infty} \|x_i\| = 0 \text{ and } \lim_{i \rightarrow \infty} t_{0,i} = \infty \text{ and } t_i \geq t_{0,i} \text{ for all } i \in \mathbb{N} \right\}.$$

The following definition of conditional total stability was given by Salvadori and Visentin, [9], [10].

Definition 5. The zero solution, $x \equiv 0$, of (2.1) is conditionally totally stable if for each $\varepsilon > 0$ and $t_0 \geq 0$ there exists $\delta_1(t_0, \varepsilon) \geq 0$ and $\delta_2(t_0, \varepsilon) \geq 0$ such that for each $|x_0| < \delta_1$ and for each $\lambda \in \Lambda$ with $|\lambda| < \delta_2$ the solution $x_p(t, t_0, x_0)$ of (2.2) satisfies

$$\|x_p(t, t_0, x_0)\| < \varepsilon \text{ for every } t \geq t_0.$$

Theorem 1. ([2]) The zero solution of (2.3) is uniformly totally stable if and only if there exists a nested family of contracting compact neighborhoods of the origin which are invariant and asymptotically stable.

A bifurcation theorem given in [6] is now presented.

Theorem 2. Suppose that $g(0, \lambda) \equiv 0$ and the origin of (2.4) is asymptotically stable for $\lambda = 0$ and completely unstable for $\lambda > 0$. Then there exists $\lambda^* > 0$ and a neighborhood T of the origin such that if $\lambda \in (0, \lambda^*)$ and M_λ is the largest invariant compact subset of $T \setminus \{0\}$, then (M_λ) is a family of asymptotically stable compact sets bifurcating from $\{0\}$.

3. PRIMARY RESULTS

We now present some properties of the radii of stability and instability. We use $|\cdot|$ to denote the norm in one dimension.

Proposition 1. Suppose that the zero solution of (2.1) is conditionally totally stable. Then for each $t_0 \in I$ and for each $\varepsilon > 0$ there exists $\delta(t_0, \varepsilon) > 0$ such that $r(t_0, \varepsilon, \lambda)$ need not be continuous in λ for each λ with $|\lambda| \in (0, \delta)$.

An example is given which depicts Proposition 1.

Example 1. Let $x' = f(x)$ where

$$f(x) = \begin{cases} (-1)^n \left(-x + \frac{1}{n+1} \right) & x \in \left(\frac{1}{n+1}, \frac{2n+1}{2n(n+1)} \right) \\ (-1)^n \left(x - \frac{1}{n} \right) & x \in \left[\frac{2n+1}{2n(n+1)}, \frac{1}{n} \right] \\ 0 & x = \frac{1}{n} \end{cases}. \quad (3.1)$$

An application of Theorem 1 implies the zero solution is uniformly totally stable and hence, conditionally uniformly stable. Fix $t_0 \geq 0$. Note that for each $x_0 = \frac{1}{n}$ the solution $x(t, t_0, x_0)$ satisfies

$$x(t, t_0, x_0) \equiv \frac{1}{n}.$$

Now let $\varepsilon = \frac{1}{2k-1}$ for some $k \in \mathbb{N}$. Fix $t_0 \geq 0$. Now consider the perturbation of (3.1) given by

$$x'_p = f(x_p) + \lambda \quad (3.2)$$

where $\lambda \in \left(0, \frac{1}{4k} \right)$. And so for each $\lambda \in \left(0, \frac{1}{4k} \right)$

$$r(t_0, \varepsilon, \lambda) \leq \frac{1}{2k}.$$

This is due to the fact that the right hand side of (3.2) is positive for all $x \in \left(\frac{1}{2k}, \frac{1}{2k-1} \right]$ and hence, implies that for any x_0 with

$$x_0 \in \left(\frac{1}{2k}, \frac{1}{2k-1} \right]$$

the solution $x(t)$ satisfies for some $T(x_0) > 0$

$$x(T, t_0, x_0) \geq \varepsilon.$$

And so

$$\lim_{\lambda \rightarrow 0^+} r(t_0, \varepsilon, \lambda) = \frac{1}{2k}.$$

Now, by uniqueness of the zero solution, it follows that for any x_0 such that $|x_0| < \frac{1}{2k-1}$ the solution $x(t)$ satisfies

$$|x(t, t_0, x_0)| < \varepsilon.$$

And so

$$r(t_0, \varepsilon) = \varepsilon = \frac{1}{2k-1}.$$

Hence,

$$\lim_{\lambda \rightarrow 0^+} r(t_0, \varepsilon, \lambda) = \frac{1}{2k} \neq \frac{1}{2k-1} = r(t_0, \varepsilon, 0) = r(t_0, \varepsilon).$$

Now this holds for each $k \in \mathbb{N}$. Thus $r(t_0, \varepsilon, \lambda)$ is not continuous in λ .

Proposition 2. Suppose that the zero solution $x \equiv 0$ of (2.3) is conditionally totally stable. Suppose also that the origin is asymptotically stable for $\lambda = 0$ and completely unstable for $\lambda > 0$. Then $r(t_0, \varepsilon, \lambda)$ is not continuous in ε .

Proof. By Theorem 2 there exists $\lambda^* > 0$ and a neighborhood ϑ of the origin such that if $\lambda \in (0, \lambda^*)$ and M_λ is the largest invariant compact subset of $\vartheta \setminus \{0\}$ then (M_λ) is a family of asymptotically stable compact sets bifurcating from $\{0\}$.

Now let $\lambda > 0$ and $t_0 \in I$ be fixed.

Then, by hypothesis, the origin is completely unstable. Hence, we have for some $\eta > 0$ that

$$R(t_0, \lambda) = \eta.$$

Note, by definition of complete instability, one has for every $\varepsilon < \eta$

$$r(t_0, \varepsilon, \lambda) = 0.$$

Now since M_λ is an invariant set, we see by the fact that $R(t_0, \lambda) = \eta$ that for some $\alpha > 0$ and for all $\varepsilon > \eta$ $r(t_0, \varepsilon, \lambda) \geq \alpha > 0$.

Hence,

$$\lim_{\varepsilon \rightarrow R(t_0, \lambda)^+} r(t_0, \varepsilon, \lambda) \geq \alpha \neq 0 = r(t_0, R(t_0, \lambda), \lambda).$$

And so $r(t_0, \varepsilon, \lambda)$ is not continuous in ε .

Proposition 3. Assume that the zero solution of (2.3) is conditionally totally stable. Suppose also that the zero solution of (2.3) is not a uniform attractor. Then $r(t_0, \varepsilon, \lambda)$ need not be continuous in ε .

The following example is used to depict Proposition 3.

Example 2. Let $x' = f(x)$ where

$$f(x) = \begin{cases} (-1)^n \left(-x + \frac{1}{n+1} \right) & x \in \left[\frac{1}{n+1}, \frac{2n+1}{2n(n+1)} \right] \\ (-1)^n \left(x - \frac{1}{n} \right) & x \in \left[\frac{2n+1}{2n(n+1)}, \frac{1}{n} \right] \\ 0 & x = \frac{1}{n} \\ -x & x \leq 0 \end{cases}. \quad (3.3)$$

Note that for each $x_0 = \frac{1}{n}$ the solution $x(t)$ satisfies

$$x(t, t_0, x_0) \equiv \frac{1}{n}$$

and so $x \equiv 0$ is not asymptotically stable. Let the perturbation (2.2) of (2.1) be given by

$$x'_p = f(x_p) + \lambda \quad (3.4)$$

where $\lambda > 0$ is chosen such that for all x_p with $x_p \in \left[-\frac{1}{8}, \frac{1}{8} \right]$ one has

$$x'_p \geq 0$$

and for $|x_p| = \frac{1}{8}$

$$x'_p = 0.$$

Moreover, there exists a neighborhood $\left(\frac{1}{8}, \bar{x} \right)$ where one has for each $x_p \in \left(\frac{1}{8}, \bar{x} \right)$

$$x'_p < 0.$$

Note that $g(0, \lambda) = \lambda \neq 0$. Now, choose $\varepsilon_1 = \frac{1}{8}$ and let $\varepsilon_2 < \frac{1}{8}$. Then by uniqueness of solutions and since the solution through $x_{p,0} = \frac{1}{8}$ satisfies

$$x_p \left(t, t_0, \frac{1}{8} \right) \equiv \frac{1}{8}$$

it follows that

$$r(t_0, \varepsilon_1, \lambda) = \frac{1}{8}.$$

Also,

$$r(t_0, \varepsilon_2, \lambda) = 0.$$

This is due to the fact that the zero solution of the perturbed differential equation (3.4) is completely unstable and the ball of radius ε_2 intersects this neighborhood of complete instability. Hence,

$$\lim_{\varepsilon_2 \rightarrow \varepsilon_1} r(t_0, \varepsilon_2, \lambda) = 0 \neq r(t_0, \varepsilon_1, \lambda).$$

Now this holds for each $t_0 > 0$ and $\lambda > 0$. Hence, $r(t_0, \varepsilon, \lambda)$ is not continuous in ε .

The next property which is presented again shows that the radius of stability need not be continuous in ε .

Proposition 4. Assume that the zero solution of (2.3) is conditionally totally stable. Suppose also that the zero solution of (2.4) is stable. Then $r(t_0, \varepsilon, \lambda)$ need not be continuous in ε .

An example is given to depict Proposition 4.

Example 3. Let

$$x' = f(x) = -x. \quad (3.5)$$

Construct for each $n \in \mathbb{N}$ a function $g_n(x)$ such that

$$g_n(x) = \begin{cases} 0 & x \leq \frac{1}{n} \\ 4\left(x - \frac{1}{n}\right) & x \in \left(\frac{1}{n}, \frac{2}{n}\right] \\ 2x & x \in \left(\frac{2}{n}, \frac{3}{n}\right] \\ -6\left(x - \frac{4}{n}\right) & x \in \left(\frac{3}{n}, \frac{4}{n}\right] \\ 0 & x > \frac{4}{n} \end{cases}$$

Then it follows that the zero solution $x_p \equiv 0$ of the perturbed differential equation given by

$$x'_p = f(x_p) + g_n(x_p)$$

is stable. Also, the zero solution $x \equiv 0$ of (3.5) is uniformly asymptotically stable.

Now, let $\varepsilon_1 = \frac{24}{7n}$ and let $\varepsilon_2 < \frac{24}{7n}$. Then

$$r(t_0, \varepsilon_2, \lambda_n) \leq \frac{1}{n}.$$

This is due to the fact that for some $x_p \in \left(\frac{1}{n}, \frac{4}{n}\right)$ one has

$$x'_p > 0.$$

Note that for $x = \frac{24}{7n}$

$$f(x_p) + g_n(x_p) \equiv 0.$$

And so, by uniqueness of the solution $x_p\left(t, t_0, \frac{24}{7n}\right) \equiv \frac{24}{7n}$, it follows that

$$r(t_0, \varepsilon_1, \lambda_n) = \frac{24}{7n}.$$

Hence, again one sees that

$$\lim_{\varepsilon_2 \rightarrow \varepsilon_1} r(t_0, \varepsilon_2, \lambda_n) = \frac{1}{n} \neq \frac{24}{7n} = r(t_0, \varepsilon_1, \lambda_n).$$

And so $r(t_0, \varepsilon, \lambda)$ is not continuous in ε for each $t_0 \geq 0$ and $\lambda \in \mathbb{R}$ fixed. And the example is complete.

It is natural to wonder whether the “strength” given heuristically of the conditional total stability of (2.1) is related to the size of the radius of stability and the strength of “attraction” of a nested, compact family of neighborhoods of the origin of the unperturbed system. Can we relate the size of allowable perturbations of a system that preserve the stability of the origin with both of these properties? The following proposition and example tell us:

Proposition 5. The “strength” of conditional total stability cannot be measured by the radius of stability.

The following example is used to depict Proposition 5. The sets of allowable perturbations are not disjoint in the following example, although one could construct an example where the sets of allowable perturbations are disjoint. Our example is nonautonomous.

Example 4. Let

$$x'_1 = f_1(t, x_1) = -\frac{2x_1}{t^2} \quad (3.6)$$

where

$$-\frac{1}{2} \leq x_1 \leq \frac{1}{2}$$

and

$$x'_2 = f_2(t, x_2) = -\frac{x_2}{t}$$

where

$$-\frac{1}{4} \leq x_2 \leq \frac{1}{4}. \quad (3.7)$$

Then the zero solution $x \equiv 0$ of (3.6) and the zero solution of (3.7) are uniformly stable. In fact, the zero solutions of (3.6) and (3.7) are asymptotically stable as well.

Now $r_1(\varepsilon) = \frac{1}{2}$ and $r_2(\varepsilon) = \frac{1}{4}$ where $r_1(\varepsilon)$ and $r_2(\varepsilon)$ are the radii of uniform

stability of (3.6) and (3.7) respectively. And so

$$r_1(\varepsilon) \geq r_2(\varepsilon).$$

Now fix $\varepsilon > 0$. Then note that a set of perturbations for which the zero solution of the perturbed differential equation of (3.6) is stable contains the class of functions given by

$$g_1(t, x, \lambda) = \lambda \frac{x}{t^p} \text{ for all } p \geq 2. \quad (3.8)$$

Note also that a set of perturbations for which the zero solution of the perturbed differential equation of (3.7) is stable contains the class of functions given by

$$g_2(t, x, \lambda) = \lambda \frac{x}{t^p} \text{ for all } p \geq 1. \quad (3.9)$$

And so by (3.8) and (3.9) the class of perturbations of (3.6) is contained in that of (3.7), yet the radius of stability of (3.6) is greater than that of (3.7).

Remark 1. The above example suggests that if we consider the property – the strength of the stability of the zero solution of (2.1) – then we need to account for both the radius of stability as well as the size of the class of allowable perturbations. One approach to obtain a measure of the size of the allowable perturbations is by using the theory of prolongations given in [2]. We plan to investigate this further.

REFERENCES

- [1] Anderson, P., Qualitative features of solutions of perturbed differential equations, Ph.D. Dissertation, The University of Texas at Arlington, 1999.
- [2] Auslander, J. and Seibert, P., Prolongations and stability in dynamical systems, Ann. Inst. Fourier, 14, 1964, 237-268.
- [3] Hubbard, J.H. and West, B.H., *Differential Equations: A Dynamical Systems Approach*, Springer-Verlag, New York, 1991.
- [4] Lakshmikantham, V. and Leela, S., *Differential and Integral Inequalities*, V.2, Academic Press, Inc., New York, 1969.
- [5] Malkin, I.G., Stability in the case of constantly acting disturbances, PMM, 8, 1944, 241-245.
- [6] Marchetti, F., Negrini, P., Salvadori, L., and Scalia, M., Liapunov direct method in approaching bifurcation problems, Ann. Mat. Pura. Appl., 108 (4), 211-226.
- [7] Massera, J.L., Contributions to stability theory, Ann. Math., (2), 64, 1956, 182-206; Correction, Ann. of Math., (2), 68, 1958, 202.
- [8] Routh, N., Habets, P., and Laloy, M., *Stability Theory by Liapunov's Direct Method*, Springer-Verlag, New York, 1977.
- [9] Salvadori, L., On the conditional total stability of equilibrium for mechanical systems, Le Mathematiche, 46, 1991, 415-427.
- [10] Salvadori, L. and Visentini, F., Sulla stabilità totale condizionata dell'equilibrio nella meccanica dei sistemi olanami, Rend. di Mat., Serie VII, 12, 1992, 475-495.
- [11] Salvadori, L. and Visentini, F., Proceedings of a conference entitled *Dynamical Systems* in Udine, Italy, 1997.
- [12] Yoshizawa, T., Stability theory by Liapunov's second method, Publications of the Mathematical Society of Japan, Japan, 1966.

2 EXTREMAL SOLUTIONS OF HEMIVARIATIONAL INEQUALITIES WITH D.C.-SUPERPOTENTIALS

Siegfried Carl

Martin-Luther-Universität Halle-Wittenberg

Fachbereich Mathematik und Informatik

Institut für Analysis

06099 Halle, Germany

1. INTRODUCTION

The variational formulation of various boundary value problems in mechanics and engineering governed by nonconvex, possibly nonsmooth energy functionals (so-called superpotentials) leads to hemivariational inequalities introduced by Panagiotopoulos, cf. e.g. [9, 12, 14], to model problems including nonmonotone, possibly multivalued constitutive laws. An abstract formulation of a hemivariational inequality reads as follows.

Let X be a reflexive Banach space and X^* its dual, let $A : X \rightarrow X^*$ be some pseudomonotone and coercive operator satisfying certain continuity conditions, and let $h \in X^*$ be some given element. Find $u \in X$ such that

$$\langle Au - h, v - u \rangle + J^0(u, v - u) \geq 0, \quad \text{for all } v \in X \quad (1.1)$$

where $J^0(u, v)$ denotes the generalized directional derivative in the sense of Clarke of a locally Lipschitz functional (superpotential) $J : X \rightarrow \mathbb{R}$, cf. e.g. [12]. An equivalent multivalued formulation of (1.1) is given by

$$Au + \partial_c J(u) \ni h, \quad \text{in } X^* \quad (1.2)$$

where $\partial_c J(u)$ denotes Clarke's generalized gradient, cf. [8, Chapter 2]. Abstract existence results for (1.1) (resp. (1.2)) can be found e.g. in [12].

In this paper we consider (1.1) (resp. (1.2)) when $A : X \rightarrow X^*$ is a nonpotential elliptic operator of Leray-Lions type acting from some Sobolev space to its dual, and the functional $J : X \rightarrow \mathbb{R}$ is given in the form of a d.c.-functional which means

that J can be represented as the difference of two convex (locally Lipschitz) functionals $J_1, J_2 : X \rightarrow \mathbb{R}$, i.e., $J(u) = J_1(u) - J_2(u)$.

Our main goal is to prove the existence of *extremal* solutions of (1.2) within a sector formed by appropriately defined upper and lower solutions when the functional J of (1.2) is an integral functional generated by a d.c.-function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$, cf. [9], of the form

$$\Phi(u) = \Phi_1(u) - \Phi_2(u), \quad (1.3)$$

where Φ_1 and Φ_2 are convex, possibly nonsmooth functions. By means of (1.3) indeed most practical engineering problems can be characterized such as nonmonotone zig-zag friction, delamination problems, plasticity with softening or semipermeability problems, cf. [14].

General existence and enclosure results have been obtained recently for both the stationary and dynamic hemivariational inequalities in [3, 4] when J is only locally Lipschitzian. However, in order to get the existence of extremal solutions some additional structure on the energy functional involved is needed. This paper extends a recent extremality result of the author obtained for dynamic hemivariational inequalities reported in a survey article [6] in that now also the operator A may be of nonpotential type. Moreover, we discuss also some relevant example for which the existence of global extremal solutions can be proved in a constructive way.

In Section 2 we give some notations and a precise formulation of the problem. Section 3 provides necessary auxiliary results, and in Section 4 we prove our main result. Finally, in Section 5 we discuss an example that covers various applications.

2. NOTATIONS AND HYPOTHESES

Let $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ be a d.c.-function, i.e.,

$$\Phi(s) = \Phi_1(s) - \Phi_2(s),$$

where $\Phi_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, 2$ are convex functions having subdifferentials $\partial\Phi_i : \mathbb{R} \rightarrow 2^{\mathbb{R}/\emptyset}$, $i = 1, 2$, defined on the whole real line. It is well known that the subdifferentials $\partial\Phi_i$ are generated by nondecreasing functions $f_i : \mathbb{R} \rightarrow \mathbb{R}$ in the following way

$$\partial\Phi_i(s) = [f_i(s-), f_i(s+)]. \quad (2.1)$$

Here $f_i(s \pm)$ denote the one-sided limits given by $f_i(s \pm) := \lim_{\varepsilon \downarrow 0} f_i(s \pm \varepsilon)$.

Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary $\partial\Omega$. We denote by $V = W^{1,p}(\Omega)$ and $V_0 = W_0^{1,p}$ the usual Sobolev spaces with $1 < p < \infty$, and V^* and $V_0^* = W^{-1,q}(\Omega)$, $1/p + 1/q = 1$, their corresponding dual spaces, respectively,

and $\langle \cdot, \cdot \rangle$ the duality pairing between them. We introduce the natural partial ordering in $L^p(\Omega)$, that is $u \leq w$ if and only if $w - u$ belongs to the set $L_+^p(\Omega)$ of all nonnegative elements of $L^p(\Omega)$, which induces also a partial ordering in the Sobolev space V . If $u, w \in V$ and $u \leq w$, then

$$[u, w] = \{v \in V \mid u \leq v \leq w\}$$

denotes the order interval formed by u and w .

Let A be a quasilinear elliptic differential operator of the form

$$Au = -\sum_{i=1}^N \frac{\partial}{\partial x_i} a_i(\cdot, u, \nabla u)$$

and assume the following conditions on the coefficients $a_i (i = 1, \dots, N)$:

(A1) Each $a_i : \Omega \times \mathbb{R} \times \mathbb{R}^N \rightarrow \mathbb{R}$ satisfies Carathéodory conditions, i.e., $a_i(x, t, \xi)$ is measurable in $x \in \Omega$ for all $(t, \xi) \in \mathbb{R} \times \mathbb{R}^N$ and continuous in (t, ξ) for almost every (a.e.) $x \in \Omega$. There exist a constant $c_0 > 0$ and a function $k_0 \in L^q(\Omega)$, $1/p + 1/q = 1$, such that

$$|a_i(x, t, \xi)| \leq k_0(x) + c_0(|t|^{p-1} + |\xi|^{p-1}),$$

for a.e. $x \in \Omega$ and for all $(t, \xi) \in \mathbb{R} \times \mathbb{R}^N$.

(A2) $\sum_{i=1}^N (a_i(x, t, \xi) - a_i(x, t, \xi'))(\xi_i - \xi'_i) \geq \mu |\xi - \xi'|^p$ for a.e. $x \in \Omega$, for all $t \in \mathbb{R}$, and for all $\xi, \xi' \in \mathbb{R}^N$ with μ being some positive constant.

(A3) $|a_i(x, t, \xi) - a_i(x, t', \xi)| \leq [k_1(x) + |t|^{p-1} + |t'|^{p-1} + |\xi|^{p-1}] \omega(|t - t'|)$, $i = 1, \dots, N$, for some function $k_1 \in L^q(\Omega)$, for a.e. $x \in \Omega$, for all $t, t' \in \mathbb{R}$ and for all $\xi \in \mathbb{R}^N$, where $\omega : [0, \infty) \rightarrow [0, \infty)$ is the *modulus of continuity* satisfying

$$\int_0^\infty \frac{dr}{\omega^q(r)} = +\infty. \quad (2.2)$$

Remark 2.1. Hypothesis (A3) is satisfied for example in case that $\omega(|t - t'|) = c|t - t'|^{1/q}$ with some positive constant c , i.e., the coefficients $a_i(x, t, \xi)$ satisfy a Hölder condition with respect to t .

Let a denote the semilinear form associated with the differential operator A by

$$\langle Au, \varphi \rangle := a(u, \varphi) = \int_{\Omega} \sum_{i=1}^N a_i(x, u, \nabla u) \frac{\partial \varphi}{\partial x_i} dx,$$

which due to (A1) is well defined on $V \times V$.

Let J be the integral functional defined by means of the d.c.-function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ as follows:

$$J(u) = \int_{\Omega} \Phi(u(x)) dx. \quad (2.3)$$

Since Φ is locally Lipschitz, its generalized gradient $\partial_c \Phi$ is well-defined. Assuming a growth condition of $\partial_c \Phi$ in the form

$$\eta \in \partial_c \Phi(s) \text{ implies } |\eta| \leq c(1 + |s|^{p-1}), \text{ for all } s \in \mathbb{R}, \quad (2.4)$$

then the integral functional (2.3) is well defined on $L^p(\Omega)$ and is locally Lipschitzian. Its generalized gradient satisfies the following relation

$$\partial_c J(u)(x) \subset \partial_c \Phi(u(x)) \quad (2.5)$$

which holds for J considered as a mapping from $L^p(\Omega)$ and V as well, [8, Chapter 2.7] and [7, 4]. Furthermore, if in addition either $\Phi(s)$ or $-\Phi(s)$ is regular for all $s \in \mathbb{R}$ in the sense of Clarke which means that cf. [8, Chapter 2.3.4]

- (i) for all $r \in \mathbb{R}$ the one-sided directional derivative $\Phi'(s; r)$ exists, and
- (ii) for all $r \in \mathbb{R}$, $\Phi'(s; r) = \Phi^0(s; r)$

is satisfied (respectively for $-\Phi$), then equality holds in (2.5), cf. [8, Chapter 2.7]. By means of generalized subdifferential calculus according to [8, Chapter 2.3] regularity of either the d.c.-function Φ or $-\Phi$ is ensured if the convex functions Φ_i in its representation satisfy the following hypothesis:

(H1) For each $s \in \mathbb{R}$ one of the subdifferentials $\partial \Phi_i$ of the convex functions $\Phi_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, 2$, is a singleton.

Thus under (H1) equality holds in (2.5) and, moreover, we have

$$\partial_c \Phi(s) = \partial \Phi_1(s) - \partial \Phi_2(s), \quad \text{for all } s \in \mathbb{R}. \quad (2.6)$$

This is because convexity of $\Phi_i : \mathbb{R} \rightarrow \mathbb{R}$ implies that Φ_i are locally Lipschitzian, and thus from [8, Proposition 2.2.4] it follows that by our assumption (H1) for each $s \in \mathbb{R}$ one of the functions Φ_i must be strictly differentiable in s . Using [8, Chapter 2.3, Corollary 2] the latter implies

$$\partial_c \Phi(s) = \partial_c \Phi_1(s) - \partial_c \Phi_2(s)$$

which yields in view of $\partial_c \Phi_i(s) = \partial \Phi_i(s)$ the relation (2.6). Thus in what follows we shall assume that the d.c.-function Φ is generated by a pair of convex functions Φ_i , $i = 1, 2$ satisfying (H1).

Let $h \in V_0^*$ be given. In this paper we consider the following hemivariational problem:

Find $u \in V_0$ such that

$$\langle Au, \varphi - u \rangle + \int_{\Omega} \Phi^0(u, \varphi - u) dx \geq \langle h, \varphi - u \rangle, \quad \text{for all } \varphi \in V_0, \quad (2.7)$$

where A is the quasilinear elliptic operator defined above, and Φ^0 denotes the generalized directional derivative in the sense of Clarke of the d.c.-function Φ . Under (H1) and (2.4) the hemivariational inequality (2.7) is equivalent with the following problem:

Find $u \in V_0$ such that

$$Au + w = h, \quad \text{in } V_0^*, \quad (2.8)$$

where $w \in L^q(\Omega) \subset V_0^*$ satisfies $w(x) \in \partial_c \Phi(u(x))$ for a.e. $x \in \Omega$. Furthermore, by (H1) and thus relation (2.6), it follows that any $w \in \partial_c \Phi(s)$ has a unique representation in the form $w = w_1 - w_2$ where $w_i \in \partial \Phi_i(s)$, $i = 1, 2$, for each $s \in \mathbb{R}$. This gives rise to the following equivalent definition of a (weak) solution of problem (2.7) and (2.8), respectively, which will be used in our investigations.

Definition 2.1. A function $u \in V_0$ is called a *solution* of problem (2.7) if there is a function triple $(u, w_1, w_2) \in V_0 \times L^q(\Omega) \times L^q(\Omega)$ such that

- (i) $w_i(x) \in \partial \Phi_i(u(x))$ for a.e. $x \in \Omega$,
- (ii) $a(u, \varphi) + \int_{\Omega} (w_1 - w_2) \varphi dx = \langle h, \varphi \rangle$ for all $\varphi \in V_0$.

Let $f_i : \mathbb{R} \rightarrow \mathbb{R}$ be the nondecreasing functions generating the subdifferentials $\partial \Phi_i$, $i = 1, 2$. By $\bar{f}_i, \underline{f}_i : \mathbb{R} \rightarrow \mathbb{R}$ we denote the one-sided limits, i.e.,

$$\bar{f}_i(s) := f_i(s+), \quad \underline{f}_i(s) := f_i(s-),$$

which exist due to the monotonicity of f_i . Denote by \bar{F}_i and \underline{F}_i the Nemytskij operators related with \bar{f}_i and \underline{f}_i , respectively. Now we introduce the notion of upper and lower solution for the hemivariational inequality (2.7) as follows.

Definition 2.2. A function $\bar{u} \in V$ is called an *upper solution* of the hemivariational inequality (2.7) if there is a pair $(\bar{u}, \bar{w}_1) \in V \times L^q(\Omega)$ such that

- (i) $\bar{u} \geq 0$ on $\partial \Omega$,
- (ii) $\bar{w}_1(x) \in \partial \Phi_1(\bar{u}(x))$ for a.e. $x \in \Omega$,
- (iii) $a(\bar{u}, \varphi) + \int_{\Omega} \bar{w}_1 \varphi dx \geq \int_{\Omega} \bar{F}_2(\bar{u}) \varphi dx + \langle h, \varphi \rangle$ for all $\varphi \in V_0 \cap L_+^p(\Omega)$.

Similarly, a function $\underline{u} \in V$ is a *lower solution* for (2.7) if there is a pair $(\underline{u}, \underline{w}_1) \in V \times L^q(\Omega)$ such that $\underline{w}_1(x) \in \partial \Phi_1(\underline{u}(x))$ and reversed inequalities hold in (i) and (iii) of Definition 2.2 with \bar{u}, \bar{w}_1 and \bar{F}_2 replaced by $\underline{u}, \underline{w}_1$ and \underline{F}_2 , respectively. Further, instead of the growth condition (2.4), it is enough to assume the following hypothesis which may be considered as a growth condition of $\partial_c \Phi$ within the interval formed by an ordered pair of upper and lower solutions.

(H2) There is a constant $\alpha > 0$ such that $\bar{F}_1(\bar{u} + \alpha), \underline{F}_1(\underline{u} - \alpha), \bar{F}_2(\bar{u}), \underline{F}_2(\underline{u}) \in L^q(\Omega)$ for given upper and lower solutions \bar{u} and \underline{u} , respectively, satisfying $\underline{u} < \bar{u}$.

Finally, we define the notion of extremal solutions with respect to an order interval.

Definition 2.3. A solution u^* is called the *greatest solution* within the order interval $[\underline{u}, \bar{u}]$ if for any solution $u \in [\underline{u}, \bar{u}]$ we have $u \leq u^*$. Similarly, u_* is the *least solution* in $[\underline{u}, \bar{u}]$ if for any solution $u \in [\underline{u}, \bar{u}]$ it holds $u_* \leq u$. The least and greatest solutions are called the *extremal* ones within the sector $[\underline{u}, \bar{u}]$.

Our main goal is to prove the existence of extremal solutions of problem (2.7) within an order interval $[\underline{u}, \bar{u}]$ of upper and lower solutions. Furthermore, it will be shown that these extremal solutions may be characterized as specific solutions of some variational inequalities involving discontinuous nonlinearities.

Remark 2.2. It should be noted that regularity of Φ or $-\Phi$ in the sense of Clarke does not mean its classical differentiability as it can be seen from (2.6), since only one of the subdifferentials $\partial\Phi_i(s)$ needs to be a singleton.

3. PRELIMINARIES

Let \bar{u} and \underline{u} be upper and lower solutions for (2.7) satisfying $\underline{u} \leq \bar{u}$. Throughout this section we shall assume that the hypotheses (A1), (A2), (A3), and (H1), (H2) are satisfied.

Let $\tilde{f}_2 : \mathbb{R} \rightarrow \mathbb{R}$ be any single valued selection of $\partial\Phi_2$, i.e., $\tilde{f}_2(s) \in \partial\Phi_2(s)$ for all $s \in \mathbb{R}$. Special selections of $\partial\Phi_2$ are $\bar{f}_2, \underline{f}_2$ and any other selection \tilde{f}_2 satisfies $\underline{f}_2(s) \leq \tilde{f}_2(s) \leq \bar{f}_2(s)$ for all $s \in \mathbb{R}$. For fixed $v \in L^p(\Omega)$ consider the following variational inequality

$$Au + \partial\Phi_1(u) \ni \tilde{F}_2(v) + h \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (3.1)$$

where \tilde{F}_2 denotes the Nemytskij operator related with \tilde{f}_2 . In just the same way as in Definition 2.2, a function $\bar{U} \in V$ is called an upper solution for problem (3.1) if there is an $\bar{W} \in L^q(\Omega)$ such that $\bar{W}(x) \in \partial\Phi_1(\bar{U}(x))$ a.e. in Ω and

4. $\bar{U} \geq 0$ on $\partial\Omega$,
5. $a(\bar{U}, \varphi) + \int_{\Omega} \bar{W} \varphi dx \geq \int_{\Omega} \tilde{F}_2(v) \varphi dx + \langle h, \varphi \rangle$, for all $\varphi \in V_0 \cap L^p_+(\Omega)$.

Similarly, a lower solution \underline{U} is defined. The following lemma is a special case of a result obtained by the author in [5].

Lemma 3.1. Let $v \in [\underline{u}, \bar{u}]$ and let \bar{U} and \underline{U} be any upper and lower solution of the BVP (3.1), respectively, satisfying $\underline{U} \leq \bar{U}$ and belonging to the interval $[\underline{u}, \bar{u}]$. Then the variational inequality (3.1) has extremal solutions $U^*, U_* \in [\underline{U}, \bar{U}]$.

In order to deduce Lemma 3.1 from the result in [5], we only need to make sure that for any $v \in [\underline{v}, \bar{v}]$ $\tilde{F}_2(v)$ is well defined. This, however, follows from hypothesis (H2), the inequality $\underline{f}_2(s) \leq \tilde{f}_2(s) \leq \bar{f}_2(s)$, and the monotonicity of the nonlinearities involved. Note also that due to the last inequality the given \bar{u} and \underline{u} are also upper and lower solutions of (3.1) for any $v \in [\underline{v}, \bar{v}]$.

By means of Lemma 3.1, we are going to prove an extremality result for the following related variational inequality involving the discontinuous nonlinear term \tilde{F}_2 , i.e.,

$$Au + \partial\Phi_1(u) \ni \tilde{F}_2(u) + h \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (3.2)$$

where the Nemytskij operator \tilde{F}_2 is generated by any single valued selection \tilde{f}_2 of $\partial\Phi_2$. In the proof of the extremality result for (3.2), which will be given in Lemma 3.3, we apply the following fixed point result for increasing mappings in partially ordered sets due to S. Heikkilä, cf. [11, Proposition 1.2.1].

Lemma 3.2. Given a nondecreasing mapping $P : Z \mapsto Z$ of a partially ordered set Z to itself and let $\bar{u} \in Z$. Then there exists a unique inversely well-ordered chain C in Z , called an i.w.o. chain of P -iterations of \bar{u} , satisfying

$$\bar{u} = \max C \text{ and if } u < \bar{u} \text{ then } u \in C \text{ iff } u = \inf P\{z \in C \mid z > u\}.$$

If $u^* = \inf P[C]$ exists and $P\bar{u} \leq \bar{u}$, then u^* is the greatest fixed point of P in

$$(\bar{u}) := \{z \in Z \mid z \leq \bar{u}\}.$$

Lemma 3.3. Let $\bar{v}, \underline{v} \in V$ be any upper and lower solutions of the BVP (3.2) satisfying $\underline{u} \leq \underline{v} \leq \bar{v} \leq \bar{u}$ where \bar{u} and \underline{u} are the given upper and lower solutions of the original problem (2.7) according to hypothesis (H2). Then for any fixed single-valued selection \tilde{f}_2 of $\partial\Phi_2$ the BVP (3.2) possesses extremal solutions within the interval $[\underline{v}, \bar{v}]$.

Proof: Note that, in particular, \bar{u} and \underline{u} are also upper and lower solutions, respectively, of problem (3.2). In the proof we focus on the existence of the greatest solution of (3.2) within $[\underline{v}, \bar{v}]$, since the existence of the least solution can be shown in a similar way. Lemma 3.1 and Lemma 3.2 will be the main tools used in the proof. Let us introduce the following partially ordered set Z given by

$$Z := \{z \in V \mid z \in [\underline{v}, \bar{v}] \text{ and } z \text{ is an upper solution of the BVP (3.2)}\},$$

and define an operator P that assigns to each $z \in Z$ the greatest solution of the BVP

$$Au + \partial\Phi_1(u) \ni \tilde{F}_2(z) + h \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega \quad (3.3)$$

within the interval $[\underline{v}, z]$. In view of the monotonicity of \tilde{F}_2 , we obtain that $\bar{U} := z$ and $\underline{U} := \underline{v}$ are upper and lower solutions of (3.3), respectively. Thus the existence of extremal solutions of (3.3) within $[\underline{v}, z]$ follows from Lemma 3.1 which proves that the operator P is well defined, and $Pz \leq z$ holds. Moreover, the monotonicity of \tilde{F}_2 implies that Pz is again an upper solution of (3.2) which shows that $P : Z \rightarrow Z$. By means of the extremality result of [5], we shall show that P is nondecreasing. To this end let $z, \hat{z} \in Z$ satisfy $z \leq \hat{z}$. We have to show that $Pz \leq P\hat{z}$ holds. By definition Pz and $P\hat{z}$ are the corresponding greatest solutions of (3.3) with respect to the interval $[\underline{v}, z]$ and $[\underline{v}, \hat{z}]$, respectively, i.e.,

$$A(Pz) + \partial\Phi_1(Pz) \ni \tilde{F}_2(z) + h \quad (\text{a})$$

$$A(P\hat{z}) + \partial\Phi_1(P\hat{z}) \ni \tilde{F}_2(\hat{z}) + h. \quad (\text{b})$$

Sine the inequality $\underline{v} \leq Pz \leq z \leq \hat{z}$ holds, it follows by the monotonicity of \tilde{F}_2 that Pz is a special lower solution for problem (b). But \hat{z} is also an upper solution of problem (b) which implies the existence of solutions of problem (b) within the interval $[Pz, \hat{z}]$. However, $P\hat{z}$ is the greatest solution of (b) within the larger interval $[\underline{v}, \hat{z}] \supset [Pz, \hat{z}]$ and thus it follows $Pz \leq P\hat{z}$.

To apply the fixed point result formulated by Lemma 3.2, we still have to show that whenever C is the inversely well ordered (i.w.o.) chain of P -iterations generated by \bar{v} as defined in Lemma 3.2, then $u^* = \inf P[C] \in Z$ exists.

Let C be the i.w.o. chain of P -iterations of \bar{v} . Because P is nondecreasing, then $P[C]$ is also an inversely well-ordered chain in $[\underline{v}, \bar{v}]$, cf. [11]. Since $P[C]$ is uniformly $L^p(\Omega)$ -bounded, we get from (A1), (A2), and (H2) that $P[C]$ is uniformly bounded also in V_0 which yields by [11, Lemma 4.1.2] the existence of a nonincreasing sequence $(u_n)_{n=0}^\infty$ in $P[C]$ converging to $u^* = \inf P[C] \in V_0$ weakly in V_0 and strongly in $L^p(\Omega)$ as $n \rightarrow \infty$. By definition, $u_n = Pz_n$ satisfies $\underline{v} \leq u_n \leq z_n$ and

$$Au_n + w_1^n = \tilde{F}_2(z_n) + h \quad \text{in } \Omega, \quad u_n = 0 \quad \text{on } \partial\Omega, \quad (3.4)$$

where $w_1^n \in \partial\Phi_1(u_n)$. Since $u_n \in [\underline{v}, \bar{v}]$, it follows from (H2) that (w_1^n) is bounded in $L^q(\Omega)$ and thus, there is a subsequence of (w_1^n) (again denoted by (w_1^n)) which is weakly convergent in $L^q(\Omega)$ to w_1^* . Since the convex function $\Phi_1 : \mathbb{R} \rightarrow \mathbb{R}$ is continuous (even locally Lipschitzian), by [1, Prop. 2.7, Chapter 2] it follows that $w_1^n(x) \in \partial\Phi_1(u_n(x))$ if and only if $w_1^n \in \partial J_1(u_n)$ where the functional $J_1 : L^p(\Omega) \rightarrow (-\infty, \infty]$ defined by

$$J_1(u) = \begin{cases} \int_{\Omega} \Phi_1(u(x,t)) dx & \text{if } \Phi_1(u) \in L^1(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

is convex and lower-semicontinuous on $L^p(\Omega)$. Thus for any $\varphi \in L^p(\Omega)$ and all n we get

$$J_1(\varphi) \geq J_1(u_n) + \int_{\Omega} w_1^n(\varphi - u_n) dx$$

which yields as $n \rightarrow \infty$

$$J_1(\varphi) \geq J(u^*) + \int_{\Omega} w_1^*(\varphi - u^*) dx$$

or $w_1^*(x) \in \partial \Phi_1(u^*(x))$, for a.e. $x \in \Omega$. From (3.4) we obtain by the monotonicity of \tilde{F}_2 along with $\underline{v} \leq u^* \leq u_n \leq z_n$ the following inequality

$$Au_n + w_1^n \geq \tilde{F}_2(u^*) + h \quad \text{in } \Omega, \quad u_n = 0 \quad \text{on } \partial\Omega, \quad (3.5)$$

for all n . Taking the special test function $\varphi = u_n - u^*$ in the weak formulation of (3.4) we obtain

$$\begin{aligned} \langle Au_n, u_n - u^* \rangle &= a(u_n, u_n - u^*) = \langle h, u_n - u^* \rangle \\ &\quad + \int_{\Omega} (\tilde{F}_2(z_n) - w_1^n)(u_n - u^*) dx. \end{aligned} \quad (3.6)$$

Since $u_n \rightarrow u^*$ strongly in $L^p(\Omega)$ and $u_n \rightarrow u^*$ weakly in V_0 , by the boundedness in $L^q(\Omega)$ of the sequences $(\tilde{F}_2(z_n))$ and (w_1^n) from (3.6) we get

$$\limsup_{n \rightarrow \infty} a(u_n, u_n - u^*) \leq 0. \quad (3.7)$$

Hypotheses (A1) and (A2) imply that the operator $A : V_0 \rightarrow V_0^*$ is pseudomonotone and, moreover, satisfies the (S_+) property, cf. e.g. [15], which yields by (3.7) the strong convergence $u_n \rightarrow u^*$ in V_0 . This allows to pass to the limit in (3.5) as $n \rightarrow \infty$ which yields

$$Au^* + w_1^* \geq \tilde{F}_2(u^*) + h \quad \text{in } \Omega, \quad u^* = 0 \quad \text{on } \partial\Omega. \quad (3.8)$$

Hence, from (3.8) we see that $u^* = \inf P[C]$ is an upper solution of (3.2), i.e., $u^* \in Z$. Now Lemma 3.2 can be applied which proves that u^* is greatest fixed point of P in $(\bar{v}) = \{z \in Z \mid z \leq \bar{v}\}$. Hence, it follows that u^* must be greatest solution of (3.2) within the interval $[\underline{v}, \bar{v}]$.

4. MAIN RESULT

The main result of this paper is given in the following theorem.

Theorem 4.1. Let \bar{u} and \underline{u} be upper and lower solutions of the BVP (2.7) satisfying $\underline{u} \leq \bar{u}$, and let hypotheses (A1), (A2), (A3), and (H1), (H2) be satisfied. Then problem (2.7) possesses extremal solutions within the sector $[\underline{u}, \bar{u}]$. Moreover, these extremal solutions can be characterized as specific solutions of some variational inequalities involving discontinuous nonlinearities.

Proof: Let \bar{u} and \underline{u} be the given upper and lower solutions of the original hemivariational inequality (2.7), respectively, according to Definition 2.2. Consider the following variational inequalities with discontinuous nonlinearities:

$$Au + \partial\Phi_1(u) \ni \bar{F}_2(u) + h \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (4.1)$$

and

$$Au + \partial\Phi_1(u) \ni \underline{F}_2(u) + h \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (4.2)$$

where the Nemytskij operators \bar{F}_2 and \underline{F}_2 are related with the special single valued selection \bar{f}_2 and \underline{f}_2 of $\partial\Phi_2$, respectively, introduced in Section 2. It can easily be seen that \bar{u} and \underline{u} are upper and lower solutions, respectively, for both problems (4.1) and (4.2). According to Lemma 3.3 there exist the least and greatest solution \bar{u}_* and \bar{u}^* of (4.1) and \underline{u}_* and \underline{u}^* of (4.2), respectively, within the order interval $[\underline{u}, \bar{u}]$. First notice that any solution of (4.1) or (4.2) is also a solution of the original problem (2.7). We shall show that the greatest solution \bar{u}^* of (4.1) and the least solution \underline{u}_* of (4.2) are the greatest and least solutions of our original problem (2.7), respectively, within the interval $[\underline{u}, \bar{u}]$, which yields also a characterization of the extremal solutions. To this end let $u \in [\underline{u}, \bar{u}]$ be any solution of (2.7), i.e., there is a triple $(u, w_1, w_2) \in V_0 \times L^q(\Omega) \times L^q(\Omega)$ satisfying

$$Au + w_1 = w_2 + h \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (4.3)$$

where $w_i \in \partial\Phi_i(u)$, $i = 1, 2$. We shall show that $u \leq \bar{u}^*$ holds.

Since $w_2 \in \partial\Phi_2(u)$, it follows $w_2 \leq \bar{F}_2(u)$, and hence, u is a lower solution for problem (4.1). Then by Lemma 3.3 there exist extremal solutions of (4.1) with respect to the interval $[\underline{u}, \bar{u}]$. However, \bar{u}^* is the greatest solution of (4.1) with respect to the bigger interval $[\underline{u}, \bar{u}] \supset [\underline{u}, \bar{u}]$ which implies that \bar{u}^* exceeds also any solution of (4.1) out of $[\underline{u}, \bar{u}]$ and hence, it follows $u \leq \bar{u}^*$.

In a similar way one can prove $\underline{u}_* \leq u$ for any solution $u \in [\underline{u}, \bar{u}]$ of (2.7). This completes the proof of our main result.

Remark 4.1. Notice that the comparison of u and \bar{u}^* as given in the proof above requires the extremality result of Lemma 3.3 which is based on extremality results for nonmonotone quasilinear elliptic operators of Leray-Lions type (Lemma 3.1)

and on an abstract fixed point theorem for monotone increasing operators in partially ordered sets (Lemma 3.2).

Remark 4.2. According to the proof of Theorem 4.1 the extremal solutions u^* and u_* of the hemivariational inequality (2.7) within the interval $[\underline{u}, \bar{u}]$ formed by upper and lower solutions are given by the greatest solution \bar{u}^* of (4.1) and the least solution \underline{u}_* of (4.2), respectively.

5. EXAMPLE

For a number of relevant applications such as for example the nonmonotone interior semipermeability problem, cf. e.g. [9, 12, 13], one can find easily upper and lower solutions by simple computations. Moreover, in some cases these upper and lower solutions can be shown to be global bounds for any solution of a particular problem, such that by our main result the existence of global extremal solutions can be ensured. To illustrate this we consider the following example:

$$-\Delta u + \partial_c \Phi(u) \ni g \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (5.1)$$

where $g \in L^2(\Omega)$ is some given function and $\partial_c \Phi(u)$ is the generalized gradient of some locally Lipschitz d.c.-function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ satisfying a growth condition in the form

$$\eta \in \partial_c \Phi(s) : |\eta| \leq c(1 + |s|) \quad \text{for all } s \in \mathbb{R}. \quad (5.2)$$

Let us assume that the generalized gradient $\partial_c \Phi(u)$ may be decomposed into the difference of two subdifferentials, i.e., $\partial_c \Phi(u) = \partial \Phi_1(u) - \partial \Phi_2(u)$ such that hypothesis (H1) of Section 2 is satisfied. Then problem (5.1) may equivalently be written in the form

$$-\Delta u + \partial \Phi_1(u) - \partial \Phi_2(u) \ni g \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega. \quad (5.3)$$

Let $V = W^{1,2}(\Omega)$, $V_0 = W_0^{1,2}(\Omega)$ and

$$a(u, \varphi) = \int_{\Omega} \nabla u \cdot \nabla \varphi \, dx$$

then $u \in V_0$ is a solution of (5.1) if and only if there are $w_i \in L^2(\Omega)$, $i = 1, 2$ such that

6. $w_i(x) \in \partial \Phi_i(u(x))$, for a.e. $x \in \Omega$,

7. $a(u, \varphi) + \int_{\Omega} w_1 \varphi \, dx = \int_{\Omega} (w_2 + g) \varphi \, dx$, for all $\varphi \in V_0$.

Let $f_i : \mathbb{R} \rightarrow \mathbb{R}$ be the nondecreasing functions which generate the subdifferentials $\partial \Phi_i$ and $\bar{f}_i, \underline{f}_i : \mathbb{R} \rightarrow \mathbb{R}$ the corresponding one-sided limits, i.e.,

$$\partial \Phi_i(s) = [\underline{f}_i(s), \bar{f}_i(s)].$$

Then according to Definition 2.2 the function $\bar{u} \in V$ is an upper solution of (5.1) if there is a $\bar{w}_1 \in L^2(\Omega)$ such that

- (i) $\bar{u} \geq 0$ on $\partial\Omega$,
- (ii) $\bar{w}_1(x) \in \partial\Phi_1(\bar{u}(x))$ for a.e. $x \in \Omega$,
- (iii) $a(\bar{u}, \varphi) + \int_{\Omega} \bar{w}_1 \varphi dx \geq \int_{\Omega} (\bar{f}_2(\bar{u}) + g) \varphi dx$ for all $\varphi \in V_0 \cap L^2_+(\Omega)$.

Similarly a lower solution of (5.1) is defined. For problem (5.1) the following global existence result holds.

Theorem 5.1. Let the subdifferential $\partial\Phi_2 : \mathbb{R} \rightarrow 2^{\mathbb{R}}/\emptyset$ be bounded, i.e., there is a constant $k > 0$ such that

$$\eta \in \partial\Phi_2(s) : |\eta| \leq k \quad \text{for all } s \in \mathbb{R}.$$

Then under the assumption on Φ made above the BVP (5.1) has global extremal solutions, which means that among all solutions, that (5.1) may have there are extremal ones with respect to the underlying natural partial ordering.

Proof: Due to the boundedness condition on $\partial\Phi_2$, one readily verifies that any upper solution of the problem

$$-\Delta u + \partial\Phi_1(u) \ni k + g(x) \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (5.4)$$

is also an upper solution of (5.1), and any lower solution of the problem

$$-\Delta u + \partial\Phi_1(u) \ni -k + g(x) \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (5.5)$$

is also a lower solution of (5.1). In particular, the uniquely defined solutions \bar{u} and \underline{u} of (5.4) and (5.5), respectively, are upper and lower solutions of (5.1) satisfying $\underline{u} \leq \bar{u}$. By applying our main result (Theorem 4.1) there exist extremal solutions of (5.1) within the order interval $[\underline{u}, \bar{u}]$.

Now we are going to show that any solution of (5.1) necessarily belongs to the interval $[\underline{u}, \bar{u}]$ which proves that the extremal solutions of (5.1) within $[\underline{u}, \bar{u}]$ are in fact global extremal ones.

Let u be any solution of (5.1) which means that $u \in V_0$ and there are $w_i \in L^2(\Omega)$ satisfying $w_i(x) \in \partial\Phi_i(u(x))$, for a.e. $x \in \Omega$ such that the following relation holds:

$$a(u, \varphi) + \int_{\Omega} w_1 \varphi dx = \int_{\Omega} (w_2 + g) \varphi dx, \quad \text{for all } \varphi \in V_0. \quad (5.6)$$

The upper solution \bar{u} of (5.1) which is the unique solution of (5.4) satisfies

$$\bar{u} \in V_0 : a(\bar{u}, \varphi) + \int_{\Omega} \bar{w}_1 \varphi dx = \int_{\Omega} (k + g) \varphi dx, \quad \text{for all } \varphi \in V_0, \quad (5.7)$$

where $\bar{w}_1(x) \in \partial\Phi_1(\bar{u}(x))$ a.e. in Ω . Subtracting (5.7) from (5.6) we get, in particular, for all $\varphi \in V_0 \cap L_+^2(\Omega)$

$$a(u - \bar{u}, \varphi) + \int_{\Omega} (w_1 - \bar{w}_1) \varphi \, dx = \int_{\Omega} (w_2 - k) \varphi \, dx \quad (5.8)$$

where $w_i(x) \in \partial\Phi_i(u(x))$ and $\bar{w}_1(x) \in \partial\Phi_1(\bar{u}(x))$ a.e. in Ω . Since the right hand side of (5.8) is nonpositive, we obtain from (5.8) by using the special test function $\varphi = (u - \bar{u})^+$

$$\left\| \nabla(u - \bar{u})^+ \right\|_{L^2(\Omega)}^2 + \int_{\Omega} (w_1 - \bar{w}_1)(u - \bar{u})^+ \, dx \leq 0. \quad (5.9)$$

The maximal monotonicity of $\partial\Phi_1$ yields

$$\int_{\Omega} (w_1 - \bar{w}_1)(u - \bar{u})^+ \, dx = \int_{\{u > \bar{u}\}} (w_1 - \bar{w}_1)(u - \bar{u}) \, dx \geq 0$$

and thus, from (5.9) it follows

$$\left\| \nabla(u - \bar{u})^+ \right\|_{L^2(\Omega)}^2 = 0$$

which due to Poincaré-Friedrichs' inequality implies $(u - \bar{u})^+ = 0$, i.e., $u \leq \bar{u}$. In just the same way, one shows that any solution u of (5.1) satisfies $u \geq \underline{u}$ where \underline{u} is the unique solution of (5.5). This completes the proof of the existence of global extremal solutions of problem (5.1).

As mentioned above, any upper solution \bar{v} of (5.4) is also an upper solution of (5.1) and any lower solution \underline{v} of (5.5) is also a lower solution of (5.1). Moreover, it can easily be seen that the inequalities

$$\underline{v} \leq \underline{u} \leq \bar{u} \leq \bar{v} \quad (5.10)$$

hold, where \bar{u} and \underline{u} are the unique solutions of (5.4) and (5.5), respectively. Determining bounds in the form \bar{v} and \underline{v} is a much easier task, and in some cases already constants can be taken for them.

To illustrate this, let us assume, in addition, that $g \in L^\infty(\Omega)$ and that there exists a positive constant \bar{s} such that

$$\partial\Phi_1(\bar{s}) \geq k + \|g\|_{L^\infty(\Omega)}$$

and a negative constant s such that

$$\partial\Phi_1(s) \leq -k - \|g\|_{L^\infty(\Omega)}.$$

Thus the constant functions $\bar{v}(x) \equiv \bar{s}$ and $\underline{v}(x) \equiv s$ are upper and lower solutions of (5.4) and (5.5), respectively, and all solutions of (5.1) belong to the interval $[s, \bar{s}]$.

Remark 5.1. According to Theorem 4.1 the greatest solution of (5.1) corresponds with the greatest solution of the variational inequality:

$$-\Delta u + \partial\Phi_1(u) \ni \bar{f}_2(u) + g \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad (5.11)$$

and the least solution of (5.1) corresponds with the least solution of

$$-\Delta u + \partial\Phi_1(u) \ni \underline{f}_2(u) + g \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega. \quad (5.12)$$

Since $\bar{f}_2 : \mathbb{R} \rightarrow \mathbb{R}$ is right-continuous and $\underline{f}_2 : \mathbb{R} \rightarrow \mathbb{R}$ is left-continuous, the greatest solution u^* of (5.11) and the least solution u_* of (5.12) can be obtained in a constructive way by a monotone iteration process, cf. [2]. For this purpose, let \bar{v} be any upper solution of (5.4) and \underline{v} be any lower solution of (5.5). The iteration

$$u_0 := \bar{v}, \quad -\Delta u_{n+1} + \partial\Phi_1(u_{n+1}) \ni \bar{f}_2(u_n) + g \quad \text{in } \Omega, \quad u_{n+1} = 0 \quad \text{on } \partial\Omega,$$

yields a sequence (u_n) converging monotonically from above to the greatest solution u^* , i.e., $u_n \rightarrow u^*$.

Similarly the iteration

$$u_0 := \underline{v}, \quad -\Delta u_{n+1} + \partial\Phi_1(u_{n+1}) \ni \underline{f}_2(u_n) + g \quad \text{in } \Omega, \quad u_{n+1} = 0 \quad \text{on } \partial\Omega,$$

yields a sequence (u_n) converging monotonically from below to the least solution u_* , i.e., $u_n \rightarrow u_*$.

REFERENCES

- [1] Barbu, V. and Precupanu, Th., *Convexity and Optimization in Banach Spaces*, Sijthoff and Noordhoff, International Publishers, 1978.
- [2] Carl, S., A combined variational-monotone iterative method for elliptic boundary value problems with discontinuous nonlinearity, *Appl. Anal.*, 43, 1992, 21-45.
- [3] Carl, S., Enclosure of solutions for quasilinear dynamic hemivariational inequalities, *Nonlinear World*, 3, 1996, 281-298.
- [4] Carl, S. and Dietrich, H., The weak upper and lower solution method for quasilinear elliptic equations with generalized subdifferentiable perturbations, *Appl. Anal.*, 56, 1995, 263-278.
- [5] Carl, S., Leray-Lions operators perturbed by state-dependent subdifferentials, *Nonlinear World*, 3, 1996, 505-518.
- [6] Carl, S., A survey of recent results on the enclosure and extremality of solutions for quasilinear hemivariational inequalities, In *From Convexity to Nonconvexity*, a volume dedicated to the memory of Professor Gaetano Fichera, Eds. R. Gilbert, P.D. Panagiotopoulos, and P. Pardalos, Kluwer Academic, in press.
- [7] Chang, K.C., Variational methods for non-differentiable functionals and their applications to partial differential equations, *J. Math. Anal. Appl.*, 80, 1981, 102-129.
- [8] Clarke, F.H., *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

- [9] Dem'yanov, F., Stavroulakis, G.E., Polyakova, L.N., and Panagiotopoulos, P.D., *Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economics*, Kluwer Academic Publishers, Dordrecht, 1996.
- [10] Duvaut, G. and Lions, J.L., *Inequalities in Mechanics and Physics*, Springer-Verlag, Berlin, 1976.
- [11] Heikkilä, S. and Lakshmikantham, V., *Monotone Iterative Techniques for Discontinuous Nonlinear Differential Equations*, Marcel Dekker, New York, 1994.
- [12] Naniewicz, Z. and Panagiotopoulos, P.D., *Mathematical Theory of Hemivariational Inequalities and Applications*, Marcel Dekker, New York, 1995.
- [13] Miettinen, M. and Panagiotopoulos, P.D., On parabolic hemivariational inequalities and applications, *Nonlinear Analysis*, 35, 1998, 885-915.
- [14] Panagiotopoulos, P.D., *Hemivariational Inequalities: Applications in Mechanics and Engineering*, Springer-Verlag, New York, 1993.
- [15] Zeidler, E., *Nonlinear Functional Analysis and Its Applications II/B: Nonlinear Monotone Operators*, Springer-Verlag, New York, 1990.

3 DEGENERATE QUASILINEAR PARABOLIC PROBLEMS WITH SLOW DIFFUSIONS

C.Y. Chan

University of Louisiana at Lafayette

and

W.Y. Chan

University of Science and Arts of Oklahoma

1. INTRODUCTION

Let $T(>0)$, $m(>1)$, $p(>0)$ and $q(\geq 0)$ denote constants, $D = (0, 1)$, $\Omega_T = D \times (0, T]$, and \overline{D} and $\overline{\Omega}_T$ be the closures of D and Ω_T respectively. We consider the following degenerate quasilinear parabolic problem,

$$\left. \begin{array}{l} x^q u_t = (u^m)_{xx} + u^p \quad \text{in } \Omega_T, \\ u(x, 0) = u_0(x) \quad \text{on } \overline{D}, \\ u(0, t) = 0 = u(1, t) \quad \text{for } t \in (0, T], \end{array} \right\} \quad (1.1)$$

where $u_0(x) \in C^{2+\alpha}(\overline{D})$ for some $\alpha \in (0, 1)$ is a positive function in D such that $u_0(0) = 0 = u_0(1)$ and

$$(u_0^m)'' + u_0^p \geq 0 \quad \text{in } D. \quad (1.2)$$

The above problem arises in plasma physics (Berryman [1] and Berryman and Holland [2]) with u denoting the particle density. The problem (1.1) describes a particle diffusion crossing a magnetic field in a toroidal octupole plasma containment device; x^q is a geometric factor, and mu^{m-1} is the diffusion coefficient. Since the term mu^{m-1} tends to zero as $u \rightarrow 0$, the problem (1.1) describes a phenomenon having a “slow diffusion”.

When $q = 0$ and $u_0(x) \geq 0$, Galaktionov [3] investigated existence of a weak solution of the n -dimensional version of the problem (1.1) in a bounded n -dimensional domain. He proved that (i) for $p < m$, a weak solution exists globally; (ii) for $p = m$, global existence of a weak solution depends on the first eigenvalue of the problem,

$$\Delta Z = -\lambda Z \text{ for } x \in \tilde{D}, \quad Z = 0 \text{ for } x \in \partial \tilde{D},$$

where Δ is the n -dimensional Laplace operator, \tilde{D} is a bounded n -dimensional domain, and $\partial \tilde{D}$ is the boundary of \tilde{D} ; (iii) for $p > m$, there exist initial conditions such that a weak solution blows up in a finite time. Similar results for more general degenerate parabolic problems with $q = 0$ were obtained by Levine and Sacks [4]. For existence of classical solutions of some degenerate parabolic problems with $q = 0$, we refer to the book of Samarskii, Galaktionov, Kurdyumov and Mikhailov [5, pp. 23 and 29-30].

For the problem (1.1) with $m = 1$, $p > 1$, and $u_0(x) \geq 0$ in D , existence and uniqueness of a classical solution were studied by Floater [6] for the case $p \leq q + 1$, and by Chan and Liu [7] for the case $p > q + 1$.

Let $v = u^m$. Then, the problem (1.1) becomes

$$\left. \begin{aligned} x^q v_t &= m v^{(m-1)/m} v_{xx} + m v^{(p+m-1)/m} && \text{in } \Omega_T, \\ v(x, 0) &= v_0(x) \quad (\equiv u_0^m(x)) && \text{on } \bar{D}, \\ v(0, t) &= 0 = v(1, t) && \text{for } t \in (0, T], \end{aligned} \right\} \quad (1.3)$$

where from (1.2),

$$v''_0 + v_0^{p/m} \geq 0 \text{ in } D.$$

Our plan in Section 2 is to show existence of a classical solution of the problem (1.3) in order to prove existence of a classical solution of the problem (1.1). In Section 3, we discuss the location of the blow-up point.

2. EXISTENCE

Let ε be a sufficiently small positive number (less than 1). We consider the following problem,

$$\left. \begin{aligned} x^q v_t &= m v^{(m-1)/m} v_{xx} + m v^{(p+m-1)/m} && \text{in } \Omega_T, \\ v(x, 0) &= v_0(x) + \varepsilon && \text{on } \bar{D}, \\ v(0, t) &= \varepsilon = v(1, t) && \text{for } t \in (0, T], \end{aligned} \right\} \quad (2.1)$$

Let v_ε denote a solution of the problem (2.1), $\delta (< 1/2)$ be a positive number,

$$D_\delta = (\delta, 1), \quad \Omega_{\delta T} = D_\delta \times (0, T],$$

and \bar{D}_δ and $\bar{\Omega}_{\delta T}$ be the closures of D_δ and $\Omega_{\delta T}$ respectively. We consider the following problem,

$$\left. \begin{aligned} x^q v_t &= m v^{(m-1)/m} v_{xx} + m v^{(p+m-1)/m} && \text{in } \Omega_{\delta T}, \\ v(x, 0) &= v_0(x) + \varepsilon && \text{on } \bar{D}_\delta, \\ v(\delta, t) &= v_0(\delta) + \varepsilon, \quad v(1, t) = \varepsilon && \text{for } t \in (0, T]. \end{aligned} \right\} \quad (2.2)$$

Let $v_{\varepsilon\delta}$ denote a solution of the problem (2.2). To establish existence of $v_{\varepsilon\delta}$, we construct a sequence $\{w_i\}$ as follows: $w_0 = v_0 + \varepsilon$, and for $i = 1, 2, 3, \dots$,

$$\left. \begin{array}{lll} x^q w_{i_i} = mw_{i-1}^{(m-1)/m} w_{i_\infty} + mw_{i-1}^{(p+m-1)/m} & \text{in} & \Omega_{\delta T}, \\ w_i(x, 0) = v_0(x) + \varepsilon & \text{on} & \overline{D}_\delta, \\ w_i(\delta, t) = v_0(\delta) + \varepsilon, \quad w_i(1, t) = \varepsilon & \text{for} & t \in (0, T]. \end{array} \right\} \quad (2.3)$$

The proofs of the results in this section are given in the paper [8].

Lemma 1. For any arbitrarily fixed δ , and any nonnegative integer i , there exists some $t_1 \in (0, T]$ such that for the problem (2.3),

- (i) $w_i \in C^{2+\alpha, 1+\alpha/2}(\overline{\Omega}_{\delta t_1})$ and is unique.
- (ii) $w_i \geq w_0$ and $w_i \geq 0$ on $\overline{\Omega}_{\delta t_1}$.
- (iii) $\{w_i\}$ is a monotone nondecreasing sequence on $\overline{\Omega}_{\delta t_1}$.

Lemma 2.

- (i) For any arbitrarily fixed δ and ε , the problem (2.2) has a solution,
 $v_{\varepsilon\delta} \in C(\overline{\Omega}_{\delta t_1}) \cap C^{2+\alpha, 1+\alpha/2}(D_\delta \times [0, t_1]).$
- (ii) $v_{\varepsilon\delta_1} \geq v_{\varepsilon\delta_2}$ on $\overline{\Omega}_{\delta t_1}$ for any arbitrarily fixed ε and any positive numbers δ_1 and δ_2 such that $\delta_1 \leq \delta_2$.
- (iii) $v_{\varepsilon_1\delta} \leq v_{\varepsilon_2\delta}$ on $\overline{\Omega}_{\delta t_1}$ for any arbitrarily fixed δ and any positive numbers ε_1 and ε_2 such that $\varepsilon_1 \leq \varepsilon_2$.

We now let δ tend to 0.

Lemma 3.

- (i) The problem (2.1) has a solution,

$$v_\varepsilon \in C(\overline{\Omega}_{t_1}) \cap C^{2+\alpha, 1+\alpha/2}(D \times [0, t_1]).$$

- (ii) $v_{\varepsilon_1} \leq v_{\varepsilon_2}$ on $\overline{\Omega}_{t_1}$ for any positive numbers ε_1 and ε_2 such that $\varepsilon_1 \leq \varepsilon_2$.

We now let ε tend to 0 to give a local existence result.

Theorem 4. The problem (1.3) has a solution,

$$v \in C(\overline{\Omega}_{t_1}) \cap C^{2+\alpha, 1+\alpha/2}(D \times [0, t_1]).$$

Using $v = u^m$, we have

$$x^q u_i = m v^{(m-1)/m} u_{xx} + (m-1) v^{-1/m} v_x u_x + v^{p/m}.$$

We note that $v^{(m-1)/m}/x^q > 0$ in Ω_{t_1} . To prove existence of u , we show the Hölder continuity of $m v^{(m-1)/m}/x^q$, $(m-1) v^{-1/m} v_x/x^q$, and $v^{p/m}/x^q$. The following result gives local existence of a solution.

Theorem 5 The problem (1.1) has a solution,

$$u \in C(\overline{\Omega}_{t_1}) \cap C^{2+\alpha, 1+\alpha/2}(D \times [0, t_1]).$$

Let t_s be the supremum of the time interval for existence of u . We modify the proof of Theorem 8 of Chan and Chan [9] to obtain the following result.

Theorem 6. The problem (1.1) has a solution,

$$u \in C(\bar{D} \times [0, t_s]) \cap C^{2+\alpha, 1+\alpha/2}(D \times [0, t_s]).$$

If $t_s < \infty$, then u is unbounded in $D \times (0, t_s)$.

3. BLOW-UP POINT

When $m = 1$, $1 < p \leq q + 1$, and $u_0(x)$ satisfies the following condition,

$$\left(\frac{u_0(x)}{x} \right)' \leq 0 \text{ for } x \in D,$$

Floater [6] proved that $x = 0$ is the only blow-up point. The proof of the next result is given in the paper [10]. It extends the result of Floater.

Theorem 7. If $1 < m < p \leq q + 1$, and

$$\left(\frac{u_0^m(x)}{x} \right)' \leq 0 \text{ for } x \in D,$$

then $x = 0$ is the only blow-up point of u .

REFERENCES

- [1] Berryman, J.G., Evolution of a stable profile for a class of nonlinear diffusion equations with fixed boundaries, *J. Math. Phys.*, 18, 1977, 2108-2115.
- [2] Berryman, J.G. and Holland C.J., Evolution of a stable profile for a class of nonlinear diffusion equations II, *J. Math. Phys.*, 19, 1978, 2476-2480.
- [3] Galaktionov, V.A., Boundary-value problem for the nonlinear parabolic equation $u_t = \Delta u^{\sigma+1} + u^\beta$, *Differential Equations*, 17, 1981, 551-555.
- [4] Levine, H.A. and Sacks, P.E., Some existence and nonexistence theorems for solutions of degenerate parabolic equations, *J. Differential Equations*, 52, 1984, 135-161.
- [5] Samarskii, A.A., Galaktionov, V.A., Kurdyumov, S.P., and Mikhailov, A.P., *Blow-up in Quasilinear Parabolic Equations*, Walter de Gruyter, New York, 1995.
- [6] Floater, M.S., Blow-up at the boundary for degenerate semilinear parabolic equations, *Arch. Rational Mech. Anal.*, 114, 1991, 57-77.
- [7] Chan, C.Y. and Liu, H.T., Global existence of solutions for degenerate semilinear parabolic problems, *Nonlinear Anal.*, 34, 1998, 617-628.
- [8] Chan, C.Y. and Chan, W.Y., Global existence of classical solutions for degenerate quasilinear parabolic problems with slow diffusions, preprint.
- [9] Chan, C.Y. and Chan, W.Y., Existence of classical solutions for degenerate semilinear parabolic problems, *Appl. Math. Comput.*, 101, 1999, 125-149.
- [10] Chan, C.Y. and Chan, W.Y., Blow-up at the boundary for degenerate quasilinear parabolic problems with slow diffusions, preprint.

4 SUPERASYMPTOTIC PERTURBATION ANALYSIS OF THE KELVIN-HELMHOLTZ INSTABILITY OF SUPERSONIC SHEAR LAYERS

S. Roy Choudhury

Department of Mathematics
University of Central Florida
Orlando, Florida 32816-1364

A global asymptotic analysis of the traveling wave Kelvin-Helmholtz instability of a supersonic, finite-width velocity shear layer is carried out. The resulting solution, comprising a composite WKBJ and boundary-layer solution, satisfying outgoing, spatially damping radiative wave boundary conditions, has important applications in elucidating the energy transfer between the fluid and the unstable traveling wave solutions. Limitations in the use of this global asymptotic solution arise from the well-known directional character of the "connection formulae" at the turning points of the potential. In order to overcome these limitations, a supersymptotic analysis is developed based on recent work of Dingle, Berry and others. The structure of the resulting traveling wave solutions agrees closely with previously computed numerical solutions. In addition, the condition for the occurrence of the traveling wave instability is derived, and the absence of this mode in compressible tangential velocity discontinuities is explained.

1. INTRODUCTION

The Kelvin-Helmholtz (K-H) instability caused by tangential velocity discontinuities in homogeneous plasma is of crucial interest in understanding the problems of space, astrophysical and geophysical situations involving sheared plasma flows. A detailed understanding of the structure and dynamics of magnetopause regions, such as the presence of the magnetospheric boundary layer and of rapid boundary motions, has been obtained from the recent satellite observations of particles and fields.

Many workers have discussed the instability of the interface between the solar wind and the magnetosphere [1-5], of coronal streamers moving through the solar

wind, the boundaries between the adjacent sectors in the solar wind, the structure of the tails of comets [6,7], and the boundaries of the jets propagating from the nuclei of extragalactic double radio sources into their lobes [8-9]. The linear K-H instability of non-magnetized shear layers has been studied for flows with a subsonic velocity change by Chandrasekhar, Syrovatskii, and Norhtrop [10], and with an abrupt velocity jump of arbitrary magnitude by Gerwin [11]. Ray and Ershkovich [12], and Miura [13] have discussed the stability of compressible, magnetized, finite width shear layers for a linear and hyperbolic tangent velocity profile. The K-H instability of a finite width, ideal magnetohydrodynamic shear layer with linear and hyperbolic tangent velocity profiles in the transition region have been discussed by Roy Choudhury and Lovelace, and Miura and Pritchett [14] and by Roy Choudhury [15] considering arbitrary magnetic field in (y, z) -plane. Uberoi [16] has investigated the finite thickness and propagation angle effects on the marginal instability by considering the three layered structure of plasma regions: the magnetosheath, the boundary layer, and the magnetosphere. Fujimoto and Terasawa [17] have carried out the study of ion inertia effects on the K-H instability of two fluid plasmas. Sharma and Shivastava [18] have presented the nonlinear analysis of the drift K-H instability for electrostatic perturbations. Malik and Singh [19] have studied chaos in the K-H instability in superposed magnetic fluids with uniform relative motion.

These studies have been generalized by various workers [20-30] to low-collision, anisotropic pressure regimes using generalized models for the anisotropic pressure.

However, a crucial feature that is still missing is a detailed understanding of the physical mechanism underlying the instability. It is clear that the traveling wave instability of finite width layers [12-15], which is clearly of great relevance in physical settings since it occurs in very large regions of the (wavenumber, Mach number) parameter space, is driven by transfer of energy from the mean fluid flow to the unstable waves. However, the detailed mechanism of this energy transfer is hard to analyze, although it is understood [14, 15] that it occurs at a “resonance” layer where the phase-speed of the waves match the flow-speed of the fluid.

In this paper, we develop a global asymptotic analysis for the short wavelength unstable modes of a supersonic finite width velocity shear layer. In particular, this elucidates the role of the wave-fluid resonance in driving the more global traveling-wave instability. It also enables us to demonstrate that this instability is absent in the absence of the wave-fluid resonance. The global asymptotic solution will also form the basis of future work considering the energy transfer by the unstable waves out of the velocity shear layer – this is considered briefly in the last section.

The remainder of this paper is organized as follows. The equations describing the shear layer are summarized in Section 2, and the role of the wave-fluid resonance is considered in preliminary fashion in Section 3. The “inner” boundary-layer or resonance layer solution is contained in Section 4. Section 5 develops the complete phase-integral (WKBJ) solution in the remainder of the shear layer, and performs the superasymptotic truncation at the least term as discussed there. Section 6 then constructs the final global solution including the imposition of the boundary conditions, while Section 7 discusses this global solution and also demonstrates that

the traveling wave instability does not occur in the absence of the wave-fluid resonance. Section 7 also briefly discusses possible future applications of the global solution to energy transport by the instability.

2. BASIC EQUATIONS

In this section we shall summarize the equations characterizing the equilibrium shear layer, as well as those describing the linear perturbation quantities. We shall follow [14], although the notation is slightly changed from that employed there to add clarity to the later discussions of the energy density and the energy flux density.

2a. Equilibrium and the Linear Eigenvalue Problem

Consider a compressible inviscid neutral fluid with an adiabatic equation of state

$$\frac{d}{dt}(p\rho^{-\gamma}) = 0$$

with $d/dt \equiv (\partial/\partial t + \mathbf{v} \cdot \nabla)$. The equilibrium configuration has a flow velocity $\mathbf{v} = \hat{z}\mathbf{v}_z(x)$, constant density ρ , pressure p , and temperature T . The adiabatic exponent is denoted by γ .

Expanding all physical variables as

$$\chi = \chi_0 + \epsilon \left[\hat{\chi}_1(x) e^{i(k_y y + k_z z - \omega t)} + c.c. \right] + O(\epsilon^2) \quad (2.1)$$

with the linear perturbation quantities being of order ϵ , and with $\hat{\chi}_1$ denoting the Fourier amplitudes of the linear perturbation quantities, we obtain from [14] the following set of first-order equations:

$$\begin{aligned} i(k_z v_z - \omega) \rho_0 \hat{v}_{1x} &= -\hat{p}'_1 \\ i(k_z v_z - \omega) \rho_0 \hat{v}_{1y} &= -ik_y \hat{p}_1 \\ i(k_z v_z - \omega) \rho_0 \hat{v}_{1z} &= -\rho_0 v'_z \hat{v}_{1x} - ik_z \hat{p}_1 \\ i(k_z v_z - \omega) \rho_1 + \rho_0 (ik_y \hat{v}_{1y} + ik_z \hat{v}_{1z} + \hat{v}'_{1x}) &= 0 \end{aligned} \quad (2.2)$$

and

$$\frac{\hat{p}_1}{p_0} = \frac{\gamma \hat{\rho}_1}{\rho_0}$$

Here, the prime denotes a derivative with respect to x . The frequency ω is assumed to have at least a small positive imaginary part, so that the solutions correspond to those of an initial value problem. Equation (2.2) may be combined into a composite equation [14] for the pressure perturbation \hat{p}_1 that, with appropriate boundary conditions external to the shear layer, is an eigenvalue equation with the complex frequency eigenvalue $\omega = \omega_r + i\omega_i$. This eigenvalue equation is [14]

$$\hat{p}_1'' - \frac{2u' \hat{p}_1'}{u} = B^2 (1 - u^2) \hat{p}_1. \quad (2.3)$$

The prime now denotes a derivative with respect to the dimensionless x variable $\bar{x} = x/L$, where L is the width of the shear layer that extends from $\bar{x} = -1/2$ to $\bar{x} = 1/2$. Henceforth, we will omit the overbar and denote the dimensionless x coordinate by x . In (2.3) we have introduced the dimensionless wave number

$$B \equiv kL \quad \text{with} \quad k = (k_y^2 + k_z^2)^{1/2} \quad (2.4a)$$

the dimensionless frequency

$$W \equiv \omega/kc_s \quad (2.4b)$$

and the dimensionless flow velocity

$$u \equiv \left(\frac{k_z v_z(x)}{kc_s} \right) - W. \quad (2.4c)$$

The adiabatic sound speed is $c_s \equiv (\gamma p_0/\rho_0)^{1/2}$.

A linear velocity profile is investigated in the present work: The structure of the unstable standing ($\omega_r = 0$) and traveling ($\omega_r \neq 0$) modes for this velocity profile was investigated earlier [14, 15] for both unmagnetized and magnetized shear layers. The results for “sinusoidal” and “hyperbolic tangent” velocity profiles [1, 7] are similar. For the linear profile

$$v_z(x \geq 1/2) = \text{constant} \equiv v_{zm} \quad (2.5a)$$

and

$$v_z(x \leq -1/2) = \text{constant} \equiv -v_{zm}. \quad (2.5b)$$

Hence,

$$u = 2Ax - W \quad (2.6a)$$

for $|x| < 1/2$, where the reduced Mach number is defined as

$$A \equiv \left(\frac{k_z M}{c_s} \right) \quad (2.6b)$$

and where

$$M \equiv \left(2 \frac{v_{zm}}{c_s} \right) \quad (2.6c)$$

is the Mach number of the shear layer.

For the external regions $|x| \geq 1/2$, $u = \pm A - W$, and hence $u' = 0$. The solutions of the pressure equation (3) in these regions are [14]

$$\hat{p}_1 = \text{constant} \exp [i(-k_- x + k_y y + k_z z - \omega t)], \quad x \leq -1/2 \quad (2.7a)$$

and

$$\hat{p}_1 = \text{constant} \exp [i(-k_+ x + k_y y + k_z z - \omega t)], \quad x \geq 1/2 \quad (2.7b)$$

where

$$\omega = \omega_r + i\omega_i \equiv kc_s(W_r + iW_i)$$

and

$$k_{\pm} \equiv B[(\pm A - W)^2 - 1]^{1/2}. \quad (2.8)$$

In (2.7) we have chosen the spatially damping outgoing solutions in the comoving frame of the fluid on either side of the shear layer [14]. Unstable modes correspond to $\omega_i > 0$, with $\omega_r = 0$ for standing modes and $\omega_r \neq 0$ for traveling modes.

2 b. Critical Points and Regions of the Shear Layer

Here, we consider the various regions and various critical locations within the shear layer. The equilibrium shear layer and the various associated points and (or) regions are shown in Figure 1.

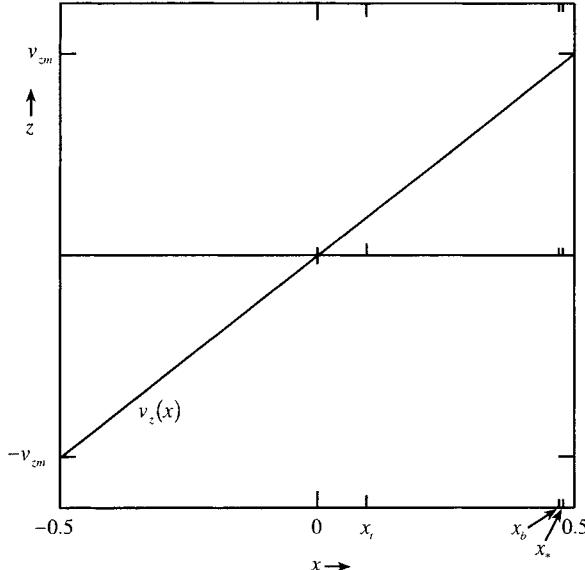


Figure 1. Geometry of the flow for a linear $v_z(x)$ profile. The typical locations of x_t , x_* , and x_b are indicated for short-wavelength modes.

The eigenvalue equation (2.3) has two turning points in the potential term at $u = \pm 1$, as well as the regular singular point at the layer where $u = 0$ for the neutrally stable case $W_i = 0$. The regular singular point at $u = 0$ is the location of the wave-fluid resonance $v_{ph} = \omega_r/k_z = v_z$. Here, we shall consider traveling modes with $\omega_r = \omega_{r1} > 0$ (with the understanding that equivalent solutions with $\omega_r = -\omega_{r1}$ and the same value of ω_i exist) [14]. The wave-fluid resonance occurs at $x = x_*$ with

$$x_* = \frac{W_r}{2A}. \quad (2.9)$$

From earlier numerical results [14, 15], we know that $W_r \rightarrow A$, and $W_i \rightarrow 0$ in the short-wavelength regime $B^2 \gg 1$. Since the instability growth rate tends to zero in this regime, the turning point $u = -1$ occurs within the velocity shear layer, and

the wave-fluid resonance at $x = x_*$ occurs very close to the right edge of the layer at $x = 1/2$. The turning point $u = 1$ falls outside the shear layer.

We shall construct a globally valid asymptotic solution to (2.3) satisfying radiative boundary conditions at $x = \pm 1/2$ obtained from (2.7). This solution will be valid in the short-wavelength regime $B^2 \gg 1$, where $W_r \rightarrow A$ and $W_i \rightarrow 0$. External to the shear layer, the solutions will be the radiative waves given by (2.7) with the complex constants c_1 and c_2 in those equations being computed self-consistently. Within the shear layer, the eigenmodes will consist of a boundary-layer solution, in the vicinity of the wave-fluid resonance, matched to a WKBJ solution with one turning point at $u = -1$ which is valid in the rest of the shear layer.

The boundary-layer solution will be applicable in a narrow region in the vicinity of $x = x_*$ and located near the right edge of the shear layer. The boundary-layer region extends from a point $x = x_b$ to the right edge $x = 1/2$. The location x_b is determined as follows. Setting $\hat{p}_1 = u\chi$ yields

$$\chi'' = \left[B^2(1-u^2) + 2\left(\frac{u'}{u}\right)^2 - \frac{u''}{u} \right] \chi, \quad (2.10)$$

This is in the Liouville-Green form [31]. For our linear velocity profile, $|u| < 2A$ when $x_* < 1/2$, or the wave-fluid resonance occurs within the shear layer. Also, $u' = 2A = \text{constant}$, and $u'' = 0$. The second term in parentheses on the right-hand side of (2.10) may clearly be neglected with respect to the first (for $B^2 \gg 1$) everywhere in the shear layer save near the wave-fluid resonance $u = 0$. In other words, away from $u = 0$, χ is given by the “related” equation [31]. This remains valid in the vicinity of the turning point $u = -1$ as will be discussed in Section 4.2. Balancing the first and second terms on the right-hand side of (2.10) to estimate the location beyond which the second term may be neglected, and defining $\delta \equiv 1/2 - x_b$, we have

$$\delta \sim 1/B \quad (2.11a)$$

$$x_b \sim (1/2 - 1/B). \quad (2.11b)$$

Clearly, $\delta \rightarrow 0$ for $B \rightarrow \infty$, so that this is a boundary-layer region. The second term on the right-hand side of (2.10) may be neglected for $x < x_b$, where the WKBJ solution will apply.

3. THE WAVE-FLUID RESONANCE

In this section we consider the role of the resonance layer in driving the traveling wave Kelvin-Helmholtz instability. Physical aspects are presented in this section, with more mathematical features being discussed in Section 7. The role of the wave-fluid resonance (referred to as the “critical layer” in the fluid mechanics literature) in the incompressible Taylor-Goldstein equation has been considered earlier. The

situation is somewhat different for compressible inviscid shear flows and much of the earlier analyses do not apply.

The unstable traveling Kelvin-Helmholtz waves act to transport energy from the mean flow to $x \rightarrow \pm\infty$. When the region where the fluid moves faster than the wave is absent, the traveling wave instability is absent. A proof of this is given in Section 7 where it is shown that, with the wave-particle resonance absent, imposing the correct outgoing radiative boundary conditions at one edge of the shear layer forces the solution at the other boundary to include an unphysical incoming wave solution. The occurrence of the traveling wave instability is thus clearly associated with the existence of the resonance layer at $x = x_*$ within the shear layer and requires

$$|W_r| \leq A \quad (3.1)$$

which agrees with the numerical solutions [14]. Physically, this means that the phase speed of the waves lies in between that of the two external regions. Notice that it is valid to interpret the function of the wave-fluid resonance as mediating the transfer of energy from the fluid flow to the waves. This is analogous to the usual interpretation in plasma physics for the Landau, or wave-particle, resonance. The nonoccurrence of the traveling modes in the compressible fluid dynamical vortex sheets is also clear now. For vortex sheets, the resonance occurs only in either external region for $W_r = \pm A$ and not within the zero-width sheet. The energy-transfer mechanism is thus absent.

Around the wave-fluid resonance $u = 0$, (2.3) may be written as

$$\hat{p}'_1 = (p''_1 - B^2 \hat{p}_1) u / 4A \quad (3.2)$$

and upon taking a derivative

$$\hat{p}''_1 = (p'''_1 - B^2 \hat{p}'_1) u / 4A + (\hat{p}''_1 - B^2 \hat{p}_1) / 2.$$

From these, we have at the resonance layer $u = 0$

$$\hat{p}'_1(x = x_*) = 0 \quad (3.3a)$$

$$\hat{p}''_1(x = x_*) = -B^2 \hat{p}_1(x = x_*). \quad (3.3b)$$

Hence, the eigenfunctions \hat{p}_1 have an extremum at the resonant layer. The sign of $\hat{p}_1(x_*)$ determines whether this is a maximum or a minimum. In that the perturbation equations are linearized, $\hat{p}_1(x_*)$ may be chosen to have either sign. We consider the case with $\hat{p}_1(x_*) < 0$, so that the \hat{p}_1 profile has a minimum at the resonant layer.

4. RESONANCE LAYER SOLUTION

In this section, we derive the boundary-layer solution around the wave-fluid resonance.

Within the boundary layer $x_b < x < 1/2$, $u^2 \ll 1$ and (2.3) may be approximated by (3.2). It is necessary to retain all three terms of this equation as may be seen by employing a stretching of the x coordinate [31]. Also, we notice that there is no

jump in \hat{p}_1 at the resonant layer, which may be seen from the discussion following (3.2) or by applying the Plemelj formula to (3.2) for $W_i \rightarrow 0_+$. Writing (3.2) in terms of the new variables

$$\begin{aligned} y &\equiv u/2A = x - W/2A \\ \zeta &\equiv (yB/2A) \end{aligned} \quad (4.1)$$

we have the inner equation in the stretched coordinate ζ

$$\left(\frac{d^2}{d\zeta^2} - \frac{2}{\zeta} \frac{d}{d\zeta} \right) \hat{p}_1 = 4A^2 \hat{p}_1.$$

The substitutions $\hat{p}_1 = \zeta^{3/2} P$ and $s = 2iA\zeta$ reduce this equation to the Bessel form

$$s^2 \frac{d^2 P}{ds^2} + s \frac{dP}{ds} + \left(s^2 - \frac{9}{4} \right) P = 0.$$

Thus, within the boundary layer, the general solution for the pressure perturbation is (with subscript IV denoting the boundary layer, see Figure 1)

$$\begin{aligned} \hat{p}_{1,IV} &= d_1 y^{3/2} J_{3/2}(iBy) + d_2 y^{3/2} J_{-3/2}(iBy) \\ &= (2/i\pi B)^{1/2} \left\{ \bar{C} \left[-2A\zeta d_1/B + i d_2/B \right] \right. \\ &\quad \left. + \bar{S} \left[d_1/B - i d_2 2A\zeta/B \right] \right\} \end{aligned} \quad (4.2)$$

where

$$\bar{C} \equiv \cosh(By) \quad (4.3a)$$

$$\bar{S} \equiv \sinh(By) \quad (4.3b)$$

and d_1 and d_2 are arbitrary constants.

5. SUPERASYMPTOTIC PHASE-INTEGRAL (WKBJ) SOLUTIONS

We shall follow the treatments of Nayfeh [31], Dingle [32] and Berry & Howls [33] to construct the solutions in the region away from the resonant layer. A superasymptotic truncation, in the sense explained later, will be employed.

Using (2.6a), (2.3) may be re-cast into the form

$$\frac{d^2 p_1}{du^2} - \frac{2}{u} \frac{dp_1}{du} - \frac{B^2}{4A^2} (1-u^2) p_1 = 0 \quad (5.1)$$

Letting

$$\tilde{u} = -u, \quad (5.2)$$

the last equation becomes

$$\frac{d^2 p_1}{d\tilde{u}^2} - \frac{2}{\tilde{u}} \frac{dp_1}{d\tilde{u}} - \frac{B^2}{4A^2} (1-\tilde{u}^2) p_1 = 0. \quad (5.3)$$

Using the Langer transformations [32]

$$\tilde{u} = e^r, \quad (5.4)$$

and

$$y = p_1 \tilde{u}^{-3/2} \quad (5.5)$$

on the independent and dependent variables respectively reduces (5.3) to the form

$$\frac{d^2y}{dr^2} = \bar{X}(r)y, \quad (5.6)$$

where the potential is

$$\bar{X}(r) = \frac{9}{4} + \frac{B^2}{4A^2} (e^{2r} - e^{4r}). \quad (5.7)$$

Defining

$$\tilde{x} = 2r \quad (5.8)$$

transforms (5.6) to

$$\frac{d^2y}{d\tilde{x}^2} = X(\tilde{x})y, \quad (5.9)$$

where

$$X(\tilde{x}) = -\frac{B^2}{4A^2} \left\{ \frac{1}{4} (e^{2\tilde{x}} - e^{4\tilde{x}}) - \frac{9}{4} \frac{A^2}{B^2} \right\}. \quad (5.10)$$

In terms of the variable \tilde{u} , the turning point of (5.6) or (5.9) which is within the shear layer is at

$$\tilde{u}_t \equiv \left[\frac{1 + \left[1 + \frac{32A^2}{B^2} \right]^{1/2}}{2} \right]^{1/2}. \quad (5.11)$$

This is most easily obtained from (2.6a) and (2.10) and may be easily expressed in terms of r and \tilde{x} using (5.4) and (5.8).

5a. Complete Solution in the Exponential Region $x_t < x < x_b$

The interval in which the solutions have an exponential form, i.e., the region to the left of x_b is thus

$$\tilde{u}(x_b) < \tilde{u} < \tilde{u}_t. \quad (5.12)$$

Using (5.4), it is thus clear that $(r - r_t) < 0$ in the region of exponential solutions, or

$$\tilde{\epsilon} = \text{sign}(r - r_t)|_{\text{exponential}} = -1. \quad (5.13)$$

Following Dingle [32], the complete phase-integral expansions on the exponential side in the short-wave-length regime are

$$(y^\pm)_{\text{exp}} = \mathbf{Y}_\pm = \exp \left\{ \pm \tilde{\epsilon} \int^{\tilde{x}} X^{1/2}(t) dt \right\} X^{-1/4} Y_\pm. \quad (5.14)$$

Here, the multipliers in (5.14) are

$$Y_\pm = \sum_{r=0}^{\infty} (\pm 1)^r Y_r, \quad Y_0 = 1, \quad (5.15a)$$

where the Y_r 's satisfy the recursion relation

$$-\tilde{\epsilon} y_{r+1} = \frac{1}{32} \int^{\tilde{x}} \chi^{-5/2} \{ 5\chi'^2 - 4\chi\chi'' - 16\chi^2\Delta \} Y_r d\tilde{x} + \frac{1}{2} \chi^{1/2} \frac{dY_r}{d\tilde{x}}. \quad (5.15b)$$

Here

$$\chi = -\frac{B^2}{16A^2} e^{\tilde{x}} (e^{\tilde{x}} - 1) \quad (5.16a)$$

and

$$\Delta \equiv X(\tilde{x}) - \chi(\tilde{x}) = \frac{9}{16}. \quad (5.16b)$$

Also, the so-called ‘singulant’ [32, 33], which is the difference of the arguments of the two exponential terms in (5.14) is

$$\begin{aligned} \tilde{\mathfrak{F}} &= 2\tilde{e} \int X^{1/2} dr \\ &= \sqrt{R} - \sqrt{a} \ln \left[\frac{2a + bz + 2\sqrt{aR}}{z} \right] \\ &\quad - \frac{bA}{B} \sin^{-1} \left[\frac{2cz + b}{\sqrt{\frac{9B^2}{4A^2} + \frac{B^4}{16A^4}}} \right]. \end{aligned} \quad (5.17)$$

Here,

$$a = 9/4, b = B^2/4A^2 = -c, z = e^{2r} = \tilde{u}^2, R = a + bz + cz^2. \quad (5.18)$$

Note that $\tilde{u} = 0$ at the resonance layer x_* , and hence the ‘singulant’ $\tilde{\mathfrak{F}}$ blows up there. Using (2.11) for an estimate of the left boundary x_b of the resonance layer together with (2.6a), (5.4), and (5.20) yields $z(x_b) = z_b = 0.003472$ at this point for the illustrative values $A = \sqrt{2}$, $B = 25$. Hence, from (5.17), we have

$$\tilde{\mathfrak{F}}(x_b) = -15.697. \quad (5.19)$$

This is a quantity which we shall subsequently need for the ‘superasymptotic’ truncation of the complete phase-integral solution (5.14).

Using (5.4), we shall write the general solution in the region (5.12) of exponential solutions (denoted by subscript III), as a linear combination

$$p_{1,III} = (\tilde{u})^{3/2} \left[GX^{-1/4} \exp \left\{ \frac{\tilde{\mathfrak{F}}}{2} \right\} Y_+ + HX^{-1/4} \exp \left\{ -\frac{\tilde{\mathfrak{F}}}{2} \right\} Y_- \right] \quad (5.20)$$

of the solutions in (5.14).

5b. Complete Solution in the Oscillatory Region $x < x_*$

Following the solutions (5.14) across the Stokes discontinuity at the turning point \tilde{u}_t in (5.11) in the manner discussed in Dingle [32], the solution on the oscillatory side (denoted by I) may be written as

$$p_{1,I} = (\tilde{u})^{3/2} \left[G y_{osc}^+ + H y_{osc}^- \right], \quad (5.21)$$

where

$$y_{osc}^+ = (-X)^{-1/4} (y_{even} \cos \psi + y_{odd} \sin \psi) \quad (5.22a)$$

$$y_{osc} = 2(-X)^{-1/4} \left(y_{even} \sin \psi - y_{odd} \cos \psi \right), \quad (5.22b)$$

with

$$y_{even} = \sum_{r=0}^{\infty} (-1)^r y_{2r}, \quad (5.22c)$$

$$y_{odd} = \sum_{r=0}^{\infty} (-1)^r y_{2r+1} \quad (5.22d)$$

$$y_r \equiv i^r Y_r \quad (5.22e)$$

$$\psi = \left| \int^{\bar{x}} (-X)^{1/2} dt \right| + \frac{\pi}{4}. \quad (5.22f)$$

For future reference, we shall need $2 \int^{\bar{x}} (-X)^{1/2} dt$ for computing ψ . This is

$$\tilde{\mathfrak{F}}_{osc} = i \left\{ \sqrt{\tilde{R}} + \frac{\tilde{a}}{\sqrt{-\tilde{a}}} \sin^{-1} \left[\frac{2\tilde{a} + \tilde{b}z}{z\sqrt{\tilde{b}^2 - 4\tilde{a}\tilde{c}}} \right] + \frac{\tilde{b}}{2\sqrt{\tilde{c}}} \ln \left[2\sqrt{\tilde{c}\tilde{R}} + 2\tilde{c}z + \tilde{b} \right] \right\}, \quad (5.23)$$

where

$$\begin{aligned} \tilde{a} &= -9/4, \tilde{b} = -B^2/4A^2 = -\tilde{c}, z = e^{2r} = \tilde{u}^2, \\ \tilde{R} &= \tilde{a} + \tilde{b}z + \tilde{c}z^2. \end{aligned} \quad (5.24)$$

5c. Solutions Near the Turning Point

Expanding the potential (24) near the turning point $r_t \equiv \ln \tilde{u}_t$, we may express (5.6) near the turning point as

$$\epsilon^2 d^2 y/dz^2 = \theta \bar{z} y \quad (5.25)$$

where

$$\bar{z} \equiv (r_t - r), \quad (5.26)$$

and

$$\theta \equiv (-2e^{2r_t} + 4e^{4r_t}) \quad (5.27a)$$

$$\epsilon \equiv 2A/B. \quad (5.27b)$$

Using a substitution $t = \epsilon^{-2/3} \theta^{1/3} \bar{z}$ to recast (5.25) into the Airy equation $d^2 y/dt^2 = ty$, and also using (5.5), the solution near the turning point is

$$p_{1,II} = (\tilde{u})^{3/2} \left[EAi(\epsilon^{-2/3} \theta^{1/3} \bar{z}) + F Bi(\epsilon^{-2/3} \theta^{1/3} \bar{z}) \right]. \quad (5.28)$$

Carrying out the asymptotic matching between $p_{1,II}$ and $p_{1,III}$ in the standard way [31, 32] yields the relations (note that $\tilde{\epsilon} = -1$ in $p_{1,III}$)

$$2G = \frac{E(\theta\pi)^{1/6}}{\sqrt{\pi}} \quad (5.29)$$

and

$$H = \frac{F(\theta\pi)^{1/6}}{\sqrt{\pi}}. \quad (5.30)$$

5d. Superasymptotic Truncation

There has been controversy surrounding the bi-directionality of connection formulas such as (5.22), (5.29) and (5.30), i.e. whether one can go from the oscillatory to the exponential region and vice-versa. Dingle [32] and Berry [33] discuss the issue, and the conflicting conclusions of various authors, and conclude that the contradictions arise from the neglect of exponentially small terms in the usual Poincare asymptotics [31 – 33]. This may be remedied, and the connection formulae rendered truly bi-directional by retaining exponentially small terms. In order to accomplish this, Dingle [32] used Darboux's Theorem to establish the expression

$$y_r(\tilde{\mathfrak{F}}) = \frac{1}{2\pi(\tilde{\mathfrak{F}})^r} \sum_{s=0}^{\infty} (r-s-1)! (-\tilde{\mathfrak{F}})^s Y_s(\tilde{\mathfrak{F}}), \quad r \gg 1 \quad (5.31)$$

for the late terms in (5.15a). Thus, in (5.15a), the later terms in Y_+ are linked to the early terms in Y_- via (5.31), with the idea being that this relation contains information on the exponentially small terms.

Next, following an observation of Stokes', the following procedure is employed to reduce the error in the asymptotic expansions (5.14) to an exponentially small level. First, the expansions are truncated at the least term $r = N_0$ in (5.15a). Next, in the terms for $r > N_0$ (these often increase, causing divergence), Y_r is replaced by the expression (5.31) for $r \gg 1$. This modified ‘divergent tail’ is re-summed into an exponentially small term called a ‘terminant’ using Borel summation [32, 33, 35]. This is the so-called ‘superasymptotic’ solution. This procedure has been iterated by Berry and Howls [33] to obtain even better or ‘hyperasymptotic’ estimates.

For our purposes, we will not require all the details of the re-summation of the tail using ‘terminants’ to obtain exponential accuracy. We shall employ the ‘superasymptotic’ approximation by truncating (5.14)/(5.15) at the least term $r = N_0$. This ensures that the error is exponentially small and that the connection formulae are bi-directional as needed for our purposes.

One may estimate the value $r = N_0$ at which the smallest term occurs by using the large r expression (5.31). The result is [32, 33]

$$N_0 = \text{Int}|\tilde{\mathfrak{F}}|, \quad (5.32)$$

where int denotes the integer part. Using (5.19), we have $N_0 = 16$, and we shall truncate (5.15a) (and (5.20)) at $r = 16$.

6. GLOBAL SOLUTION

To obtain a global solution valid everywhere within the shear layer, solutions $\hat{p}_{1,m}$ and $\hat{p}_{1,nv}$ must be asymptotically matched, and $\hat{p}_{1,i}$ and $\hat{p}_{1,nv}$ must be matched, respectively, to the appropriate outgoing spatially damping solutions [14] at $x = -1/2$ and $x = 1/2$. For the illustrative case $A = 2^{1/2}$, $B = 25$ we obtain the dimensionless frequency W and construct the pressure eigenfunction profile. It is

found that both the eigenfrequency W and the structure of the eigenfunction $\hat{p}_1(x)$ agree closely with previously computed numerical solutions at this point.

6.1. Boundary Conditions

The solutions to the eigenvalue equation (2.3), external to the shear layer, are given by (2.7). These correspond to spatially damping outgoing wave solutions in the comoving reference frame of the fluid on either side of the shear layer [14]. Suppressing the y , z , and t dependences, the boundary conditions on the solutions are

$$\hat{p}_1(x = -1/2) = c_1 e^{-ik_-(x+1/2)} \Big|_{x=-1/2} \quad (6.1)$$

$$\hat{p}_1(x = 1/2) = c_2 e^{-ik_+(x-1/2)} \Big|_{x=1/2}. \quad (6.2)$$

For our linear analysis, only the ratio (c_2/c_1) is important. Either of c_1 and c_2 may be set equal to unity, with the other being self-consistently determined. However, we retain both c_1 and c_2 to facilitate the comparison with previously obtained numerical results.

The wavenumbers in the external regions, k_+ and k_- , are given by (2.8) with $k_{i+} < 0$, and k_{r+} , k_{i-} , and $k_{r-} > 0$. For the short-wavelength region we are considering, with $B^2 \gg 1$, $W_i \rightarrow 0$, and $W_r \rightarrow A$, (2.8) implies that

$$\begin{aligned} k_+ &\approx -iB \left[1 - (A - W_r)^2 \right]^{1/2} \\ &\equiv ik_{i+} \end{aligned} \quad (6.3)$$

and

$$\begin{aligned} k_- &\approx B \left[(A + W_r)^2 - 1 \right]^{1/2} + \frac{iBW_r(A + W_r)}{\left[(A + W_r)^2 - 1 \right]^{1/2}} \\ &\equiv k_{r-} + ik_{i-}. \end{aligned} \quad (6.4)$$

Here, we consider $W_r > 0$. Analogous considerations apply for $W_r < 0$. We have also ignored the small real part of k , because of the strong damping of the solution from its large imaginary part, i.e., for $W_r > 0$, the wave

$$\hat{p}_1 = c_2 \exp - |k_{i+}| x$$

to the right of the shear layer is evanescent. On the other hand, to the left of the layer ($x < -1/2$), the pressure perturbation

$$\hat{p}_1 = c_1 e^{-ik_{r-}x} e^{|k_{i-}|x}$$

corresponds to a propagating, slowly damped wave.

6.2. Matching

Matching \hat{p}_1 and \hat{p}'_1 at $x = 1/2$ using (4.2) and (6.2) together with $k_+ \approx -iB$ for $W_r \rightarrow A$, we have

$$d_1 = \frac{c_2(\delta + \beta B)}{(\alpha\delta - \gamma\beta)} \quad (6.5a)$$

$$d_2 = \frac{-c_2(\alpha B + \gamma)}{(\alpha\delta - \gamma\beta)} \quad (6.5b)$$

where

$$\alpha = \left(\frac{2}{i\pi B} \right)^{1/2} \left(\frac{\tilde{S}_+}{B} - y_+ \tilde{C}_+ \right) \quad (6.5c)$$

$$\beta = i \left(\frac{2}{i\pi B} \right)^{1/2} \left(\frac{\tilde{C}_+}{B} - y_+ \tilde{S}_+ \right) \quad (6.5d)$$

$$\gamma = - \left(\frac{2B}{i\pi} \right)^{1/2} y_+ \tilde{S}_+ \quad (6.5e)$$

$$\delta = -i \left(\frac{2B}{i\pi} \right)^{1/2} y_+ \tilde{C}_+ \quad (6.5f)$$

and the + subscripts indicate quantities evaluated at $x = 1/2$.

Examining (5.20) (using (5.17)), it is easy to check (by taking the derivative of each of the two terms) that the H component has $k_{r-} < 0$ and thus corresponds to a rightward going wave for $-\frac{1}{2} \leq x < x_b$. This is clearly unphysical, since this would correspond to incoming waves at the left boundary $x = -1/2$. Since we require outgoing solutions at $x = -1/2$, we take $H = 0$.

Next, matching $p_{1,III}$ and $p_{1,IV}$ at $x = x_b$ we have

$$G = \left(\frac{2}{i\pi B} \right)^{1/2} \left\{ \cosh(By_b) \left[-\frac{2A\zeta_b d_1}{B} + id_2/B \right] \right. \\ \left. + \sinh(By_b) \left[d_1/B - \frac{id_2 2A\zeta_b}{B} \right] (\tilde{u}_b)^{-3/2} X_b^{1/4} \exp\left(-\frac{i\tilde{\delta}_b}{2}\right) (y_+)_b \right\}, \quad (6.6)$$

where the subscript b denotes quantities evaluated at x_b . Using (6.5), this equation gives G in terms of the amplitude c_2 of the solution at the right edge $x = 1/2$ of the shear layer. Hence, E and F may be obtained in terms of c_2 from (5.29) and (5.30). This completes the solution in all four regions, viz. $p_{1,I}$ in the oscillatory region $-0.5 \leq x < x_i$; $p_{1,II}$ near the turning point x_i ; $p_{1,III}$ in the exponential region $x_i < x \leq x_b$; and the boundary layer or resonance layer solution $p_{1,IV}$ for $x_b \leq x \leq 0.5$. Note that, as discussed in Section 5d, Y_\pm of (32a) is truncated at the supersasymptotic level $r = 16$ in (5.20), (5.21) and (6.6).

7. RESULTS AND DISCUSSION

We now use our solutions to determine the eigenvalues and eigenfunctions for the representative point $A = 2^{1/2}$, $B = 25$. Using (2.9) and (2.11), we have $x_* = 1/2 - 1/2 B$. Hence,

$$W_r = 1.36 \text{ and } W_i \rightarrow 0. \quad (7.1)$$

Using (6.5) and (6.6)

$$\begin{aligned} d_1 &= 307.4(2i/\pi)^{1/2} c_2 \\ d_2 &= 233.8(2/i\pi)^{1/2} c_2 \\ G &= 0.019632 c_2. \end{aligned} \quad (7.2)$$

We calibrate our linear solution amplitude against the numerical solutions of ref. 1, $\hat{p}_1(x = 1/2) = c_2 = -48 - i33$. Hence, for $A = 2^{1/2}$ and $B = 25$, we obtain

$$\hat{p}_1(x_*) = -57.2 - i39.3 \quad (7.3)$$

and

$$\hat{p}_1(x_t) = -0.879 - i0.604$$

where (4.2), (5.28) and the values $Ai(0) = Bi(0)/3^{1/2} = 0.355$ have been used. Also, from (2.8)

$$k_+ = -25i, \text{ and } k_- = 64.7 - i0.43.$$

All the above results for $\hat{p}_1(x_*)$, $\hat{p}_1(x_t)$, k_+ , k_- , and the eigenvalues W_r and W_i agree closely with the exact numerical values shown in Figure 9 of ref. 14 at the same point of the (B, A) plane. Notice that k_- is incorrectly given as 45 in Figure 9 of ref 14, although the figure is otherwise correct. The global solution for the \hat{p}_1 eigenfunction given by our combined phase integral and or boundary-layer solution is shown in Figure 2 for the illustrative point $A = 2^{1/2}$, $B = 25$. Clearly, the agreement with the numerically obtained eigenfunction in Figure 9 of ref. 14 is very good.

Physically, it is clear that the wave-fluid resonance acts to convert the energy of the mean, or equilibrium, fluid flow to wave energy that is radiated or carried out from both edges of the shear layer. In mathematical terms, in the presence of the wave-fluid resonance the resonant denominator necessitates the introduction of a boundary-layer solution. As usual, this introduces an additional degree of freedom mathematically, via a rapid spatial variation, thereby enabling the radiative boundary condition at the second boundary to be simultaneously satisfied. Since the boundary-layer solution is matched to the outgoing wave solution at $x = 1/2$, we may ensure outgoing wave solutions at $x = -1/2$ simultaneously.

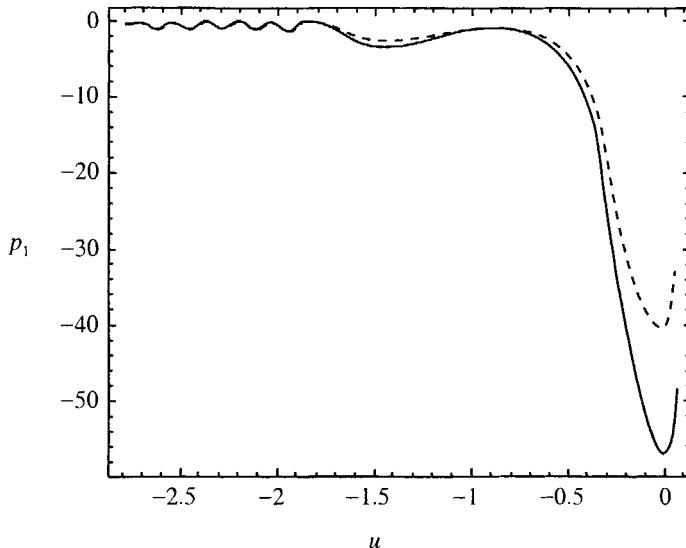


Figure 2. The global solution for the real (—) and imaginary (---) parts of the pressure eigenfunction \hat{p}_1 at $A = 2^{1/2}$, $B = 25$. The change in the nature of the eigenfunction from the oscillatory solution $\hat{p}_{1,I}$ to the exponential solution $\hat{p}_{1,M}$ occurs in crossing the turning point at $x = x_t$. The crossover is effected by the inner solution $\hat{p}_{1,N}$. The oscillatory solution $\hat{p}_{1,I}$ represents an outgoing, spatially damping solution at $x = -1/2$, in a frame of reference moving at the local fluid speed $-v_m$. The boundary-layer solution $\hat{p}_{1,IV}$ in the vicinity of the wave-fluid resonance at $x = x_*$, connects the exponential solution $\hat{p}_{1,M}$ to the outgoing, spatially damping wave solution in the comoving frame of reference of the fluid at $x = 1/2$.

The detailed aspects of the energy transfer will be considered in subsequent work by employing the supersymptotic solution developed here. Clearly, one will need to consider the energy density as well, and show that the unstable traveling waves indeed act to carry the energy of the mean flow out to $x \rightarrow \pm\infty$, i.e., the net flux of energy is outwards from the shear layer. This will involve the extension of earlier work for the incompressible case [36-44].

REFERENCES

- [1] Sen, A.K., Stability of the magnetosphere boundary, *Planetary and Space Science* 13, 1965, 131-141.
- [2] McKenzie, J.F., Hydromagnetic oscillations of the geomagnetic tail and plasma sheet, *J. Geophysical Res.* 75, 1970, 5331-5339.

- [3] Southwood, D.J., Some features of the field line resonances in the magnetosphere, Planetary and Space Science 22, 1974, 1024-1032.
- [4] Chen, L. and Hasegawa, A., A theory of long-period magnetic pulsations, J. Geophysical Res. 79, 1974, 1024-1032.
- [5] Scarf, F.L., Kurth, W.S., Gurnett, D.A., Bridge, H.S., and Sullivan, J.D., Jupiter tail phenomena upstream from Saturn, Nature 292, 1981, 585-586.
- [6] D'borowolny, H. and D'Angelo, N., Wave motion in type I comet tails, in Cosmic Plasma Physics K. Schindler, ed., Plenum, New York, 1972.
- [7] Ershkovich, A.I., Nusnov, A.A., and Chernikov, A.A., Oscillations of type I comet tails, Planetary and Space Science 20, 1972, 1235-1243; and, Nonlinear waves in type I comet tails 21, 1973, 663-673.
- [8] Turland, B.D. and Scheuer, P.A.G., Instabilities of Kelvin-Helmholtz type for relativistic streaming, Monthly Notices Roy. Astron. Soc. 176, 1976, 421-441.
- [9] Blandford, R.D. and Pringle, J.E., Kelvin-Helmholtz instability of relativistic beams, Monthly Notices Roy. Astron. Soc. 176, 1976, 443-454.
- [10] Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability*, Dover originally published 1961, Oxford, Clarendon, New York, 1981; Syrovatskii, A., The Helmholtz instability, Soviet Physics Uspekhi 62, 1957, 247-253; Northrop, T.G., Helmholtz instability of a plasma, Physical Review Second Series, 103, 1956, 1150-1154.
- [11] Gerwin, R.A., Stability of the interface between two fluids in relative motion, Rev. Modern Phys. 40, 1968, 652-658.
- [12] Ray, T.P. and Ershkovich, A.I., Kelvin-Helmholtz instabilities of magnetized shear layers, Monthly Notices Roy. Astron. Soc. 204, 1983, 821-826.
- [13] Miura, A., Anomalous transport by magnetohydrodynamic Kelvin-Helmholtz instabilities in the solar wind-magnetosphere interaction, J. Geophysical Res. 89, 1984, 801-818.
- [14] Choudhury, S. Roy, and Lovelace, R.V., On the Kelvin-Helmholtz instabilities of supersonic shear layers, Astrophysical J. 283, 1984, 331-342; and, On the Kelvin-Helmholtz instabilities of high-velocity magnetized shear layers 302, 1986, 188-199; Miura, A. and Pritchett, P.L., Nonlinear stability analysis of the MHD Kelvin-Helmholtz instability in a compressible plasma, J. Geophysical Res. 87, 1982, 7431-7444.
- [15] Choudhury, S. Roy, Kelvin-Helmholtz instabilities of supersonic, magnetized shear layers, J. Plasma Phys. 35, 1986, 375-392.
- [16] Uberoi, C., On the Kelvin-Helmholtz instabilities of structured plasma layers in the magnetosphere, Planetary and Space Science 34, 1985, 1223-1227.
- [17] Fujimoto, M. and Terasawa, T., Ion inertia effect on the Kelvin-Helmholtz instability, J. Geophysical Res. 96, 1991, 15725-15734.
- [18] Sharma, A.C. and Shrivastava, K.M., Magnetospheric plasma waves, Astrophys. Space Sci. 200, 1993, 107-115.
- [19] Malik, S.K. and Singh, M., Chaos in Kelvin-Helmholtz instability in magnetic fluids, Phys. Fluids A 4, 1992, 2915-2922.
- [20] Choudhury, S. Roy, and Patel, V.L., Kelvin-Helmholtz instabilities of high-velocity, magnetized anisotropic shear layers, Phys. Fluids 28, 1985, 3292-3301.
- [21] Duhamel, S., Gratton, F., and Gratton, J., Hydromagnetic oscillations of a tangential discontinuity in the CGL approximation, Phys. Fluids 13, 1970, 1503-1509.
- [22] Duhamel, S., Gratton, F., and Gratton, J., Radiation of hydromagnetic waves from a tangential velocity discontinuity, Phys. Fluids 14, 1971, 2067-2071.

- [23] Duhamel, S. and Gratton, J., Effect of compressibility on the stability of a vortex sheet in an ideal magnetofluid, *Phys. Fluids* 16, 1972, 150-152.
- [24] Rajaram, R., Kalra, G.L., and Tandon, J.N., Discontinuities and the magnetosphere phenomena, *J. Atm. Terr. Phys.* 40, 1978, 991-1000.
- [25] Rajaram, R., Kalra, G.L., and Tandon, J.N., Discontinuities in the magnetosphere, *Astrophys. Space Sci.* 67, 1980, 137-150.
- [26] Talwar, S.P., Hydromagnetic stability of the magnetospheric boundary, *J. Geophysical Res.* 69, 1964, 2707-2713.
- [27] Talwar, S.P., Kelvin-Helmholtz instability in an anisotropic plasma, *Phys. Fluids* 8, 1965, 1295-1299.
- [28] Pu, Zu-Yin, Kelvin-Helmholtz instability in collisionless space plasmas, *Phys. Fluids B* 1, 1989, 440-447.
- [29] Brown, K. and Choudhury, S. Roy, Kelvin-Helmholtz Instabilities of High-Velocity Magnetized Shear Layers with Generalized Polytrope Laws, *Quart. Appl. Math.*, in press.
- [30] Brown, K. and Choudhury, S. Roy, Quasiperiodicity and Chaos in the Nonlinear Evolution of the Kelvin-Helmholtz Instability of Supersonic Anisotropic Tangential Velocity Discontinuities, *J. Nonlin. Sci.*, submitted.
- [31] Nayfeh, A.H., *Perturbation Methods*, John Wiley & Sons, New York, 1973, 315-318 and 335-342; Pearson, C.E., *Handbook of Applied Mathematics*. Van Nostrand Reinhold, New York, 1983, 667.
- [32] Dingle, R.B., *Asymptotic Expansions: Their Derivation and Interpretation*, Academic, London, 1973.
- [33] Berry, M.V. and Howls, C.J., Hyperasymptotics, *Proc. Roy. Soc. London A* 430, 1990, 657-667.
- [34] Berry, M.V., *Asymptotics, Superasymptotics, Hyperasymptotics, in Asymptotics Beyond All Orders*, H. Segur, S. Tanveer and H. Levine Eds., Plenum, New York, 1991.
- [35] Boyd, J.P., *Weakly Nonlocal Solitary Waves and Beyond-All-Orders Asymptotics*, Kluwer, Dordrecht, 1998.
- [36] Acheson, D.J., On over-reflexion, *J. Fluid Mech.* 77, 1976, 433-472.
- [37] Craik, A.D.D., *Wave Interactions and Fluid Flows*, Cambridge University Press, London, 1985.
- [38] Bretherton, F.P., Wave action and energy, *Q.J.R. Meteorol. Soc.* 92, 1966, 466-471.
- [39] Booker, J.R. and Bretherton, The critical layer for internal gravity waves in a shear flow, *F.P., J. Fluid Mech.* 27, 1967, 513-539.
- [40] Eltayeb, I.A. and McKenzie, Critical-layer behavior and wave amplification of a gravity wave incident upon a shear layer, *J.F., J. Fluid Mech.* 72, 1975, 661-671.
- [41] Van Duin, C.A. and Kelder, H., Reflection properties of internal gravity waves, *J. Fluid Mech.* 120, 1982, 505-521.
- [42] Barston, E.M., Electrostatic oscillations in inhomogeneous cold plasmas, *Ann. Phys. N.Y.* 29, 1964, 282-303; Sedlacek, Z., Electrostatic normal modes, *J. Plasma Phys.* 6, 1971, 187; Sedlacek, Z., *Cesk. Cas. Fyz. B* 23, 1973, 892-901.
- [43] Landau, L.D. and Lifshitz, E.M., *Fluid Mechanics*, Pergamon Press, Oxford, 1959; Lighthill, M.J., *Waves in Fluids*, Cambridge University Press, London, 1978.
- [44] Whitham, G.B., *Linear and Nonlinear Waves*, John Wiley & Sons, New York, 1974.

5 SOLITARY WAVES WITH GALILEAN INVARIANCE IN DISPERSIVE SHALLOW-WATER FLOWS

C.I. Christov

Department of Mathematics
University of Louisiana at Lafayette
Lafayette, LA 70504-1010

ABSTRACT

The present work deals with a recently derived nonlinear dispersive system for shallow-water flows. Unlike the classical Boussinesq models, the new one possesses Galilean invariance. It is investigated numerically by means of a conservative difference scheme. In order to understand the intrinsic physical mechanisms behind the balance between nonlinearity and dispersion larger phase speeds of the solitary waves are considered which are formally beyond the applicability of the weakly nonlinear approximation.

The pseudo-particle behavior of the solitary waves is interrogated. It is shown that the system with Galilean invariance is, in a sense, more “elastic” than the classical Boussinesq model. Snap-shots of the interactions of the localized waves are presented graphically. The phase shifts experienced by the pseudo-particles are shown to be of the opposite sign to these for systems without Galilean invariance.

INTRODUCTION

After John Scott-Russell discovered the “great wave” there were different attempts to explain its existence and to find its appropriate model. Boussinesq [4, 2, 3] introduced the fundamental idea of balance between the nonlinearity and dispersion and derived the first approximate expression for the dispersion in the case of weakly

nonlinear long waves. We call this balance “Boussinesq Paradigm”. During the years different Boussinesq equations have been derived under the assumption of balance between weak nonlinearity and weak dispersion (the latter taking place for long waves). They are *generalized wave equations* which offer the opportunity to investigate the generic features of dispersive wave models, such as head-on collisions of localized structures (solitary waves/quasi-particles) even beyond the framework of the long-wave weakly-nonlinear assumptions. Boussinesq equations are not always fully integrable. As a rule they possess at least three conservation/balance laws – for mass, energy, and momentum.

Recently a more general form (preserving the Galilean invariance) of the dispersive shallow water equations has been derived [8, 9]. It has been shown to possess a solitary-wave solution of *sech* type which makes it very useful in paradigmatic sense for investigation of solitonic (pseudo-particle) behavior of localized solutions.

In the present work we investigate the properties of the new model as a dynamical system. To this end a special conservative difference scheme is constructed which generalizes to the case of Galilean invariant systems, the schemes previously developed by the author. A number of cases of soliton interactions are treated ranging from weakly nonlinear case to a strongly nonlinear case with nonlinear blow-up of the solution in finite time.

The model is expected to provide additional basis for soliton research especially as far as systems with Galilean invariance are concerned.

1. DISPERSIVE SHALLOW WATER SYSTEM (DSWS)

In the recent author's works the Boussinesq's derivation is revisited with the purpose of making the model conserve the total energy of the wave system. In order to make the present note self-contained we repeat part of the derivations from [8, 9].

Consider the inviscid flow in a thin layer with free surface represented by a single-valued shape function $h(x, y, t)$. The motion in the bulk is governed by the Laplace equation for the potential Φ .

Let H be the scale for the vertical spatial coordinate (the thickness of the shallow layer) and L is the characteristic wave length in longitudinal direction. We introduce dimensionless variables according to the scheme

$$\Phi = UL\phi, \quad h = H\eta, \quad z = Hz', \quad x = Lx', \quad y = Ly', \quad t = LU^{-1}t',$$

where $U = \sqrt{gH}$ is the characteristic scale for the velocity. Henceforth, the primes will be omitted without fear of confusion. In terms of dimensionless variables the Laplace equation takes the form

$$\beta \Delta \phi + \frac{\partial^2 \phi}{\partial z^2} = 0. \quad (1.1)$$

Here $\beta \equiv H^2 L^{-2}$ is called dispersion parameter. It is a small quantity for horizontal length scales L which are long compared to the depth of the layer H . The free surface in dimensionless form is given by $z = 1 + \eta$. The kinematic and dynamic conditions read

$$\frac{\partial \eta}{\partial t} + \nabla \phi \cdot \nabla \eta = \frac{1}{\beta} \frac{\partial \phi}{\partial z}, \quad (1.2)$$

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} (\nabla \phi)^2 + \frac{1}{2\beta} \left(\frac{\partial \phi}{\partial z} \right)^2 + \eta = 0. \quad (1.3)$$

Boussinesq expanded the solution of Laplace equation (1.1) into a power series with respect to β . Acknowledging the non-flux condition at the bottom of the layer he showed that the series contain only the even powers of the coordinate z , namely

$$\phi(x, y, z, t) = \sum_0^{\infty} (-\beta \Delta)^m f(x, y, t) \frac{z^{2m}}{(2m)!}, \quad (1.4)$$

where $f(x, y, t) \stackrel{\text{def}}{=} \phi(x, y, z = 0, t)$ is the unknown function representing the value of potential at the bottom of the layer.

Now the derivatives entering the surface conditions (1.2), (1.3) can be identified. Upon introducing the relevant expressions into the governing system for the surface motion and neglecting the terms proportional to β^m ($m \geq 2$) one arrives at the following approximate system containing only the surface variables η, f :

$$\frac{\partial \eta}{\partial t} + \left[\nabla f - \frac{\beta}{2} \nabla [(1 + \eta)^2 \Delta f] \right] \cdot \nabla \eta = -(1 + \eta) \Delta f + \frac{\beta}{6} (1 + \eta)^3 \Delta^2 f, \quad (1.5)$$

$$\begin{aligned} \frac{\partial f}{\partial t} - \frac{\beta}{2} \frac{\partial}{\partial t} [(1 + \eta)^2 \Delta f] + \frac{1}{2} (\nabla f)^2 + \eta - \frac{\beta}{2} \nabla f \cdot \nabla [(1 + \eta)^2 \Delta f] \\ + \frac{\beta}{2} [(1 + \eta) \Delta f]^2 = 0. \end{aligned} \quad (1.6)$$

For small values of the dispersion parameter β the main idea of Boussinesq was to look for weakly nonlinear waves of amplitudes of the order of the small parameter, namely $|\eta|, |f| \approx O(\beta)$. Then within the leading asymptotic order $O(\beta)$ the system reduces to a linear hyperbolic equation for the wave propagation. In the next order $O(\beta^2)$, two small effects – nonlinearity and dispersion – take place. The famous *sech* solution discovered by Boussinesq demonstrates that these two effects can be balanced pointwise making a wave propagate as a linear disturbance

according to the linear wave equation from the leading asymptotic order. Thus a wave retains its shape unchanged if left alone (without collisions with other waves). This is what we call “Boussinesq Paradigm”. Formally speaking one can seek for a solution $\eta = \beta \bar{\eta} + O(\beta^2)$, $f = \beta \bar{f} + O(\beta^2)$. Within the adopted asymptotic order $O(\beta^2)$ one gets

$$\frac{\partial \bar{\eta}}{\partial t} + \beta \nabla \cdot \bar{\eta} \nabla \bar{f} = -\Delta f + \frac{\beta}{6} \Delta^2 \bar{f} + O(\beta^2), \quad (1.7)$$

$$\frac{\partial \bar{f}}{\partial t} - \frac{\beta}{2} \frac{\partial \Delta \bar{f}}{\partial t} + \frac{\beta}{2} (\nabla \bar{f})^2 + \bar{\eta} = 0 + O(\beta^2), \quad (1.8)$$

and the overbars will be omitted without fear of confusion.

Although the above derived system is a straightforward asymptotic reduction of the system (1.1), (1.2), (1.3) it differs qualitatively from the latter because (1.7), (1.8) does not bring about the conservation of energy. In this sense, it is not a consistent asymptotic approximation of the original system. It means that in the asymptotic reduction a quality has been lost.

In order to find the correct energy-conserving form of the system, we introduce a new variable

$$\chi = \eta - \frac{\beta}{2} \frac{\partial \Delta f}{\partial t}$$

and upon substituting it in (1.7), (1.8) we get

$$\begin{aligned} \frac{\partial \chi}{\partial t} + \beta \nabla \cdot \chi \nabla f + \frac{\beta^2}{2} \nabla \cdot \left(\frac{\partial \Delta f}{\partial t} \nabla f \right) &= -\Delta f + \frac{\beta}{6} \Delta^2 f - \frac{\beta}{2} \frac{\partial^2 \Delta f}{\partial t^2}, \\ \frac{\partial f}{\partial t} &= -\frac{\beta}{2} (\nabla f)^2 - \chi. \end{aligned}$$

The term $-\frac{\beta^2}{2} \nabla \cdot \left(\frac{\partial \Delta f}{\partial t} \nabla f \right)$, can be neglected within the asymptotic order $O(\beta^2)$ and the system adopts the form

$$\frac{\partial \chi}{\partial t} = -\beta \nabla \cdot \chi \nabla f - \Delta f + \frac{\beta}{6} \Delta^2 f - \frac{\beta}{2} \frac{\partial^2 \Delta f}{\partial t^2}, \quad (1.9)$$

$$\frac{\partial f}{\partial t} = -\frac{\beta}{2} (\nabla f)^2 - \chi. \quad (1.10)$$

The following energy balance law holds

$$\frac{dE}{dt} = \oint_{\partial D} \left[(1 + \beta \chi) f_t \frac{\partial f}{\partial n} + \frac{\beta}{2} f_t \frac{\partial f_u}{\partial n} + \frac{\beta}{2} f_t \frac{\partial \Delta f}{\partial n} - \frac{\beta}{2} \Delta f \frac{\partial f_t}{\partial n} \right] ds, \quad (1.11)$$

$$E = \frac{1}{2} \int_D \left[\chi^2 + (1 + \beta \chi)(\nabla f)^2 + \frac{\beta}{6} (\Delta f)^2 + \frac{\beta}{2} (\nabla f_t)^2 \right] dx,$$

which allows us to call the system (1.9), (1.10) “Energy Consistent Boussinesq Paradigm”. We believe that this is the system that fulfills the Boussinesq program without unnecessary deficiencies stemming from the oversimplifications in the moving frame.

The most suitable set of boundary conditions are those that bring about the conservation of the total energy stem from the requirement that the right-hand side of (1.11) be equal to zero. There are three sets of conditions compatible with that requirement. We select the Dirichlet set of b.c.

$$f_t = 0 \rightarrow f = f_b(x, y) \text{ and } \frac{\partial f}{\partial n} = 0 \quad \text{for } (x, y) \in \partial D \quad (1.12)$$

which also secures the balance law for the wave momentum (see below).

Here is to be mentioned that both η and χ are implicit functions of the respective systems (functions for which no boundary conditions are posed) and there are no mathematical reasons to prefer one formulation over the other. Hence, we will not use the original variables.

Another balance (or conservation) law holds for the wave momentum (called also *pseudomomentum* [12, 13]) which is defined as

$$\mathbf{P} = - \int_D \eta \nabla f \, dy \equiv - \int_D \left[\chi \nabla f + \frac{\beta}{2} \frac{\partial \Delta f}{\partial t} \nabla f \right] dx \, dy. \quad (1.13)$$

In [8, 9] the following balance law for the total pseudomomentum is derived

$$\begin{aligned} -\frac{d\mathbf{P}}{dt} &= \int_D \left\{ -\left[\nabla \cdot (\nabla f \nabla f) - \frac{1}{2} \nabla(\nabla f)^2 \right] - \frac{1}{2} \nabla \chi^2 - \beta \left[\nabla f \nabla \cdot (\chi \nabla f) + \frac{\chi}{2} \nabla(\nabla f)^2 \right] \right. \\ &\quad \left. + \frac{\beta}{6} \left[\nabla \cdot (\nabla \Delta f \nabla f) - \frac{1}{2} \nabla(\nabla \Delta f \cdot \nabla f) \right] + \frac{\beta}{2} \Delta f_t \nabla f_t \right\} dx \, dy \\ &= \oint_{\partial D} \left[-\frac{\beta}{2} \left(\frac{\partial f_t}{\partial n} \nabla f_t - \frac{1}{2} (\nabla f_t)^2 \right) \mathbf{n} - \frac{1}{2} \chi^2 \mathbf{n} - \left(\frac{\partial f}{\partial n} \nabla f - \frac{1}{2} (\nabla f)^2 \mathbf{n} \right) \right. \\ &\quad \left. + \frac{\beta}{6} \left(\frac{\partial \Delta f}{\partial n} \nabla f - \frac{1}{2} \nabla \Delta f \cdot \nabla f \right) \mathbf{n} + \beta \chi \nabla f \frac{\partial f}{\partial n} \right] ds. \end{aligned}$$

Most of the terms in the balance law for the pseudomomentum cancel for boundary conditions (1.12) and hence,

$$\frac{d\mathbf{P}}{dt} = \oint_{\partial D} \left\{ -\frac{\beta}{4} (\nabla f_t)^2 - \frac{1}{4} (\nabla f)^2 + \frac{\beta}{12} (\nabla \Delta f \cdot \nabla f) + \frac{1}{2} \chi^2 \right\} \mathbf{n} \, ds. \quad (1.14)$$

2. SINGLE-EQUATION FORMULATION

Upon introducing (1.10) into (1.9) the function χ is readily excluded to obtain a single equation for the potential f , namely

$$f_{tt} + 2\beta \nabla f \cdot \nabla f_t + \beta f_t \Delta f + \frac{3\beta^2}{2} (\nabla f)^2 \Delta f - \Delta f + \frac{\beta}{6} \Delta^2 f - \frac{\beta}{2} \frac{\partial^2 \Delta f}{\partial t^2} = 0, \quad (2.1)$$

with Hamiltonian density

$$\mathcal{H} = \frac{1}{2} \left[f_t^2 + (\nabla f)^2 - \frac{\beta^2}{4} (\nabla f)^4 + \frac{\beta}{6} (\Delta f)^2 + \frac{\beta}{2} (\nabla f_t)^2 \right]. \quad (2.2)$$

Note that the nonlinearity of the dynamic condition (1.10) is responsible for the cubic nonlinearity of equation (2.1). The latter is of higher order in β , but it cannot be neglected without destroying the Galilean invariance.

The system described by equation (2.1) can be re-interpreted in a field-theoretic framework (very much in the same vein as in our previous work concerning the 6th order Boussinesq equation [10]) by introducing a Lagrangian density

$$\begin{aligned} \mathcal{L} &= \int_D \mathcal{L} dx dy, \quad P^\omega = - \int_D \nabla f \frac{\delta \mathcal{L}}{\delta f_t} dx dy, \\ \mathcal{L} &= \mathcal{K} - \mathcal{W}, \quad \mathcal{H} = \mathcal{K} + \mathcal{W}, \quad \mathcal{K} = \frac{1}{2} f_t^2 + \frac{\beta}{4} f_t (\nabla f)^2, \\ \mathcal{W} &= \frac{1}{2} \left[(\nabla f)^2 - \frac{\beta}{2} f_t (\nabla f)^2 - \frac{\beta^2}{4} (\nabla f)^4 - \frac{\beta}{3} (\Delta f)^2 + \frac{\beta}{2} (\nabla f_t)^2 \right]. \end{aligned} \quad (2.3)$$

In (2.3) one easily recognizes the pseudomomentum defined in (1.14), with an additional contribution probably due to the transport of the finite domain D when the Eulerian-Lagrangian passage is duly taken into account (see equation (4.42) from [13]). The appropriate boundary conditions yield the balance of *wave momentum*.

3. ONE-DIMENSIONAL VERSION

For two-dimensional flows the velocity potential and the surface elevation do not depend on the coordinate y . Naturally it is 1D for the surface variables. Then the system (1.9), (1.10) reduces to the following

$$\frac{\partial \chi}{\partial t} + \beta \frac{\partial}{\partial x} \left(\chi \frac{\partial f}{\partial x} \right) = - \frac{\partial^2 f}{\partial x^2} + \frac{\beta}{6} \frac{\partial^2 f}{\partial x^4} - \frac{\beta}{2} \frac{\partial^4 f}{\partial t^2 \partial x^2}, \quad (3.1)$$

$$\frac{\partial f}{\partial t} + \frac{\beta}{2} \left(\frac{\partial f}{\partial x} \right)^2 = -\chi. \quad (3.2)$$

Being reminded that f has the meaning of velocity potential taken at the bottom of the layer one can introduce new variables $u \stackrel{\text{def}}{=} f_x$ and $q_x \stackrel{\text{def}}{=} -\chi$. When the region under consideration is a finite interval, say $x \in [-L_1, L_2]$, then the boundary conditions (1.12) read

$$u = 0, \quad q_x = 0, \quad x = -L_1, L_2. \quad (3.3)$$

Upon integrating equation (3.1) once and acknowledging the boundary conditions one obtains

$$\frac{\partial q}{\partial t} + \beta u \frac{\partial q}{\partial x} = u - \frac{\beta}{6} \frac{\partial^2 u}{\partial x^2} + \frac{\beta}{2} \frac{\partial^2 u}{\partial t^2}, \quad (3.4)$$

$$\frac{\partial u}{\partial t} + \beta u \frac{\partial u}{\partial x} = \frac{\partial^2 q}{\partial x^2}. \quad (3.5)$$

4. BOUSSINESQ PARADIGM EQUATION

Boussinesq attempted to describe the nearly quasi-stationary wave phenomena in the moving frame [2, 3, 4]. He argued that for motions that evolve slowly in the moving coordinate frame the time derivatives are reasonably well approximated and can be replaced by the spatial ones.

Then two major ways of simplifications of the original system are possible. The first one is to simplify the convective nonlinear terms neglecting the nonlinearity in (1.10) or the cubic term in (2.1).

If f_t is replaced by f_x in the quadratic nonlinear term one arrives at

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2}{\partial x^2} \left[u + \frac{3\beta}{2} u^2 + \frac{\beta}{2} \frac{\partial^2 u}{\partial t^2} - \frac{\beta}{6} \frac{\partial^2 u}{\partial x^2} \right], \quad (4.1)$$

which was called in [7] “Boussinesq Paradigm Equation” (BPE). Note that it is not the equation derived by Boussinesq himself. The above simplification destroys the Galilean invariance of the system. It should be mentioned that the equation (4.1) appears also in the theory of longitudinal (acoustic) vibrations of rods (see, e.g., [15]) and in continuum limit for lattices (see, e.g., [16, 17]) where the lack of Galilean invariance can be an actual physical property.

The second simplification consists in changing the temporal derivatives to spatial ones in the linear dispersion terms. It led Boussinesq to an equation which was incorrect in the sense of Hadamard. Later on Boussinesq equation was regularized in [1] where an equation similar to (4.1) was derived (called “Regularized Long Wave Equation”, or RLWE). More detailed discussion on this matter can be found in ([8, 9]). Note that RLWE is linearly stable [14, 1, 20]). It is

not fully integrable (just as the original DSWS system is not), which is compliant with the physical nature of the problem.

BPE equation (4.1) can be rewritten as a system (see, [7])

$$u_t = q_{xx}, \quad q_t = u + \frac{3\beta}{2} u^2 + \frac{\beta}{2} u_{tt} - \frac{\beta}{6} u_{xx}. \quad (4.2)$$

The Hamiltonian structure of the above system is found in [7] and shown that for the b.c. from (3.3) one has

$$\frac{dM}{dt} = 0, \quad M = \int_{-L_1}^{L_2} u \, dx \quad (4.3)$$

$$\frac{dE}{dt} = 0, \quad E = \frac{1}{2} \int_{-L_1}^{L_2} \left[q_x^2 + u^2 + \frac{\beta}{2} u^3 + \frac{\beta}{2} u_t^2 + \frac{\beta}{6} u_x^2 \right] dx \quad (4.4)$$

$$\frac{dP}{dt} = \left[\frac{u^2}{2} + \beta u^3 - \frac{\beta}{4} u_t^2 - \frac{\beta}{12} u_x^2 \right]_{-L_1}^{L_2} = -\frac{\beta_2}{2} u_x^2 \Big|_{-L_1}^{L_2}, \quad (4.5)$$

$$P = \int_{-L_1}^{L_2} u \left(q_x + \frac{\beta}{2} u_{xt} \right) dx = \int_{-L_1}^{L_2} \left(u q_x - \frac{\beta}{2} u_t u_x \right) dx.$$

BPE is preferable over RLWE because for the former the *wave mass* is conserved alongside with the *energy*. In other words, the presence of the spatial fourth derivative requires boundary conditions whose satisfaction in turn brings about the conservation of the *wave mass* which is also a property of the original hydrodynamic problem.

It is to be mentioned that though the system (4.2) looks rather similar to the original DSWS (1.9), (1.10), there is a significant difference due to the fact that the latter is Galilean invariant, while the former is not. Respectively, the Lagrangian and Hamiltonian densities for the two systems are different. BPE is our choice for a dynamical system without Galilean invariance for which the balance between nonlinearity and dispersion holds.

Note that the derivations of the present section are not restricted to 1D surface motions. One-dimensional Boussinesq equations are considered only for the sake of comparison with the classical works.

5. THE SOLITARY WAVE OF DSWS

The *sech* solitons of BPE are given by (see [7]):

$$u = \frac{a}{\cosh^2[b(x-ct)]}, \quad a = \frac{c^2-1}{\beta}, \quad b = \sqrt{\frac{(c^2-1)}{2\beta\left(c^2-\frac{1}{3}\right)}}, \quad (5.1)$$

where c is the phase speed or *celerity* of wave. The *sech*-es exist either for supersonic celerities $c > 1$ or for $c < \sqrt{1/3}$. Only the supersonic *sech*-es are of physical relevance to shallow-water flows because for small β the subsonic ones are not long-length waves.

Although more complex than any of the Boussinesq equations, the DSWS system (3.4), (3.5) shares with them the existence of localized solution which is stationary in the moving frame $x - ct$. In [8, 9] the following *sech*-like solution is found

$$u = \frac{a \operatorname{sign}(c)}{\frac{|c|-1}{2} + \cosh^2[b(x-ct)]}, \quad a = \frac{c^2 - 1}{\beta}, \quad b = \sqrt{\frac{(c^2 - 1)}{2\beta\left(c^2 - \frac{1}{3}\right)}}, \quad (5.2)$$

which exists in the same range as the BPE solitons (see, [7]). The fact that DSWS admits a *sech*-like solution renders unnecessary the Boussinesq simplifications in the moving frame. The *sech*-like solution (5.2) is another candidate for the John Scott Russell's "Great (Permanent) Wave".

Comparing (5.1) and (5.2) reveals that the only difference is the term $\frac{1}{2}(|c|-1)$

in the denominator. This means that in the limit of weakly-nonlinear case $|c-1| \sim O(\beta) \ll 1$ they will be quantitatively very close. For arbitrary c the difference is small in the "tails" of the waves. It is significant only near the origin of coordinate system where is the smallest value of the denominator ($\cosh \approx 1$) and then only for significantly supercritical c .

In order to keep within the long-wave approximation, we have taken the supercritical celerities $c^2 = 1 + \beta$ and calculated the shapes of the solitary waves of DSWS and BPE. Figure 1 shows the comparison between the two solutions. The DSWS wave is always of smaller amplitude than the BPE one. For really small β the differences are quantitatively very small and it is hard to distinguish which one corresponds better to the experiment. A good case for comparison with the experiments of John Scott Russell could be $\beta = 0.2$.

The subcritical case is formally overboard the shallow-water theories since for $|c| < \sqrt{1/3}$ the solitary waves are not long waves. Yet, knowing the shapes of the subcritical solitary waves is of crucial importance for understanding the results of the numerical investigation of the evolution of colliding waves which is presented in the sections to follow. The most conspicuous feature of the subcritical waves is that they are depressions.

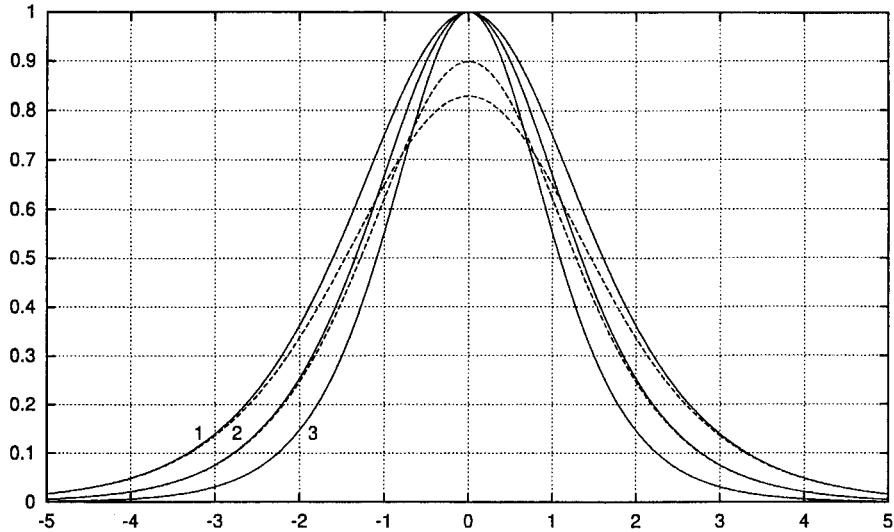


Figure 1. Comparison between BPE and DSWS solitary waves for supercritical phase speeds $c = \sqrt{1 + \beta}$: 1) $\beta = 1$; 2) $\beta = 0.5$; 3) $\beta = 0.1$.

Figure 2 shows that for $c = 0$, the depressions are less steep, but of largest amplitude. With the increase of c , they become narrower and their amplitudes

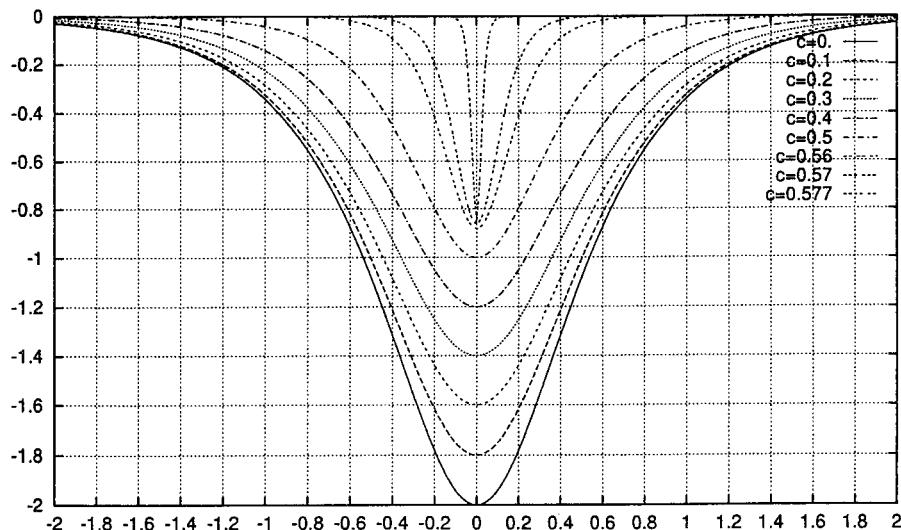


Figure 2. DSWS seches for subcritical phase speeds and $\beta = 1$.

somewhat decrease. For the limiting case $c \rightarrow \sqrt{1/3}$, the support of the localized solution vanishes and it becomes a solution of infinitely short length. The support for $\beta = 0.1$ is about three times shorter than for $\beta = 1$, which means that even for $c = 0$, it is significantly lesser than unity. This means that one cannot speak about long-wave solutions in this case.

In the present work, we consider DSWS in somewhat more paradigmatic way as a toy-object allowing the investigation of the interaction of pseudo-particles in a system with Galilean invariance. For this reason, we select $\beta = 0.6 \sim O(1)$ and consider the whole range of waves, e.g., strongly nonlinear short waves. The shapes of the solitary waves of DSWS and BPE are shown in Figure 3 for variety of supercritical phase speeds.

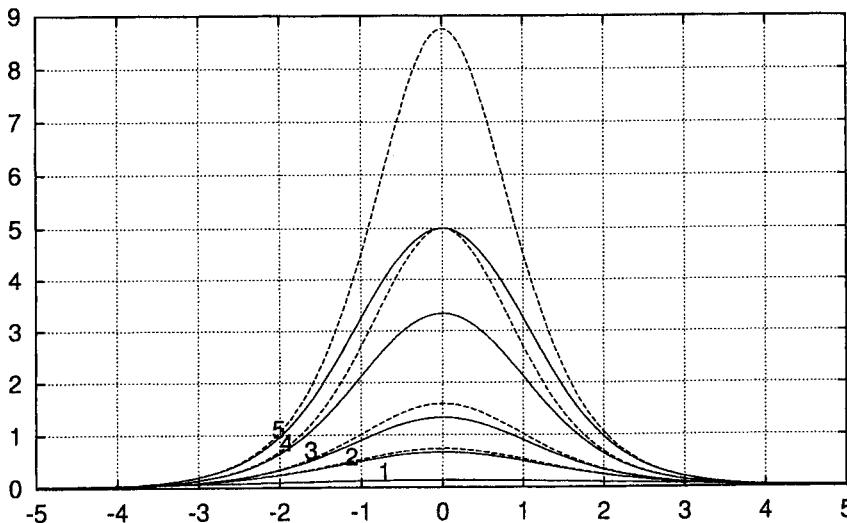


Figure 3. Solitary waves of DSWS ——— and BPE - - - for $\beta = 0.6$ and different phase speeds: 1) $c = 1.04$; 2) $c = 1.2$; 3) $c = 1.4$; 4) $c = 2.$; 5) $c = 2.5$.

Important characteristic of the solitary waves shown in Figure 3 is that the supercritical BPE *sech*-es are taller than the DSWS *sech*-es for the same magnitude of the phase speed.

6. CONSERVATIVE DIFFERENCE SCHEME

In previous author's papers [6, 11, 7], the way to construct conservative schemes for the Boussinesq Paradigm was outlined and their efficiency was demonstrated.

Following [7], we construct a conservative scheme for the Galilean invariant case treated here. We introduce a regular mesh in the interval $[-L_1, L_2]$, $x_i = -L_1 + (i-1) H$, $h = (L_1 + L_2)/(N-1)$, where N is the total number of grid points. We use a simplest linearization combined with an internal iteration (referred to by the composite superscript k). It appears to be robust enough and economical.

$$\frac{u_i^{n+1,k} - u_i^n}{\tau} = \frac{q_{i+1}^{\frac{n+1}{2},k} - 2q_i^{\frac{n+1}{2},k} + q_{i-1}^{\frac{n+1}{2},k}}{h^2} - \frac{\beta}{8h} \left[(u_{i+1}^{n+1,k-1})^2 - (u_{i-1}^{n+1,k-1})^2 + (u_{i+1}^n)^2 - (u_{i-1}^n)^2 \right] \quad (6.1)$$

$$\begin{aligned} \frac{q_i^{\frac{n+1}{2},k} - q_i^{\frac{n-1}{2}}}{\tau} &= -\frac{\beta}{8h} \left[q_{i+1}^{\frac{n+1}{2},k-1} - q_{i-1}^{\frac{n+1}{2},k-1} + q_{i+1}^{\frac{n-1}{2}} - q_{i-1}^{\frac{n-1}{2}} \right] (u_i^{n+1,k} + u_i^{n-1}) \\ &\quad - \frac{\beta}{12} \left[\frac{u_{i+1}^{n+1,k} - 2u_i^{n+1,k} + u_{i-1}^{n+1,k}}{h^2} + \frac{u_{i+1}^{n-1} - 2u_i^{n-1} + u_{i-1}^{n-1}}{h^2} \right] \\ &\quad + \frac{\beta}{2} \frac{u_i^{n+1,k} - 2u_i^n + u_i^{n-1}}{\tau^2} + \frac{u_i^{n+1,k} + u_i^{n-1}}{2}, \end{aligned} \quad (6.2)$$

with b.c.

$$u_N^{n+1,k} = u_1^{n+1,k} = 0, \quad q_N^{\frac{n+1}{2},k} - q_{N-1}^{\frac{n+1}{2},k} = q_2^{\frac{n+1}{2},k} - q_1^{\frac{n+1}{2},k} = 0. \quad (6.3)$$

The inner iterations start from the functions obtained at the previous time stage $u_i^{n+1,0} = u_i^n$ and $q_i^{\frac{n+1}{2},0} = q_i^n$, and are terminated at certain $k = K$ when

$$\max |u_i^{n+1,K} - u_i^{n,K-1}| \leq 10^{-13} \max |u_i^{n+1,K}|.$$

The value 10^{-13} is selected to be large enough in comparison with the round-off error 10^{-14} . In general, the number of iterations K (in our calculations we keep them around six to eight by means of adjusting the time increment to the "swiftness" of the motion) depends on the size of time increment. After the inner iterations converge one obtains, in fact, the solution for the new time stage $n+1$ of the nonlinear conservative difference scheme, namely $u_i^{n+1} \stackrel{\text{def}}{=} u_i^{n+1,K}$, $q_i^{\frac{n+1}{2}} \stackrel{\text{def}}{=} q_i^{\frac{n+1}{2},K}$.

From now on we shall not refer any more to the internal iterations (hence, omitting the composite index k), but rather consider the general properties of the scheme (6.1), (6.2) where the iterations are considered as accomplished.

Generalizing the derivation from [6], we prove that the above approximation secures the conservation of *energy* on difference level for arbitrary potential $U(u)$, namely the difference approximations of the *mass* and *energy*

$$\begin{aligned} E^{\frac{n+1}{2}} &= \frac{h}{2} \sum_{i=1}^{N-1} \frac{(u_i^{n+1})^2 + (u_i^n)^2}{2} - U(u_i^{n+1}) - U(u_i^n) + \frac{\beta}{2} \left(\frac{u_i^{n+1} - u_i^n}{\tau} \right)^2 \\ &\quad + \frac{1}{2h} \sum_{i=1}^{N-1} \frac{\beta}{12} \left[(u_{i+1}^{n+1} - u_i^{n+1})^2 + (u_{i+1}^n - u_i^n)^2 \right] + \left(q_{i+1}^{\frac{n+1}{2}} - q_i^{\frac{n+1}{2}} \right)^2, \\ M^{n+1} &= \sum_{i=2}^{N-1} u_i^{n+1} h, \end{aligned}$$

are conserved by the difference scheme (6.2), (6.1) in the sense that $M^{n+1} = M^n$ and $E^{\frac{n+1}{2}} = E^{\frac{n-1}{2}}$. As far as the satisfaction of the conservation and balance laws does not depend on the truncation error, we call it “strict in numerical sense”.

The scheme (6.1), (6.2) consists of two conjugated tridiagonal systems. We render them to a single five-diagonal system and apply the specialized solver for Gaussian elimination with pivoting [5].

7. NUMERICAL EXPERIMENTS ON SOLITON DYNAMICS

In this section we present the calculations obtained for different wave systems as well as the comparison to the BPE results. We chose $\beta = 0.6 \sim O(1)$ which is consistent with our aim to investigate the system beyond the range of its formal relevance to shallow-water flows.

In the figures to follow the material is organized to show snap-shots of the wave system in the upper panel and the trajectories of the centers of the solitary waves – in the lower panel. The time interval is the same for the two panels, but the panel with trajectories is appropriately zoomed for convenience. The scale for vertical axis pertains to the amplitude of the waves in the initial moment of time. In the lower panel, the solid lines represent BPE solution, the dashed lines – DSWS, and the dotted lines are the trajectories, which the solitons could have followed were they not to interact with each other. Since numerically a center of a “hump” is defined as the local maximum of the absolute value of the wave profile, it is impossible to locate it precisely during the collision of the two localized waves. For this reason the trajectories of the DSWS “humps” look somewhat erratic during the collision. Yet after the two main “humps” resume their identity, the results for the trajectories are smoother.

First we begin with the weakly-nonlinear case $c = 1.04$ and $c = 1.02$. It is depicted in Figure 4. One sees that the interaction is virtually elastic with no residual signals in the place of the collision, and no significant phase shift. As already found in [9], the phase shift for the Galilean invariant system is approximately twice as small and positive while the BPE predicts negative phase shift.

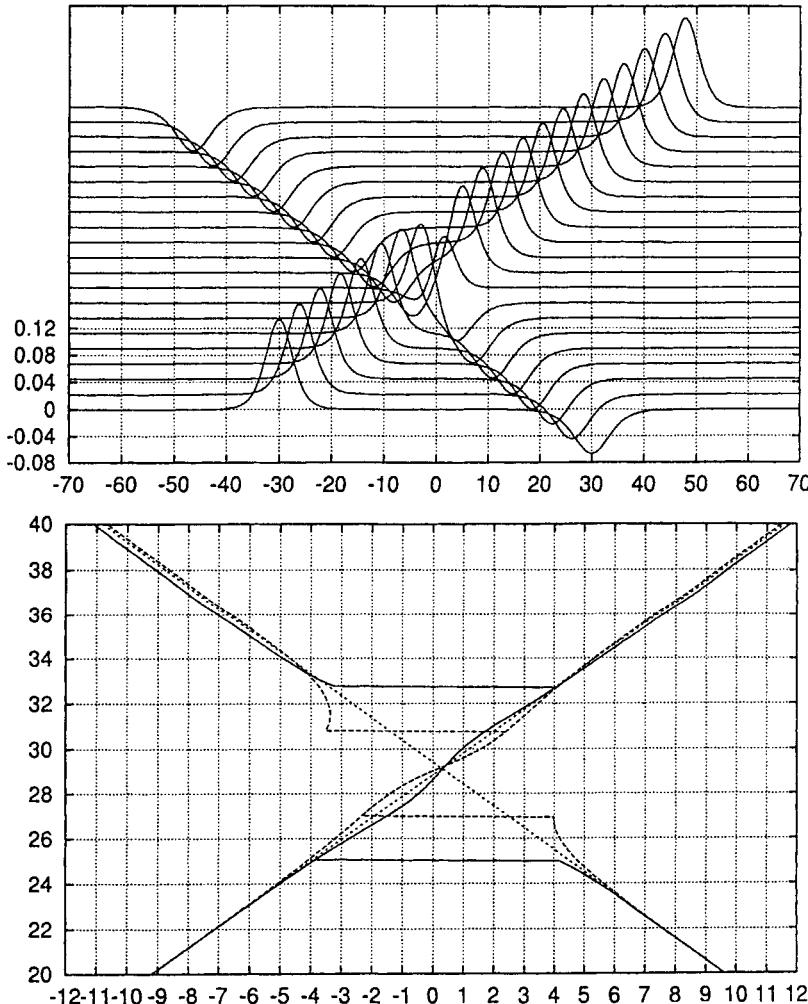


Figure 4. The weakly nonlinear case for $\beta = 0.6$, $c_l = 1.04$, and $c_r = -1.02$. Upper panel: Snapshots of the wave system for different times in the interval $0 \leq t \leq 75$. Lower panel: Trajectories of the centers of solitary waves. BPE: —, DSWS: - - -, trajectories of free solitons: · · · · ·.

Next we move to moderately nonlinear case presented in Figure 5. The general tendency in the phase shift is the same, but now it is almost twice as large as in the weakly nonlinear case. In addition, after the collision, some wriggles appear propagating with the characteristic speed (they are linear waves of small amplitudes) and trail the two main humps.

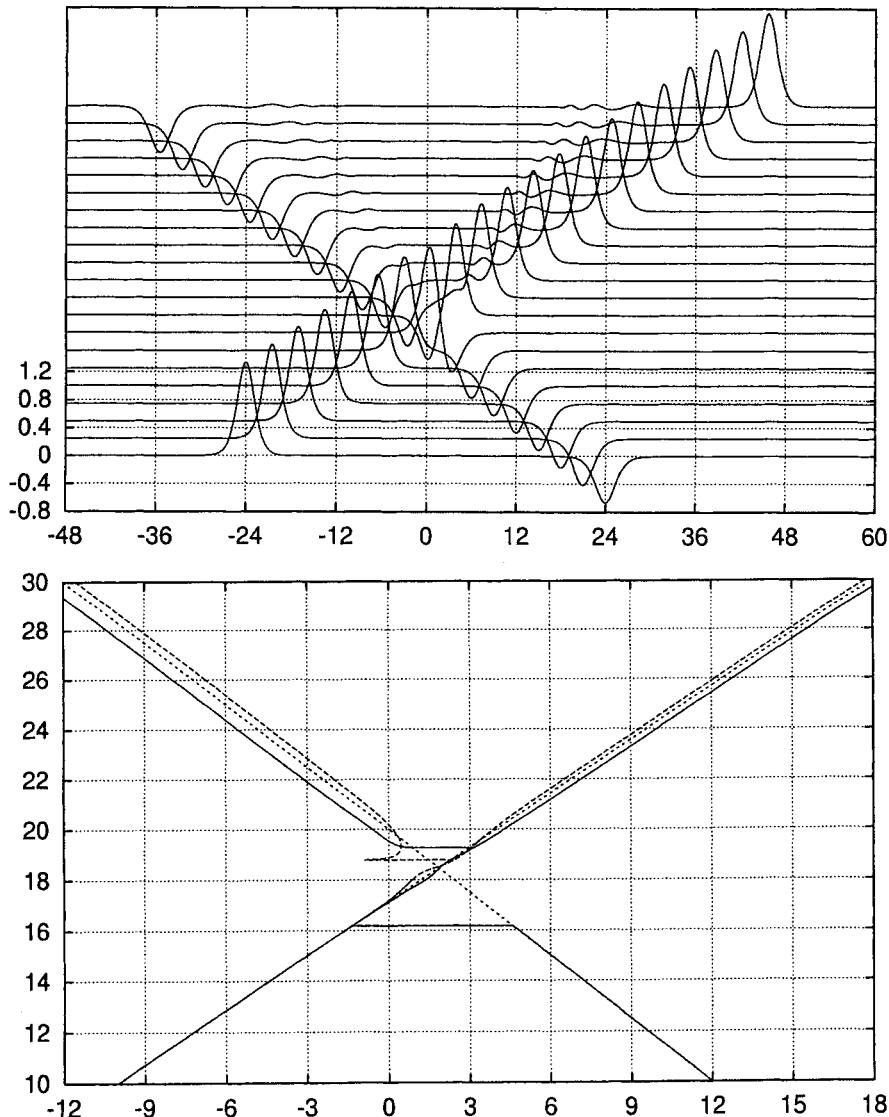


Figure 5. A moderately nonlinear case for $\beta = 0.6$, $c_l = 1.4$, and $c_r = -1.2$. Upper panel: Snapshots of the wave system for different times in the interval $0 \leq t \leq 50$.

Lower panel: Trajectories of the centers of solitary waves. BPE: —, DSWS: - - -, trajectories of free solitons: · · · · ·.

A further increase of the phase speeds (as shown in Figure 6) doubles the magnitude of the phase shift, but now the phase shifts for DSWS and BPE have almost the same absolute values remaining of opposite signs.

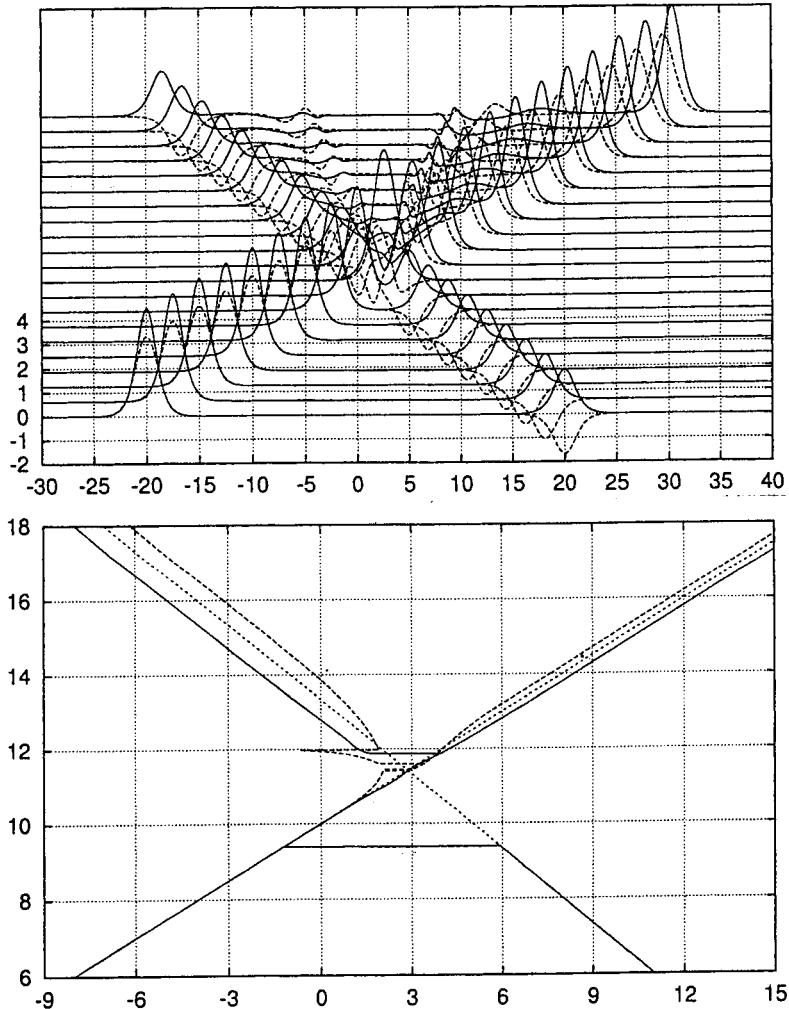


Figure 6. A nonlinear case for $\beta = 0.6$, $c_l = 2$, and $c_r = -1.5$. Upper panel: Snapshots of the wave system for different times in the interval $0 \leq t \leq 25$. Lower panel: Trajectories of the centers of solitary waves. BPE: —, DSWS: - - -, trajectories of free solitons: · · · · ·.

In order to give some more tangible information on the behavior of the two systems, we juxtapose directly the snapshots of the wave systems for the two systems under consideration. One sees that the profiles are similar, save the fact that the left-going wave in DSWS is negative (which is the physically correct case) while the same wave for BPE is positive (one of the deficiencies of the simplifications in the moving frame). The positions of the wriggles excited by the collision are roughly the same for the two systems, but their amplitudes are larger in DSWS rather than in BPE. In turn, the DSWS-wriggles are smoother which implies stronger elasticity of the system. In Figure 6 the difference in the signs of the phase shifts is also well seen.

Finally, we treat a case with very strong nonlinearity (shown in Figure 7). For

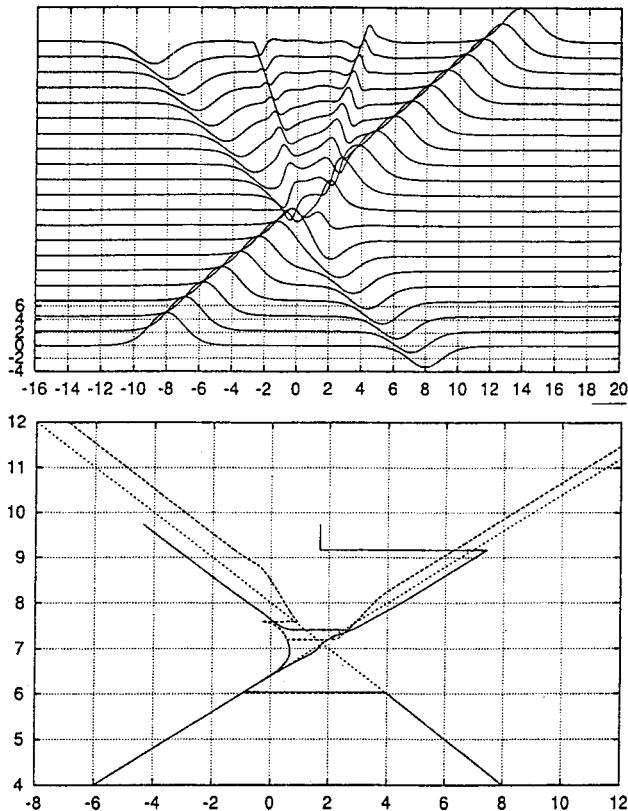


Figure 7. Strongly nonlinear case for $\beta = 0.6$, $c_l = 2.5$, and $c_r = -2.0$. Upper panel: Snapshots of the wave system for different times in the interval $0 \leq t \leq 18$. Lower panel: Trajectories of the centers of solitary waves. BPE: —, DSWS: - - -, trajectories of free solitons: · · · · .

this set of parameters, the BPE solution ends up in finite-time nonlinear blow-up (see, e.g., [18, 19]) while the DSWS solution survives. This is another sign of the stronger elasticity of the system with Galilean invariance (DSWS).

Figure 7 shows that (curiously enough) in BPE the solitary wave of smaller amplitude (the left-going one in this case) blows-up earlier than the wave of larger amplitude and larger phase speed.

8. CONCLUSIONS

A recently derived (see [8, 9]) dispersive shallow-water system is considered which originally appears in the long-wave models of the flow of inviscid liquid with free surface. The new system preserves the Galilean invariance of the original problem which is its main advantage over the Boussinesq equations. Its Hamiltonian structure is derived and new analytical solution of type of solitary wave is obtained. A conservative difference scheme is constructed and an algorithm for its implementation is developed.

The interactions of solitary localized waves are investigated for the case of significantly supercritical phase speeds of the solitary waves. This case is formally outside the region of applicability of long-wave weakly-nonlinear Boussinesq derivations. Hence, the new model is considered in a paradigmatic fashion and the intrinsic mechanisms of interactions of the pseudo-particles (solitary waves) are interrogated for the first time in the literature for a system with Galilean invariance. It is shown that for moderate nonlinearities, the behaviour is qualitatively similar to the weakly nonlinear case. The interactions are fairly elastic and the signals excited after the collision of two solitary waves are small. The phase shift is smaller than those for a system without Galilean invariance (so-called Boussinesq Paradigm Equation) and of opposite sign.

The strongly nonlinear cases reveal the same relationship between these signs of the phase shifts, but the system with Galilean invariance DSWS tends to be more robust in the sense that its solution exists for phase speeds for which a nonlinear blow-up takes place for BPE. Together with the smaller phase shift for moderate phase speeds, the last property underscores somewhat stronger elasticity of the system with Galilean invariance.

Acknowledgments. This work is supported by Grant LEQSF (1999-2002)-RD-A-49 from the Louisiana Board of Regents and partly by Grant NZ-611/96 of the National Science Foundation of Bulgaria.

REFERENCES

- [1] Benjamin, T.B., Bona, J.L., and Mahony, J.J., Model equation for long waves in nonlinear dispersive systems, *Phil. Trans. Roy. Soc., London*, A272, 1972, 47-78.
- [2] Boussinesq, J.V., Théorie de l'intumescence liquide appelée onde solitaire ou de translation, se propageant dans un canal rectangulaire, *Comp. Rend. Hebd. Des séances de l'Acad. Des Sci.*, 72, 1871, 755-759.
- [3] Boussinesq, J.V., Théorie générale des mouvements qui sont propagés dans un canal rectangulaire horizontal, *Comp. Rend. Hebd. Des Séances de l'Acad. Des Sci.*, 73, 1871, 256-260.
- [4] Boussinesq, J.V., Théorie des ondes et des remous qui se propagent le long d'un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond, *Journal de Mathématiques Pures et Appliquées*, 17, 1872, 55-108.
- [5] Christov, C.I., *Gaussian Elimination with Pivoting for Multidiagonal Systems*, Internal Report 4, University of Reading, 1994.
- [6] Christov, C.I., Numerical investigation of the long-time evolution and interaction of localized waves, In *Fluid Physics, Proceedings of Summer Schools*, world Scientific, Singapore, M.G. Velarde and C.I. Christov, editors, 403-422, 1995.
- [7] Christov, C.I., Conservative difference scheme for Boussinesq model of surface waves, In *Proc. ICFD V*, Oxford University Press, W.K. Morton and M.J. Baines, editors, 343-349, 1995.
- [8] Christov, C.I., Soliton-supporting model for dispersive shallow-water flows, In *Application of Mathematics in Technology*, Publishing House of the Technical University of Sofia, Sofia, 78-87, 1996.
- [9] Christov, C.I., An energy-consistent Galilean-invariant dispersive shallow-water model, *Wave Motion*, 2000, submitted.
- [10] Christov, C.I., Maugin, G.A., and Velarde, M.G., On the well-posed Boussinesq Paradigm with purely spatial higher-order derivatives, *Phys. Rev. E*, 54, 1996, 3621-3637.
- [11] Christov, C.I. and Velarde, M.G., Inelastic interaction of Boussinesq solitons, *J. Bifurcation and Chaos*, 4, 1994, 1095-1112.
- [12] Maugin, G.A., Application of an energy-momentum tensor in nonlinear elastodynamics, *J. Mech. Phys. of Solids*, 29, 1992, 1543-1558.
- [13] Maugin, G.A., *Material Inhomogeneities in Elasticity*, Chapman & Hall, London, 1993.
- [14] Peregrine, D.H., Calculations of the development of an undular bore, *J. Fluid Mech.*, 25, 1966, 321-330.

- [15] Soerensen, M.P., Christiansen, P.L., and Lohmdahl, P.S., Solitary waves on nonlinear elastic rods I, J. Acoust. Sos. Am., 76, 1984, 871-879.
- [16] Toda, M., Wave propagation in harmonic lattices, J. Phys. Soc. Japan, 23, 1967, 501-506.
- [17] Toda, M., *Theory of Nonlinear Lattice*, 2nd Edition, Springer-Verlag, New York, 1989.
- [18] Turitzyn, S.K., Nonstable solitons and sharp criteria for wave collapse, Phys. Rev. E, 47, 1993, R13-R16.
- [19] Turitzyn, S.K., On Toda lattice model with a transversal degree of freedom sufficient criterion of blow-up in the continuum limit, Phys. Letters A, 143, 1993, 267-269.
- [20] Whitham, G.B., *Linear and Nonlinear Waves*, J. Wiley, New York, 1974.

6 DISCRETE DYNAMICAL SYSTEMS DESCRIBED BY NEUTRAL EQUATIONS

Constantin Corduneanu
University of Texas at Arlington

1. INTRODUCTION

This paper is concerned with discrete dynamical systems described by neutral difference equations of the form

$$\Delta f(n, x_n) = g(n, x_n), \quad x_n = x(n), \quad (1.1)$$

where Δ stands for the first order difference, i.e.,

$$\Delta u \ln = u_{n+1} - u_n \quad (1.2)$$

with $n \in \mathbb{Z}$ or $n \in \mathbb{Z}_+$.

The equation (1.1), in which f, g and x stand for vectors in a certain vector space, together with an initial condition of the form

$$x(0) = x_0, \quad (1.3)$$

will determine, under conditions to be specified below, a unique solution $x = \{x(n) : n \in \mathbb{Z}_+\}$.

Sometimes, when interested in solutions of (1.1) defined on \mathbb{Z} , we may not rely on an initial condition like (1.3). The solution will be determined by imposing on it a certain qualitative condition.

We shall be interested, particularly, in a special case of (1.1), namely

$$\Delta f(x_n) = g(n, x_n), \quad (1.4)$$

which is equivalent to

$$f(x_{n+1}) = f(x_n) + g(n, x_n). \quad (1.5)$$

One can see from (1.5) that in order to be able to construct a solution satisfying (1.3), the invertibility of f is a natural hypothesis. In this case (1.5) becomes

$$x_{n+1} = f^{-1}(f(x_n) + g(n, x_n)), \quad (1.6)$$

which constitutes an usual first order difference equation of the form

$$x_{n+1} = F(n, x_n). \quad (1.7)$$

Under the same hypothesis of invertibility of f , one can reduce (1.4) to the explicit form (1.7) by means of the substitution

$$y_n = f(x_n). \quad (1.8)$$

Indeed, (1.4) becomes

$$\Delta y_n = g\left(n, f^{-1}(y_n)\right) = G\left(n, y_n\right), \quad (1.9)$$

which is precisely of the form (1.7).

Let us precise now the meaning of the variables involved in the above equations. Throughout this paper we will make the following assumptions:

- a) $x = \{x_n\}$ takes its value in the Euclidian space \mathbb{R}^m ,
- b) $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a homeomorphic map,
- c) $g : \mathbb{Z}_+ \times \mathbb{R}^m \rightarrow \mathbb{R}^m$,
- or
- c') $g : \mathbb{Z} \times \mathbb{R}^m \rightarrow \mathbb{R}^m$.

2. CONVERGENT SOLUTIONS

We will first consider the problem of existence of solutions of (1.4) on \mathbb{Z}_+ , such that

$$\lim_{n \rightarrow \infty} x_n = x_\infty \quad (2.1)$$

exists in \mathbb{R}^m . In other words, we seek solutions x such that $x \in c(\mathbb{R}^m)$ = the space of convergent sequences in \mathbb{R}^m .

If we write (1.5) for $n = 0, 1, 2, \dots, N$, then by addition of these equations one obtains

$$f(x_{N+1}) = f(x_0) + \sum_{k=0}^N g(k, x_k). \quad (2.2)$$

The following proposition is an immediate consequence of the formula (2.2):

The necessary and sufficient condition for convergence (i.e. $x \in c(\mathbb{R}^m)$) of the solution of (1.4), under initial condition (1.3), is the convergence of the series

$$\sum_{k=0}^{\infty} g(k, x_k). \quad (2.3)$$

There are several cases that may occur in connection with the requirement of convergence of the series (2.3).

First, it is possible (2.3) diverges for any $x_0 \in \mathbb{R}^m$. In this case there is no convergent solution to equation (1.4).

Second, it is possible to have convergence of (2.3) only for some $x_0 \in \mathbb{R}^m$. This means there exist convergent solutions to (1.4), but not all of them are convergent.

Third, it is possible to have convergence of (2.3), regardless of the choice of $x_0 \in \mathbb{R}^m$.

It is easy to illustrate each of the above mentioned situations. For instance, if we take $m = 1$ and $g(k, x) = (k+1)^{-1}$, $k \geq 0$, then whatever initial value x_0 we choose, the series (2.3) is divergent.

The second case is illustrated by the scalar equation $x_{n+1} = x_n(1 - x_n)$. It is obvious that $x_0 = 0$ or $x_0 = 1$ provide convergent solutions. If $0 < x_0 < 1$, one obtains a decreasing sequence solution (because $x_{n+1} - x_n = -x_n^2$). Since all x_n are positive, it follows that $x_\infty = \lim x_n$ as $n \rightarrow \infty$ does exist, and $x_\infty \geq 0$. But x_∞ satisfies $x_\infty = x_\infty(1 - x_\infty)$, which implies $x_\infty = 0$. Hence, $0 \leq x_0 \leq 1$ provides only convergent solutions. On the other hand, if we take $x_0 = -1$, one obtains an unbounded solution.

Finally, the third situation, when all solutions are convergent, can be simply illustrated by taking $g(k, x) = \theta$ for all $k \geq K_0$, and any $x \in \mathbb{R}^m$. Then, (2.3) reduces to a finite sum, regardless of the choice of the initial condition. Obviously, the terms of the sequence $\{x_n\}$ will coincide, starting with large enough k .

It is an interesting problem to provide conditions on the map g , such that the series (2.3) is convergent for any initial choice. In other words, all solutions of equation (1.4) will be convergent. A careful examination of formula (2.2) is helpful in tailoring such conditions.

On behalf of assumption b) above, (2.2) can be also written in the form

$$x_{N+1} = f^{-1} \left(f(x_0) + \sum_{k=0}^N g(k, x_k) \right). \quad (2.4)$$

Let us make now an assumption on f^{-1} , in order to be able to use the Gronwall's inequality for discrete variables. Namely,

$$\|f^{-1}(x)\| \leq \alpha \|x\| + \beta, \quad x \in \mathbb{R}^m, \quad (2.5)$$

with α and β positive numbers. In other words, we assume that f^{-1} , has linear growth.

Based on (2.5), from (2.4) we derive the following inequality

$$\|x_{N+1}\| \leq \alpha \|f(x_0)\| + \beta + \alpha \sum_{k=0}^N \|g(k, x_k)\|. \quad (2.6)$$

This last inequality can be processed easily if we make the following hypothesis on the map g :

$$\|g(k, x)\| \leq c_k \|x\|, \quad k = 0, 1, 2, \dots, \quad (2.7)$$

where the series (with positive terms)

$$\sum_{k=0}^{\infty} c_k \quad (2.8)$$

is convergent.

From (2.6), (2.7) we obtain

$$\|x_{N+1}\| \leq \alpha \|f(x_0)\| + \beta + \alpha \sum_{k=0}^N c_k \|x_k\|. \quad (2.9)$$

The inequality (2.9) is of Gronwall's type [5], and we obtain

$$\|x_{N+1}\| \leq (\alpha \|f(x_0)\| + \beta) \exp \left\{ \alpha \sum_{k=0}^N c_k \right\}. \quad (2.10)$$

From (2.10) one finds out that, for any $n \in \mathbb{Z}_+$,

$$\|x_{N+1}\| \leq (\alpha \|f(x_0)\| + \beta) \exp \left\{ \alpha \sum_{k=0}^{\infty} c_k \right\}, \quad (2.11)$$

which shows that, under conditions (2.5) and (2.7) for the maps f and g , the sequence $\{x_n\}$ is bounded in \mathbb{R}^m .

The boundedness of $\{x_n\}$ implies the convergence of the series (2.3), if we take into account (2.7) and (2.8).

Hence, $\{x_n\}$ converges as $n \rightarrow \infty$, and denoting $x_\infty = \lim x_n$ as $n \rightarrow \infty$, we obtain from (2.10)

$$\|x_\infty\| \leq (\alpha \|f(x_0)\| + \beta) \exp \left\{ \alpha \sum_{k=0}^{\infty} c_k \right\}. \quad (2.12)$$

We summarize the discussion above in the following result.

Theorem 1. Consider the difference equation (1.4), under conditions a), b), c), (2.5), (2.7), and (2.8). Then, the solution of (1.4), satisfying (1.3), is convergent, for any $x_0 \in \mathbb{R}^m$.

3. ALMOST PERIODIC SOLUTIONS

Let us turn now our attention to the form (1.9) of equation (1.4), which has been obtained by means of the substitution (1.8). Since this equation is of the form (1.7), for which a rich literature is available [1], [5], we will apply one of the results in [2] (see also [3], in a different approach) in order to derive a criterion of almost periodicity for bounded solutions to (1.4).

In [2], Theorem 1 states the almost periodicity of a bounded solution of (1.4) is guaranteed by the following conditions:

- 1) $G(n, y)$ is almost periodic on \mathbb{Z} , uniformly with respect to y in any bounded set of \mathbb{R}^m ;

2) $G(n, y)$ is monotone in y , i.e.

$$\langle G(n, x) - G(n, y), x - y \rangle \geq \lambda |x - y|^2, \quad (3.1)$$

for any $x, y \in \mathbb{R}^m$, with $\lambda > 1$ a constant.

We have now to transfer from G , to g and f , the above properties.

Theorem 2. Consider the equation (1.4), and assume the following hypotheses: f satisfies condition b); g satisfies condition c) and is almost periodic from \mathbb{Z} to \mathbb{R}^m , in the first argument uniformly with the second argument in bounded sets of \mathbb{R}^m ; f and g are satisfying the condition

$$\langle g(n, u) - g(n, v), f(u) - f(v) \rangle \geq \lambda |f(u) - f(v)|^2, \quad (3.2)$$

for any $u, v \in \mathbb{R}^m$, and fixed $\lambda > 0$.

Then, if (1.4) has a bounded solution (on \mathbb{Z}), this solution is almost periodic.

The proof of Theorem 2 requires only to show that condition (3.2) is the same as condition (3.1), taking into account that $G(n, y) = y + g(n, f^{-1}(y))$. This is easily obtained if we rely on the fact that f is a homeomorphism of \mathbb{R}^m .

Many other results can be obtained for equation (1.4), using readily available results for first order difference equations.

4. THE CASE OF EQUATION (1.1)

Before we conclude this note, we want to make a remark on the more general (than (1.4)) equation (1.1). From (1.1) one derives the formula

$$f(N+1, x_{N+1}) = f(0, x_0) + \sum_{k=0}^N g(k, x_k). \quad (4.1)$$

From (4.1) one can develop the same type of argument as in the special case $f(n, x) = f(x)$. We will assume that an inequality of the form

$$|f(n, x)| \geq g(|x|), \quad n \in \mathbb{Z}_+, \quad (4.2)$$

holds, where $g(u)$ is a scalar function defined for $u \in \mathbb{R}_+$.

If we impose on g the condition to possess an inverse function g^{-1} which has linear growth (see inequality (2.10)), then we can proceed with (1.1) in the same manner we proceeded in the case $f(n, x) = f(x)$.

We leave to the reader the task to carry out the details.

REFERENCES

- [1] Agarwal, R.P., *Difference Equations and Inequalities*. M. Dekker, New York, 1992.
- [2] Corduneanu, C., Almost periodic discrete processes, *Libertas Mathematics*, II (1982), 159-169.

- [3] Corduneanu, C., Discrete qualitative inequalities applications, Nonlinear Analysis, TMA, Vol. XXV (1995), 933-939.
- [4] Elaydi, S., Graef, J., Ladas, G., Peterson, A. (eds), *Proceedings of the First International Conference on Difference Equations*, Gordon and Breach Publ., 1995.
- [5] Lakshmikantham, V. and Trigiante, D., *Theory of Difference Equations, Numerical Methods and Applications*, Academic Press, Boston-San Diego, 1988.

7 OSCILLATION OF THIRD ORDER DIFFERENTIAL EQUATIONS WITH AND WITHOUT DELAY

R.S. Dahiya
Department of Mathematics
Iowa State University
Ames, Iowa 50011

ABSTRACT

Sufficient conditions are obtained for the solutions of third order differential equations to be oscillatory or to approach zero monotonically as $t \rightarrow \infty$.

1. INTRODUCTION

There is a large literature on the behavior of solutions of third order linear differential equations [1]-[10], [12], [14], and many of these papers deal with the oscillatory and/or nonoscillatory properties of their solutions. V.S. Rao and R.S. Dahiya [13] discussed the behavior of the solutions of the equation

$$\left(r(y')' \right)' + (qy)' + q_2 y' = 0. \quad (1.1)$$

In Section 2, we shall be interested in the generalized equation

$$\left(b(t)(a(t)y')' \right)' + (q_1 y)' + q_2 y' = f(t). \quad (1.2)$$

Section 3 contains third order differential equations with delay. A non-trivial solution of (1.2) is said to be oscillatory on the interval I if it has infinitely many zeroes on I, otherwise, nonoscillatory. The equation (1.2) is said to be oscillatory or nonoscillatory, respectively, depending on the existence or nonexistence of an oscillatory solution.

2. OSCILLATION OF DIFFERENTIAL EQUATIONS WITHOUT DELAY

Assume that

$$a, b \in C^1[t_0, \infty), a, b > 0 \text{ and} \quad (2.1)$$

$$q_i \in C^1[t_0, \infty), i = 1, 2. \quad (2.2)$$

Theorem 1. Let

$$(i) \quad (q_1 + q_2) \geq 0,$$

$$(ii) \quad (q_2 - q_1)' \leq 0,$$

$$(iii) \quad (a'b - ab') \geq 0, \text{ and}$$

$$(iv) \quad \lim_{t \rightarrow \infty} \frac{a'(t)}{a(t)} < \infty.$$

The conditions (i), (ii), and (iii) hold on $[t_0, \infty)$ but not identically zero on any subinterval of $[t_0, \infty)$. Let $f(t) = 0$.

If

$$\left\{ A + B \int_{t_0}^t b^{-1}(s) ds - \int_{t_0}^t b^{-1}(s) q_1(s) ds \right\} < 0 \quad (2.3)$$

for t sufficiently large and any constants A and B , then equation (1.2) is oscillatory.

Proof. Let t_1 be the last zero of y , then without loss of generality we can assume that there exists $t_2 \geq t_1$ such that $y(t) > 0$ for $t \in [t_2, \infty)$ and $y'(t) > 0$ for

$$t \in (t_2, \alpha) \text{ with some } \alpha \text{ close to } t_2. \text{ Let } t_3 \in (t_2, \alpha).$$

Dividing equation (1.2) by y and integrating from t_3 to t , we get

$$\int_{t_3}^t \left(\frac{(b(ay'))'}{y} \right) (s) ds + \int_{t_3}^t \left(\frac{(q_1 y')}{y} \right) (s) ds + \int_{t_3}^t \left(\frac{(q_2 y')}{y} \right) (s) ds = 0.$$

Now integrating by parts, we obtain

$$\left(\frac{b(ay')}{y} \right) (t) - \left(\frac{b(ay')}{y} \right) (t_3) + \int_{t_3}^t \left(\frac{(b(ay'))'}{y^2} \right) (s) y'(s) ds$$

$$+ q_1(t) - q_1(t_3) + \int_{t_3}^t \left(\frac{q_1 y'}{y} \right) (s) ds + \int_{t_3}^t \left(\frac{q_2 y'}{y} \right) (s) ds = 0.$$

This implies

$$\left(\frac{b(ay')'}{y} \right) (t) + \int_{t_3}^t \frac{(q_1 + q_2)y'}{y} ds + \int_{t_3}^t \frac{a'by'^2}{y^2} ds + \int_{t_3}^t \frac{aby'y''}{y^2} ds = k_1 - q_1(t), \quad (2.4)$$

where

$$k_1 = \left(\frac{b(ay')'}{y} \right) (t_3) + q_1(t_3).$$

The last integral (2.4) can be written as follows:

$$\int_{t_3}^t \frac{ab(y'^2)}{2y^2} ds = \left(\frac{aby'^2}{2y^2} \right) (t) - \left(\frac{aby'^2}{2y^2} \right) (t_3) - \int_{t_3}^t \left(\frac{ab}{2y^2} \right)' y'^2 ds,$$

or,

$$\int_{t_3}^t \left(\frac{ab}{2y^2} \right)' (y'^2) ds = \left(\frac{aby'^2}{2y^2} \right) (t) - \int_{t_3}^t \frac{(ab)' y'^2}{2y^2} ds + \int_{t_3}^t \frac{aby'^3}{y^3} ds - k_2 \quad (2.5)$$

where

$$k_2 = \left(\frac{aby'^2}{2y^2} \right) (t_3).$$

Substituting (2.5) in (2.4), we get

$$\begin{aligned} & \left(\frac{b(ay')'}{y} \right) (t) + \int_{t_3}^t \frac{(q_1 + q_2)y'}{y} ds + \int_{t_3}^t \frac{a'by'^2}{y^2} ds + \left(\frac{aby'^2}{2y^2} \right) (t) \\ & - \int_{t_3}^t \frac{(ab)' y'^2}{2y^2} ds + \int_{t_3}^t \frac{aby'^3}{y^3} ds = k_1 + k_2 - q_1(t). \end{aligned}$$

Dividing by $b(t)$, it follows

$$\begin{aligned} & \left(\frac{(ay')'}{y} \right) (t) + \frac{1}{b(t)} \int_{t_3}^t \frac{(q_1 + q_2)y'}{y} ds + \frac{1}{b(t)} \int_{t_3}^t \frac{a'by'^2}{y^2} ds + \left(\frac{ay'^2}{2y^2} \right) (t) \\ & - \frac{1}{b(t)} \int_{t_3}^t \frac{(ab)' y'^2}{2y^2} ds + \frac{1}{b(t)} \int_{t_3}^t \frac{aby'^3}{y^3} ds = \frac{k}{b(t)} - \frac{q_1(t)}{b(t)}, \end{aligned}$$

or,

$$\begin{aligned} & \left(\frac{(ay')'}{y} \right) (t) + b^{-1}(t) \int_{t_3}^t (q_1 + q_2) \frac{y'}{y} ds + b^{-1}(t) \int_{t_3}^t (a'b - ab') \frac{y'^2}{2y^2} ds \\ & + b^{-1}(t) \int_{t_3}^t ab \frac{y'^3}{y^3} ds + \left(\frac{ay'^2}{2y^2} \right) (t) = kb^{-1}(t) - b^{-1}(t) q_1(t), \end{aligned} \quad (2.6)$$

where

$$k = k_1 + k_2.$$

Now integrating (2.6) from t_3 to t , we obtain

$$\begin{aligned} & \int_{t_3}^t \frac{(ay')'}{y} ds + \int_{t_3}^t b^{-1}(s) \int_{t_3}^s (q_1 + q_2) \frac{y'}{y} du ds + \\ & \int_{t_3}^t b^{-1}(s) \int_{t_3}^s (a'b - ab') \frac{y'^2}{2y^2} du ds + \int_{t_3}^t b^{-1}(s) \int_{t_3}^s \frac{aby'^3}{y^3} du ds + \int_{t_3}^t \frac{ay'^2}{2y^2} ds \\ &= k \int_{t_3}^t b^{-1}(s) ds - \int_{t_3}^t b^{-1}(s) q_1(s) ds, \end{aligned}$$

or,

$$\begin{aligned} & \left(\frac{ay'}{y} \right)(t) + \frac{3}{2} \int_{t_3}^t \frac{ay'^2}{y^2} ds + \int_{t_3}^t b^{-1}(s) \int_{t_3}^s (q_1 + q_2) \frac{y'}{y} du ds \\ & + \int_{t_3}^t b^{-1}(s) \int_{t_3}^s (a'b - ab') \frac{y'^2}{2y^2} du ds + \int_{t_3}^t b^{-1}(s) \int_{t_3}^s \frac{aby'^3}{y^3} du ds \\ &= k \int_{t_3}^t b^{-1}(s) ds - \int_{t_3}^t b^{-1}(s) q_1(s) ds + M, \end{aligned} \quad (2.7)$$

where

$$M = \left(\frac{ay'}{y} \right)(t_3).$$

Suppose $y'(t) > 0$ for $t \in [t_3, \infty)$, then by using (i), (iii), and (2.3) in (2.7), we obtain

$$\frac{a(t) y'(t)}{y(t)} < 0. \quad (2.8)$$

Since $a(t) > 0$ and $y(t) > 0$, then from (2.8) we have $y'(t) < 0$ for sufficiently large t and this contradicts our assumption. So it is clear that there exists $t_4 \geq t_3$ such that $y'(t_4) = 0$.

Now, we shall conclude the theorem by showing that

$$y(t_1) = y'(t_4) = 0$$

which contradicts $y(t) > 0$ for $t \geq t_1$.

Multiplying equation (1.2) by y and integrating from t_1 to t , it follows

$$\int_{t_1}^t \left(b(ay')' \right)' \cdot y ds + \int_{t_1}^t (q_1 y)' \cdot y ds + \int_{t_1}^t q_2 y' \cdot y ds = 0,$$

or,

$$\left(b(ay')' \cdot y \right)(t) - \left(b(ay')' \cdot y \right)(t_1) - \int_{t_1}^t b(ay')' \cdot y' ds +$$

$$(q_1 y^2)(t) - (q_1 y^2)(t_1) - \int_{t_1}^t q_1(y^2)' ds + \int_{t_1}^t q_1 y' y ds + \int_{t_1}^t q_2 y' y ds = 0,$$

or,

$$\begin{aligned} & \left(b(ay')' \cdot y \right)(t) - \left(b(ay')' \cdot y \right)(t_1) - \int_{t_1}^t (aby''y' + a'b'y'^2) ds \\ & + (q_1 y^2)(t) - (q_1 y^2)(t_1) - \int_{t_1}^t q_1(y^2)' ds \\ & + \frac{1}{2} \int_{t_1}^t q_1(y^2)' ds + \frac{1}{2} \int_{t_1}^t q_2(y^2)' ds = 0. \end{aligned} \quad (2.9)$$

Since $y(t_1) = 0$, then (2.9) becomes

$$\left(b(ay')' \cdot y \right)(t) - \int_{t_1}^t (aby''y' + a'b'y'^2) ds + (q_1 y^2)(t) + \frac{1}{2} \int_{t_1}^t (q_2 - q_1)(y^2)' ds = 0,$$

or

$$\begin{aligned} & (aby''y + a'b'y'^2)(t) - \int_{t_1}^t a'b'y'^2 ds - \frac{1}{2} \int_{t_1}^t ab(y'^2)' ds \\ & + \frac{1}{2} \int_{t_1}^t (q_2 - q_1)(y^2)' ds + (q_1 y^2)(t) = 0. \end{aligned} \quad (2.10)$$

Integrating by parts and using again $y(t_1) = 0$, we get

$$\begin{aligned} & (abyy'' + a'byy')(t) - \int_{t_1}^t a'b'y'^2 ds - \frac{1}{2} ((aby'^2)(t) - (aby'^2)(t_1)) + \frac{1}{2} \int_{t_1}^t (ab)' y'^2 ds \\ & + \frac{1}{2} ((q_2 - q_1)y^2)(t) - ((q_2 - q_1)y^2)(t_1) - \frac{1}{2} \int_{t_1}^t (q_2 - q_1)' y^2 ds + (q_1 y^2)(t) = 0, \end{aligned}$$

or,

$$\begin{aligned} & (abyy'') + (a'byy')(t) - \frac{1}{2} (aby'^2)(t) + \frac{1}{2} (aby'^2)(t_1) \\ & + \frac{1}{2} ((q_2 - q_1)y^2)(t) + \frac{1}{2} \int_{t_1}^t (ab' - a'b) y'^2 ds - \frac{1}{2} \int_{t_1}^t (q_2 - q_1)' y^2 ds = 0. \end{aligned} \quad (2.11)$$

Define

$$F(y(t)) = \frac{1}{2} aby'^2 - abyy'' - a'byy' - \frac{1}{2} (q_2 - q_1) y^2. \quad (2.12)$$

Using (2.12) in (2.11), we have

$$F(y(t)) = F(y(t_1)) + \frac{1}{2} \int_{t_1}^t (ab' - a'b) y'^2 ds - \frac{1}{2} \int_{t_1}^t (q_2 - q_1) y^2 ds, \quad (2.13)$$

where

$$F(y(t_1)) = \frac{1}{2} (aby'^2)(t_1) \geq 0. \quad (2.14)$$

Using (ii), (iii), and (2.14) in (2.13), we prove that $F(y(t))$ is a strictly increasing function, which vanishes whenever y has a double zero, i.e., $y = y' = 0$.

Since $y(t_4) > 0$ and $y'(t_4) = 0$ then, we have

$$F(y(t_4)) = -(abyy'')(t_4) - \frac{1}{2} (q_1 + q_2) y^2(t_4) > 0. \quad (2.15)$$

From (2.15) we see that

$$y''(t_4) < 0. \quad (2.16)$$

It is clear that $y'(t)$ cannot vanish more than once in $[t_4, \infty)$. Hence,

$$y'(t) < 0 \text{ for } t > t_4 \text{ and}$$

$$\lim_{t \rightarrow \infty} y(t) \text{ exists.}$$

Now we divide the rest of the argument into three cases depending on the sign of $y''(t)$.

Case 1. Let $y''(t) \leq 0$ eventually.

Then $y(t)$ becomes eventually negative and this contradicts

$$y(t) > 0 \text{ for all } t \geq t_1.$$

Case 2. Let $y''(t) \geq 0$ eventually. Then, since $y'(t) < 0$ for $t > t_4$, we would have

$$\lim_{t \rightarrow \infty} y'(t) = 0$$

and consequently,

$$\lim_{t \rightarrow \infty} \frac{F(y(t))}{a(t) b(t)} \leq \lim_{t \rightarrow \infty} \left(\frac{1}{2} y'^2 - yy'' - \frac{a'}{a} yy' \right)(t). \quad (2.17)$$

Since $y(t) > 0$, $y''(t) \geq 0$ eventually and $\lim_{t \rightarrow \infty} y(t)$ exists and by using (iv), we get from (2.17)

$$\lim_{t \rightarrow \infty} \frac{F(y(t))}{a(t) b(t)} \leq 0$$

which contradicts the fact that F is strictly increasing.

Case 3. Suppose y'' changes its sign for arbitrary large t . Then for any $\varepsilon > 0$, there exists a large t such that

$$0 > y'(t) > -\varepsilon$$

and a relative maximum of $y(t)$ at \bar{t} such that

$$0 > y'(\bar{t}) > -\varepsilon \text{ and } y''(\bar{t}) = 0.$$

Then we have

$$F(y(\bar{t})) \leq a(\bar{t}) b(\bar{t}) \frac{\varepsilon^2}{2} + |a'(\bar{t})| b(\bar{t}) y(\bar{t}) \varepsilon$$

for arbitrary large \bar{t} and this implies that

$$\lim_{t \rightarrow \infty} F(y(t)) \leq 0.$$

This is a contradiction to the fact that $F(y(t))$ is strictly increasing. Therefore, the proof is complete.

Example 1. Consider the equation

$$\left(e^{-t} (e^t y')' \right)' + (ty)' + (1-t) y' = 0 \quad (2.18)$$

where

$$a(t) = e^t, b(t) = e^{-t}, q_1(t) = t \text{ and } q_2(t) = 1-t.$$

It is clear that

$$q_1 + q_2 = t + 1 - t > 0, (q_2 - q_1)' = (1 - t - t)' = -2 < 0, a'b - ab' = 2$$

$$\text{and } \lim_{t \rightarrow \infty} \frac{a'}{a} = \lim_{t \rightarrow \infty} \frac{e^t}{e^t} = 1 < \infty.$$

All conditions of Theorem 1 are satisfied. Hence, all solutions of equation (2.18) are oscillatory.

In fact $y = \sin(t)$ is a solution of equation (2.18).

3. OSCILLATION OF DIFFERENTIAL EQUATIONS WITH DELAY

In this section we study the effect of delay on the properties of solutions of equation of the form

$$\left(b(ay')' \right)' (t) + q_1 y(t) + q_2 y(\tau(t)) = 0, \quad (3.1)$$

where a, b, q_1 and q_2 are the same as in (2.1) and (2.2) and $\tau(t)$ satisfies the following:

$$\tau(t) \leq t, \tau'(t) > 0 \text{ and } \tau(t) \rightarrow \infty \text{ as } t \rightarrow \infty. \quad (3.2)$$

Theorem 2. Assume that

$$q_1 \geq 0 \text{ and } q_2 \geq 0 \text{ for } t \in [t_0, \infty), \quad (3.3)$$

$$b'(\tau(t)) \leq 0 \text{ for } t \in [t_0, \infty), \text{ and} \quad (3.4)$$

there exists a differentiable function

$$\delta : [t_1, \infty) \rightarrow (0, \infty) \text{ such that} \quad (3.5)$$

$\delta' < 0$ and $\delta'' > 0$ for $t \geq t_0$ and

$$\delta' \leq \frac{b'}{2b} \delta \text{ for } t \geq t_0. \quad (3.6)$$

If

$$\int_{t_1}^{\infty} \left\{ (q_1 + q_2) \delta - \frac{a(\tau) b \delta'^2}{4B\tau\delta} \right\} dt = \infty, \quad (3.7)$$

$$\int_{t_1}^{\infty} \frac{t}{a\delta} dt < \infty \quad (3.8)$$

and

$$\int_{t_1}^{\infty} \frac{1}{a\delta} \left(\int_{t_1}^s \frac{(q_1 + q_2)}{b} \delta du \right) ds dt = \infty \quad (3.9)$$

then every solution $y(t)$ of equation (3.1) is either oscillatory or

$$\lim_{t \rightarrow \infty} y(t) = 0.$$

Proof. Let $y(t)$ be a nonoscillatory solution of equation (3.1). We may assume without loss of generality that $y(t) > 0$ for $t \geq t_0$, then there exists $t_1 \geq t_0$ such that $y(\tau(t)) > 0$ for $t \geq t_1$. From equation (3.1) we have

$$\left(b(ay')' \right)' = -q_1 y(t) - q_2 y(\tau(t)). \quad (3.10)$$

It is clear that $-\left(b(ay')' \right)'$ is positive for $t \geq t_1$, hence, $y(t)$ is monotone and one-signed. $y'(t)$ and $y''(t)$ are also monotone and one-signed for sufficiently large t .

Claim 1:

$$(ay')' > 0 \text{ for } t \geq t_1. \quad (3.11)$$

From equation (3.10) we have

$$b(ay'')' + b'(ay')' \leq 0,$$

or,

$$(ay'')' \leq -\frac{b'}{b} (ay')'. \quad (3.12)$$

If $(ay')' \leq 0$ then ay' is decreasing and concave down, therefore ay' is eventually negative which is a contradiction. Thus (3.11) is true.

Claim 2:

$$y'(t) < 0 \text{ for } t \geq t_1. \quad (3.13)$$

If $y' \geq 0$ for $t \geq t_1$, then $y(t)$ is increasing and positive. Define the function

$$\omega(t) = \frac{\left(b(ay')' \delta \right)(t)}{y(\tau(t))}. \quad (3.14)$$

By differentiating (3.14), we have

$$\omega'(t) = \frac{y(\tau(t)) \cdot \left(b(ay')' \right)' \delta + y(\tau(t)) \left(b(ay')' \right) \delta' - \left(b(ay')' \delta \right) y'(\tau(t)) \tau'}{y^2(\tau(t))},$$

or,

$$\omega'(t) = \left(b(ay')' \right)' \frac{\delta}{y(\tau(t))} + b(ay')' \frac{\delta'}{y(\tau(t))} - \frac{\left(b(ay')' \delta \right)}{y(\tau(t))} \frac{y'(\tau)}{y(\tau)} \tau'. \quad (3.15)$$

Using (3.10) and (3.14) in (3.15), we obtain

$$\omega'(t) = -q_1 \delta \frac{y(t)}{y(\tau)} - \delta q_2 + \frac{\delta'}{\delta} \omega - \omega \frac{y'(\tau)}{y(\tau)} \tau'. \quad (3.16)$$

Since $y' \geq 0$ and $(ay')' > 0$, then by Kiguradze's Lemma [11], we have

$$ay' \geq B t (ay')' \text{ for some } B > 0 \quad (3.17)$$

and

$$\left((ay')' \right)(t) \leq \left(a(\tau) y'(\tau) \right)' \text{ for } t \geq t_2 \geq t_1. \quad (3.18)$$

Since $y' \geq 0$ for $t \geq t_1$, then $y(\tau) \leq y(t)$, or,

$$\frac{y(t)}{y(\tau)} \geq 1.$$

Now (3.16) can be written as

$$\omega'(t) \leq -(q_1 + q_2) \delta + \frac{\delta'}{\delta} \omega - \frac{a(\tau) y'(\tau)}{a(\tau) y(\tau)} \tau' \omega. \quad (3.19)$$

By using (3.17) in (3.19), it follows

$$\omega'(t) \leq -(q_1 + q_2) \delta + \frac{\delta'}{\delta} \omega - B \frac{\left(a(\tau) y'(\tau) \right)'}{a(\tau) y(\tau)} \tau \tau' \omega. \quad (3.20)$$

Now from (3.18) and (3.20), we get

$$\omega'(t) \leq -(q_1 + q_2) \delta + \frac{\delta'}{\delta} \omega - B \frac{\left(a(t) y'(t) \right)'}{a(t) y(t)} \tau \tau' \omega.$$

Using (3.14) again, we obtain

$$\omega'(t) \leq -(q_1 + q_2)\delta + \frac{\delta'}{\delta} \omega - B \frac{\tau\tau'}{a(\tau) b(t) \delta} \omega^2,$$

or

$$\omega'(t) \leq -(q_1 + q_2)\delta - \left\{ \left(\frac{B\tau\tau'}{a(\tau) b\delta} \right) \omega^2 - \frac{\delta'}{\delta} \omega \right\}.$$

Now completing the square on the right hand side, it follows

$$\omega'(t) \leq -(q_1 + q_2)\delta - \left\{ \left(\frac{B\tau\tau'}{a(\tau) b\delta} \right)^{\frac{1}{2}} \omega - \frac{\delta'/2\delta}{\left(\frac{B\tau\tau'}{a(\tau) b\delta} \right)^{\frac{1}{2}}} \right\}^2 + \frac{\delta'^2 a(\tau)b}{4B\tau\tau'\delta},$$

or,

$$\omega'(t) \leq -(q_1 + q_2)\delta + \delta'^2 \frac{a(\tau)b}{4B\tau\tau'\delta}. \quad (3.21)$$

Integrating (3.21) from t_2 to t , we get

$$\omega(t) - \omega(t_2) \leq - \int_{t_2}^t \left\{ (q_1 + q_2)\delta - \delta'^2 \frac{a(\tau)b}{4B\tau\tau'\delta} \right\} ds,$$

or,

$$\int_{t_2}^t \left\{ (q_1 + q_2)\delta - \delta'^2 \frac{a(\tau)b}{4B\tau\tau'\delta} \right\} ds \leq \omega(t_2) - \omega(t). \quad (3.22)$$

Because of (3.11) and (3.14), we would claim that

$$\omega(t) > 0 \text{ for } t \geq t_1$$

and the (3.22) could be written as

$$\int_{t_2}^t \left\{ (q_1 + q_2)\delta - \delta'^2 \frac{a(\tau)b}{4B\tau\tau'\delta} \right\} ds \leq \omega(t_2),$$

or

$$\int_{t_2}^t \left\{ (q_1 + q_2)\delta - \delta'^2 \frac{a(\tau)b}{4B\tau\tau'\delta} \right\} ds < \infty.$$

This contradicts (3.7). Hence, (3.13) is true.

From (3.13) and the fact that $y(t) > 0$ for $t \geq t_1$, there exists a constant $c \geq 0$ such that

$$\lim_{t \rightarrow \infty} y(t) = c.$$

We will prove that $c = 0$.

Let $c > 0$, then there exists $t^* \geq t_1$ such that

$$y(\tau(t)) > \frac{c}{2} \text{ for } t \geq t^*. \quad (3.23)$$

Define the function

$$G(t) = ay'\delta \text{ for } t \geq t^*. \quad (3.24)$$

Differentiating (3.24), we get

$$G'(t) = (ay')' \delta + ay'\delta'. \quad (3.25)$$

Multiplying (3.25) by $b(t)$ and differentiating again, we obtain

$$bG'' + b'G' = \left(b(ay')' \right)' \delta + 2b(ay')' \delta' + b'(ay')' \delta' + b(ay')\delta''. \quad (3.26)$$

Using (3.25) in (3.26), we have

$$G''(t) = \frac{\delta}{b} \left(b(ay')' \right)' + 2(ay')' \delta' - \frac{b'}{b} (ay')' \delta + (ay') \delta''. \quad (3.27)$$

Now if we use (3.10) in (3.27), we get

$$G''(t) = -\frac{\delta}{b} (q_1 y(t) + q_2 y(\tau)) + 2(ay')' \left(\delta' - \frac{b'}{2b} \delta \right) + (ay') \delta''. \quad (3.28)$$

Using (3.5) and (3.13) in (3.28), will give

$$G''(t) \leq -\frac{\delta}{b} (q_1 y(t) + q_2 y(\tau)) + 2(ay')' \left(\delta' - \frac{b'}{2b} \delta \right). \quad (3.29)$$

By using (3.6) and (3.11) in (3.29), we obtain

$$G''(t) \leq -\frac{\delta}{b} (q_1 y(t) + q_2 y(\tau)). \quad (3.30)$$

Using (3.23) in (3.20), we deduce that

$$G''(t) \leq -\frac{c}{2} \frac{\delta}{b} (q_1 + q_2) \text{ for } t \geq t^*. \quad (3.31)$$

Integrating (3.31) from t^* to t , we get

$$G'(t) \leq G'(t^*) - \frac{c}{2} \int_{t^*}^t (q_1 + q_2) \frac{\delta}{b} ds.$$

Integrating again, will lead to

$$G(t) \leq G(t^*) + G'(t^*)(t - t^*) - \frac{c}{2} \int_{t^*}^t \int_s^t (q_1 + q_2) \frac{\delta}{b} du ds. \quad (3.32)$$

Now if we use (3.13) in (3.24), we conclude that

$$G(t) < 0 \text{ for } t \geq t^*. \quad (3.33)$$

Using (3.5), (3.11), and (3.13) in (3.25), we get

$$G(t) > 0 \text{ for } t \geq t^*. \quad (3.34)$$

Now using (3.33) and (3.34) in (3.32), we obtain

$$G(t) \leq G'(t^*)t - \frac{c}{2} \int_{t^*}^t \int_s^t (q_1 + q_2) \frac{\delta}{b} du ds. \quad (3.35)$$

From (3.24) and (3.35), it follows

$$y'(t) \leq G'(t^*) \frac{t}{a\delta} - \frac{c}{2} \frac{1}{a\delta} \int_{t^*}^t \int_{t^*}^s (q_1 + q_2) \frac{\delta}{b} du ds. \quad (3.36)$$

Integrating (3.36) also from t^* to t , we have

$$y(t) \leq y(t^*) + G'(t^*) \int_{t^*}^t \frac{s}{a\delta} ds - \frac{c}{2} \int_{t^*}^t \frac{1}{a\delta} \int_{t^*}^s \int_{t^*}^u (q_1 + q_2) \frac{\delta}{b} dv du ds. \quad (3.37)$$

Taking the limit as $t \rightarrow \infty$, we get

$$\lim_{t \rightarrow \infty} y(t) \leq y(t^*) + G'(t^*) \int_{t^*}^{\infty} \frac{s}{a\delta} ds - \frac{c}{2} \int_{t^*}^{\infty} \frac{1}{a\delta} \int_{t^*}^s \int_{t^*}^u (q_1 + q_2) \frac{\delta}{b} dv du ds. \quad (3.38)$$

Now using (3.8) and (3.9) in (3.38), will give a contradiction.

Hence, $c = 0$ and the proof is complete.

Example 2. Consider the equation

$$\left(e^{-t/4} \left(e^{t/2} y' \right)' \right)' + \frac{1}{8} e^{t/4} y(t) + \frac{1}{4} e^{(t/4-\pi)} y(t-\pi) = 0 \quad (3.39)$$

where

$$a(t) = e^{t/2}, b(t) = e^{-t/4}, q_1(t) = \frac{1}{8} e^{t/4}, q_2(t) = \frac{1}{4} e^{(t/4-\pi)}, \delta(t) = e^{-t/8},$$

and $\tau(t) = t - \pi$.

It is clear that

$$\tau(t) = t - \pi < t, \tau' = 1 > 0, \tau(t) \rightarrow \infty \text{ as } t \rightarrow \infty, \delta' = -\frac{1}{8} e^{-t/8} < 0,$$

$$\delta'' = \frac{1}{64} e^{-t/8} > 0, \delta' = \frac{b'}{2b} \delta = -\frac{1}{8} e^{-t/8} < 0,$$

$$\int_{t^*}^{\infty} \left((q_1 + q_2) \delta - \frac{a(\tau) b \delta'^2}{4B\tau' \delta} \right) dt = \int_{t^*}^{\infty} \left(\frac{1}{8} e^{t/8} (1 + 2e^{-\pi}) - \frac{e^{-\pi/2} e^{t/8}}{256B(t-\pi)} \right) dt = \infty,$$

$$\int_{t^*}^{\infty} \frac{2}{a\delta} dt = \int_{t^*}^{\infty} t e^{3t/8} dt = \infty \text{ and } \int_{t^*}^{\infty} \frac{1}{a\delta} \int_{t^*}^s \int_{t^*}^u (q_1 + q_2) \frac{\delta}{b} du ds dt =$$

$$\int_{t^*}^{\infty} e^{3t/8} \int_{t^*}^s \frac{1}{8} e^{3u/8} (1 + 2e^{-\pi}) du ds dt = \infty.$$

All the conditions of Theorem 2 are satisfied, then the conclusion of the theorem holds.

In fact, $y(t) = e^{-t}$ is a solution of equation (3.39).

REFERENCES

- [1] Ahmed, S. and Benharbit, A., Some oscillation properties of third order linear homogeneous differential equation, Ann. Polon. Math., 31, 1975, 15-21.
- [2] Atkinson, F.V., On second order nonlinear oscillation, Pacific J. Math., 5, 1955, 643-647.
- [3] Azbelev, N.V. and Caljuk, Z.B., On the question of the differential equations, Mat. Sb. [N.S.], 51 (93), 1960, 475-486.
- [4] Barrett, J.H., Canonical forms for third order linear differential equations, Ann. Mat. Pure Appl., 65, 1964, 253-274.
- [5] Bellman, R., *Stability Theory of Differential Equations*, McGraw Hill, New York, 1953.
- [6] Bhatia, N.P., Some oscillation theorems for second order differential equations, J. Math. Anal. Appl., 15, 1966, 442-446.
- [7] Dahiya, R.S. and Singh, B., On the oscillation of a second order delay equation, Journal of Mathematical Analysis and Applications, 48, 1974, No. 2, 610-617.
- [8] Erbe, L., Integral comparison theorems for third order linear differential equations, Pacific J. Math., 85, 1979, 35-38.
- [9] Grove, E.A., Ladas, G., Schinas, J., Sufficient conditions for the oscillation of delay and neutral delay equations, Canad. Math. Bull., 31, 1988, 459-466.
- [10] Hanan, M., Oscillation criteria for third order linear differential equations, Pacific J. Math., 11, 1961, 919-944.
- [11] Kiguradz, I.T., On the oscillation of solutions of the equation $d^m u/dt^m + a(t) |u|^n \operatorname{sign} u = 0$, Mat. Sab., 65, 1964, 172-187.
- [12] Lade, G.S., Lakshmikantham, and Zhang, B.G., *Oscillation Theory of Differential Equations with Deviating Arguments*, Marcel Dekker, Inc., New York, 1987.
- [13] Sree Hari Rao, V. and Dahiya, R.S., Properties of solutions of a class of third order linear differential equations, Periodic Math. Hungary, 20, 1989, 177-184.
- [14] Waltman, P., Oscillation criteria for third order nonlinear differential equations, Pacific J. Math., 18, 1966, 385-389.

8 NUMERICAL TECHNIQUES FOR SOLVING A BIHARMONIC EQUATION IN A SECTORIAL REGION

Elias Deeba

University of Houston-Downtown
Houston, TX 77002

and

Suheil A. Khuri
American University of Sharjah
Sharjah, UAE

and

Shishen Xie
University of Houston-Downtown
Houston, TX 77002

1. INTRODUCTION

The behavior of fluid in a cavity when subjected to movement of one of its surrounding walls is modeled by a version of the Navier-Stokes equations. The problem to be discussed in this paper can be described as follows: A two-dimensional sectorial cavity

$$G = \{(r, \theta) | 1 < r < a, -\alpha < \theta < \alpha\}$$

is filled with incompressible fluid. The cavity is covered on the top with a flat plate. The steady plane motion is generated by the uniform translation of the plate with a constant unit velocity in the direction of increasing r . In the absence of inertial terms (i.e., for zero Reynolds numbers) we have creeping flow or well-known as Stokes flow which can be obtained from a stream function satisfying the biharmonic equation,

$$\nabla^4 u = 0,$$

(see [1], [2]).

The purpose of this paper is to seek a numerical solution to the biharmonic boundary value problem governing Stokes flow in a sectorial cavity. In Section 2, we will describe the boundary value problem in details. In Sections 3 and 4, the boundary value problem is decomposed into a coupled system of Poisson equations, and the convergence for an iterative scheme for the coupled system is discussed. A numerical algorithm is developed in Section 5 to find the stream function of the flow by solving the resulting linear system iteratively. Finally, based on the numerical data from the algorithm, contour lines of the stream function reflecting the eddies in the cavity are shown in Section 6.

2. A BIHARMONIC BOUNDARY VALUE PROBLEM GOVERNING STOKES FLOW IN A SECTORIAL CAVITY

A two-dimensional sectorial cavity

$$G = \{(r, \theta) | 1 < r < a, -\alpha < \theta < \alpha\}$$

is filled with incompressible fluid. Its boundary Γ consists of the top edge $\theta = \alpha$, the bottom edge $\theta = -\alpha$, the inner arc $r = 1$ and the outer arc $r = a$. See Figure 1.

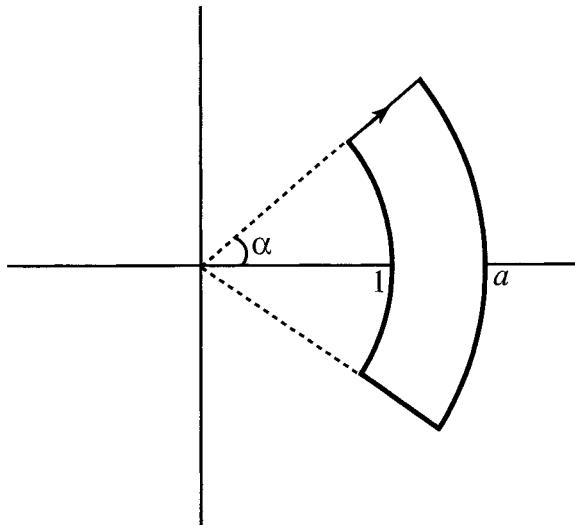


Figure 1. Fluid fills the sectorial region G .

The top edge $\theta = \alpha (1 \leq r \leq a)$ is a flat plate, which translates with a constant unit velocity in the radial direction, thus setting the fluid into motion. In the absence of inertial terms (i.e., Reynolds numbers = 0) there is a steady creeping flow which can be obtained from a stream function u .

The derivatives of u give the peripheral velocity component

$$v_\theta = \frac{\partial u}{\partial r},$$

and the radial velocity component

$$v_r = -\frac{1}{r} \frac{\partial u}{\partial \theta}.$$

The Stokes flow in G is governed by the biharmonic equation

$$\nabla^4 u = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right)^2 u(r, \theta) = 0, \quad (2.1)$$

with the boundary conditions posed on the inner and outer arcs as

$$u(1, \theta) = u(a, \theta) = \frac{\partial u}{\partial r}(1, \theta) = \frac{\partial u}{\partial r}(a, \theta) = 0, \quad (2.2)$$

and on the top and bottom edges as

$$u(r, \pm\alpha) = 0, \quad -\frac{1}{r} \frac{\partial u}{\partial \theta}(r, \alpha) = 1, \quad \frac{\partial u}{\partial \theta}(r, -\alpha) = 0. \quad (2.3)$$

Notice that ∇^2 is the conventional notation for the Laplace operator (in polar coordinates for this particular case) hence, $\nabla^4 = (\nabla^2)^2$ denotes the biharmonic operator.

3 . AN EQUIVALENT FORMULATION OF THE BIHARMONIC PROBLEM

In this section, we will discuss the coupled equation approach to solve a biharmonic boundary value problem. In this approach, the biharmonic problem is reduced to a system of coupled Poisson equations with appropriate boundary conditions.

Let G denote a bounded domain in the plane with boundary Γ . An equivalent formulation of the biharmonic problem for the domain G

$$\begin{aligned} \nabla^4 u(p) &= \phi(p), & p \in G \\ u(p) &= f(p), & p \in \Gamma \\ \frac{\partial u(p)}{\partial n} &= g(p) & p \in \Gamma \end{aligned} \quad (3.1)$$

is the coupled system, [5]:

$$\begin{aligned} \nabla^2 u(p) &= w(p), & p \in G \\ u(p) &= f(p), & p \in \Gamma \end{aligned} \quad (3.2)$$

where ∇^2 is the Laplace operator in polar coordinates, and

$$\begin{aligned} \nabla^2 w(p) &= \phi(p), & p \in G \\ w(p) &= \nabla^2 u(p) - c \left[\frac{\partial u(p)}{\partial n} - g(p) \right], & p \in \Gamma \end{aligned} \quad (3.3)$$

where c is an arbitrary constant.

The following theorem confirms the equivalence relation between problem (3.1) and the coupled system (3.2) and (3.3), (see [3]).

Theorem 3.1 Let $u \in C^4(G) \cap C^1(\bar{G})$, and assume that u has piecewise continuous second derivatives on Γ .

1. If u is a solution of (3.1), then u is a solution of (3.2) and (3.3) for every constant c .
2. If (u, w) is a solution of (3.2) and (3.3) for any $c \neq 0$, then u is a solution of (3.1).

To solve the coupled system, we need to introduce an iterative scheme, which converges for an essentially arbitrary starting function $w^{(0)}$.

Let $w^{(0)}$ be a given value such that

$$\nabla^2 w^{(0)}(p) = \phi(p), \quad p \in G.$$

Then the sequences $\{u^{(k)}\}$ and $\{w^{(k)}\}$ are defined by an iterative scheme

$$\begin{aligned} \nabla^2 u^{(k)}(p) &= w^{(k-1)}(p), & p \in G \\ u^{(k)}(p) &= f(p), & p \in \Gamma \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} \nabla^2 w^{(k)}(p) &= \phi(p), & p \in G \\ w^{(k)}(p) &= \nabla^2 u^{(k)}(p) - c \left[\frac{\partial u^{(k)}(p)}{\partial n} - g(p) \right], & p \in \Gamma \end{aligned} \tag{3.5}$$

$k = 1, 2, \dots$. This scheme alternately determines $u^{(1)}, w^{(1)}, u^{(2)}, w^{(2)}, \dots$, by solving Poisson boundary value problems (3.4) and (3.5) recursively.

The convergence conditions of the iterative scheme is stated in the following theorem (see [3]).

Theorem 3.2 The iterative scheme in (3.4) and (3.5) converges for arbitrary $w^{(0)}$ such that $\nabla^2 w^{(0)} = \phi$ if and only if $0 < c < 2\lambda_1$, where λ_1 is the smallest eigenvalue of the Dirichlet eigenvalue problem

$$\nabla^4 u(p) = 0, \quad p \in G$$

$$u(p) = 0, \quad p \in \Gamma.$$

$$\nabla^2 u(p) = \lambda_1 \frac{\partial u(p)}{\partial n}, \quad p \in \Gamma$$

We are now in a position to set up the iterative scheme for Problems (2.1), (2.2), and (2.3) via the coupled systems (3.4) and (3.5).

As before, we denote G to be the sectorial cavity $\{(r, \theta) | 1 < r < a, -\alpha < \theta < \alpha\}$, and Γ the boundary of $G : r = 1, r = a, \theta = -\alpha$ and $\theta = \alpha$.

Let $w^{(0)}$ be given such that

$$\nabla^2 w^{(0)}(r, \theta) = 0, \quad (r, \theta) \in G,$$

since $\phi = 0$ in our problem. In the algorithm, we simply set $w^{(0)} = 1$.

Substituting the boundary values (2.2) and (2.3) into (3.4) and (3.5), we can obtain the coupled system

$$\begin{aligned} \nabla^2 u^{(k)}(r, \theta) &= w^{(k-1)}(r, \theta), & (r, \theta) \in G \\ u^{(k)}(r, \theta) &= 0, & (r, \theta) \in \Gamma \end{aligned} \quad (3.6)$$

and

$$\begin{aligned} \nabla^2 w^{(k)}(r, \theta) &= 0, & (r, \theta) \in G \\ w^{(k)}(r, \theta) &= \nabla^2 u^{(k)}(r, \theta) - c \left[\frac{\partial u^{(k)}(r, \theta)}{\partial n} - g(r, \theta) \right], & (r, \theta) \in \Gamma \end{aligned} \quad (3.7)$$

where

$$\begin{aligned} \frac{\partial u^{(k)}}{\partial n} &= \frac{\partial u^{(k)}}{\partial r}, & \text{for } r = 1 \text{ and } r = a, \\ \frac{\partial u^{(k)}}{\partial n} &= \frac{\partial u^{(k)}}{\partial \theta}, & \text{for } \theta = \pm\alpha, \end{aligned}$$

and

$$g(r, \theta) = \begin{cases} 0, & r = 1, r = a, \text{ or } \theta = -\alpha, \\ -r, & \theta = \alpha. \end{cases}$$

4. AVERAGING TO PRODUCE A CONVERGENT SCHEME

To improve the convergence, we also introduce an averaging scheme into our numerical algorithm.

Let $0 \leq \varepsilon \leq 1$ and $0 \leq \delta \leq 1$ be two constants, and let $u^{(0)}$ and $w^{(0)}$ be given such that $\nabla^4 u^{(0)} = \nabla^2 w^{(0)} = \phi$ in G and $u^{(0)} = f$ on Γ . For $k = 1, 2, 3, \dots$,

$$\begin{aligned} \nabla^2 \bar{u}^{(k)}(p) &= w^{(k-1)}(p), & p \in G \\ \bar{u}^{(k)}(p) &= f(p), & p \in \Gamma \\ u^{(k)}(p) &= \varepsilon u^{(k-1)}(p) + (1 - \varepsilon) \bar{u}^{(k)}(p), & p \in G \end{aligned} \quad (4.1)$$

and

$$\begin{aligned}\nabla^2 \bar{w}^{(k)}(p) &= \phi(p), & p \in G \\ \bar{w}^{(k)}(p) &= \nabla^2 u^{(k)}(p) - c \left[\frac{\partial u^{(k)}(p)}{\partial n} - g(p) \right], & p \in \Gamma. \\ w^{(k)}(p) &= \delta w^{(k-1)}(p) + (1-\delta) \bar{w}^{(k)}(p), & p \in G\end{aligned}\quad (4.2)$$

We notice that if $\varepsilon = \delta = 0$, (4.1) and (4.2) are the same as the original coupled system (3.4) and (3.5). For properly chosen ε and δ , the scheme (4.1) and (4.2) converges. Furthermore, the averaging scheme speeds up the convergence, (see [3]).

5. THE FINITE DIFFERENCE SCHEME

In this section, we introduce a finite difference scheme to solve the coupled systems (3.6) and (3.7).

We first superimpose a grid system over the domain $G \cup \Gamma = \{(r, \theta) | 1 \leq r \leq a, -\alpha \leq \theta \leq \alpha\}$ with mesh sizes

$$h_r = \frac{a-1}{I}, \text{ and } h_\theta = \frac{\alpha - (-\alpha)}{J} = \frac{2\alpha}{J},$$

where I and J are positive integers representing the total number of mesh points in r and θ directions, respectively.

Without loss of generality, we assume that $a = 3$ and $\alpha = \frac{\pi}{4}$, [2]. Then the

two-dimension sectorial cavity is specified as

$$G = \left\{ (r, \theta) \mid 1 < r < 3, -\frac{\pi}{4} < \theta < \frac{\pi}{4} \right\}.$$

Let G_h denote the set of mesh points on G , and Γ_h the set of mesh points on the boundary Γ . Each mesh point $(r_i, \theta_j) \in G_h$ can be computed by $r_i = 1 + ih_r$, for $i = 1, 2, \dots, I-1$, and $\theta_j = -\frac{\pi}{4} + j h_\theta$, for $j = 1, 2, \dots, J-1$.

Using these notations we can discretized Problems (3.6) and (3.7) to be:

$$\begin{aligned}\nabla_h^2 u^{(k)}(r_i, \theta_j) &= w^{(k-1)}(r_i, \theta_j), & (r_i, \theta_j) \in G_h \\ u^{(k)}(r_i, \theta_j) &= 0, & (r_i, \theta_j) \in \Gamma_h\end{aligned}\quad (5.1)$$

and

$$\begin{aligned}\nabla_h^2 w^{(k)}(r_i, \theta_j) &= 0, & (r_i, \theta_j) \in G_h \\ w^{(k)}(r_i, \theta_j) &= \nabla_h^2 u^{(k)}(r_i, \theta_j) - c \left[\frac{\partial u^{(k)}(r_i, \theta_j)}{\partial n} - g(r_i, \theta_j) \right], & (r_i, \theta_j) \in \Gamma_h\end{aligned}\quad (5.2)$$

with $k = 1, 2, \dots$.

To shorten the notation, we use subscripts $i j$ to represent (r_i, θ_j) in the future expressions. For example, $w_{ij}^{(k)}$ is the shortened notation for $w^{(k)}(r_i, \theta_j)$.

The finite difference scheme for the Laplace operator ∇_h^2 in polar coordinates can be written as

$$\nabla_h^2 u_{ij}^{(k)} = \frac{1}{r_i} \frac{r_{\frac{i+1}{2}} u_{i+1,j}^{(k)} - \left(r_{\frac{i+1}{2}} + r_{\frac{i-1}{2}} \right) u_{ij}^{(k)} + r_{\frac{i-1}{2}} u_{i-1,j}^{(k)}}{h_r^2} + \frac{1}{r_i^2} \frac{u_{i,j+1}^{(k)} - 2u_{ij}^{(k)} + u_{i,j-1}^{(k)}}{h_\theta^2}. \quad (5.3)$$

Let

$$a_i = \frac{r_{\frac{i+1}{2}}}{r_i h_r^2}, \quad b_i = \frac{r_{\frac{i-1}{2}}}{r_i h_r^2}, \quad c_i = d_i = \frac{1}{r_i h_\theta^2},$$

and

$$e_i = -\left(\frac{r_{\frac{i+1}{2}} + r_{\frac{i-1}{2}}}{r_i h_r^2} + \frac{2}{r_i^2 h_\theta^2} \right) = -2 \left(\frac{1}{h_r^2} + \frac{1}{r_i^2 h_\theta^2} \right),$$

$$\text{since } r_{\frac{i+1}{2}} + r_{\frac{i-1}{2}} = \left(r_i + \frac{h_r}{2} \right) + \left(r_i - \frac{h_r}{2} \right) = 2r_i.$$

Using these notations and the difference scheme (5.3), we can write the equations in (5.1) and (5.2) as

$$\nabla_h^2 u_{ij}^{(k)} = a_i u_{i+1,j}^{(k)} + b_i u_{i-1,j}^{(k)} + c_i u_{i,j+1}^{(k)} + d_i u_{i,j-1}^{(k)} + e_i u_{ij}^{(k)} = w_{ij}^{(k-1)}, \quad (5.4)$$

and,

$$\nabla_h^2 w_{ij}^{(k)} = a_i w_{i+1,j}^{(k)} + b_i w_{i-1,j}^{(k)} + c_i w_{i,j+1}^{(k)} + d_i w_{i,j-1}^{(k)} + e_i w_{ij}^{(k)} = 0, \quad (5.5)$$

respectively, for $1 \leq i \leq I-1$ and $1 \leq j \leq J-1$.

In equations (5.4) and (5.5), some terms of $u_{ij}^{(k)}$ and $w_{ij}^{(k)}$ may be known (at boundary points) while others are unknown (at the interior points). This means that we shall solve a nonhomogeneous system of linear equations with all known values being transferred to the right. Iterative procedures, such as successive overrelaxation algorithm (SOR), are quite effective for solving the system. To use SOR, we write equations (5.4) and (5.5) in the form

$$u_{ij}^{(k)} = \frac{1}{e_i} \left(w_{ij}^{(k-1)} - a_i u_{i+1,j}^{(k)} - b_i u_{i-1,j}^{(k)} - c_i u_{i,j+1}^{(k)} - d_i u_{i,j-1}^{(k)} \right) \quad (5.6)$$

and

$$w_{ij}^{(k)} = \frac{1}{e_i} \left(-a_i w_{i+1,j}^{(k)} - b_i w_{i-1,j}^{(k)} - c_i w_{i,j+1}^{(k)} - d_i w_{i,j-1}^{(k)} \right), \quad (5.7)$$

respectively.

Next we need to discretize the boundary conditions in Problems (5.1) and (5.2). The boundary condition in (5.1) is trivial: $u_{ij}^{(k)} = 0$ or more explicitly

$$\begin{aligned} u_{0,j}^{(k)} &= 0, \quad u_{I,j}^{(k)} = 0 \quad \text{for } j = 1, 2, \dots, J \\ u_{i,0}^{(k)} &= 0, \quad u_{i,I}^{(k)} = 0 \quad \text{for } i = 1, 2, \dots, I \end{aligned} \quad . \quad (5.8)$$

On the other hand, the discretization of the boundary condition $w_{ij}^{(k)}$ in (5.2) requires lengthy computation. The partial derivative $\frac{\partial u^{(k)}(r_i, \theta_j)}{\partial n}$ for $(r_i, \theta_j) \in \Gamma$ (that is, $i = 0, i = I, j = 0$, and $j = J$) are discretized using the central difference formula. We explain this process by showing the work on the top edge $\theta = \frac{\pi}{4}$.

On the top edge $\left(\theta = \frac{\pi}{4}\right)$, the index j equals to J , and i ranges from 1 to $I - 1$.

The boundary condition (5.8) determines that $u_{i,J}^{(k)} = u^{(k)}\left(r_i, \frac{\pi}{4}\right) = 0$, and

$$\frac{\partial u}{\partial n} \Big|_{\theta=\frac{\pi}{4}} = \frac{\partial u}{\partial \theta} \Big|_{\theta=\frac{\pi}{4}} = -r.$$

Therefore, the function g in Equation (5.2) satisfies

$$g_{i,J} = -r_i. \quad (5.9)$$

Applying these results and the central difference formula to the boundary condition (5.2), we obtain

$$\begin{aligned} w_{i,J}^{(k)} &= \frac{u_{i,J+1}^{(k)} + u_{i,J-1}^{(k)}}{r_i^2 h_\theta^2} - \frac{c}{2h_\theta} \left(u_{i,J+1}^{(k)} - u_{i,J-1}^{(k)} \right) + cg_{i,J} \\ &= \frac{2u_{i,J-1}^{(k)}}{r_i^2 h_\theta^2} + \left(\frac{2}{r_i^2 h_\theta^2} - \frac{c}{h_\theta} \right) \frac{u_{i,J+1}^{(k)} - u_{i,J-1}^{(k)}}{2} + cg_{i,J}. \end{aligned} \quad (5.10)$$

Since $u_{i,J+1}^{(k)}$ is not located within the region $G \cup \Gamma$, it is necessary to express it in terms of the values of u at interior grids.

If $u_{i,J+1}^{(k)}$ and $u_{i,J-1}^{(k)}$ have Taylor expansions

$$u_{i,J+1}^{(k)} = u_{i,J}^{(k)} + h_\theta \frac{\partial u}{\partial \theta} \Big|_{ij} + \frac{h_\theta^2}{2} \frac{\partial^2 u}{\partial \theta^2} \Big|_{ij} + O(h_\theta^3) \quad (5.11)$$

$$u_{i,J-1}^{(k)} = u_{i,J}^{(k)} - h_\theta \frac{\partial u}{\partial \theta} \Big|_{ij} + \frac{h_\theta^2}{2} \frac{\partial^2 u}{\partial \theta^2} \Big|_{ij} + O(h_\theta^3), \quad (5.12)$$

where $\frac{\partial u}{\partial \theta} \Big|_{i,j} = \frac{\partial u(x_i, t_j)}{\partial \theta}$.

Subtracting (5.12) from (5.11) yields

$$u_{i,J+1}^{(k)} - u_{i,J-1}^{(k)} = 2h_\theta \frac{\partial u}{\partial \theta} \Big|_{ij} + O(h_\theta^2),$$

and hence,

$$\begin{aligned} \frac{u_{i,J+1}^{(k)} + u_{i,J-1}^{(k)}}{2} &\approx 2h_\theta \left. \frac{\partial u}{\partial \theta} \right|_{ij} \\ &= h_\theta \cdot \frac{u_{i,J}^{(k)} - u_{i,J-1}^{(k)}}{h_\theta} = -u_{i,J-1}^{(k)} \end{aligned} \quad (5.13)$$

since $u_{i,J}^{(k)} = 0$ as the given boundary condition.

We substitute (5.13) into the last expression of (5.10), and simplify using the relation (5.9) to obtain the discretized boundary condition on the top edge $\theta = \frac{\pi}{4}$,

$$\begin{aligned} w_{iJ}^{(k)} &= \frac{2u_{i,J-1}^{(k)}}{r_i^2 h_\theta^2} + \left(\frac{2}{r_i^2 h_\theta^2} - \frac{c}{h_\theta} \right) (-u_{i,J-1}^{(k)}) + cg_{i,J} \\ &= \frac{c}{h_\theta} u_{i,J-1}^{(k)} - cr_i = c \left(\frac{u_{i,J-1}^{(k)}}{h_\theta} - r_i \right). \end{aligned} \quad (5.14)$$

Similarly, we can derive the boundary conditions on the other three edges,

$$\begin{aligned} w_{i0}^{(k)} &= \frac{cu_{i1}^{(k)}}{h_\theta}, \quad \text{for } \theta = -\frac{\pi}{4}; \\ w_{0j}^{(k)} &= \left(\frac{1}{r_0 h_r} + \frac{c}{h_r} \right) u_{1j}^{(k)}, \quad \text{for } r = 1; \\ w_{Ij}^{(k)} &= \left(\frac{-1}{r_I h_r} + \frac{c}{h_r} \right) u_{Ij}^{(k)}, \quad \text{for } r = 3. \end{aligned} \quad (5.15)$$

By the analysis above, we have completed the discretization process. In the next section, we will briefly discuss the execution of the numerical algorithm.

6. THE NUMERICAL COMPUTATION

In this section, we discuss the numerical scheme that is developed to solve the Stokes flow problem (5.1) and (5.2). The C++ program that generates the numerical data of u at the mesh points is available upon request. In the computation, we use equal number of mesh point in the r direction as that in θ direction, that is, $I = J$. We let $M = I = J$. The constant c in the algorithm is equal to 5.0.

From Theorem 3.2, we understand that the initial values of $w_{ij}^{(0)}$ can be arbitrary as long as it satisfies the equation

$$\nabla^2 w^{(0)}(r, \theta) = 0, \quad (r, \theta) \in G.$$

As a natural and simple choice, we start with the initial values of $w^{(0)}(r_i, \theta_j) = 1$ for $i, j = 1, 2, \dots, M-1$.

We next apply (5.4) with $k = 1$ to each interior mesh point $(r_i, \theta_j) \in G_h$ ($i, j = 1, 2, \dots, M-1$), along with the boundary conditions in (5.8), to

form a system of $(M - 1)^2$ linear equations. Writing the system of linear equations in matrix form, we use the Successive Overrelaxation (SOR) method (see [4]) to solve the resulting (sparse) matrix equation. To accelerate the convergence, we also incorporate the averaging scheme (4.1) in our algorithm with $\varepsilon = \delta = 0.5$.

After the matrix equation is solved, we substitute the solutions $u_{ij}^{(1)}$ into the boundary conditions (5.14) and (5.15) to find the values of $w^{(1)}$ at the grid points along the boundary. Inserting the boundary values of $w^{(1)}$ into the finite difference equation (5.5), we establish a matrix equation, which can be solved by SOR methods again. The solutions $w_{ij}^{(1)}$ of the equation are used on the right hand side of equation (5.4) to determine $u_{ij}^{(2)}$. We average the solution again to improve the convergence.

Repeating these steps, we can find $w_{ij}^{(2)}, u_{ij}^{(3)}, w_{ij}^{(3)}, \dots, u_{ij}^{(n-1)}, w_{ij}^{(n-1)}, u_{ij}^{(n)}$ ($i, j = 1, 2, \dots, M - 1$), until the desired accuracy is reached.

Before closing this section, two more things should be mentioned regarding the algorithm. The first is about the order in which the mesh points are processed. An effective technique is to divide the mesh into odd and even meshes, like the white and black square on a chess board. This process takes advantage of Equations (5.6) and (5.7) which implies that the odd points depend only on the even mesh values and vice versa. This way we can carry out one half sweep updating the odd values, and then another half sweep to update the even points with the new odd values.

The second is about the use of the Chebyshev acceleration, [4], in the algorithm. It is not unusual for the error to grow before convergence is achieved. With the Chebyshev acceleration the norm of the error always decreases with each iteration. Therefore, the Chebyshev acceleration is used in the algorithm to reduce the total number of iterations required.

This algorithm is implemented with Borland C++ first on 5×5 ($M = 5$) mesh, and then the mesh size is gradually increased. The largest mesh size we implement is 45×45 . Figure 2 shows the graphs generated by Maple when it is fed with the numerical data obtained from various runs of the program. We can see contour lines of the stream function reflecting the eddies in the cavity.

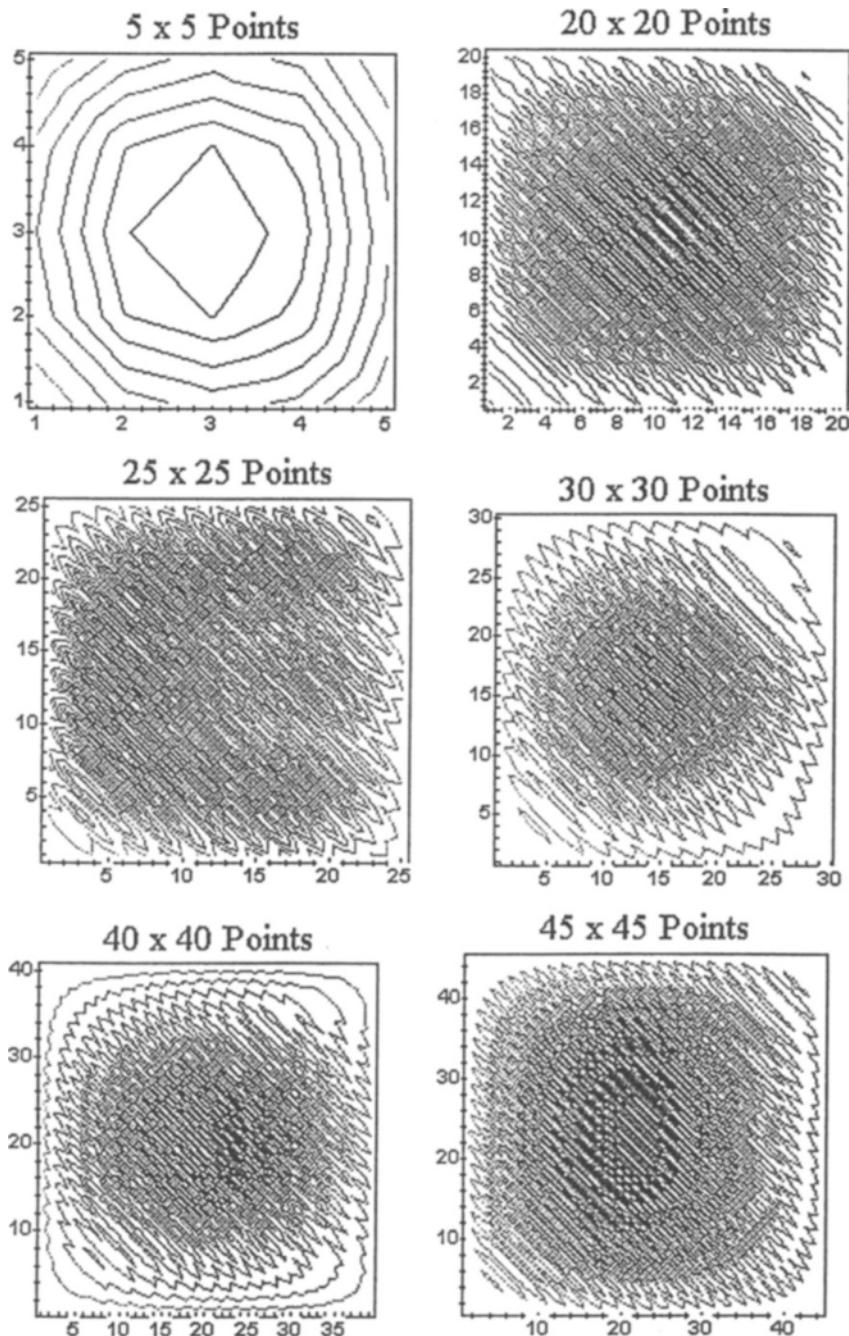


Figure 2. The eddies in the cavity.

REFERENCES

- [1] Burggraf, O.R., Analytical and numerical studies of the structure of the steady separated flows, J. Fluid Mech., Vol. 24, 1966, 113-151.
- [2] Khuri, S.A., Biorthogonal series solution of Stokes flow problem in sectorial regions, SIAM J. Appl. Math., vol. 56, No. 1, 1996, 19-39.
- [3] McLaurin, J.W., A general coupled equation approach for solving the biharmonic boundary value problem, SIAM J. Numer. Anal., Vol. 11, 1974, 14-33.
- [4] Press, W.H. and et. al., *Numerical Recipes – The Art of Scientific Computing*, Second Edition, Cambridge University Press, 1992.
- [5] Smith, J., The coupled equation approach to the numerical solution of the biharmonic equation, Part I, II, SIAM J. Numer. Anal., Vol. 5, 1968, 323-339; Vol. 7, 1970, 104-111.

9 THE SEAMOUNT ON A SLOPING SEABED PROBLEM

Robert P. Gilbert
and

Miao Ou

Department of Mathematical Sciences
University of Delaware
Newark, DE 19716

and

Yongzhi S. Xu
Department of Mathematics
University of Tennessee
Chatanooga, TN

ABSTRACT

This paper deals with an inverse acoustics problem in the ocean. The problem we investigate is the location of a non-homogeneity caused by a sea-mount or some object lying on a sloping seabed. This problem is solved by constructing an acoustic Green's function for the wedge. This is done by using the method of images. The inversion procedure is motivated by our earlier work on the seamount problem for a shallow ocean of uniform depth [17].

1. FORMULATION OF THE DIRECT PROBLEM

Buchanan, Gilbert, Wirgin, and Xu have investigated, in detail, inverse problems for uniform, finite depth oceans with a completely reflecting basement. These results were reported on in a sequence of papers [2, 3, 12, 13, 16, 9, 10]. The methodology used in these papers was first to obtain an operator which produced the far field from an incident ray scattered off the target. Then the inverse problem was formulated as an extremal problem. With a suitable fundamental singular solution

for the wedge domain, this method is applicable to an ocean with an inclined seabed.

Across-section in the $x - y$ plane of a buried object on a sloping seabed is visualized in Figure 1. The z axis is assumed to lie perpendicular to the page. It is assumed that the basement is completely reflecting. Then the acoustic pressure, generated by a point source at the given location $\vec{x} := (x_0, y_0, z_0)$, satisfies

$$\Delta p + k^2 p = -\delta(\vec{x} - \vec{x}_0), \quad \vec{x} \in R_h^3 \setminus \bar{\Omega}, \quad (1.1)$$

$$p = 0 \quad \text{at } y = 0 \quad (1.2)$$

$$\frac{\partial p}{\partial z} = 0 \quad \text{at } x = y \tan(\theta_0), \quad (x, y) \notin \mathcal{M} \quad (1.3)$$

$$\frac{\partial p}{\partial v} = 0 \quad \text{on } \mathcal{M}, \quad (1.4)$$

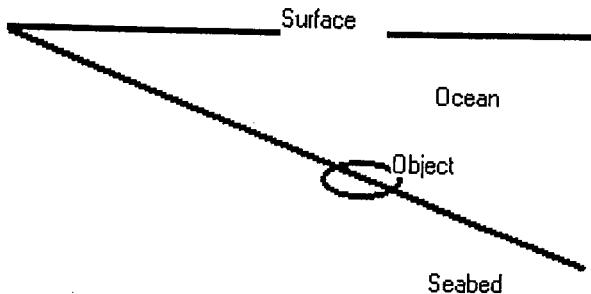


Figure 1. Cross-section of an object partially buried on a inclined seabed.

And the Sommerfeld out-going radiation condition. The wedge region we consider is

$$R_{\theta_0}^3 = \{(x, y, z) : 0 \leq x < \infty, 0 \leq y \leq x \tan(\theta_0), |z| < \infty\}$$

or in cylindrical coordinates as

$$\{(r, \theta, z) : 0 \leq r < \infty, 0 \leq \theta \leq \theta_0, |z| < \infty\};$$

Ω is the sea-mount, and \mathcal{M} is the surface of the sea-mount,

$$\mathcal{M} := \{(x, y, z) : y := f(x, z), \text{ where } (x, z) \in D_0\}.$$

Here D_0 is the projection of the sea-mount D onto the plane $y = 0$. The sea-mount is denoted by $D := \{(x, y, z) : x \tan(\theta_0) < y < f(x, z), \text{ where } (x, z) \in D_0\}.$

For an ocean with a sloping seabed without a sea-mount, the solution to (1.1), (1.2), (1.3), and (1.4) is the Green's function for the Helmholtz equation in $R_{\theta_0}^3$.

Buchingham [4, 5] has constructed this Green's function using integral transforms;

however, for our purposes, we need to exhibit clearly the singular behavior at the source point. To this end we construct the Green's function by the method of images. Suppressing the z variable, which is perpendicular to the wedge, we begin with a source at the point (x_0, y_0) . There are two sequences of images that we obtain. The first sequence begins by a reflection through the line $x = y \tan(\theta_0)$ and then proceeds with a reflection through $y = 0$ another through $x = y \tan(\theta_0)$, etc. We designate these source points as $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots$. The second sequence begins with a reflection through $y = 0$, the next through $x = y \tan(\theta_0)$, the third reflection through $y = 0$, etc. We designate these source points as $(x_0, y_0), (\tilde{x}_1, \tilde{y}_1), (\tilde{x}_2, \tilde{y}_2), \dots$. We construct these image points by successive reflections and rotations as follows. Let $T_{\theta_0} x \rightarrow x'$ be the transformation from the (x, y) coordinate frame to the (x', y') coordinate frame. It is represented by the rotation through the angle θ_0

$$T_{\theta_0} \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} \cos(\theta_0) & -\sin(\theta_0) \\ \sin(\theta_0) & \cos(\theta_0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

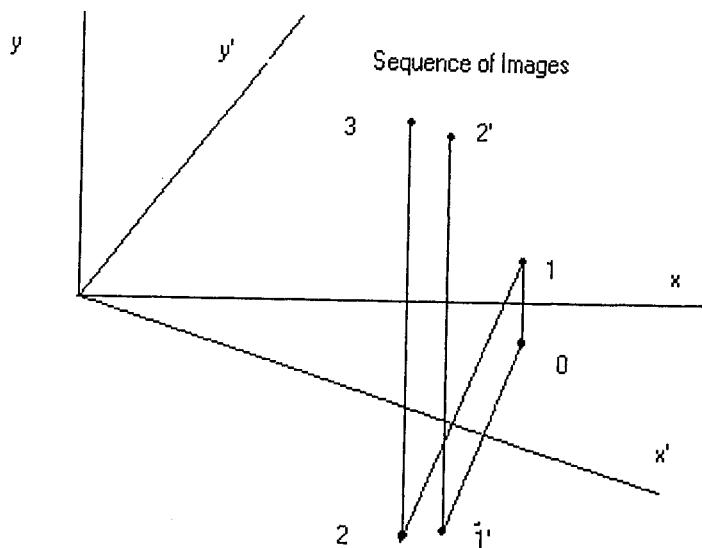


Figure 2. The sequence of image points to compute the Green's function.

The inverse of $\mathbf{T}_{\theta_0} \vec{x}$ is the transformation from the (x', y') coordinate frame to the (x, y) coordinate frame. It is represented by $\mathbf{T}_{-\theta_0} \vec{x}' \rightarrow \vec{x}$, that is the rotation through the angle $-\theta_0$

$$\mathbf{T}' \begin{pmatrix} x' \\ y' \end{pmatrix} := \begin{pmatrix} \cos(\theta_0) & \sin(\theta_0) \\ -\sin(\theta_0) & \cos(\theta_0) \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

Let \mathbf{E} be the reflection matrix

$$\mathbf{E} := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The image point (x_1, y_1) is then computed by the scheme

$$\begin{aligned} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} &= \begin{pmatrix} \cos(\theta_0) & \sin(\theta_0) \\ -\sin(\theta_0) & \cos(\theta_0) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos(\theta_0) & -\sin(\theta_0) \\ \sin(\theta_0) & \cos(\theta_0) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= \begin{pmatrix} \cos(2\theta_0) & -\sin(2\theta_0) \\ -\sin(2\theta_0) & \cos(2\theta_0) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \mathbf{E} \mathbf{T}_{2\theta_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}. \end{aligned}$$

In a similar manner we compute (x_2, y_2) to be

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \mathbf{E} \mathbf{E} \mathbf{T}_{2\theta_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \mathbf{T}_{2\theta_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

The image point (x_3, y_3) is also found by this procedure to be

$$\begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = \mathbf{E} \mathbf{T}_{4\theta_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

It may be shown inductively that the image points (x_n, y_n) , $n = 1, 2, 3, \dots$ are given for even values $n = 2k$ by

$$\begin{pmatrix} x_{2k} \\ y_{2k} \end{pmatrix} = \mathbf{T}_{2k\theta_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix};$$

whereas, for odd values $n = 2k+1$ they are

$$\begin{pmatrix} x_{2k+1} \\ y_{2k+1} \end{pmatrix} = \mathbf{E} \mathbf{T}_{(2k+2)\theta_0} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

We next consider the image points of the sequence $(\tilde{x}_n, \tilde{y}_n)$. It may be shown that these points are given for $n = 2k$ by

$$\begin{pmatrix} \tilde{x}_{2k} \\ \tilde{y}_{2k} \end{pmatrix} = \begin{pmatrix} \cos(2k\theta_0) & \sin(2k\theta_0) \\ -\sin(2k\theta_0) & \cos(2k\theta_0) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix},$$

and for $n = 2k+1$

$$\begin{pmatrix} \tilde{x}_{2k+1} \\ \tilde{y}_{2k+1} \end{pmatrix} = \mathbf{E} \begin{pmatrix} \tilde{x}_{2k} \\ \tilde{y}_{2k} \end{pmatrix} = \begin{pmatrix} \cos(2k\theta_0) & \sin(2k\theta_0) \\ \sin(2k\theta_0) & -\cos(2k\theta_0) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Consequently, the Green's function has the representation

$$G(\vec{x}, \vec{x}_0) + E(\vec{x}, \vec{x}_0) + \sum_{n=0}^{\infty} (-1)^n [E(\vec{x}, \vec{x}_n) - E(\vec{x}, \vec{x}_n)], \quad (1.5)$$

where

$$\vec{x} := \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \text{and} \quad \vec{\tilde{x}} := \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix}$$

and

$$E(\vec{x}, \vec{y}) := \frac{\exp\{ik(|\vec{x} - \vec{y}|)\}}{4\pi|\vec{x} - \vec{y}|}. \quad (1.6)$$

The solution of problem (1.4) may be represented using Green's formula as

$$p(\vec{x}, \vec{x}_0) = G(\vec{x}, \vec{x}_0) + \int_{\mathcal{M}} \left\{ G(\vec{x}, \vec{y}) \frac{\partial p_{sc}(\vec{y})}{\partial v_g} - p_{sc}(\vec{y}) \frac{\partial G(\vec{x}, \vec{y})}{\partial V_y} \right\} ds_y, \quad (1.7)$$

for $\vec{x} \in R_{\theta_0}^3 \setminus \overline{\Omega}$; here $p_{sc}(\vec{y})$ is the unique solution of the integral equaiton

$$p_{sc}(\vec{y}) + 2 \int_{\mathcal{M}} p_{sc}(\vec{x}) \frac{\partial G(\vec{y}, \vec{x})}{\partial v_y} ds_x = -2 \int_{\mathcal{M}} G(\vec{x}, \vec{y}) \frac{\partial}{\partial v_y} G(\vec{y}, \vec{x}_0) ds, \quad \vec{y} \in \mathcal{M} \quad (1.8)$$

and

$$\frac{\partial}{\partial v} p_{sc}(\vec{y}) = -\frac{\partial}{\partial v_y} G(\vec{y}, \vec{x}_0), \quad \vec{y} \in \mathcal{M}. \quad (1.9)$$

The inverse problem consists of determining the sea-mount \mathcal{M} when $p(\vec{x}, \vec{x}_0)$ is given for all $\vec{x} \in \Gamma_1 \cap R_{\theta_0}^3$, $\Gamma_1 := \{(x, y, z) : z = d_1 = \text{constant}\}$, and $\vec{x}_0 \in \Gamma_2 \cap R_{\theta_0}^3$, $\Gamma_2 := \{(x, y, z) : z = d_2 = \text{constant}\}$.

Here we assume that Γ_1 and Γ_2 are strictly above the sea-mount, i.e., $\max_{x, y} \{z | z = f(x, y)\} < \min \{d_1, d_2\}$.

1.1 Uniqueness of the sea-mount problem

We assume that both $\Gamma_1 \cap R_{\theta_0}^3$ (the receiving plane) and $\Gamma_2 \cap R_{\theta_0}^3$ (the source location plane) are above the sea-mount. That is, \mathcal{M} is disjoint with $\Gamma_j \cap R_{\theta_0}^3$, $j = 1, 2$. The proofs of the following theorems are similar in structure to the approach we used in [17]; hence, we refer the reader to that work for further details.

Theorem 1: Assume that D_1 and D_2 are two sea-mounts with rigid boundaries M_1 and M_2 respectively. Furthermore, suppose that the corresponding solutions of problem (1.4) coincide on $\Gamma_1 \cap R_{\theta_0}^3$ for all $\vec{x}_0 \in \Omega$, where Ω is the unbounded component of $R_{\theta_0}^3 (\overline{D}_1 \cup \overline{D}_2)$, then $D_1 = D_2$.

In general, one sends in incident waves from all directions. In order to simplify this requirement, we need the following lemmas:

Lemma 1: Let $D \subset R_{\theta_0}^3$ be a bounded domain with C^2 boundary and assume that $R_{\theta_0}^3 \setminus D$ is connected. D is located strictly below $\Gamma_1 \cap R_{\theta_0}^3$, i.e., $\min \{y \mid (x, y, z) \in \overline{D}\} > d_1$. Let $G(\cdot, \vec{x}_0)$ be the Green's function for the wedge with source at \vec{x}_0 ,

$$H := \left\{ \frac{\partial G}{\partial \nu} (\cdot, \vec{x}_0) - iG(\cdot, \vec{x}_0) : x_0 \in \Gamma_1 \right\}. \quad (1.10)$$

Then H is complete in $L^2(\partial D)$.

Proof: Assume $\varphi \in L^2(\partial D)$ satisfies

$$\int_{\partial D} \overline{\varphi(\vec{y})} \left\{ \frac{\partial}{\partial \nu_y} G(\vec{y}, \vec{x}_0) - iG(\vec{y}, \vec{x}_0) \right\} ds(\vec{y}) = 0, \quad (1.11)$$

for all $\vec{x}_0 \in \Gamma_1$. Then the combined single- and double-layer potential

$$u(\vec{x}) := \int_{\partial D} \overline{\varphi(\vec{y})} \left\{ \frac{\partial}{\partial \nu_y} G(\vec{y}, \vec{x}) - iG(\vec{y}, \vec{x}) \right\} ds(\vec{y}), \quad x \in R_{\theta_0}^3 \setminus \partial D, \quad (1.12)$$

satisfies the Helmholtz equation in $R_{\theta_0}^3 \setminus \partial D$, the out-going radiation condition as $r \rightarrow \infty$, vanishing Neumann data on $x = y \tan \theta_0$, $y > d_1$, and

$$u(\vec{x}) \Big|_{\Gamma_1 \cap R_{\theta_0}^3} = u(\vec{x}) \Big|_{z=0} = 0. \quad (1.13)$$

It implies that $u = 0$ in $R_{\theta_0}^3 \setminus \overline{D}$, as shown in [17]. Because the singularity of the Green's function is due to the first term in its series expansion we may group the Green's function as

$$G(\vec{x}, \vec{y}) = \frac{e^{ik|\vec{x}-\vec{y}|}}{4\pi|\vec{x}-\vec{y}|} + \Phi_1(\vec{x}, \vec{y}), \quad (1.14)$$

where $\Phi_1(\vec{x}, \vec{y})$ is continuous at $\vec{x} \rightarrow \vec{y}$, we obtain the boundary integral equation

$$\varphi + \mathbf{K}\varphi - i\mathbf{S}\varphi = 0 \quad \text{on } \partial D. \quad (1.15)$$

Here

$$\mathbf{K}\varphi(\vec{x}) := 2 \int_{\partial D} \frac{\partial G}{\partial \nu_y}(\vec{y}, \vec{x}) \varphi(\vec{y}) ds(\vec{y}), \quad (1.16)$$

$$\mathbf{S}\varphi(\vec{x}) := 2 \int_{\partial D} G(\vec{y}, \vec{x}) \varphi(\vec{y}) ds(\vec{y}). \quad (1.17)$$

The operator $\mathbf{I} + \mathbf{K} - i\mathbf{S}$ is invertible in the wedge and its inverse is a bounded linear operator in $L^2(\partial D)$. Hence, we have from (1.9) $\varphi = 0$ on ∂D and the completeness of H is proved.

Lemma 2: Let D be a bounded domain with C^2 boundary ∂D such that $R_{\theta_0}^3 \setminus \overline{D}$ is connected. D is located strictly below Γ_1 . Let $u \in C^2(D) \cap C^1(\overline{D})$ be a solution of the Helmholtz equation. Then there exists a sequence v_n in the span of

$$V := \text{span} \left\{ G(\cdot, \vec{x}_0) : \vec{x}_0 \in \Gamma_1 \right\}$$

such that

$$v_n \rightarrow u, \quad \nabla v_n \rightarrow \nabla u, \quad \text{as } n \rightarrow \infty, \quad (1.14)$$

uniformly on compact subsets of D .

Theorem 2: Assume that D_1 and D_2 are two sea-mounts with rigid boundaries M_1 and M_2 , such that the corresponding solutions of (1.4) coincide on Γ_1 for all $x_0 \in \Gamma_2$, then $D_1 = D_2$.

In view of the v_n being linear combinations of point-source waves from sources on Γ_1 , are the assumption of the Theorem it follows that the corresponding solutions $v_{n,1}^s$ and $v_{n,2}^s$ for the sea-mounts D_1 and D_2 coincide in Γ_1 . Using the same argument as is used in the proof of Theorem (1) (see [17]) it follows that

$$v_n^s := v_{n,1}^s = v_{n,2}^s \quad \text{in } \Omega. \quad (1.15)$$

Moreover,

$$\frac{\partial v_n^2}{\partial v} + \frac{\partial v_n}{\partial v} = 0 \quad \text{on } \partial D_j \cap \partial \Omega, j = 1, 2. \quad (1.16)$$

As a consequence of the continuous dependence of the solution to the exterior Neumann problem on the boundary condition, along with the boundary condition (1.24) and the convergence (1.22), it follows that

$$v_n^s \rightarrow p_j^s, \quad n \rightarrow \infty, \quad (1.17)$$

uniformly on compact subsets of Ω for $j = 1, 2$. Therefore, it must hold that $p_1^s = p_2^s$ in Ω . By Theorem 1, we conclude that $D_1 = D_2$.

1.2 A linearized algorithm for construction of the sea-mount

Let us consider the following linearized algorithm to find the shape of the sea-mount. Let $f_0(x, y)$ be an initial guess for the shape function $f(x, y)$. Then we propose the following recursion scheme to determine the shape function:

$$\delta p_n = p - p_n, \quad \text{and } \delta f_n = f - f_n, \quad n = 0, 1, 2, \dots, \quad (1.18)$$

and where the corresponding sequence of the partially submerged object and its surface are given by

$$\mathcal{D}_n := \left\{ (x, y, z) : x \tan(\theta_0) > y > f_n(x, z), (x, y) \in D_n \right\}$$

$$\mathcal{M}_n := \left\{ (x, y, z) : y = f_n(x, z), (x, z) \in D_n \right\}.$$

Substituting (1.26) into (1.1-1.4) and neglecting terms of $O(\delta^2)$ and higher we have

$$\Delta p_n + k^2 p_n = -\delta(\vec{x} - \vec{x}_0), \quad \text{where } \vec{x} \in R_{\theta_0}^3 \setminus \bar{D}_n \quad (1.19)$$

$$p_n = 0 \quad \text{at } y = 0 \quad (1.20)$$

$$\frac{\partial p_n}{\partial v} = 0 \quad \text{at } y = x \tan(\theta_0), (x, y) \notin \mathcal{M}_n \quad (1.21)$$

$$\frac{\partial p_n}{\partial v} = 0 \quad \text{on } \mathcal{M}_n, n = 0, 1, \dots, \quad (1.22)$$

and

$$\Delta \delta p_n + k^2 \delta p_n = 0, \quad \text{where } \vec{x} \in R_{\theta_0}^3 \setminus \bar{D}_n \quad (1.23)$$

$$\delta p_n = 0 \quad \text{at } y = 0 \quad (1.24)$$

$$\frac{\partial \delta p_n}{\partial v} = 0 \quad \text{at } x = y \tan(\theta_0), (x, y) \notin \mathcal{M}_n \quad (1.25)$$

$$\frac{\partial \delta p_n}{\partial v} = -\left(\frac{\partial^2}{\partial v^2} p_n \right) \delta f_n, \quad \text{on } \mathcal{M}_n. \quad (1.26)$$

We can now use single-layer potentials to obtain a relation between δp_n and δf_n in (1.32) – (1.34). Let us represent δp_n as

$$\delta p_n(\vec{x}) := \int_{\mathcal{M}_n} G(\vec{x}, \vec{y}) \phi(\vec{y}) ds_y, \quad \vec{x} \in R_{\theta_0}^3 \setminus \bar{D}_n. \quad (1.27)$$

Then $\phi(\vec{y})$ satisfies

$$\phi(\vec{x}) - 2 \int_{\mathcal{M}_n} \frac{\partial G(\vec{x}, \vec{y})}{\partial v_x} \phi(\vec{y}) ds_y = -2 \frac{\partial^2}{\partial v^2} p_n \delta f_n \quad \text{on } \mathcal{M}_n, \quad (1.28)$$

and

$$\int_{\mathcal{M}_n} G(\vec{x}, \vec{y}) \phi(\vec{y}) ds_y = \delta p_n(\vec{x}) := p(\vec{x}) - p_n(\vec{x}), \quad \text{for } \vec{x} \in \Gamma_1. \quad (1.29)$$

This suggests an iterative algorithm for solving the inverse problem:

1. Make an initial guess for the shape function $f_0(x, z)$.
2. At the n^{th} stage, solve for $p_n(\vec{x})$ using (1.27) – (1.30).
3. Using $p(\vec{x}) - p_n(\vec{x})$ for $\delta p_n(\vec{x})$, solve $\phi(\vec{y}) = \phi_n(\vec{y})$ for $\vec{y} \in \mathcal{M}_n$ using (1.37).
4. For a chosen accuracy, given by $\varepsilon_n > 0$, define

$$\delta f_n := \min \left\{ \varepsilon_n, \left[\phi(\vec{x}) - 2 \int_{\mathcal{M}_n} \frac{\partial G(\vec{x}, \vec{y})}{\partial v_x} \phi(\vec{y}) ds_y \right] \left[-2 \frac{\partial^2}{\partial v^2} p_n \right]^{-1} \right\}.$$

5. Now upgrade $f_n + \delta f_n \rightarrow f_{n+1}$.
6. We repeat the above steps for $n = 1, 2, \dots$ solving for $p_n, \delta p_n, \phi_n, \delta f_n$ respectively until $|\delta f_n| < \varepsilon$ for some chosen ε .

Step 3 in the above algorithm solves an ill-posed integral equation, inherited from the original ill-posedness of the inverse problem. A proper regularization method must be adapted in order to solve (1.37). With this in mind, we first discuss some properties of the integral operators \mathbf{T} and \mathbf{T}_n defined by

$$\mathbf{T}\phi(\vec{x}) := \int_{\mathcal{M}} G(\vec{x}, \vec{y}) \phi(\vec{y}) ds_y, \quad v \in \Gamma_1, \quad (1.30)$$

$$\mathbf{T}_n \phi(\vec{x}) = \int_{\mathcal{M}_n} G(\vec{x}, \vec{y}) \phi(\vec{y}) ds_y, \quad \vec{x} \in \Gamma, \quad (n = 0, 1, 2, \dots). \quad (1.31)$$

We will need the following spaces that are weighted in $x^1 = (x_1, x_2) \in R^2$,

$$L^{2,s}(R^2) := \left\{ u : \left(1 + |x^1|^2\right)^{s/2} u \in L^2(R^2) \right\},$$

$$H^{1,s}(R^2) := \left\{ u : D^\alpha u \in L^{2,s}(R^2), |\alpha| < 1 \right\},$$

where we use the multi-index notation $\alpha = (\alpha_1, \alpha_2)$, $|\alpha| = |\alpha_1| + |\alpha_2|$ and $D^\alpha = \frac{\partial^\alpha}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}}$; L^2 denote the space of square-integrable functions. We use $L^2(\mathcal{M})$, $H^1(\mathcal{M})$, $L^2(\mathcal{M}_n)$ and $H^1(\mathcal{M}_n)$ to denote the usual Hilbert spaces and Sobolev spaces.

In view of the normal mode integral [4, 5] representation of $G(\vec{x}, \vec{y})$

$$G(\vec{x}, \vec{y}) = \sum_{m=0}^{\infty} I_v(r, r', z) \sin(v\theta) \sin(v\theta'), \quad (1.32)$$

where the modal integral is defined by

$$I_v := \frac{1}{k} \int_0^\infty p \frac{\exp(-\eta|z|)}{\eta} J_v(pr) J_v(pr') dp, \quad (1.33)$$

where $\eta = \sqrt{p^2 - k^2}$, $v = \frac{(m+1/2)\pi}{\theta_0}$ and $J_v(\zeta)$ is a Bessel function of the first kind and order v . From this representation, we know that $G(\vec{x}, \vec{y})$ is real analytic in \vec{x} for any $\vec{y} \in \mathcal{M}$; and for some constant C ,

$$|G(\vec{x}, \vec{y})| < C|x_1|^{-1/2},$$

$$|D^\alpha G(\vec{x}, \vec{y})| < C|x^1|^{-1/2}, \quad |\theta_0| \leq 2,$$

uniformly for $\vec{y} \in \mathcal{M}$ as $|x^1| \rightarrow \infty$. From this it follows:

Theorem 3:

1. The operator \mathbf{T} is compact from $L^2(\mathcal{M})$ into $H^{1,-s}(\Gamma_1)$ for $s > 1/2$.
2. The operator \mathbf{T}_n is compact from $L^2(\mathcal{M}_n)$ into $H^{1,-s}(\Gamma_1)$ for $s > 1/2$.

Theorem 4: The operator \mathbf{T} is injective and has dense range provided that the mixed boundary valued problem

$$\Delta u + k^2 u = 0, \quad \vec{x} \in \overline{D}, \quad (1.34)$$

$$u = 0 \quad \text{on } \mathcal{M}, \quad (1.35)$$

$$\frac{\partial u}{\partial \nu} = 0 \quad \text{at } \left\{ x = y \tan(\theta_0) \right\} = : \gamma_{\theta_0} \quad 0 \leq x \leq a, \quad (1.36)$$

$$(1.37)$$

has no nontrivial solution.

Proof: We first prove that from $\mathbf{T}\phi = 0$ it follows that $\phi = 0$. Consider

$$u(\vec{x}) = \int_{\mathcal{M}} g(\vec{x}, \vec{y}) \phi(\vec{y}) ds_y, \quad \vec{x} \in R_{\theta_0}^3. \quad (1.38)$$

$u(\vec{x})$ satisfies (1.23) in $R_{\theta_0}^3 \setminus \overline{D}$, $u(\vec{x}) = 0$ for $\vec{x} \in \Gamma_1 \cup \Gamma_\alpha$, $\frac{\partial u}{\partial \nu}(\vec{x}) = 0$ on γ_{θ_0} and $u(\vec{x})$ satisfies the out-going radiation condition. Then $u(\vec{x}) = 0$ for $\vec{x} \in \{(r, \theta, z) : \Gamma_0 := \{y = 0\}\}$; hence, $u(\vec{x}) = 0$ for $\vec{x} \in R_{\theta_0}^3 \setminus D$.

Define

$$\mathbf{K}\phi(\vec{x}) := 2 \int_{\mathcal{M}} \frac{\partial}{\partial \nu_x} G(\vec{x}, \vec{y}) \phi(\vec{y}) ds_y, \quad \vec{x} \in \mathcal{M}. \quad (1.39)$$

The jump relation of the normal derivative of $u(\vec{x})$ on \mathcal{M} implies

$$\phi - \mathbf{K}\phi = 0, \quad \text{on } \mathcal{M}. \quad (1.40)$$

A typical way similar to that in ([7], p128) show that $\phi = 0$ on \mathcal{M} , provided the problem (1.34)-(1.37) has no nontrivial solution. We conclude that \mathbf{T} is injective.

Now we show that if $(\psi, \mathbf{T}\phi)_{L^{2,-s}(\Gamma_1)} = 0$ for all $\phi \in L^2(\mathcal{M})$, then $\psi = 0$. That is, we need to show that from $\mathbf{T}^*\psi = 0$ on \mathcal{M} it follows that $\psi = 0$ on Γ ; here $\mathbf{T}^* : L^{2,-s}(\Gamma_1) \rightarrow L^2(\mathcal{M})$ is the adjoint operator of \mathbf{T} :

$$\mathbf{T}^*\psi(\vec{y}) = \int_{\Gamma} G(\vec{x}, \vec{y}) \psi(\vec{x}) \left(|x^1|^2 + 1 \right)^{-s/2} ds_x, \quad \vec{y} \in \mathcal{M}. \quad (1.41)$$

Now consider

$$v(\vec{y}) := \int_{\Gamma} G(\vec{x}, \vec{y}) \psi(\vec{x}) \left(|x^1|^2 + 1 \right)^{-s/2} ds_x, \quad \vec{y} \in R_{\theta_0}^3. \quad (1.42)$$

$v(\vec{y})$ is a solution of the problem (1.34)-(1.37); hence, $v(\vec{y}) = 0$ in Ω . But $v(\vec{y})$ satisfies the Helmholtz equation in $R_{\theta_0}^3 \setminus \Gamma_1$. So $v(\vec{y}) = 0$ on $\{(r, \theta, z) : 0 \leq z \leq d\} \cap R_{\theta_0}^3$. Define

$$S\psi(\vec{y}) := 2 \int_{\Gamma} \frac{\partial}{\partial \nu_y} G(\vec{x}, \vec{y}) \psi(\vec{x}) \left(|x^1|^2 + 1 \right)^{-s/2} ds_x, \quad \vec{y} \in \Gamma_1. \quad (1.43)$$

The jump relation of the normal derivative of $v(\vec{y})$ on Γ_1 implies

$$\psi + \mathbf{S}\psi = 0, \quad \text{on } \Gamma_1. \quad (1.44)$$

Now we can conclude similar to the discussion for $u(\vec{x})$ that $\psi = 0$ on Γ_1 .

Based on Theorems 3 and 4, we may apply the Tikhonov regularization to step 3, that is, we solve

$$\alpha \phi_\alpha + \mathbf{T}_n^* \mathbf{T}_n \phi_\alpha = \mathbf{T}_n^* (p - p_n) \quad (1.45)$$

with some regularization parameter $\alpha > 0$ instead of (1.29). The regularity of discrepancy principle for the Tikhonov regularization (see, for example, [7], Th. 4.16, p99) follows.

Theorem 5: if $\delta p_n \in \mathbf{T}(L^2(\mathcal{M}))$, then

$$\phi_\alpha = (\alpha \mathbf{I} + \mathbf{T}_n^* \mathbf{T}_n)^{-1} \mathbf{T}_n^* (p - p_n) \quad (1.46)$$

approaches $\mathbf{T}_n^{-1}(\delta p_n)$ as $\alpha \rightarrow 0$.

Acknowledgement: This research was supported in part by the National Science Foundation through grant BES-9820813.

REFERENCES

- [1] Ahluwalia, D. and Keller, J., *Exact and Asymptotic Representations of the Sound Field in a Stratified Ocean*, Wave Propagation and Underwater Acoustics, Lecture Notes in Physics, 70, 1977, Springer, Berlin.
- [2] Buchanan, J.L., Gilbert, R.P., and Wirgin, A., Finding an inclusion in a shallow ocean using a canonical domain method, Proc. Fourth European Conf. Underwater Acoustics, (Eds. A. Alippi and G.B. Cappelli), 1998, Rome, 389-394.
- [3] Buchanan, J.L., Gilbert, R.P., Wirgin, A., and Xu, Y., Identification of acoustically soft solids of revolution in a wave guide using the ICBA method, Analytical and Computational Methods in Scattering and Applied Mathematics, (Eds. Santosa and Stakgold), 1999, CRC Press.
- [4] Buckingham, M.J., Acoustic propagation in a wedge shaped with perfectly reflecting boundaries, Hybrid Formulation of Wave Propagation and Scattering, (Ed. L.B. Felsen), 1984.
- [5] Buckingham, M.J., Theory of acoustic radiation in corners with homogeneous and mixed perfectly reflecting boundaries, J. Acoust. Soc. Am., 86(6), 1989, 2273-2291.
- [6] Burton, A.J. and Miller, G.F., The application of integral equation methods to the numerical solutions of some exterior boundary-value problems, Proc. Royal Soc. London Ser. A, 323, 1971, 201-210.
- [7] Colton, D. and Kress, R., *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer-Verlag, 1993.

- [8] Coronas, J. and Sun, Z., *Transient Reflection and Transmission Problem for Fluid-Saturated Porous Media*, Invariant Imbedding and Inverse Problems, SIAM, Philadelphia.
- [9] Gilbert, R.P., Scotti, T., Wirgin, A., and Xu, Y.S., Identification of a 3d object in a shallow ocean from scattered sound, C.R. Acad. Sci. Paris lib, 325, 1997, 1320-1327.
- [10] Gilbert, R.P., Scotti, T., Wirgin, A., and Xu, Y.S., The unidentified object problem in a shallow ocean, J. Acoust. Soc. Am., 103, 1998, 1320-1327.
- [11] Gilbert, R.P. and Xu, Y., An inverse problem for harmonic acoustics in stratified oceans, J. Math. Anal. Appl., 17(1), 1993, 121-137.
- [12] Gilbert, R.P. and Xu, Y., Starting fields and far fields in ocean acoustics, Wave Motion, 11, 1989, 507-524.
- [13] Gilbert, R.P. and Xu, Y., Dense sets and the projection theorem for acoustic harmonic waves in homogeneous finite depth oceans, Math. Methods Appl. Scis., 12, 1989, 69-76.
- [14] Gilbert, R.P. and Xu, Y., Acoustic waves and far-field patterns in two dimensional oceans with poros-elastic seabed, Proceeding of Their IMACS Symposium on Computational Acoustics, Harvard University, Massachusetts, 1991.
- [15] Gilbert, R.P. and Xu, Y., The propagation problem and far-field patterns in a stratified finite-depth ocean, Math. Methods in the Appl. Sciences, 12, 1990, 199-208.
- [16] Gilbert, R.P. and Xu, Y., *Generalized Herglotz Functions and Inverse Scattering Problems in Finite Depth Oceans*, Inverse Problems, SIAM, 1992.
- [17] Gilbert, R.P. and Xu, Y., The seamount problem, Nonlinear Problems in Applied Mathematics, (Eds. T.S. Angell, L.P. Cook, R.E. Kleinman, and W.E. Ohnstead), SIAM, Philadelphia, 1996, 140-149.
- [18] Kirsch, A. and Kress, R., Uniqueness in inverse obstacle scattering, Inverse Problems, 9, 1993, 285-299.
- [19] Schenck, H.A., Improved integral formulation for acoustic radiation problems, J. Acoust. Soc. Amer., 44, 1968, 41-58.
- [20] Xu, Y., *Direct and Inverse Scattering in Shallow Oceans*, Ph.D. Thesis, University of Delaware, 1990.

10 DISCRETE SIMULATION IN NONLINEAR DYNAMICS WITH APPLICATIONS

Donald Greenspan

Department of Mathematics

The University of Texas at Arlington

Arlington, TX 76019-0408

1. INTRODUCTION

Contemporary science teaches us that:

- (a) All things change with time.
- (b) All material bodies consist of atoms and/or molecules.

In this paper we will discuss computer simulation which is consistent with both the above principles. Applications will be directed primarily to phenomena in science and engineering, although the ideas and methods extend to other disciplines. The computers used in the examples to be discussed are the Digital AXP275 personal computer and the CRAY YMP/8 supercomputer.

The fundamental mathematical problem will be a multibody problem called the general N -body problem, which is formulated as follows.

In cgs units, and for $i = 1, 2, 3, \dots, N$, let P_i of mass m_i be at $\vec{r}_i = (x_i, y_i, z_i)$ with velocity $\vec{v}_i = (v_{x,i}, v_{y,i}, v_{z,i})$ at any time $t > 0$. For generality we call P_i a set of particles. Let the positive distance between distinct particles P_i and P_j be $r_{ij} = r_{ji}$. Let the force on P_i due to P_j be $\vec{F}_{ij} = \vec{F}_{ij}(r_{ij})$. Then the general N -body problem is to determine the motion of the system of particles when each particle acts on all other particles and the initial positions and velocities are given.

Our first problem is to select a dynamical equation from one of the three major areas of physics, that is, Newtonian Mechanics, Quantum Mechanics, and Relativity. In Quantum Mechanics the N -body for the time dependent Schrödinger

equation requires $3N+1$ dimensions. Thus, for example, simulation of the solar system in Quantum Mechanics requires 31 dimensions. In Relativity, because simultaneity is denied, N -body problems are limited to $N=1$, so that solar system simulation is not possible at all. For simplicity, then, we turn first, by default, to Newtonian Mechanics.

In Newtonian Mechanics the N -body problem is governed by the deterministic system of second order, ordinary differential equations

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \sum_{\substack{j=1 \\ j \neq i}}^N \vec{F}_{ij}, \quad i = 1, 2, 3, \dots, N. \quad (1.1)$$

We will consider the following three cases:

- (a) N small ($1 < N \leq 200$);
- (b) N large ($200 < N \leq 10000$);
- (c) N very large ($10000 < N < 10^{20}$).

We begin with case (b), that is, N large.

2. N LARGE

For any N we will have to solve numerically a system of $2N$ second order differential equations in 2 dimensions and $3N$ second order equations in 3 dimensions. Any of the available numerical methods can be used for this purpose.

Let us then review first some fundamental aspects of classical molecular mechanics [1]. Qualitatively, two molecules interact only locally, that is, when in close proximity, and their interaction is of the following general nature:

- (a) when pushed together, the molecules repel;
- (b) when pulled apart, the molecules attract; and
- (c) repulsion is of a higher order of magnitude than is attraction.

As a simplistic example, consider $F(r) = -\frac{1}{r^7} + \frac{1}{r^{13}}$. Then, for $r = 1$, $F = 0$, so

that there is no force, which is called equilibrium. For $r > 1$, say, $r = 2$, $F(2)$ is negative, which will yield attraction. For $r < 1$, say, $r = 0.1$, $F(0.1)$ is positive, which will yield repulsion. Moreover, as r converges to zero, the magnitude of the repulsion becomes unbounded.

Let us then consider now a problem, which is of interest in meteorology and chemical engineering, that is, the collision modes of fluid microdrops. In particular, we consider the collision modes of microdrops of water. For water, a classical molecular potential given by Rowlinson is

$$\phi(r) = 1.9647 \left(10^{-13}\right) \left[\left(\frac{2.725}{r}\right)^{12} - \left(\frac{2.725}{r}\right)^6 \right] \text{erg}, \quad (r \text{ in Angstroms}).$$

Differentiating the negative of the potential to obtain the force \vec{F} in dynes yields readily

$$F(r) = (4.325809) 10^{-5} \left[2 \left(\frac{2.725}{r}\right)^{13} - \left(\frac{2.725}{r}\right)^7 \right].$$

In order to study how two water drops can collide, we will first generate a single water drop. We will do this in detail in 2 dimensions in order to demonstrate a fundamental physical property of the methodology, which is not seen readily in 3 dimensions. The discussion will then proceed in 3 dimensions. Hence, in the XY plane construct a regular, triangular grid using a triangular edge length of 2.725 \AA . Consider then only those points of the grid which lie interior to the circle

$$x^2 + y^2 = 2320.$$

At each such grid point place a water molecule, thus yielding a system of 1128 molecules arranged in a circular configuration. All initial velocities are set to zero and all molecules are allowed to interact with all other molecules. The leap-frog formulas [2] are applied numerically. The dynamical interaction is simulated with $T = 10^{14} t$ and $\Delta T = 0.0002$. The system contracts until $T = 11.2$, at which time its energy is mostly potential. At this time all velocities are set to zero. The simulation is then continued until $T = 14.0$, at which time all velocities are again reset to zero. Thereafter, the system no longer shows large vibrations, as shown in Figure 1 at the indicated times. What is to be observed is that in each water microdrop in Figure 1, the outer molecules show maximum separation, which results in large attractive forces. This is the mechanism of surface tension, which is not a consequence of the Navier-Stokes equations.

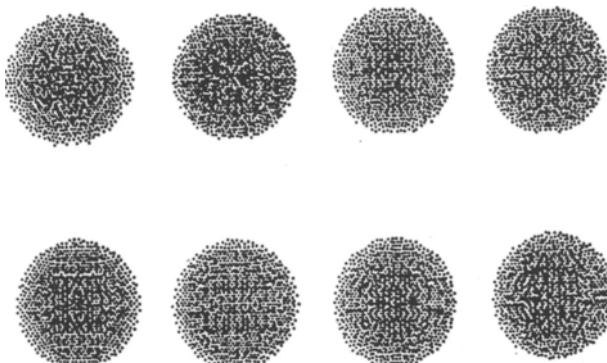


Figure 1. Surface tension.

Proceeding now in the same spirit but in 3 dimensions [3], Figure 2 shows a 3 dimensional water drop with 2051 molecules. Figure 3 is the result of mirror imaging of Figure 2 to construct two microdrops of water which are 3\AA apart. In order to study drop collisions, each molecule in the unshaded drop has its velocity increased by \bar{v} while each molecule in the shaded drop has its velocity increased by $-\bar{v}$. For various choices of \bar{v} [3], Figure 4 shows an oscillating oblateness mode, Figure 5 shows a raindrop mode, Figure 6 shows a dumbbell mode, and Figure 7 shows a soft collision, teardrop mode. Each of these modes has been produced in the laboratory. In all cases, surface tension, in time, forces a spherical steady state drop.



Figure 2. A 3-dimensional water drop with $N = 2051$.

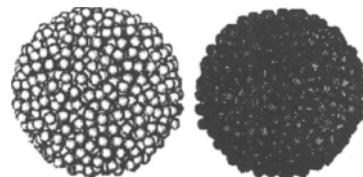


Figure 3. Two 3-dimensional water drop 3\AA apart.

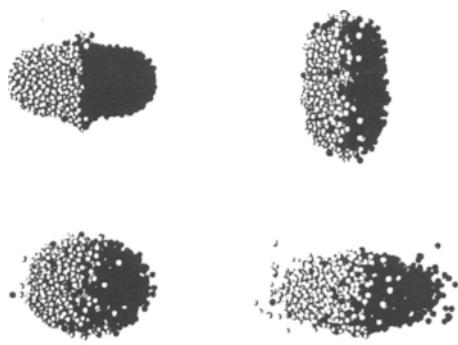


Figure 4. Oscillating oblateness mode.

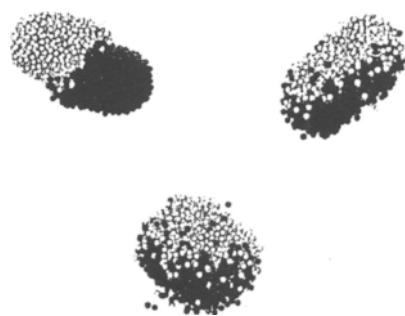


Figure 5. Raindrop mode.

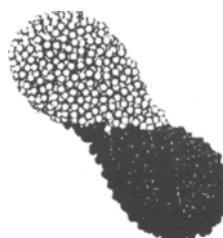


Figure 6. Dumbbell mode.



Figure 7. Soft collision, teardrop mode.

3. N VERY LARGE

Traditionally, cases for N very large have been studied through statistical mechanics, with results restricted largely to steady state. Since our interests are in dynamics, we will develop an alternate method which begins with the engineering approach called the "lumped mass" technique. When confronted with a very large set of molecules, we will aggregate them into a finite set of N particles, for which, as in Section 2, N is large. The total molecular mass will be distributed over the N particles. A molecular type force whose magnitude F is given by

$$F = -\frac{G}{r^p} + \frac{H}{r^q} \quad (3.1)$$

will be determined as follows for the particles. The parameters p and q will be selected to be approximately 3 and 5, respectively. This will prevent the massive particle aggregates from the volatile interactions which are characteristic of molecules. The parameters G and H will be determined by equating the total potential energy of the molecular system to that of the particle system. To assure that (3.1) is truly local, a parameter D , called the local interaction distance parameter will also be introduced.

We consider first an example in the study of cracks and fractures. This area is of interest in geology, nuclear reactor design, and aircraft integrity. The problem we

will consider is that of determining where a stressed copper plate with a slot will first crack [4].

A Lennard-Jones potential for copper is

$$\phi(r) = -(1.398) 10^{-10}/r^6 + (1.551) 10^{-8}/r^{12} \text{ erg}.$$

Consider then a rectangular, copper plate in 2 dimensions, which is 8 cm by 11.5 cm, as shown in Figure 8. If we assume the copper atoms in the plate are arranged at the vertices of a regular triangular grid with edge length equal to the equilibrium length, then there are $(1.745) 10^{17}$ atoms in the plate. The mass of an individual atom is $(1.054) 10^{-22} \text{ g}$, the total mass is $(1.840) 10^{-5} \text{ g}$, and the total potential energy is $-(1.6493) 10^5 \text{ erg}$. The copper atoms are now aggregated into 2713 particles, which are set in the plate on a regular triangular grid with edge length 0.2 cm. Distributing the total atomic mass over the particles yields an individual particle mass of $(1.840) 10^{-5}/2713 = (6.782) 10^{-9} \text{ g}$. Assuming the total potential energies are equal then results in an interatomic force whose magnitude is

$$F = -\frac{3.2422}{R^3} + \frac{0.1297}{R^5}, \quad (R \text{ in cm}).$$

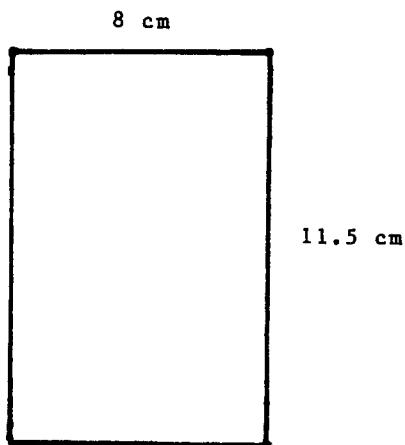


Figure 8. The copper plate.

There results then a system of 2713 second order differential equations, which determine the dynamics from given initial data.

We assume at present that each particle interacts only with its nearest neighbors and that the elastic limit is defined by dF/dR becoming negative.

Next, we introduce a slot into the plate by removing the 15 particles shown in Figure 9.

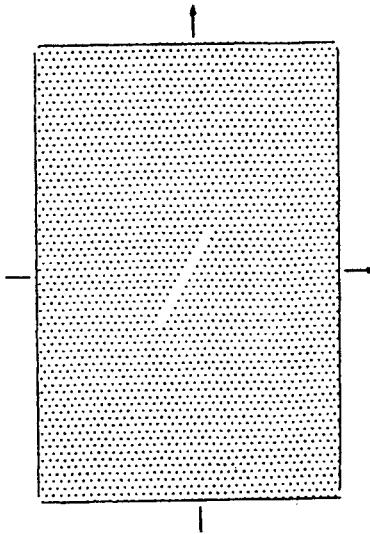


Figure 9. The slotted plate.

The plate is now stressed by moving the uppermost row upward and the lower most row downward at each time step [4]. We seek to find where the plate will first crack.

Figure 10 shows the force field development with time in the lower half of the plate. The force is transmitted to the interior in waves. At $T = 12.8$ a crack develops in the lower left corner of the slot, which is in complete agreement with experiment. Of course, by symmetry, a crack also develops in the upper right corner of the slot.

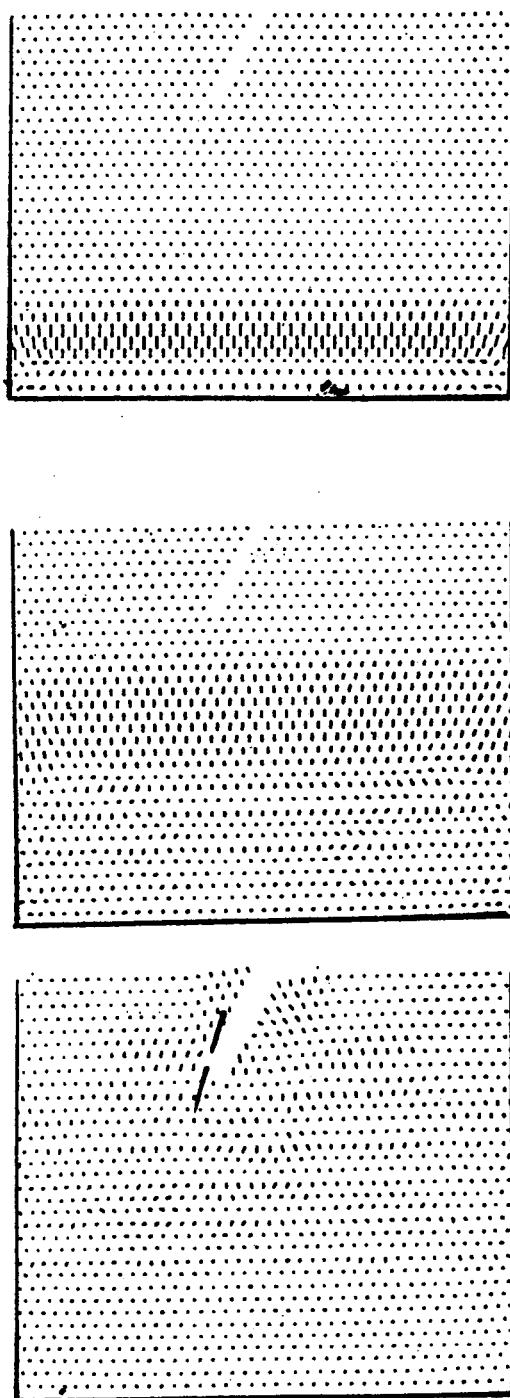


Figure 10. Crack development.

In our next three examples we include gravity, which acts on all particles uniformly.

Figure 11 shows the fall of a pendant water drop from a ceiling [5]. At T_{6000} a neck forms. At T_{9000} the neck breaks, resulting in a new small pendant drop one the ceiling and the falling of the major portion of the drop. From T_{10000} to T_{13000} one sees drop oscillation and convergence toward a spherical steady state.

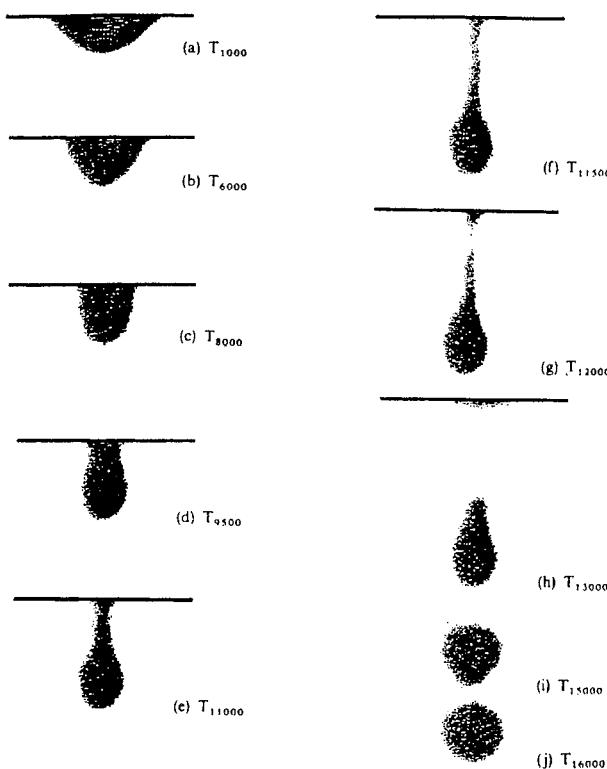


Figure 11. Fall of a pendant water drop.

Figure 12 shows the emergence of large carbon dioxide bubbles from a water basin [6]. In this example, one must use the chemists rule of thumb called the law of empirical bonding to determine the force between the CO_2 particles and the H_2O particles, each of which is known, while the interaction between the different species is not known.

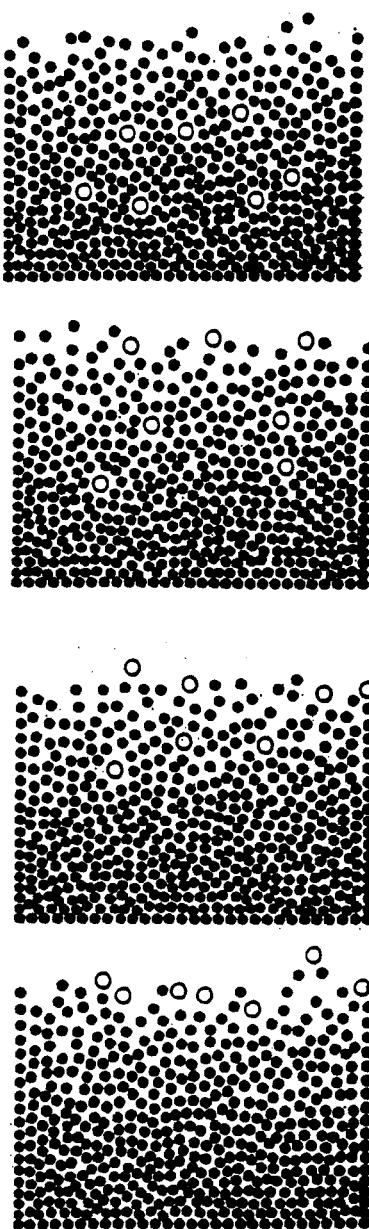


Figure 12. Emergence of CO_2 bubbles from water.

Figure 13 shows the adhesion of a water drop on a graphite surface. In materials science the estimate of the angle of adhesion is a constant of basic interest. For the parameter choices in the figure, the resulting angle is 60.1° [7], which compares most favorably with the experimental result of 60° .

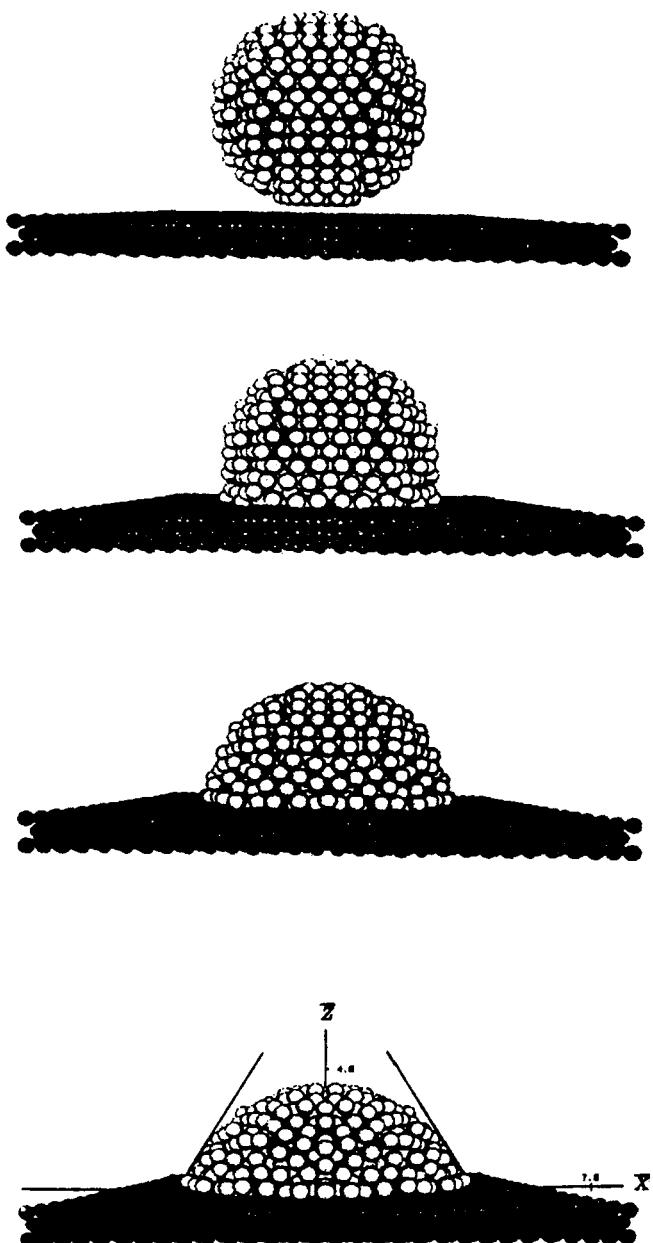


Figure 13. Formation of a sessile water drop on a graphite surface.

Next let us address the problem of how one develops intuition for the simulation being discussed. Intuition is, of course, one of the major assets which one would like in any field of endeavor. Our approach to the development of intuition is to study qualitative particle models. In such models one can vary all the parameters to see the effect of each. The next two examples then are strictly qualitative and in each we describe the intuition derived.

Figure 14 shows the development of vortex motion for the classical cavity problem [8]. At T_{1500} one sees the onset of vortex motion. Particles are compressed in the upper left corner, which results in repulsion downward. Particles in the upper right corner are dragged to the left, creating a partial vacuum, which is filled by particles below it by their repulsion. Thus, the counterclockwise vortex development is the direct result of the large particle repulsion. If one assumes, as in classical hydrodynamics, that liquids are incompressible, these mechanisms cannot be deduced. At T_{100000} one sees motion down the left wall, arms penetrating the fluid below the primary vortex, and, because no effective vortex motion has occurred during the previous 50000 steps, that one has reached a steady state.

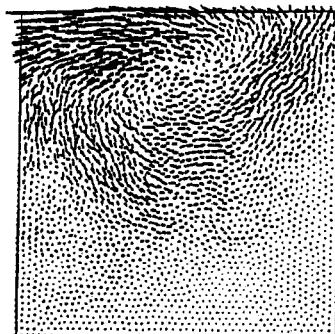
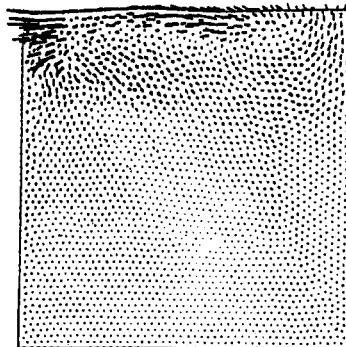


Figure 14. The cavity problem

Figure 15 shows 1052 particles of three different types. There are 38 particles, which have strong attraction, 246, which have medium attraction, and 768, which have weak attraction [9]. Our aim now is to simulate the biological results of Holtfretter, in which he separated normal tissue into endoderm, mesoderm and ectoderm and observed that the tissue self reorganized into its original form. This process is called morphogenesis. Figure 16 shows the self reorganization of the “endoderm” particles, Figure 17 shows the self reorganization of the “mesoderm” particles, and Figure 18 show the final self reorganization of the entire system. Thus, we see that simulation with particles allows one to model self reorganization processes.

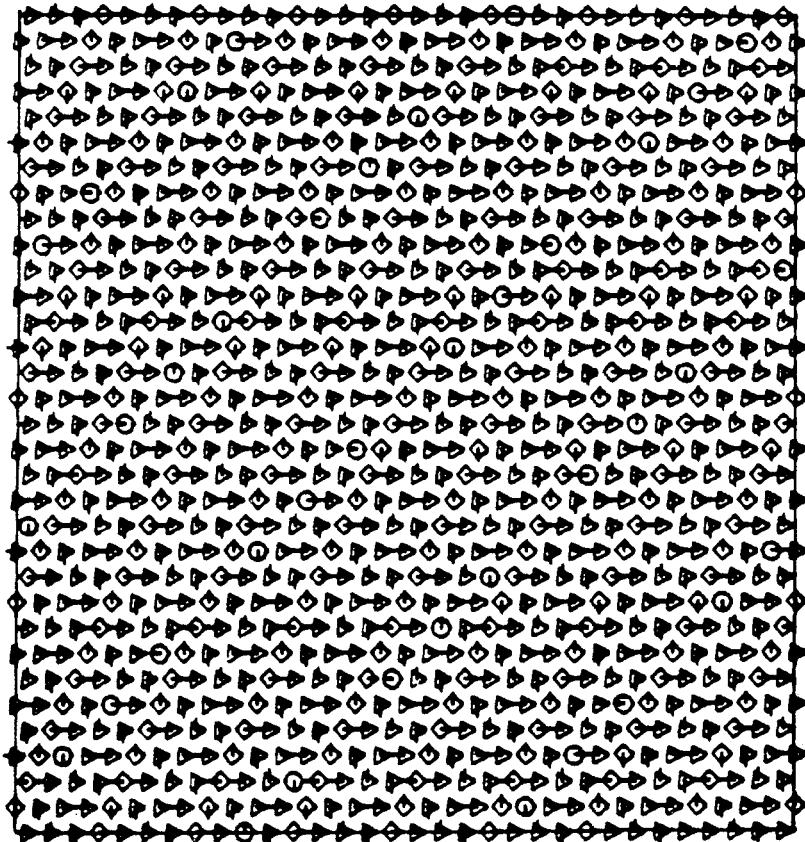


Figure 15. Particle separation.

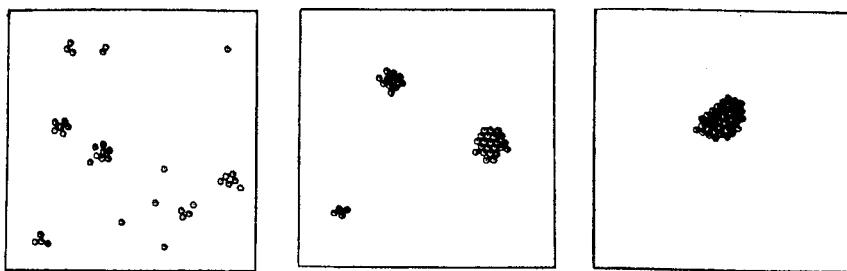


Figure 16. "Endoderm" self reorganization.

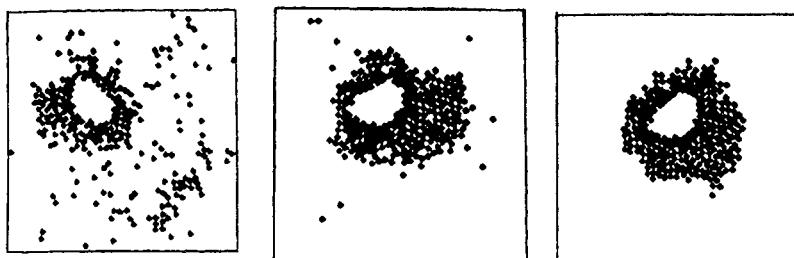


Figure 17. "Mesoderm" self reorganization.

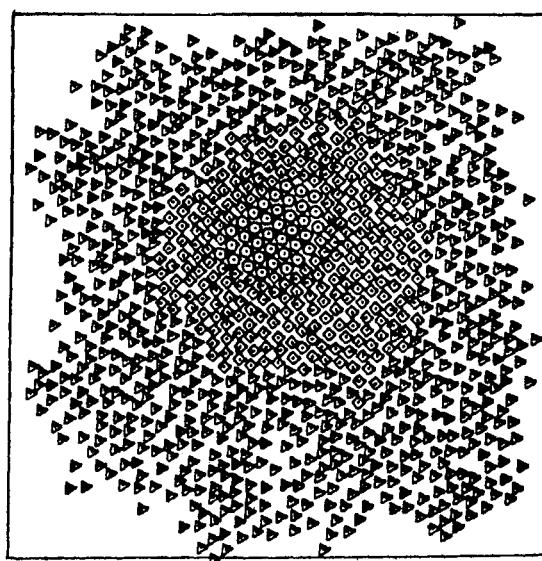


Figure 18. Complete self reorganization.

Particle simulations of other phenomena also are available [10-15]. These include formation of minimal surfaces, snap through, soliton interaction, capillary action, flame development, and turbulent vortex motion.

4. N SMALL

If N is small, we would like to do a very good job in solving the N -body problem. By this we mean that we would like not only to solve the problem with accuracy, but we would also like to preserve numerically any basic physical invariants of the system. To do this in detail, we concentrate on the 3-body problem, which is the prototype problem because it contains all the difficulties of the general N -body problem. The entire discussion, which follows, extends in a natural way to the general problem.

For $i = 1, 2, 3$, let P_i of mass m_i be at $\vec{r}_i = (x_i, y_i, z_i)$ at time t . Let the positive distance between P_i and P_j , $i \neq j$, be $r_{ij} = r_{ji}$. Let $\phi = \phi_{ij} = \phi(r_{ij})$, given in ergs, be the potential for the pair P_i, P_j . The force on P_i due to P_j is

$$\vec{F}_{ij} = -\frac{\partial \phi}{\partial r_{ij}} \frac{\vec{r}_i - \vec{r}_j}{r_{ij}}.$$

Then the Newtonian dynamical equations for the three-body problem are

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = -\frac{\partial \phi}{\partial r_{ij}} \frac{\vec{r}_i - \vec{r}_j}{r_{ij}} - \frac{\partial \phi}{\partial r_{ik}} \frac{\vec{r}_i - \vec{r}_k}{r_{ik}}, \quad i = 1, 2, 3 \quad (4.1)$$

where $i = 1$ implies $j = 2, k = 3; i = 2$ implies $j = 1, k = 3; i = 3$ implies $j = 1, k = 2$.

Theorem. System (2.1) conserves energy, linear momentum, and angular momentum. It is also covariant under translation, rotation, and uniform relative motion of coordinate frames [16].

In order to solve an initial value problem for (4.1) numerically, we will first rewrite it as the system

$$\frac{d\vec{r}_i}{dt} = \vec{v}_i, \quad i = 1, 2, 3, \quad (4.2)$$

$$m_i \frac{d\vec{v}_i}{dt} = -\frac{\partial \phi}{\partial r_{ij}} \frac{\vec{r}_i - \vec{r}_j}{r_{ij}} - \frac{\partial \phi}{\partial r_{ik}} \frac{\vec{r}_i - \vec{r}_k}{r_{ik}}, \quad i = 1, 2, 3. \quad (4.3)$$

We proceed numerically as follows. For $\Delta t > 0$, let $t_n = n(\Delta t)$, $n = 0, 1, 2, \dots$. At time t_n , let P_i be at $\vec{r}_{i,n} = (x_{i,n}, y_{i,n}, z_{i,n})$ with velocity $\vec{v}_{i,n} = (v_{i,x,n}, v_{i,y,n}, v_{i,z,n})$. Denote the distances $\|P_1 P_2\|, \|P_1 P_3\|, \|P_2 P_3\|$ by $r_{12,n}, r_{13,n}, r_{23,n}$, respectively.

Differential equations (4.2), (4.3) are now approximated by the difference equations

$$\frac{\vec{r}_{i,n+1} - \vec{r}_{i,n}}{\Delta t} = \frac{\vec{v}_{i,n+1} + \vec{v}_{i,n}}{2} \quad (4.4)$$

$$m_i \frac{\vec{v}_{i,n+1} - \vec{v}_{i,n}}{\Delta t} = -\frac{\phi(r_{ij,n+1}) - \phi(r_{ij,n})}{r_{ij,n+1} - r_{ij,n}} \cdot \frac{\vec{r}_{i,n+1} + \vec{r}_{i,n} - \vec{r}_{j,n+1} - \vec{r}_{j,n}}{r_{ij,n+1} + r_{ij,n}} \\ - \frac{\phi(r_{ik,n+1}) - \phi(r_{ik,n})}{r_{ik,n+1} - r_{ik,n}} \cdot \frac{\vec{r}_{i,n+1} + \vec{r}_{i,n} - \vec{r}_{k,n+1} - \vec{r}_{k,n}}{r_{ik,n+1} + r_{ik,n}}. \quad (4.5)$$

Note that the force is approximated, not the potential. Consistency follows immediately as $\Delta t \rightarrow 0$.

System (4.4), (4.5) constitutes 18 implicit recursion equations for the unknowns, $x_{i,n+1}, y_{i,n+1}, z_{i,n+1}, v_{i,x,n+1}, v_{i,y,n+1}, v_{i,z,n+1}$ in the 18 knowns $x_{i,n}, y_{i,n}, z_{i,n}, v_{i,x,n}, v_{i,y,n}, v_{i,z,n}$, $i = 1, 2, 3$. These are solved readily by the following Newton's method.

For the system

$$f_1(x_1, x_2, \dots, x_k) = 0$$

$$f_2(x_1, x_2, \dots, x_k) = 0$$

⋮

$$f_k(x_1, x_2, \dots, x_k) = 0$$

use the iteration formulas

$$x_1^{(n+1)} = x_1^{(n)} - \frac{f_1(x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})}{\frac{\partial f_1}{\partial x_1}(x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})}$$

⋮

$$x_2^{(n+1)} = x_2^{(n)} - \frac{f_2(x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})}{\frac{\partial f_2}{\partial x_2}(x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})}$$

$$x_k^{(n+1)} = x_k^{(n)} - \frac{f_k(x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})}{\frac{\partial f_k}{\partial x_k}(x_1^{(n)}, x_2^{(n)}, \dots, x_k^{(n)})}.$$

Theorem. The numerical method conserves exactly the same energy, linear momentum, and angular momentum as system (4.1) and does so independently of

the choice of time step. In addition, the difference equations are covariant under translation, rotation, and uniform relative motion of coordinate frames [17].

To illustrate the methodology, let us give the proof for conservation of energy.

Proof. Define

$$W_N = \sum \sum m_i (\vec{r}_{i,n+1} - \vec{r}_{i,n}) \cdot (\vec{v}_{i,n+1} - \vec{v}_{i,n}).$$

With the aid of (4.4), then,

$$\begin{aligned} W_N &= \sum_{n=0}^{N-1} \sum_{i=1}^3 m_i \frac{(\vec{r}_{i,n+1} - \vec{r}_{i,n})}{\Delta t} \cdot (\vec{v}_{i,n+1} - \vec{v}_{i,n}) \\ &= \sum_{n=0}^{N-1} \sum_{i=1}^3 m_i \frac{(\vec{v}_{i,n+1} + \vec{v}_{i,n})}{2} \cdot (\vec{v}_{i,n+1} - \vec{v}_{i,n}) \\ &= \sum_{n=0}^{N-1} \sum_{i=1}^3 m_i \left(\frac{v_{i,n+1}^2}{2} - \frac{v_{i,n}^2}{2} \right) \\ &= \sum_{i=1}^3 m_i \left[\left(\frac{v_{i,1}^2}{2} - \frac{v_{i,0}^2}{2} \right) + \left(\frac{v_{i,2}^2}{2} - \frac{v_{i,1}^2}{2} \right) + \left(\frac{v_{i,3}^2}{2} - \frac{v_{i,2}^2}{2} \right) + \dots + \left(\frac{v_{i,N}^2}{2} - \frac{v_{i,N-1}^2}{2} \right) \right] \\ &= \frac{1}{2} m_1 v_{1,N}^2 + \frac{1}{2} m_2 v_{2,N}^2 + \frac{1}{2} m_3 v_{3,N}^2 - \frac{1}{2} m_1 v_{1,0}^2 - \frac{1}{2} m_2 v_{2,0}^2 - \frac{1}{2} m_3 v_{3,0}^2 \end{aligned}$$

so that

$$W_N = K_N - K_0.$$

Next, with the aid of (4.5) and the observation that $\vec{r}_i - \vec{r}_j = \vec{r}_{ji}$, one finds

$$\begin{aligned} W_N &= \sum_{n=0}^{N-1} \left[-\frac{\phi(r_{12,n+1}) - \phi(r_{12,n})}{r_{12,n+1} - r_{12,n}} \cdot \frac{r_{12,n+1}^2 - r_{12,n}^2}{r_{12,n+1} + r_{12,n}} \right. \\ &\quad - \frac{\phi(r_{13,n+1}) - \phi(r_{13,n})}{r_{13,n+1} - r_{13,n}} \cdot \frac{r_{13,n+1}^2 - r_{13,n}^2}{r_{13,n+1} + r_{13,n}} \\ &\quad \left. - \frac{\phi(r_{23,n+1}) - \phi(r_{23,n})}{r_{23,n+1} - r_{23,n}} \cdot \frac{r_{23,n+1}^2 - r_{23,n}^2}{r_{23,n+1} + r_{23,n}} \right] \\ &= \sum_{n=0}^{N-1} (-\phi_{12,n+1} - \phi_{13,n+1} - \phi_{23,n+1} + \phi_{12,n} + \phi_{13,n} + \phi_{23,n}) \\ &= -\phi_{12,N} - \phi_{13,N} - \phi_{23,N} + \phi_{12,0} + \phi_{13,0} + \phi_{23,0}. \end{aligned}$$

Thus,

$$W_N = -\phi_N + \phi_0.$$

Hence,

$$K_N - K_0 = -\phi_N + \phi_0$$

$$K_N + \phi_N = K_0 + \phi_0, \quad N = 0, 1, 2, \dots \quad \text{QED.}$$

Moreover, since K_0 and ϕ_0 depend only on $\vec{r}_{i,0}$ and $\vec{v}_{i,0}$, it follows that $K_0 + \phi_0$ is the same in both the continuous and the discrete cases, so the energy conserved is exactly that of the continuous system. Note also that the proof was independent of Δt .

As an example, colleagues at Stanford [18] have simulated in a completely conservative fashion nonlinear satellite motion shown in Figure 19.

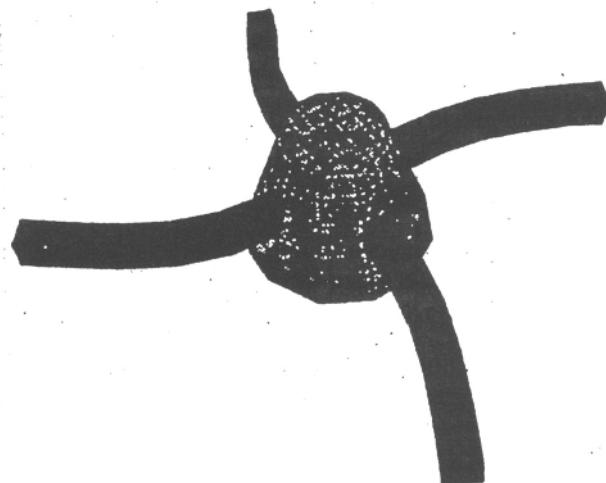


Figure 19. Nonlinear satellite rotation.

As a second example, let us consider the simplest molecule, H_2^+ , which consists of two protons and one electron. The ground state energy is invariant. The variation of the density of the electron's positions with time along the internuclear axis is shown quantum mechanically in Figure 20. Using the energy conserving

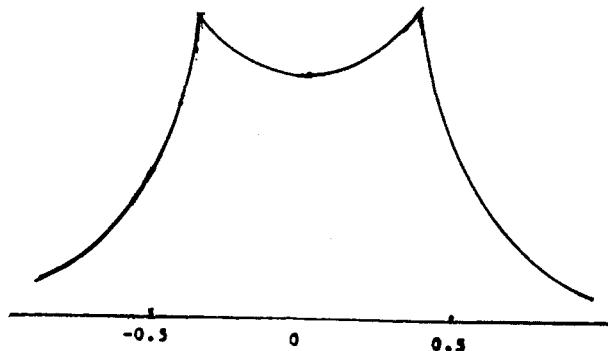


Figure 20. Electron density distribution.

methodology of this section, the projection of the electron's motion in the XY plane is shown in Figure 21. The numerical results shown in Figure 21 are in complete agreement with the quantum mechanical results shown in Figure 20.

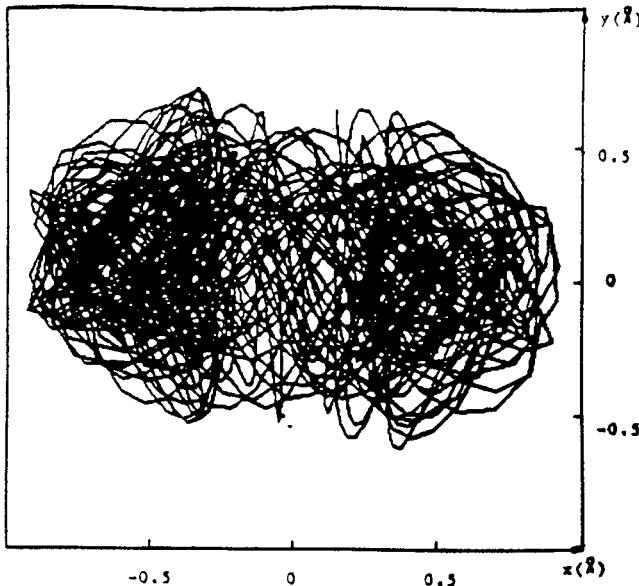


Figure 21. Projection of electron motion.

The legitimacy of using classical mechanics in this example is based on recent theoretical results of Gell-Mann and Hartle.

5. REMARKS

The 1-body problem is of interest in Special Relativity. This is valid when an oscillator oscillates near the speed of light. In Special Relativity, however, even an harmonic oscillator equation, which is easily solved by Newtonian mechanics, is no longer trivial. Indeed, its equation in the laboratory frame is

$$\ddot{x} + \left(1 - \dot{x}^2\right)^{\frac{3}{2}} x = 0.$$

This equation is not solvable using the Einstein dynamical equation

$$F = \frac{d}{dt} (mv), m = \frac{cm_0}{(c^2 - v^2)^{\frac{1}{2}}}.$$

Thus, Special Relativity must be modified now to include identical computers in the Lab and Rocket frames. For such problems we have developed methodology [19]

which not only solves oscillator problems numerically but produces numerical results which themselves are related by the fundamental Lorentz transformation.

Our present research revolves around living cell processes. We are concentrating on chemical bonding and the dynamics of the water molecule, both of which are fundamental in life processes. We have developed a new model of the bonding process [20], which has shown to be applicable to all diatomic molecules through O_2 and has now enabled us to reproduce the esoteric 104.5° bond angle of the water molecule [21].

REFERENCES

- [1] Hirschfelder, J., Curtiss, C., and Bird, R., *Molecular Theory of Gases and Liquids*, Wiley, New York, 1965.
- [2] Greenspan, D., *Arithmetic Applied Mathematics*, Pergamon, Oxford, 1980.
- [3] Greenspan, D. and Heath, L., Supercomputer simulation of the modes of colliding microdrops of water, *J. Phys. D*, 24, 1991, 2121.
- [4] Greenspan, D., Supercomputer simulation of cracks and fractures by quasimolecular dynamics, *J. Phys. Chem. Solids*, 50, 1989, 1245.
- [5] Greenspan, D., Quasimolecular simulation of large liquid drops, *J. Phys. D*, 22, 1989, 1415.
- [6] Greenspan, D., Particle simulation of large carbon dioxide bubbles in water, *Appl. Math. Mod.*, 19, 1995, 738.
- [7] Korlie, M., Ph.D. thesis, Mathematics, UT Arlington, 1996.
- [8] Greenspan, D., Particle modelling of cavity flow on a vector computer, *Comp. Meth. Appl. Math. Eng.*, 66, 1988, 291.
- [9] Greenspan, D., Particle simulation of biological sorting on a supercomputer, *Comp. Math. Applic.*, 18, 1989, 823.
- [10] Greenspan, D., Mechanisms of capillarity via supercomputer simulation, *Comp. Math. Applic.*, 16, 1988, 141.
- [11] Greenspan, D. and Casulli, V., Particle modelling of an elastic arch, *Appl. Math. Mod.*, 9, 1985, 215.
- [12] Greenspan, D., Computer-oriented n-body modelling of minimal surfaces, *Appl. Math. Mod.*, 7, 1983, 423.

- [13] Coppin, C. and Greenspan, D., A contribution to the modelling of soap films, *Appl. Math. Comp.*, 26, 1988, 315.
- [14] Greenspan, D., TR 167, Comp. Sci. Dept., UW Madison, 1972.
- [15] Greenspan, D., Quasimolecular channel and vortex modelling on a supercomputer, *Comp. Math. Applic.*, 15, 1988, 331.
- [16] Goldstine, H., *Classical Mechanics*, 2nd edition, A.-W., Reading, 1980.
- [17] Greenspan, D., Completely conservative, covariant numerical methodology, *Comp. Math. Applic.*, 29, 1995, 37.
- [18] Simo, J. and Tarnow, N., The discrete energy-momentum method. Conserving algorithms for nonlinear elasodynamics, *ZAMP*, 43, 1992, 757.
- [19] Greenspan, D., Covariant computation in special relativistic dynamics, *Physica Scripta*, 52, 1995, 353.
- [20] Greenspan, D., Electron attraction as a mechanism for the chemical bond of ground state H_2 , *Physica Scripta*, 52, 1995, 267.
- [21] Greenspan, D., A semiclassical, dynamical model of the water molecule, *Physica Scripta*, 54, 1996, 458.

11 ERGODIC TYPE SOLUTIONS OF SOME DIFFERENTIAL EQUATIONS

Jialin Hong

Institute of Computational Mathematics and
Scientific/Engineering Computing
Chinese Academy of Sciences
P.O. Box 2719
Beijing 100080, P.R.China

and

Rafael Obaya
Departamento de Matematica
Aplicada a la Ingenieria
Universidad de Valladolid
Valladolid 47011, Spain

1. INTRODUCTION

The existence of ergodic solutions, which have mean values of differential equations has received much attention because of its obvious practical importance in many fields of applied sciences. As it is well-known, the theory of almost periodic solutions including the periodic and quasi-periodic cases, has been researched in a lot of references by mathematicians (see [10, 12, 18, 21, 23, 32, 36, 37, 41-43]). Recently, some new ergodic type functions have been introduced and applied to differential equations in [5 ,6, 11, 44-46]. In this summary we pay our attention to the new ergodic type solutions of some differential equations.

1.1 Definitions of Ergodic Type Functions.

Let $C(R, R^d)$ (respectively $C(R \times \Omega, R^d)$, where $\Omega \subset R^d$) denote the Banach space of bounded continuous functions $\varphi(t)$ (resp., $\varphi(t, x)$) from R (resp., $R \times \Omega$) to R^d

with norm $\|\varphi\| = \sup_{t \in R} |\varphi(t)|$ (resp., $\|\varphi\| = \sup_{t \in R, x \in \Omega} |\varphi(t, x)|$). Let $L(R, R^d)$ (resp., $L(R \times \Omega, R^d)$) denote the space of all Lebesgue measurable and bounded functions $f(t)$ (resp., $f(t, x)$) from R (resp., $R \times \Omega$) to R^d .

Definition 1.1 [18,21]. A continuous function $f : R \rightarrow R^d$ is called almost periodic if the ε -translation set of f

$$T(f, \varepsilon) = \left\{ \tau \in R : |f(t + \tau) - f(t)| < \varepsilon \text{ for all } t \in R \right\}$$

is a relatively dense set in R . τ is called the ε -period for f . Denote by $AP(R, R^d)$ the set of all such functions.

Definition 1.2 [18, 21]. A continuous function $g : R \times R^d \times R^d \rightarrow R^d$ is called an almost periodic function for t uniformly on $R^d \times R^d$, if for any compact subset $W \subset R^d \times R^d$, the ε -translation set of g

$$T(g, \varepsilon, W) = \left\{ \tau \in R : |g(t + \tau, x, y) - g(t, x, y)| < \varepsilon \text{ for all } (t, x, y) \in R \times W \right\}$$

is a relatively dense set in R . τ is called the ε -period for g .

Definition 1.3 [5, 6, 11, 44-46]. A function $f \in C(R, R^d) \cap C(R \times \Omega, R^d)$ is called pseudo almost periodic if $f = f_1 + f_0$, where f_1 is almost periodic in $t \in R$ (almost periodic in $t \in R$, uniformly in $x \in \Omega$), and $f_0 \in PAP_0(R, R^d) \cap PAP_0(R \times \Omega, R^d)$, where

$$PAP_0(R, R^d) = \left\{ \varphi \in C(R, R^d) : m(|\varphi|) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\varphi(t)| dt = 0 \right\},$$

$$PAP_0(R \times \Omega, R^d) = \left\{ \varphi \in C(R \times \Omega, R^d) : m(|\varphi|) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\varphi(t, x)| dt = 0 \right. \\ \left. \text{uniformly in } x \in \Omega \right\}.$$

$PAP_0(R, R^d)$ is a translation invariant closed ideal of $C(R, R^d)$. f_1 and f_0 (uniquely determined, see [6, p. 1143]), are called the almost periodic component and the ergodic perturbation, respectively, of the function f . Denote by $PAP_0(R, R^d) \cap PAP_0(R \times \Omega, R^d)$ the set of all such functions f .

Definition 1.4 [5, 6]. A Lebesgue measurable function f from R to R^d (resp., from $R \times \Omega$ to R^d) is called generalized pseudo almost periodic if $f = f_1 + f_0$, where f_1 is almost periodic in $t \in R$ (almost periodic in $t \in R$, uniformly in $x \in \Omega$), and

$f_0 \in \tilde{PAP}_0(R, R^d) \left(\tilde{PAP}_0(R \times \Omega, R^d) \right)$, where $\tilde{PAP}_0(R, R^d) = \left\{ \varphi : R \rightarrow R^d \text{ Lebesgue measurable and } m(|\varphi|) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T |\varphi(t)| dt = 0 \right\}$,

$$\tilde{PAP}_0(R \times \Omega, R^d) = \left\{ \begin{array}{l} \varphi : R \times \Omega \rightarrow R^d, \text{ such that } \varphi(., x) \in \tilde{PAP}_0(R, R^d), \text{ for each } x \in \Omega \\ m(|\varphi|) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T |\varphi(t, x)| dt = 0 \text{ uniformly in } x \in \Omega \end{array} \right\}$$

f_1 and f_0 are called the almost periodic component and the ergodic perturbation, respectively, of the function f . Denote by $\tilde{PAP}(R, R^d) \left(\tilde{PAP}(R \times \Omega, R^d) \right)$ the set of all such functions f .

Definition 1.5 [28]. A function $f \in L(R \times R^d)$ is said to be ergodic if the following limit exists:

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T f(t) dt = M(f).$$

Analogously, and in accordance with [5, 6, 11, 44-46], we say that a function $f \in L(R \times \Omega, R^d)$ is ergodic if for every compact subset $K \subset \Omega$ the following limit exists uniformly for $x \in K$:

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T f(t, x) dt = M(f, x).$$

We will write $\varepsilon(R, R^d)$ and $\varepsilon(R \times \Omega, R^d)$ for the sets of ergodic functions defined on R and $R \times \Omega$ respectively.

We have

$$PAP(R, R^d) \subset \tilde{PAP}(R, R^d),$$

$$P(R, R^d) \subset QP(R, R^d) \subset AP(R, R^d) \subset PAP(R, R^d) \subset \varepsilon(R, R^d),$$

where $P(R, R^d) = \{f \in C(R, R^d) : f \text{ is a periodic function on } R\}$ and $QP(R, R^d) = \{f \in C(R, R^d) : f \text{ is a quasiperiodic function on } R\}$.

As it is well-known, there are uniformly continuous bounded functions on R which are not ergodic.

For instance, consider

$$f(t) = \begin{cases} 1 - t^2, & |t| < 1 \\ \sin \ln\left(\frac{1}{t^2}\right), & |t| \geq 1. \end{cases}$$

Obviously $f \in C(R, [-1, 1])$. Since the derivative of f is bounded (and uniformly continuous) on R , f is uniformly continuous on R . However, the following equality

$$\frac{1}{2T} \int_{-T}^T f(t) dt = \frac{1}{5} \left(\sin\left(\ln \frac{1}{T^2}\right) + 2 \cos\left(\ln \frac{1}{T^2}\right) \right) + \frac{4}{15T}, \text{ for every } T > 1,$$

implies that f is not ergodic.

In order to study ergodic solutions of differential equations via ergodic sequences, we introduce the concept of a T -ergodic function. We will write M for the Lebesgue measure on $[-1, 0]$. Let T be the translation operator $T : [-1, 0] \rightarrow [0, 1]$ given by $T(\tau) = 1 + \tau$.

Definition 1.6. A function $f \in L(R, R^d)$ is called T -ergodic if the following limit exists in lebesgue-measure on $[-1, 0]$:

$$\lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{k=-n}^n f(T^k \tau),$$

where T^k denote the composition operator, $T^0 \tau = \tau$ and $T^{-1} \tau = \tau - 1$. For a function $f \in L(R \times \Omega, R^d)$ we say that it is T -ergodic if for each compact subset K of Ω , the limit

$$\lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{k=-n}^n f(T^k \tau, x) = M(\tau, x)$$

exists in Lebesgue-measure on $[-1, 0]$, uniformly for $x \in K$.

The property of T -ergodicity implies ergodicity but the converse is not true.

1.2 An Example.

An example which shows that the function $\varphi(t) = t |\sin \pi t|^{t^N}$, $N > 6$, is unbounded and $\varphi \in \tilde{PAP}_0(R, R)$, is given in [6, pp.1143-1146] with a long and interesting proof.

Now we give another example of an unbounded continuous function $f \in \tilde{PAP}_0(R, R)$.

Example 1.1. Let

$$f(t) = \begin{cases} 2k(t - 2^{k^2} + 1), & 2^{k^2} - 1 \leq t \leq 2^{k^2} - \frac{1}{2}, \\ -2k(t - 2^{k^2}), & 2^{k^2} - \frac{1}{2} \leq t \leq 2^{k^2}, \\ 0, & \text{otherwise,} \end{cases}$$

where $k \in Z^+ = \{1, 2, 3, \dots\}$. Then $f \in \tilde{PAP}_0(R, R)$ and f is unbounded.

2. EXPONENTIAL TRICHOTOMY AND ERGODIC SOLUTIONS FOR ODE

Consider the linear differential equations

$$x'(t) = A(t)x(t), \quad t \in R, \quad (2.1)$$

$$x'(t) = A(t)x(t) + f(t), \quad t \in R, \quad (2.2)$$

where $A(t)$ is a Lebesgue measurable and bounded $n \times n$ matrix function on R ([22, pp. 28-30]), and $f \in \tilde{PAP}_0$. In order to give our main results, we need the following definition.

Definition 2.1. The equation (2.1) is said to admit an exponential trichotomy on R (see [20, 25]) if there exist linear projection P, Q such that $PQ = QP, P + Q - PQ = I$ (2.3) and constants $K \geq 1, \alpha > 0$ such that

$$\begin{cases} |X(t)PX^{-1}(s)| \leq Ke^{-\alpha(t-s)} & \text{for } 0 \leq s \leq t \\ |X(t)(I-P)X^{-1}(s)| \leq Ke^{-\alpha(s-t)} & \text{for } t \leq s, s \geq 0 \\ |X(t)QX^{-1}(s)| \leq Ke^{-\alpha(s-t)} & \text{for } t \leq s \leq 0 \\ |X(t)(I-Q)X^{-1}(s)| \leq Ke^{-\alpha(t-s)} & \text{for } s \leq t, s \leq 0, \end{cases} \quad (2.4)$$

where $X(t)$ is the fundamental matrix of (2.1) with $X(0) = I$, I the $n \times n$ unit matrix, $|\cdot|$ the Euclidean norm.

Remarks.

- (1) In the above definition, if $Q = I - P$, then the equation (2.1) admits exponential dichotomy on R ([19—24]). but no exponential dichotomy on
- (2) The equation (2.1) admits an exponential trichotomy on R with projections P and Q if and only if it has an exponential dichotomy on R^\pm (that is, exponential dichotomy on R^+ and on R^-) with projections $P_+ = P$ and $P_- = I - Q$, respectively, such that $P_+P_- = P_-P_+ = P_-$.

Proposition 2.1 [5]. Let $A(t)$ be an almost periodic matrix. Suppose that the system (2.1) satisfies an exponential dichotomy and suppose that $f \in \tilde{PAP}(R, R^n)$. Then (2.2) has a generalized pseudo almost periodic solution. Furthermore if $f \in PAP(R, R^n)$, then the pseudo almost periodic solution is unique, and one has

$$\|x\| \leq (2K/\alpha)\|f\| \text{ and } \text{mod}(x) \subset \text{mod}(A) + \text{md}(f).$$

Theorem 2.1 [25]. If the equation (2.1) admits an exponential trichotomy on R with projections P, Q and constants $K \geq 1$ and $\alpha > 0$, then the equation (2.2) has at least one ergodic solution $x \in \tilde{P}AP_0(PAP_0)$ for every $f \in \tilde{P}AP_0(PAP_0)$ for every $f \in \tilde{P}AP_0(PAP_0 \text{ or } \tilde{P}AP_0 \cap M_b)$ where $M_b = \{\varphi : R \rightarrow R^n, \text{ Lebesgue measurable and bounded}\}$.

Theorem 2.2 [26]. If equation (2.2) has at least one ergodic solution $x \in PAP_0(R)$ for every $f \in PAP_0(R)$, then equation (2.1) has an exponential trichotomy on R .

Now we consider the nonlinear differential equation

$$x'(t) = A(t)x(t) + f(t, x), \quad t \in R, \quad (2.5)$$

where $A(t)$ is a Lebesgue measurable and bounded $n \times n$ matrix function on R .

Theorem 2.3 [25]. Let the system (2.1) admit an exponential trichotomy on R with projections P, Q and constants $K \geq 1$ and $\alpha > 0$. Suppose $f(t, 0) \in \tilde{P}AP_0$, $f(t, x)$ is a bounded function from $R \times R^n$ to R^n , Lebesgue measurable in t for each fixed $x \in R^n$, and satisfies a Lipschitz condition in x , that is, there exists a constant $L > 0$ such that $|f(t, x_1) - f(t, x_2)| \leq L|x_1 - x_2|$ for every $(t, x_1), (t, x_2) \in R \times R^n$. If $0 < L < \frac{1}{2K}(1 - e^{-\alpha})$, then the equation (2.5) has a solution $x \in PAP_0$ with $|x(t)| \leq \frac{2MK}{1 - e^{-\alpha}}$, where $M = \sup_{t \in R, x \in R^n} |f(t, x)|$.

Now we apply Theorem 2.1 to Hill equation with $\tilde{P}AP_0$ forcing function

$$x''(t) + (a + bp(t))x(x) = f(t), \quad t \in R, \quad (2.6)$$

where $p(t)$ is a Lebesgue measurable and bounded function from R to R , $f \in \tilde{P}AP_0(PAP_0)$. The equation (2.6) and the corresponding homogeneous equation are respectively equivalent to the following two equations

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -(a + bp(t)) & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ f(t) \end{pmatrix} \quad (2.7)$$

and

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -(a + bp(t)) & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (2.8)$$

Theorem 2.4 [25]. If one of the following conditions holds: (1) $b \geq 0, a + p_s b < 0$; (2) $b < 0, a + p_l b > 0$, then the equation (2.7) equivalent to the Hill equation (2.6) has a unstable solution $(x_0(t), y_0(t))^T \in \tilde{P}AP_0(PAP_0)$ for every $f \in \tilde{P}AP(PAP_0)$, where $p_s = \sup_{t \in R} p(t), p_l = \inf_{t \in R} p(t)$.

3. SOME ERGODIC TYPE SEQUENCES

3.1 Almost Periodic Sequences.

Definition 3.1 [42, p. 1442]. A sequence $x : \mathbb{Z} \rightarrow \mathbb{R}^d$ is called an almost periodic sequence if the ε -translation set of x

$$T(x, \varepsilon) := \left\{ \tau \in \mathbb{Z} : |x(n + \tau) - x(n)| < \varepsilon \text{ for all } n \in \mathbb{Z} \right\}$$

is a relatively dense set in \mathbb{Z} . τ is called the ε -period for x .

Proposition 3.1 [32, 42, 43].

- (i) If a_n is an almost periodic sequence, then there exists an almost periodic function $f(t)$ such that $f(n) = a_n, \forall n \in \mathbb{Z}$;
- (ii) If f is an almost periodic function, then $\{f(n)\}$ is an almost periodic sequence.

3.2 Pseudo Almost Periodic and Generalized Pseudo Almost Periodic Sequences.

Definition 3.2 [7, 8, 24, 27].

- (1) A sequence $x : \mathbb{Z} \rightarrow \mathbb{R}^d$ is said to be a $\tilde{PAP}_0(PAP_0)$ sequence if it satisfies
$$\lim_{n \rightarrow +\infty} \frac{1}{2n} \sum_{k=-n}^n |x(k)| = 0 \quad \left(\lim_{n \rightarrow +\infty} \frac{1}{2n} \sum_{k=-n}^n |x(k)| = 0 \text{ and it is bounded} \right).$$
- (2) A sequence $x : \mathbb{Z} \rightarrow \mathbb{R}^d$ is said to be a $\tilde{PAP}(PAP)$ sequence if $x = x_1 + x_0$, where x_1 is an almost periodic sequence and x_0 is a $\tilde{PAP}_0(PAP_0)$ sequence. x_1 and x_0 (uniquely determined, see Proposition 2.4) are called the almost periodic component and the ergodic perturbation, respectively, of x .

Remarks 3.1.

- (1) A sequence vanishing at infinity is a PAP_0 sequence. The PAP_0 sequence $x(n) = \begin{cases} 1 & n = 2^k, \\ 0 & \text{otherwise} \end{cases}$ show that a PAP_0 sequence is, in general, not a sequence vanishing at infinity.
- (2) The sequence $x(n) = \begin{cases} k & n = 2^{k^2}, \\ 0 & \text{otherwise} \end{cases}, \quad k \in \mathbb{Z}^+$, is an example of an unbounded \tilde{PAP}_0 sequence.

Proposition 3.2 [7, 8, 24, 27]. Suppose that $\{x(n)\}_{n \in \mathbb{Z}}$ is a $\tilde{PAP}_0(PAP_0)$ sequence. Then there exists a function $f \in \tilde{PAP}_0(R, R^d)(PAP_0(R, R^d))$ such that $f(n) = x(n)$, $n \in \mathbb{Z}$.

Remark 3.2. The converse proposition of the above proposition is not true. We consider the function

$$f(t) = \begin{cases} 2^{|k|}(t-k)+1, & t \in [k-2^{-|k|}, k], \\ -2^{|k|}(t-k)+1, & t \in [k, k+2^{-|k|}], \\ 0 & \text{otherwise} \end{cases}$$

where $k \neq 0$. Clearly, $f \in C(R, R)$ and $\int_{-\infty}^{+\infty} |f(t)| dt = \int_{-\infty}^{+\infty} f(t) dt = 2$, thus $m(\|f\|) = 0$,

that is $f \in PAP_0(R, R)$. But $f(k) = 1$, for every $k \in \mathbb{Z} - \{0\}$. It follows that the sequence $\{f(k)\}$ is not a PAP_0 sequence.

The following proposition are proved in [7, 8, 24, 27].

Proposition 3.3. If $\{x(n)\}_{n \in \mathbb{Z}}$ is a $\tilde{PAP}(PAP)$ sequence, then there exists $f \in \tilde{PAP}(R, R^d)(PAP(R, R^d))$ such that $f(n) = x(n)$, $n \in \mathbb{Z}$.

Proposition 3.4. Let $x = \{x(n)\}_{n \in \mathbb{Z}}$ be a \tilde{PAP} sequence. Then its almost periodic component x_1 and its ergodic perturbation x_0 are uniquely determined in terms of the sequence x .

Proposition 3.5. Let $x = \{x(n)\}_{n \in \mathbb{Z}}$ be a \tilde{PAP}_0 sequence. Then the series

$$\sum_{k \in \mathbb{Z}} |x(k)| e^{-\alpha|k|}$$

converges for every $\alpha > 0$.

3.3 Ergodic Sequences.

Definition 3.3 [28]. A bounded sequence $\{x(n)\}_{n \in \mathbb{Z}}$ is said to be ergodic if the limit

$$\lim_{n \rightarrow +\infty} \frac{1}{2n} \sum_{k=-n}^n x(k)$$

exists. Denote by $\varepsilon(\mathbb{Z})$ the set of all such sequences.

Proposition 3.6 [28].

- (i) If $\{x(n)\}_{n \in \mathbb{Z}} \in \mathcal{E}(Z)$, then there exists an ergodic function $f \in \mathcal{E}(R, R^d)$ such that $f(n) = x(n)$ for every $n \in \mathbb{Z}$.
- (ii) There exists functions $f \in \mathcal{E}(R, R^d)$ such that the corresponding sequences $\{f(n)\}_{n \in \mathbb{Z}}$ are not ergodic.

4. ERGODIC TYPE SOLUTIONS VIA ERGODIC TYPE SEQUENCES

G.H. Meisters [32] showed that the existence of almost periodic solutions of ordinary differential equations (in short, ODE) is equivalent to the fact that the restriction of a bounded solution to some discrete subgroup of real is almost periodic. This is improved by Opial (see Fink [21,pp.164-169]). A natural question is:

Are there similar results on the existence of almost periodic, asymptotically almost periodic, pseudo almost periodic and ergodic solutions of differential equations with PCA?

In this part we give an affirmative answer to this question under similar conditions to those of G.H. Meisters, Z. Opial and A.M. Fink.

Consider the differential equation with PCA

$$\frac{dx}{dt} = f(t, x(t), x([t]), x([t-1]), \dots, x([t-k])), \quad t \in J, \quad (4.1)$$

where k is a positive integer, $f \in C(J \times \Omega, R^d)$, $[\cdot]$ denotes the greatest integer function. Obviously, if the right hand side of equation (4.1) depends only on the terms t and $x(t)$, then it is an ODE. A function $x : J \rightarrow R^d$ is called a solution of equation (4.1) if the following conditions are satisfied (see [3, 4, 13-16, 34, 35, 39-43]):

- (1) x is continuous on J ;
- (2) the derivative $x'(t)$ of $x(t)$ exists everywhere, with possible exception of the point $|t|$, where one-sided derivatives exist;
- (3) x satisfies equation (4.1) on each interval $[n, n+1] \subset J$.

Our main results in this section are the following.

Theorem 4.1 [7]. Let $f \in AP(R \times \Omega)$ in equation (4.1) for a compact subset $\Omega_0 \subset \Omega$. If all equations

$$\frac{dx}{dt} = g(t, x(t), x([t]), x([t-1]), \dots, x([t-k])), \quad t \in R, \quad (4.2)$$

with $g \in H(f)$ have unique solutions to initial value problems, where the initial value condition is $x(j) = x_j, j = 0, -1, -2, \dots, -k$, and $\varphi(t)$ is a solution of (4.1) with $\varphi(R)^{k+2} \subset \Omega_0$ then $\varphi \in AP(R)$ if and only if $\{\varphi(n)\}_{n \in \mathbb{Z}} \in AP(Z)$.

Theorem 4.2 [7]. Let $f \in PAP(R \times \Omega_0)$ ($AAP(R^+ \times \Omega_0)$) in equation (4.1) for a compact subset $\Omega_0 \subset \Omega$, and suppose f and its almost periodic component f_t satisfy a Lipschitz condition on Ω_0 with Lipschitz constant L . If $\varphi(t)$ is a solution of equation (4.1) with $\varphi(R)^{k+2} \subset \Omega_0$, then $\varphi \in PAP(R)$ ($AAP(R^+)$) if and only if $\{\varphi(n)\}_{n \in \mathbb{Z}} \in PAP(Z)$ ($AAP(Z^+)$).

For the ordinary differential equation

$$x'(t) = f(t, x(t)), \quad t \in R, \quad (4.3)$$

we have the following results.

Theorem 4.3 [28]. Suppose that for the function $f \in L(R \times \Omega, R^d)$ satisfy the Lipschitz condition

$$|f(t, y_1) - f(t, y_2)| \leq L(t)|y_1 - y_2|, \quad y_1, y_2 \in \Omega, \quad t \in R,$$

where the nonnegative function $L(\cdot) \in \tilde{P}AP_0(R, R)$ and suppose that for at least one point $y \in \Omega$, $\{f(\cdot, y)\}$ is τ -ergodic. Then a solution $x(t)$ of the equation (4.3) is ergodic if and only if the sequence $\{x(n)\}_{n \in \mathbb{Z}}$ is ergodic.

Theorem 4.4 [28]. Let $f(t, x) \in L(R \times \Omega, R^d)$ be uniformly continuous in t for x in compact subsets of Ω and suppose it satisfies a Lipschitz condition with Lipschitz constant $L > 0$. Furthermore assume f is such that for every Lebesgue measurable set $E \subset R$, the limit

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{[-T, T] \cap E} f(t, y) dt$$

exists for each $y \in \Omega$. Then if $x(t)$ is a solution of equation (4.3), $x \in \varepsilon(R, R^d)$ if and only if the sequence $\{x(n)\}_{n \in \mathbb{Z}} \in \varepsilon(Z)$.

An Example [7]. Consider the equation $\frac{dx}{dt} = x(t)(a(t) - b(t)x([t]))$, where

$a(t)$ and $b(t)$ are positive and continuous bounded functions on R^+ , and satisfy $\int_n^{n+1} a(t) dt = \int_n^{n+1} b(t) dt$ for every $n \in \mathbb{Z}^+$. Now we investigate the existence of almost

periodic type solutions of this equation. It follows from

$$x(n+1) = x(n) e^{\int_{a(t)}^{n+1} e^{-x(s)} \int_n^{s+1} b(t) dt}, \quad \text{for } n \in \mathbb{Z}^+$$

that if we take $x(0) = 1$, then for every $n \in \mathbb{Z}^+$, we get $x(n) = 1$. According to Theorem 4.1 and Theorem 4.2, it is concluded that the equation has a solution $x \in AP(R^+)$ (resp., $x \in PAP(R^+)$) with $x(0) = 1$ if the functions $a, b \in AP(R^+)$ (resp., $a, b \in PAP(R^+)$). The above equation is analogous to the famous logistic differential equation, but t in one argument has been replaced by $[t]$. As a result, the equation has solutions that display complicated dynamics even if $a(t) = b(t) = \text{cons.}$ (see [7, 40]).

5. SOME ERGODIC TYPE SOLUTIONS OF A CLASS OF DE WITH PCA

In this part we study the existence of pseudo almost periodic solutions and generalized pseudo almost periodic solutions for the following differential equations with pca

$$x'(t) = ax(t) + \sum_{i=-N}^N a_i x([t+i]) + f(t), \quad N \geq 2, \quad (5.1)$$

$$x'(t) = ax(t) + \sum_{i=-N}^N a_i x([t+i]) + g(t, x(t), x([t])), \quad N \geq 2 \quad (5.2)$$

where a, a_i are constants, $f \in \tilde{PAP}(R, R)$ ($PAP(R, R)$), $g \in \tilde{PAP}(R \times R^2, R)$ ($PAP(R \times R^2, R)$) is bounded, and there exists a constant $\eta > 0$ such that $|g(t, x_1, y_1) - g(t, x_2, y_2)| \leq \eta(|x_1 - x_2| + |y_1 - y_2|)$, for every $(t, x_1, y_1), (t, x_2, y_2) \in R \times R^2$. $\quad (5.3)$

Obviously, if $x(t)$ is a solution of equation (5.1) on R , then

$$x(t) = e^{a(t-n)} c_n + (e^{a(t-n)} - 1) \sum_{i=-N}^N a^{-1} a_i c_{n+i} + \int_n^t e^{a(t-s)} f(s) ds, \quad n \leq t < n+, \quad (5.4)$$

where $x(n+i) = c_{n+i}$, $-N \leq i \leq N$ (the discussion is the same for the case $a = 0$).

By using the continuity of solution at any point, we get the following difference equation

$$c_{n+1} = e^a c_n + \sum_{i=-N}^N (e^a - 1) a^{-1} a_i c_{n+i} + \int_n^{n+1} e^{a(n+1-s)} f(s) ds, \quad n \in \mathbb{Z}. \quad (5.5)$$

Let

$$\begin{aligned} b_0 &= e^a + a^{-1} a_0 (e^a - 1), & b_1 &= a^{-1} a^1 (e^a - 1) - 1, \\ b_i &= a^{-1} a_i (e^a - 1), & i &= -1, \pm 2, \dots, \pm N, \end{aligned}$$

$$h_n = - \int_{-n}^{n+1} e^{a(n+1-s)} f(s) ds.$$

Then (5.5) becomes

$$\sum_{i=-N}^N b_i c_{n+i} = h_n, \quad (5.6)$$

The corresponding homogeneous equation to equation(5.6) is

$$\sum_{i=-N}^N b_i c_{n+i} = 0, \quad (5.7)$$

According to [41], we can get the particular solutions as $c_n = \lambda^n$ for the homogeneous difference equation (5.7). At this time, λ will satisfy the following equation

$$\sum_{i=-N}^N b_i \lambda^{n+i} = 0. \quad (5.8)$$

Our main results are as follows.

Theorem 5.1 [8]. Suppose that all roots of equation (5.8) are simple (denoted by $\lambda_1, \dots, \lambda_{2N}$) and $|\lambda_i| \neq 1$, $1 \leq i \leq 2N$. Then

- (1) for any $f \in \tilde{PAP}_0(R, R)$ ($PAP_0(R, R)$ or $\tilde{PAP}_0(R, R) \cap M_b(R, R)$), equation (5.1) has a solution $x \in \tilde{PAP}_0(R, R) \setminus PAP_0(R, R)$, furthermore, x is unique if $f \in PAP_0(R, R)$ or $f \in \tilde{PAP}_0(R, R) \cap M_b(R, R)$, where $M_b(R, R) = \{\varphi : R \rightarrow R$ is Lebesgue measurable and bounded on $R\}$;
- (2) for any $f \in \tilde{PAP}(R, R)$ ($PAP(R, R)$ or $\tilde{PAP}(R, R) \cap M_b(R, R)$), equation (5.1) has a solution $x \in \tilde{PAP}(R, R) \setminus PAP(R, R)$, and x is unique if $f \in PAP(R, R)$ or $f \in \tilde{PAP}(R, R) \cap M_b(R, R)$.

Theorem 5.2 [8]. Suppose that all roots of equation (5.8) are simple (denoted by $\lambda_1, \dots, \lambda_{2N}$) and $|\lambda_i| \neq 1$, $1 \leq i \leq 2N$. Then there exists $\eta_* > 0$, such that

- (1) when $0 \leq \eta < \eta_*$, equation (5.2) has a unique solution $x \in PAP_0(R, R)$ if $g \in \tilde{PAP}_0(R \times R^2, R)$ is bounded and satisfies the Lipschitz condition (5.3);
- (2) when $0 \leq \eta < \eta_*$, equation (5.2) has a unique solution $x \in PAP(R, R)$ if $g \in \tilde{PAP}(R \times R^2, R)$ is bounded and satisfies the Lipschitz condition (5.3).

6. ERGODIC TYPE DIFFERENCE EQUATIONS AND EXPONENTIAL DICHOTOMY

Consider the following difference equation

$$y(n+1) = C(n)y(n) \quad n \in \mathbb{Z}, \quad (6.1)$$

where $C(n)$ is an invertible $d \times d$ matrix on \mathbb{Z} .

Equation (6.1) is said to admit an exponential dichotomy on \mathbb{Z} if there exist positive constants $K \geq 1$, $\alpha > 0$ and a projection $P (P^2 = P)$ such that

$$\begin{cases} |Y(n)PY^{-1}(m)| \leq Ke^{-\alpha(n-m)} & n \geq m \\ |Y(n)(I-P)Y^{-1}(m)| \leq Ke^{-\alpha(m-n)} & m \geq n, \end{cases} \quad (6.2)$$

where $Y(n)$ is the fundamental matrix solution of equation (6.1) with $Y(0) = I$ (see [33, 42, 43]).

6.1 Almost Periodic Difference Equations.

Proposition 6.1 [33]. Suppose that the linear difference equation (6.1) has an exponential dichotomy on \mathbb{Z} . Then equation (6.1) has no nontrivial solution bounded on \mathbb{Z} .

Proposition 6.2 [33]. Suppose that the linear difference equation (6.1) has an exponential dichotomy on \mathbb{Z} with constants K and α . Let $\{h(n)\}_{n \in \mathbb{Z}}$ be a bounded sequence. Then the inhomogeneous difference equation

$$y(n+1) = C(n)y(n) + h(n) \quad n \in \mathbb{Z}, \quad (6.3)$$

has a unique solution $\{y(n)\}$ bounded on \mathbb{Z} . Moreover for all n

$$|y(n)| \leq K(1 + e^{-\alpha})(1 - e^{-\alpha})^{-1} \sup_{n \in \mathbb{Z}} |h(n)|. \quad (6.4)$$

Proposition 6.3 [42, 43]. Suppose that the linear difference equation (6.1) has an exponential dichotomy on \mathbb{Z} with projection P and constants K, α . Let $k' = \{k'_i\} \subset \mathbb{Z}$ be a sequence such that $T_{k'} C(n) = D(n)$ uniformly on \mathbb{Z} . Then there exists a subsequence $k = \{k_i\}$ such that $Y(k_i)PY^{-1}(k_i) \rightarrow Q$ and the linear difference equation

$$z(n+1) = D(n)z(n) \quad (6.5)$$

has also an exponential dichotomy on \mathbb{Z} with projection Q and the same constants K, α .

Proposition 6.4 [42, 43]. Suppose that $\{C(n)\}$, $\{h(n)\}$ are almost periodic sequences and the linear difference equation (6.1) has an exponential dichotomy on Z . Then the inhomogeneous difference equation (6.3) has a unique almost periodic sequence solution.

6.2 PAP and $\tilde{P}AP$ Difference Equations and Exponential Dichotomy.

Proposition 6.5 [24, 27]. Suppose that the linear difference equation (6.1) has an exponential dichotomy on Z . Then

- (1) for any $\tilde{P}AP_0(PAP_0)$ sequence $\{h(n)\}$, equation (6.3) has a $\tilde{P}AP_0(PAP_0)$ solution $\{y(n)\}$, furthermore, $\{y(n)\}$ is unique if $\{h(n)\}$ is a PAP_0 sequence;
- (2) for any almost periodic matrix sequence $\{C(n)\}$, where $C(n)$ is invertible, and any $\tilde{P}AP(PAP)$ sequence $\{h(n)\}$, equation (6.3) has a $\tilde{P}AP_0(PAP)$ solution $\{y(n)\}$, and $\{y(n)\}$ is unique if $\{h(n)\}$ is a PAP sequence.

More results can be found in [1, 2, 9, 19, 27, 29, 30, 33].

7 . NEUTRAL DIFFERENTIAL EQUATIONS WITH PIECEWISE CONSTANT ARGUMENT

Consider the inhomogeneous neutral differential equations with piecewise constant argument of the form

$$y'(t) = A(t)y(t) + B(t)y([t]) + A_0(t)y(t-[t]) + A_1(t)y'(t-[t]) + f(t), \quad t \in R, \quad (7.1)$$

and the nonlinear neutral differential equations with piecewise constant argument of the form

$$y'(t) = A(t)y(t) + B(t)y([t]) + A_0(t)y(t-[t]) + A_1(t)y'(t-[t]) + g(t, y(t), y([t])), \quad t \in R, \quad (7.2)$$

where $A, B, A_0, A_1 : R \rightarrow R^{d \times d}$, $f : R \rightarrow R^d$, $g : R \times R^d \times R^d \rightarrow R^d$ are Lebesgue measurable.

Let $X(t)$ be the fundamental matrix solution of

$$x'(t) = A(t)x(t), \quad t \in R, \quad (7.3)$$

such that $X(0) = I$, I is the identity matrix.

Obviously, if $y(t)$ is a solution of equation (7.1), then $\{y(n)\}_{n \in Z}$ satisfies the inhomogeneous difference equation

$$y(n+1) = C(n)y(n) + h(n), \quad n \in Z, \quad (7.4)$$

with

$$C(n) = X(n+1) \left[X^{-1}(n) + \int_n^{n+1} X^{-1}(u) B(u) du \right],$$

$$h(n) = X(n+1) \int_n^{n+1} X^{-1}(u) [A_0(u)y_0(u-n) + A_1(u)y'_0(u-n) + f(u)] du,$$

where $y_0(t)$ is the solution of the following equation

$$y'_0(t) = A(t)y_0(t) + B(t)y_0(0) + A_0(t)y_0(t) + A_1(t)y'_0(t) + f(t). \quad (7.5)$$

We assume that for each $n \in \mathbb{Z}$

$$C(n) = X(n+1) \left[X^{-1}(n) + \int_n^{n+1} X^{-1}(u) B(u) du \right]$$

is an invertible $d \times d$ matrix.

Definition 7.1 [24, 40, 42, 43]. The linear differential equation with piecewise constant argument

$$y'(t) = A(t)y(t) + B(t)y([t]) \quad (7.6)$$

is said to have an exponential dichotomy if the difference equation

$$y(n+1) = C(n)y(n) \quad (7.7)$$

has an exponential dichotomy.

Throughout this section, we assume that there exists $\eta > 0$, such that

$$|g(t, x_1, y_1) - g(t, x_2, y_2)| \leq \eta [|x_1 - x_2| + |y_1 - y_2|], \quad x_i, y_i \in R. \quad (7.8)$$

No we give our main results.

Theorem 7.1 [24]. Suppose that $A(t), B(t)$ are Lebesgue measurable and bounded, and equation (7.6) admits an exponential dichotomy. If

$A_0, A_1 \in \tilde{PAP}_0(R, R^{d \times d}) \cap M_b$, $f \in \tilde{PAP}_0(R, R^d) \cap M_b$ and $|A_1| = \sup_{t \in R} \|A_{(t)}\| < 1$, then there exists $\beta_0 > 0$ such that when $|A_0| + |A_1| < \beta_0$, equation (7.1) has a unique solution $x \in PAP_0(R, R^d)$, where $M_b(R, R^d) = \{f/f : R \rightarrow R^d \text{ is Lebesgue measurable and bounded}\}$.

Theorem 7.2 [24]. Suppose that $A(t)$ and $B(t)$ are Lebesgue measurable and bounded and equation (7.6) admits an exponential dichotomy. If $A_0, A_1 \in \tilde{PAP}_0(R, R^{d \times d}) \cap M_b$, $|A_1| < 1$, and $g \in \tilde{PAP}_0(R \times R^d \times R^d, R^d)$ is bounded and satisfies the Lipschitz condition (7.8), then exists $\beta_1 > 0$ and $\eta > 0$ such that when $|A_1| + |A_0| < \beta_1$ and $0 \leq \eta \leq \eta_1$, equation (7.2) has a unique solution $y \in PAP_0(R, R^d)$.

Theorem 7.3 [24]. Suppose that $A(t), B(t)$ are almost periodic, $A_1(t) = A_0(t) \equiv 0$ and equation (7.6) admits an exponential dichotomy. Then the following facts hold,

- (1) for any $f \in \tilde{PAP}_0(R, R^d), (PAP_0(R, R^d))$, equation (7.1) has a solution $x \in \tilde{PAP}_0(R < R^d) (PAP_0(R, R^d))$, furthermore, x is unique if $f \in PAP_0(R, R^d)$ or $\tilde{PAP}(R, R^d) \cap M_b(R, R^d)$.
- (2) for any $f \in \tilde{PAP}_0(R, R^d) (PAP(R, R^d))$, equation (7.1) has a solution $x \in \tilde{PAP}(R, R^d) (PAP(R, R^d))$, and x is unique if $f \in PAP(R, R^d)$ or $\tilde{PAP}(R, R^d) \cap M_b(R, R^d)$.

Theorem 7.4 [24]. Suppose that $A(t)$, $B(t)$ are almost periodic, $A_1(t) = A_0(t) \equiv 0$ and equation (7.6) admits an exponential dichotomy. Then there exists $\eta_* > 0$ such that the following facts hold,

- (1) when $0 \leq \eta < \eta_*$, equation (7.2) has a unique solution $y \in PAP_0(R, R^d)$ if $g \in \tilde{PAP}_0(R \times R^d \times R^d, R^d)$ is bounded and satisfies the Lipschitz condition (7.8);
- (2) when $0 \leq \eta < \eta_*$, equation (7.2) has a unique solution $y \in PAP(R, R^d)$ if $g \in \tilde{PAP}(R \times R^d \times R^d, R^d)$ is bounded and satisfies the Lipschitz condition (7.8).

REFERENCES

- [1] Agarwal, R.P., *Difference Equations and Inequalities*, Marcel Dekker, New York, 1992.
- [2] Agarwal, R.P. and Wong, E.J.Y., *Advanced Topics in Difference Equations*, Kluwer, Dordrecht, 1997.
- [3] Aftabizadeh, A.R., Wiener, J. and Xu, J.M., Oscillatory and periodic solutions of delay differential equations with piecewise constant argument, Proc. Amer. Math. Soc., 99 (1987), 673-679.
- [4] Aftabizadeh, A.R. and Wiener, J., Oscillatory and periodic solutions for systems of two first order linear differential equations with piecewise constant argument, Applicable Analysis, 26 (1988), 327-338.
- [5] Ait Dads, E. and Arino, O., Exponential dichotomy and existence of pseudo almost periodic solutions of some differential equations, Nonlinear Analysis, TMA, 27 (4) (1996), 369-386.
- [6] Ait Dads, E., Ezzinbi, K. and Arino, O., Pseudo almost periodic solutions for some differential equations in a Banach space, Nonlinear Analysis, TMA, 28 (7) (1996), 1141-1155.
- [7] Alonso, A.I., Hong, J. and Obaya, R., Almost periodic type solutions of differential equations with piecewise constant argument via almost periodic type sequences, Applied Mathematics Letters, (In Press), 1998.
- [8] Alonso, A.I., Hong, J. and Rojo, J., A class of ergodic solutions of differential equations with piecewise constant arguments, Dynamics Systems and Applications, 7 (4) (1998), 561-574.
- [9] Alonso, A.I., Hong, J. and Obaya, R., Exponential Dichotomy and Trichotomy for difference equations, Computer Math. Appl., 38, (1999), 41-49.
- [10] Arendt, W. and Batty, C.J.K., Almost periodic solutions of first- and second-order Cauchy problems, J. Differential Equations, 137 (1997), 363--383.

- [11] Basit, B. and Zhang, C., New almost periodic type functions and solutions of differential equations, *Can. J. Math.* 48 (6), (1996), 1138-1153.
- [12] Belley, J.M., Fournier, G. and Hayes, J., Existence of almost periodic weak type solutions for the conservative forced pendulum equation, *J. Differential Equations*, 124 (1996), 205-224.
- [13] Busenberg, S. and Cooke, K.L., Models of vertically transmitted diseases with sequential-continuous dynamics, in *Nonlinear Phenomena in Mathematical Sciences* (V. Lakshmikantham. Ed.), pp.179-187, Academic Press, New York, 1982.
- [14] Cooke, K.L. and Wiener, J., Retarded differential equations with piecewise constant delays, *J. Math. Anal. Appl.*, 99 (1984), 265-297.
- [15] Cooke, K.L. and Wiener, J., A survey of differential equation with piecewise continuous argument, in *Lecture Notes in Mathematics*, Vol.1475, Springer-Verlag, Berlin, pp.1-15, 1991.
- [16] Cooke, K.L. and Wiener, J., Neutral differential equations with piecewise constant argument, *Bollettino Unione Matematica Italiana*, 7 (1987), 321-346.
- [17] Coppel, W.A., *Dichotomies in stability theory*, Lecture Notes in Mathematics, Vol. 629, Springer-Verlag, Berlin, 1978.
- [18] Corduneanu, C., *Almost Periodic Functions*, Chelsea Publishing Company, New York, N.Y., 1989.
- [19] Elaydi, S., *An Introduction to Difference Equations*, Springer-Verlag, New York, 1996.
- [20] Elaydi, S. and Hajek, O., Exponential trichotomy of differential systems, *J. Math. Anal. App.*, 129 (1988), 362-374.
- [21] Fink, A.M., *Almost Periodic Differential Equations*, Lecture Notes in Mathematics, Vol. 377, Springer-Verlag, Berlin, 1974.
- [22] Hale, J.K., *Ordinary Differential Equations*, Robert Krieger Publishing Company, New York, 1980.
- [23] Haraux, A., Generalized almost periodic solutions and ergodic properties of quasi-autonomous dissipative systems, *J. Differential Equations*, 48 (1983), 269-279.
- [24] Hong, J., Obaya, R. and Sanz, A.M., Almost periodic type solutions of some differential equations with piecewise constant argument, *Nonlinear Analysis TMA*, (In Press), 1998.
- [25] Hong, J., Obaya, R. and Sanz, A.M., Exponential trichotomy and a class of ergodic solutions of differential equations with ergodic perturbations, *Applied Mathematics Letters*, 12 (1), (1999), 7-13.
- [26] Hong, J., Obaya, R. and Sanz, A.M., Existence of a class of ergodic solutions implies exponential trichotomy, *Applied Mathematics Letters*, 12, (1999), 43-45.
- [27] Hong, J. and Nunez, C., On the almost periodic type difference equations, *Math. Comput. Modelling*, 28 (12), (1998), 21-31.
- [28] Hong, J., Obaya, R. and Sanz, A.M., Ergodic solutions via ergodic sequences, *Nonlinear Analysis TMA*, (In Press), 1999.
- [29] Lakshmikantham, V., Leela, S. and Martynyuk, A., *Stability Analysis of Nonlinear Systems*, Marcel Dekker, New York, 1988.
- [30] Lakshmikantham, V. and Trigiante, D., *Theory of Difference Equations: Numerical Methods and Applications*, Academic Press, 1988.
- [31] Layton, W., Existence of almost periodic solutions to delay differential equations with Lipschitz nonlinearities, *J. Differential Equations*, 55 (1984), 151-164.
- [32] Meisters, G.H., On almost periodic solutions of a class of differential equations, *Proc. Amer. Math. Soc.*, 10 (1959), 113-119.
- [33] Palmer, K.J., Exponential dichotomies, the shadowing-lemma and transversal homoclinic points, *Dynamics Reported*, (1988);Vol. 1, 265-306.
- [34] Papaschinopoulos, G., Some results concerning a class of differential equations with piecewise constant argument, *Math. Nachr.*, 166 (1994), 193-206.
- [35] Papaschinopoulos, G., On asymptotic behavior of the solutions of a class of perturbed differential equations with piecewise constant argument, *J. Math. Anal. Appl.*, 185 (1994), 490-500.

- [36] Rüss, W.M. and Summers, W.H., Minimal sets of almost periodic motions, *Math. Ann.*, 276 (1986), 145--158.
- [37] Seifert, G., Almost periodic solutions for delay-differential equations with infinite delays, *J. Differential Equations*, 41 (1981), 416--425.
- [38] Shah, S.M. and Wiener, J., Advanced differential equations with piecewise constant argument deviations, *Internat.J. Math. Math.Sci.*, 6 (1983), 671-703.
- [39] Wiener, J. and Cooke, K.L., Oscillations in systems of differential equations with piecewise constant argument, *J. Math. Anal. Appl.*, 137 (1989), 221-239.
- [40] Wiener, J., *Generalized Solutions of Functional Differential Equations*, World Scientific, 1993.
- [41] Yuan, R. and Hong, J., Almost periodic solutions of differential equations with piecewise constant argument, *Analysis*, 16 (1996), 171-180.
- [42] Yuan, R. and Hong, J., The existence of almost periodic solutions for a class of differential equations with piecewise constant argument, *Nonlinear Analysis, TMA*, 28 (8), (1997), 1439-1450.
- [43] Yuan, R. and Hong, J., Existence of almost periodic solutions of neutral differential equations with piecewise constant argument, *Science in China (Series A)*, 39 (11), (1996), 1164-1177.
- [44] Zhang, C., Pseudo almost-periodic solutions of some differential equations, *J. Math. Anal. Appl.*, 181 (199), 167-174.
- [45] Zhang, C., *Pseudo Almost-Periodic Functions and Their Applications*, Thesis, University of Western Ontario, 1992.
- [46] Zhang, C., Integration of vector-valued pseudo-almost periodic functions, *Proc. Amer. Math. Soc.*, 121 (1994), 167-174.

12 SYNCHRONOUS SOLUTIONS OF DELAYED NEURAL NETWORKS

Ying Sue Huang

Department of Mathematics

Pace University

Pleasantville, NY 10570

ABSTRACT

We consider a network of identical neurons with delayed nearest neighborhood inhibitory interaction. We study conditions on the parameters that there exist synchronous equilibrium solutions. Over certain range of parameters, the trivial solution is the only synchronous equilibrium and every solution converges to it. Over some other ranges of parameters, there are three synchronous equilibria. Stabilities of these solutions are also investigated.

1. INTRODUCTION

There has been increasing interest in the study of mathematical models of neural networks due to their richness in dynamics and broad applications. In 1984, Hopfield introduced a continuous model for a network of n neurons [12]. In this model, it is assumed that updating and propagation occur instantaneously. Then in 1989, Marcus and Westervelt [16] modified the model by introducing a time delay in the model. In fact, time delay does occur due to the finite switching speeds of the amplifiers.

We consider a network of identical neurons interconnected through nearest neighborhoods. The mathematical model can be derived from the Kirchhoff's law. With some rescaling and reparametrization, the dynamics of the neurons is governed by the following system of differential delay equations

$$\begin{aligned}\dot{x}_1(t) &= -x_1(t) + \alpha f(x_1(t-\tau)) + \beta [f(x_n(t-\tau)) + f(x_2(t-\tau))], \\ \dot{x}_2(t) &= -x_2(t) + \alpha f(x_2(t-\tau)) + \beta [f(x_1(t-\tau)) + f(x_3(t-\tau))], \\ &\dots \quad \dots \\ \dot{x}_n(t) &= -x_n(t) + \alpha f(x_n(t-\tau)) + \beta [f(x_{n-1}(t-\tau)) + f(x_1(t-\tau))].\end{aligned}$$

We can also simply express the system as follows

$$\dot{x}_i(t) = -x_i(t) + \alpha f(x_i(t-\tau)) + \beta [f(x_{i-1}(t-\tau)) + f(x_{i+1}(t-\tau))] \quad (1.1)$$

where $i \pmod{n}$. The parameters α and β measures respectively the normalized synaptic strength of self-connection and neighborhood-interaction. The activation function $f : \mathbb{R} \rightarrow \mathbb{R}$ has sigmoid form. A commonly used activation function is

$$f(x) = \tanh(\gamma x) = \frac{e^{\gamma x} - e^{-\gamma x}}{e^{\gamma x} + e^{-\gamma x}}$$

with $f'(0) = \gamma$. When rescaling the unknown x by γx , we can always reduce System (1.1) to the case where

$$\gamma = f'(0) = 1.$$

We assume throughout the paper that $f(0) = 0$, $f'(x) > 0$ and $f''(x)x < 0$ for all $x \neq 0$. Furthermore, f is bounded. One can see that $f'(x)$ has the largest value at 0 and $x \neq 0$. τ is the delay to account for the finite switching speed of amplifiers. It should be mentioned that the constant τ here is not the absolute size of the time lag required for the communication and response among neurons. In fact, System (1.1) is obtained after some rescaling and reparametrization, and the constant τ represents the ratio of the absolute size of the delay over the relaxation time of the system (see, for example, Belair, Campbell and van den Driessche [2], Marcus and Westervelt [16] and Wu [17]). Hence, this constant can be relatively large, and in such a case the dynamics of System (1.1) can be significantly different from that of the corresponding ordinary differential equation model. There has been very extensive studies of System (1.1) because of its richness as a theoretical model of collective dynamics. Among them, in the case that the delay $\tau = 0$, the convergence to the set of equilibria was studied by Hopfield [12], Cohen and Grossberg [5], Belair [1], Campbell [3], Gopalsamy and He [10]. The synchronization and stable phase-locking of System (1.1) has been studied in the case when $n = 3$ by Wu, Faria and Huang [19]. Synchronous periodic solutions and desynchronization of the system induced by the large scale of the network was investigated by Chen, Huang and Wu [6]. In this paper, we study the existence and stability of the synchronous solutions of System (1.1).

For System (1.1), we say that a solution $x = (x_1, \dots, x_n)^T : [-r, \infty) \rightarrow \mathbb{R}^n$ is *synchronous* if $x_1(t) = \dots = x_n(t)$ for all $t \in [-r, \infty)$, and *asynchronous* if otherwise.

A synchronous solution is completely characterized by the following scalar delay differential equation

$$\dot{z}(t) = -z(t) + (\alpha + 2\beta) f(z(t - \tau)). \quad (1.2)$$

Equation (1.2) may have equilibrium solutions and periodic solutions, which will lead to the synchronous equilibrium solutions and synchronous periodic solutions to System (1.1).

2. SYNCHRONOUS EQUILIBRIUM SOLUTIONS

Let $x = (x_1, \dots, x_n)^T : [-r, \infty) \rightarrow \mathbb{R}^n$ be an equilibrium solution. That is $x_j = x_j^*$ are all constants for all j . From System (1.1), we get the following equations

$$x_i^* = \alpha f(x_i^*) + \beta [f(x_{i-1}^*) + f(x_{i+1}^*)], \quad (2.1)$$

where $i \pmod{n}$.

It is very difficult to list all the solutions to System (2.1). In fact, for some values of α and β , there could be many different solutions to System (1.1). However, when $\alpha - \beta < 1$, we can show that every equilibrium of System (1.1) is synchronous. We have the following result.

Theorem 1. *If $\alpha - \beta < 1$, then every equilibrium of Systems (1.1) is synchronous.*

Proof: Let $x^* = (x_1^*, \dots, x_n^*)^T$ be an equilibrium. We then get equations (2.1). For $i \neq j$, every equilibrium x^* must satisfy

$$x_i^* - x_j^* = (\alpha - \beta) [f(x_i^*) - f(x_j^*)].$$

If $x_i^* \neq x_j^*$ for some i, j , using the mean value theorem, there is a c in between x_i^* and x_j^* such that

$$\frac{1}{\alpha - \beta} = \frac{f(x_i^*) - f(x_j^*)}{x_i^* - x_j^*} = f'(c).$$

Since $0 < f'(c) \leq 1$, we get $\alpha - \beta \geq 1$. Thus, when $\alpha - \beta < 1$, we have $x_i^* = x_j^*$ for all i and j . That is, all equilibrium solutions are synchronous.

Next, we study the synchronous equilibrium solutions.

Theorem 2. When $\alpha + 2\beta < 1$, the trivial solution $(0, 0, \dots)^T$ is the only synchronous equilibrium solution. When $\alpha + 2\beta > 1$, then there are exactly three synchronous equilibrium solutions, namely, the trivial solution $(0, 0, \dots)^T$, $x_+ = (u_+, u_+, \dots, u_+)^T$ and $x_- = (u_-, u_-, \dots, u_-)^T$ and $u = u_\pm$ satisfies that

$$u = (\alpha + 2\beta) f(u). \quad (2.2)$$

Furthermore,

$$(\alpha + 2\beta) f'(u) < 1. \quad (2.3)$$

Proof: Let $x^* = (x_1^*, \dots, x_n^*)^T$ be a synchronous equilibrium solution. Then $x_i^* = x_j^* \equiv u^*$ for all $1 \leq i, j \leq n$ and u^* satisfies that

$$u^* = (\alpha + 2\beta) f(u^*),$$

which gives equation (2.2). Because $f'(0) = 1, 0 < f'(x) \leq 1$ and $x f''(x) < 0$ for $x \neq 0$, when $\alpha + 2\beta > 1$, equation (2.2) has three solutions, namely, u_+, u_- and 0. As a result, System (1.1) admits exactly three synchronous equilibrium solutions. With direct computation, one can show that (2.3) holds at $u = u_{\pm}$.

When $\alpha + 2\beta < 1$, the only solution to equation (2.2) is the trivial solution $u = 0$. Thus the trivial solution is the only synchronous equilibrium solution to System (1.1) in this case.

3. STABILITY OF THE SYNCHRONOUS EQUILIBRIUM SOLUTIONS

Let $y_{ij} = x_i - x_j$. Then from System (1.1), we have

$$\begin{aligned} y'_{ij}(t) &= -y_{ij}(t) + (\alpha - \beta)[f(x_i(t - \tau)) - f(x_j(t - \tau))] \\ &= -y_{ij}(t) + (\alpha - \beta) p_{ij}(t) y(t - \tau), \end{aligned}$$

where

$$p_{ij}(t) = \int_0^1 f'(s x_i(t - \tau) + (1-s)x_j(t - \tau)) ds.$$

Using the essentially the same method as in Wu, Faria and Huang [19], by setting the Lyapunov functions, the following result can be obtained.

Theorem 3. If $|\alpha - \beta| < 1$, then every solution of System (1.1) with arbitrary given τ is asymptotically synchronous. If in addition, $|\alpha + 2\beta| < 1$, then every solution of (1.1) converges to the trivial solution as $t \rightarrow \infty$.

From Theorem 2, we know that when $\alpha + 2\beta < 1$, the trivial solution is the only synchronous equilibrium solution. With Theorem 3, we know that when $|\alpha - \beta| < 1$ and $|\alpha + 2\beta| < 1$, the trivial solution is in fact stable. When $\alpha - \beta < -1$ and $\alpha + 2\beta < -1$, the stability of the trivial solution depends on the size of the delay τ and the number of neurons n . In fact, bifurcation will happen which will lead to other types of synchronous solutions.

When $\alpha + 2\beta > 1$, there are three synchronous equilibrium solutions. We shall use the characteristic equations to study the stability of these solutions.

Let $(x^*, x^*, \dots, x^*)^T$ be a synchronous equilibrium solution. The linearization of System (1.1) at $(x^*, x^*, \dots, x^*)^T$ is given by

$$\dot{y}_i(t) = -y_i(t) + \alpha f'(x^*) y_i(t-\tau) + \beta f'(x^*) [y_{i-1}(t-\tau) + y_{i+1}(t-\tau)], \quad (3.1)$$

where $I \text{ (mod } n)$. The stability of solution $(x^*, x^*, \dots, x^*)^T$ is determined by the solutions of the characteristic equations. The characteristic equation is

$$\det \Delta(x^*, \lambda) = \det [(\lambda + 1) Id - f'(x^*) e^{-\lambda\tau} \delta] = 0, \quad (3.2)$$

where

$$\delta = \begin{bmatrix} \alpha & \beta & 0 & \cdots & 0 & \beta \\ \beta & \alpha & \beta & 0 & \vdots & 0 \\ 0 & \beta & \alpha & \beta & \vdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \beta & 0 & \cdots & 0 & \beta & \alpha \end{bmatrix}. \quad (3.3)$$

One can see that $\lambda + 1 - (\alpha + 2\beta) f'(x^*) e^{-\lambda\tau}$ is a factor of $\det \Delta(x^*, \lambda)$. From

$$\lambda + 1 - (\alpha + 2\beta) f'(x^*) e^{-\lambda\tau} = 0. \quad (3.4)$$

At the trivial solution $x^* = 0$ and $f'(x^*) = 1$. If $\alpha + 2\beta > 1$, (3.4) has a positive real solution. Therefore, the trivial solution is unstable. We thus obtain the following result.

Theorem 4. When $\alpha + 2\beta > 1$, the trivial equilibrium solution to System (1.1) is unstable.

Stability of the other two synchronous equilibria depends on the size of n , the delay τ , the parameters α, β , and the value of $f'(x^*)$.

When $n = 3$, the characteristic equation is

$$\det \Delta_\tau(x^*, \lambda) = [\lambda + 1 - (\alpha + 2\beta) f'(x^*) e^{-\lambda\tau}] [\lambda + 1 - (\alpha - \beta) f'(x^*) e^{-\lambda\tau}]^2. \quad (3.5)$$

We know that (2.3) holds at the solutions $x^* = u_\pm$, i.e.,

$$(\alpha + 2\beta) f'(x^*) < 1$$

even though $\alpha + 2\beta > 1$. Therefore solutions of equation

$$\lambda + 1 - (\alpha + 2\beta) f'(x^*) e^{-\lambda\tau} = 0$$

will all have negative real parts. When $|\alpha - \beta| < 1$, solutions of

$$\lambda + 1 - (\alpha - \beta) f'(x^*) e^{-\lambda\tau} = 0$$

will also have negative real parts. Thus the following result holds which is obtained in Wu, Faria and Huang [19].

Theorem 5. Let $n = 3$. If $|\alpha| < 1$, $|\alpha - 2\beta| < 1$ and $\alpha + 2\beta > 1$, then the trivial equilibrium is unstable and the other two synchronous equilibria are asymptotically stable.

When $n > 4$, it is hard to get all the factors of $\det \Delta_\tau(x^*, \lambda)$. We have to use a different approach to study the stability.

Theorem 7. Suppose α and β are both nonnegative and $\alpha + 2\beta > 1$, then the trivial equilibrium is unstable and the other two synchronous equilibria are both asymptotically stable.

Proof: We know that λ is a Floquet exponent of the linearized system (1.1) if and only if there is a solution $y = e^{\lambda t} (p_1, p_2, \dots, p_n)^T$, where p_j 's are all constants.

From System (1.1), we then get

$$(\lambda + 1) p_j - (\alpha p_j + \beta p_{j-1} + \beta p_{j+1}) f'(x^*) e^{-\lambda \tau} = 0, \quad (3.6)$$

where $j \pmod n$. Without loss of generalities, we can assume that $p_1 = 1$ and $|p_j| \leq 1$ because we can always pick the one with the largest absolute value. Therefore, let $j = 1$, we get

$$(\lambda + 1) - (\alpha + \beta p_2 + \beta p_n) f'(x^*) e^{-\lambda \tau} = 0. \quad (3.7)$$

We know that at the two equilibria, (2.3) holds. That is

$$(\alpha + 2\beta) f'(x^*) < 1.$$

When α and β have the same sign, since $|p_j| \leq 1$, we know that

$$|(\alpha + \beta p_2 + \beta p_n)| \leq \alpha + 2\beta.$$

Therefore,

$$|(\alpha + \beta p_2 + \beta p_n)| f'(x^*) \leq (\alpha + 2\beta) f'(x^*) < 1.$$

From (3.7), if there is a λ with $\operatorname{Re} \lambda > 0$, then

$$0 < \operatorname{Re} \lambda = \operatorname{Re} [(\alpha + \beta p_2 + \beta p_n) f'(x^*) e^{-\lambda \tau}] - 1 < 1 - 1 = 0,$$

a contradiction. Therefore, all the λ 's have negative real parts. The theorem is thus proved.

4. FURTHER DISCUSSION

We have discussed the existence and stability of the synchronous equilibrium solutions. In particular, we showed that when $\alpha - \beta < 1$, then every equilibrium of System (1.1) is synchronous. We further proved that if $\alpha + 2\beta > 1$, then there are

exactly three synchronous equilibria and if $\alpha + 2\beta < 1$, then the trivial solution $(0, 0, \dots)^T$ is the only synchronous equilibrium solution.

We also showed that when $|\alpha - \beta| < 1$, then every solution of System (1.1) with arbitrary given τ is asymptotically synchronous. If in addition, $|\alpha + 2\beta| < 1$, then every solution of (1.1) converges to the trivial solution as $t \rightarrow \infty$, that is, the trivial solution is stable. However, when $\alpha + 2\beta > 1$, the trivial equilibrium solution to (1.1) is unstable. Using the characteristic equations, we obtained conditions under which the two nontrivial synchronous equilibrium solutions are stable when there are three or four neurons. When there are more than four neurons, the characteristic equation becomes very hard to analyze. By estimating the Floquet exponents, showed that when α and β are both nonnegative and $\alpha + 2\beta > 1$ the two nontrivial synchronous equilibria are stable.

In general, the stabilities of the synchronous equilibria depend on the size of the delay τ also. In fact, bifurcation will happen which will lead to other types of synchronous solutions.

A synchronous solution is completely characterized by the scalar delay differential equation (1.2). When $\alpha + 2\beta < -1$, equation (1.2) is a system with negative feedback which has been extensively investigated in the literature. In particular, it is shown that when τ is in a certain range, equation (1.2) has a slowly oscillatory periodic solution. Here and in what follows, a *slowly oscillatory periodic solution* of equation (1.2) is a periodic solution $p : \mathbb{R} \rightarrow \mathbb{R}$ of (1.2) such that distances of consecutive zeros are larger than τ , and the minimal period ω is the distance of 3 consecutive zeros. When p is a slowly oscillating periodic solution, $X = (p, p, \dots, p)^T$ is a synchronous periodic solution to System (1.1). This synchronous periodic solution could actually be unstable even though the slowly oscillating periodic solution of equation is stable. This will lead to desynchronization in the network. In fact, this phenomena has been exclusively studied in the case that $\alpha = 0$ and $\beta = -\frac{1}{2}$ in Chen, Huang and Wu [6].

REFERENCES

- [1] Bélair, J., Stability in a model of a delayed neural network, *J. Dynamics and Differential Equations*, 5 (1993), 607-623.
- [2] Bélair, J., Campbell, S.A. and van den Driessche, P., Frustration, stability, and delay-induced oscillations in a neural network model, *SIAM J. Appl. Math.*, 56 (1996), 245-255.
- [3] Campbell, S.A., Stability and bifurcation of a simple neural network with multiple time delays, *Fields Institute Communication Series*, (1999), 65-81.

- [4] Cohen, M.A. and Grossberg, S., Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Trans. SMC*, 13 (1983), 815-826.
- [5] Cohen, M.A. and Grossberg, S., Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Trans. SMC*, 13 (1983), 815-826.
- [6] Chen, Y., Huang, Y. and Wu, J., Desynchronization of large scale delayed neural networks, preprint, (1999).
- [7] Chen, Y. and Wu, J., Existence and attraction of a phase-locked oscillation in a delayed network of two neurons, to appear in *Advances in Differential Equations*, (1999).
- [8] Diekmann, O., van Gils, S.A., Verduyn Lunel, S.M. and Walther, H.-O., *Delay Equations, Functional-, Complex-, and Nonlinear Analysis*, Springer-Verlag, New York, 1995.
- [9] van den Driessche, P. and Zou, X.F., Global attractivity in delayed Hopfield neural network models, *SIAM J. Math. Appl.*, (1998).
- [10] Gopalsamy, K. and He, X., Stability in asymmetric Hopfield nets with transmission delays, *Physica D*, 76 (1994), 344-358.
- [11] Hale, J.K. and Verduyn Lunel, S.M., *Introduction to Functional Differential Equations*, Applied Mathematical Sciences, Vol. 99, Springer-Verlag, New York, 1993.
- [12] Hopfield, J.J., Neurons with graded response have collective computational properties like two-stage neurons, *Proc. Nat. Acad. Sci. U.S.A.*, 81 (1984), 3088-3092.
- [13] Ivanov, A., Lani-Wayda, B. and Walther, H.-O., Unstable hyperbolic periodic solutions of differential delay equations, *WSSIAA*, 1 (1992), 301-316.
- [14] Koksal, S. and Sivasundaram, S., Stability properties of the Hopfield-type neural networks, *Dynamics and Stability of Systems*, 8 (1993), 181-187.
- [15] Krisztin, T., Walther, H.-O. and Wu, J., *Smoothness and Invariant Stratification of an Attracting Set for Delayed Monotone Positive Feedback*, Fields Institute Monographs, Vol. 11, American Mathematical Society, Providence, 1999.
- [16] Marcus, C.M. and Westervelt, R.M., Stability of analog neural networks with delay, *Phys. Rev. A* (3), 39 (1989), 347-359.
- [17] Wu, J., Symmetric functional-differential equations and neural networks with memory, *Trans. Amer. Math. Soc.*, 350 (1998), 4799-4838.
- [18] Xie, X.W., Uniqueness and stability of slowly oscillating periodic solutions of delay equations with bounded nonlinearity, *J. Dyna. Diff. Eqns.*, 3 (1991), 515-540.

- [19] Wu, J., Faria, T. and Huang, Y.S., Absolute synchronization and stable phase-locking in a network of neurons with memory, to appear in Mathematical and Computer Modeling, 1999.

13 COHERENT STRUCTURES AND STATISTICAL EQUILIBRIUM STATES IN A MODEL OF DISPERSIVE WAVE TURBULENCE

Richard Jordan

Department of Mathematical Sciences

Worcester Polytecnic Institute

Worcester, MA 01609-2280

and

Christophe Josserand

Laboratoire de Modélisation en Mécanique

UMR 7607 Université Pierre et Marie Curie and CNRS

Case 162, 4 place Jussieu, Tour 66, 75252 Paris
Cedex 05

ABSTRACT

We review a recent statistical equilibrium model of self-organization in a generic class of focusing, nonintegrable nonlinear Schrödinger (NLS) equations. Such equations provide natural prototypes for nonlinear dispersive wave turbulence. The primary result is that the statistically preferred state for such a system is a macroscopic solitary wave coupled with fine-scale turbulent fluctuations. The

coherent solitary wave is a minimizer the Hamiltonian for a fixed particle number (or L^2 norm squared). The predictions of the statistical model are compared with direct numerical simulations of the NLS equation, and it is demonstrated that the model describes the long-time average behavior of solutions remarkably well. In particular, the statistical theory accurately captures both the coherent structure and the spectrum of the solution of the NLS system in the long-time state.

1. INTRODUCTION: NLS AND SOLITON TURBULENCE

Turbulence in nonlinear media is often accompanied by the formation and persistence of large-scale coherent structures. A well-known example is the formation of macroscopic quasi-steady vortices in a turbulent large Reynolds number two dimensional fluid [1, 2, 3, 4]. Such phenomena also occur for many classical Hamiltonian systems, even though the dynamics of these systems is formally reversible [5]. In the present work, we shall focus our attention on a class of dispersive nonlinear wave equations whose solutions exhibit the tendency to form persistent coherent structures in the midst of small-scale turbulent fluctuations. This is the class of one-dimensional nonlinear Schrödinger (NLS) equations of the form

$$i\psi_t + \psi_{xx} + f(|\psi|^2)\psi = 0, \quad (1.1)$$

where $\psi(x,t)$ is a complex field. It is our primary purpose to develop a statistical model to characterize both the coherent structures and the turbulent fluctuations that emerge under the dynamics (1.1).

The NLS equation (1.1) describes the slowly-varying envelope of a wave train in a dispersive conservative system. Depending on the nonlinearity f , it models, among other things, gravity waves on deep water [6], Langmuir waves in plasmas [7], and pulse propagation along optical fibers [8]. When $f(|\psi|^2) = \pm|\psi|^2$ and equation (1.1) is posed on the whole real line or on a bounded interval with periodic boundary conditions, the equation is completely integrable [9], but for other nonlinearities and/or boundary conditions, it is nonintegrable.

We shall assume throughout that equation (1.1) is posed in a bounded one dimensional interval with either periodic or homogeneous Dirichlet boundary conditions. We restrict our attention to attractive, or focusing, nonlinearities $f(f(a) \geq 0, f'(a) > 0)$ such that the dynamics described by (1.1) is nonintegrable, free of wave collapse, and admits stable solitary-wave solutions. The dynamics under these conditions has been referred to as *soliton turbulence* [10]. Such is the

case for the important power law nonlinearities, $f(|\psi|^2) = |\psi|^s$, with $0 < s < 4$ (in the periodic case, $s \neq 2$ for nonintegrability) [11, 12], and also for the physically relevant saturated nonlinearities $f(|\psi|^2) = |\psi|^2 / (1 + |\psi|^2)$ and $f(|\psi|^2) = 1 - \exp(-|\psi|^2)$, which arise as corrections to the cubic nonlinearity for large wave amplitudes [13].

The NLS equation (1.1) may be cast in Hamiltonian form $i\psi_t = \delta H / \delta \psi^*$, where ψ^* is the complex conjugate of the field ψ , and H is the Hamiltonian:

$$H(\psi) = \int \left(|\psi_x|^2 - F(|\psi|^2) \right) dx. \quad (1.2)$$

The *potential* F is defined via the relation $F(a) = \int_0^a f(y) dy$. In addition to the

Hamiltonian, the dynamics (1.1) conserves, the particle number integral

$$N(\psi) = \int |\psi|^2 dx. \quad (1.3)$$

Equation (1.1) in one spatial dimension has solitary wave solutions of the form $\psi(x, t) = \phi(x) \exp(i\lambda^2 t)$, where ϕ satisfies the nonlinear eigenvalue equation:

$$\phi_{xx} + f(|\phi|^2) \phi - \lambda^2 \phi = 0. \quad (1.4)$$

It has been argued [10, 14] that the solitary wave solutions play a prominent role in the long-time dynamics of (1.1), in that they act as *statistical attractors* to which the system relaxes. The numerical simulations in [10, 20, 15], as well as the simulations we shall present here, support this conclusion. Indeed, it is seen that for rather generic initial conditions the field ψ evolves, after a sufficiently long time, into a state consisting of a spatially localized coherent structure, which agrees quite closely with a solution of (1.4), coupled with small-scale turbulent fluctuations. At intermediate times the solution typically consists of a collection of these soliton-like structures, but as time evolves, the solitons undergo a succession of collisions in which the smaller soliton decreases in amplitude, while the larger one increases in amplitude. When solitons collide or interact, they shed radiation, or small-scale fluctuations. This process continues until eventually a single soliton of large amplitude survives amidst the turbulent background radiation. Figure 1 illustrates the evolution of the solution of (1.1) for the particular nonlinearity $f(|\psi|^2) = |\psi|$ and with periodic boundary conditions on the spatial interval $[0, 256]$.

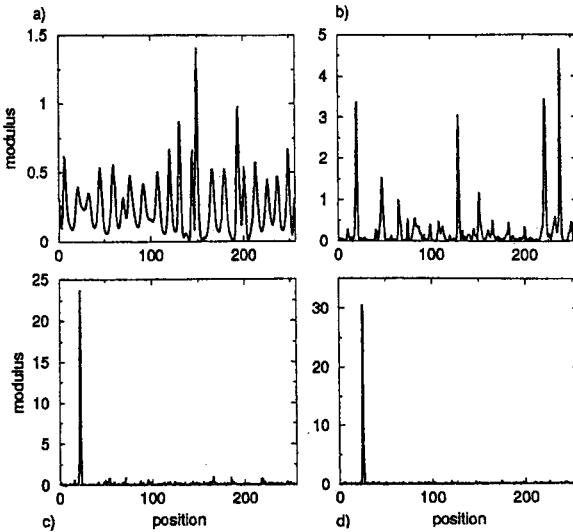


Figure 1. Profile of the squared modulus $|\psi|^2$ at four different times for the system (1.1) with nonlinearity $f(|\psi|^2) = |\psi|$ and periodic boundary conditions on the interval $[0, 256]$. The initial condition is $\psi(x, 0) = A$, with $A = 0.5$, plus a small random perturbation. The numerical scheme used to approximate the solution is the split-step Fourier method. The grid size is $dx = 0.125$, and the number of modes is $n = 2048$. A) $t = 50$ unit time: Due to the modulational instability, an array of soliton-like structures separated by the typical distance $l_i = 2\pi/\sqrt{A/2} = 4\pi$ is created; b) $t = 1050$ unit time: The solitons interact and coalesce, giving rise to a smaller number of solitons of larger amplitude; c) $t = 15050$: The coalescence process has ended. A single layer soliton remains in a background of small-amplitude radiation. Notice that at $t = 55050$ unit time (d), the amplitude of the fluctuations has diminished while the amplitude of the soliton has increased.

2. A STATISTICAL MECHANICS MODEL

In modeling the long-time behavior of a nonintegrable Hamiltonian system such as NLS, it seems natural to appeal to the methods of equilibrium statistical mechanics. That such an approach may be relevant for understanding the asymptotic-time state for NLS has already been suggested in [10], although the thermodynamic arguments presented by these authors are formal and incomplete. Recently, Jordan et al. [15] have constructed a mean-field statistical theory to characterize the large-scale structure and the statistics of the small-scale fluctuations inherent in the asymptotic-time state of the NLS system (1.1), and we shall review this theory and its predictions in the present article. The main conclusion of the theory is that the

coherent state that emerges in the long-time limit is the ground state solution of equation (1.4). In other words, it is the solitary wave that minimizes the Hamiltonian H given the constraint $N = N^0$, where N^0 is the initial and conserved value of the particle number integral. This prediction is in accord with previous theories [10, 14], but the approach taken in [15] is different. Furthermore, as we shall see, the theory of [15] provides a definite interpretation to the notion set forth in the earlier works that it is “thermodynamically advantageous” for the NLS system to approach a coherent solitary wave structure that minimizes the Hamiltonian subject to fixed particle number. We now proceed to outline the statistical model developed in [15].

In order to develop a meaningful statistical theory, we begin by introducing a finite-dimensional approximation of the NLS equation (1.1). For ease of presentation, we will consider the NLS system with homogeneous Dirichlet boundary conditions on an interval Ω of length L . Our methods are readily modified to accommodate other boundary conditions, as well. We remark that the techniques can easily be extended to higher dimensional NLS systems. Also, while we shall use a spectral truncation here, other discretization schemes would suit our purposes just as well.

Let $e_j(x) = \sqrt{2/L} \sin(k_j x)$ with $k_j = \pi j/L$, and for any function $g(x)$ on Ω denote by $g_j = \int_{\Omega} g(x) e_j(x) dx$ its j th Fourier coefficient with respect to the

orthonormal basis e_j , $j = 1, 2, \dots$. Define the functions $u^{(n)}(x, t) = \sum_{j=1}^n u_j(t) e_j(x)$ and $v^{(n)}(x, t) = \sum_{j=1}^n v_j(t) e_j(x)$, where the real coefficients $u_j, v_j, j = 1, \dots, n$,

satisfy the coupled system of ordinary differential equations

$$\begin{aligned} \dot{u}_j - k_j^2 v_j + \left(f \left(\left(u^{(n)} \right)^2 + \left(v^{(n)} \right)^2 \right) v^{(n)} \right)_j &= 0 \\ \dot{v}_j + k_j^2 u_j - \left(f \left(\left(u^{(n)} \right)^2 + \left(v^{(n)} \right)^2 \right) u^{(n)} \right)_j &= 0. \end{aligned} \tag{2.1}$$

Then the complex function $\psi^{(n)} = u^{(n)} + i v^{(n)}$ satisfies the equation

$$i \psi_t^{(n)} + \psi_{xx}^{(n)} + P^n \left(f \left(|\psi^{(n)}|^2 \right) \psi^{(n)} \right) = 0,$$

where P^n is the projection onto the span of the eigenfunctions e_1, \dots, e_n . This equation is a natural spectral approximation of the NLS equation (1.1), and it may be shown that its solutions converge as $n \rightarrow \infty$ to solutions of (1.1) [11, 16].

For given n , the system of equations (2.1) defines a dynamics on the $2n$ -dimensional phase space \mathbf{R}^{2n} . This finite-dimensional dynamical system is a Hamiltonian system, with conjugate variables u_j and v_j , and with Hamiltonian

$$H_n = K_n + \Theta_n, \quad (2.2)$$

where

$$K_n = \frac{1}{2} \int_{\Omega} \left((u_x^{(n)})^2 + (v_x^{(n)})^2 \right) dx = \frac{1}{2} \sum_{j=1}^n k_j^2 (u_j^2 + v_j^2), \quad (2.3)$$

is the kinetic energy, and

$$\Theta_n = -\frac{1}{2} \int_{\Omega} F \left((u^{(n)})^2 + (v^{(n)})^2 \right) dx, \quad (2.4)$$

is the potential energy. The Hamiltonian H_n is, of course, an invariant of the dynamics. The truncated version of the particle number

$$N_n = \frac{1}{2} \int_{\Omega} \left((u^{(n)})^2 + (v^{(n)})^2 \right) dx = \frac{1}{2} \sum_{j=1}^n (u_j^2 + v_j^2), \quad (2.5)$$

is also invariant under the dynamics (2.1). The factor $1/2$ is included in the definition of the particle number for mathematical convenience later on.

The Hamiltonian system (2.1) satisfies the Liouville property, which is to say that the measure $\prod_{j=1}^n du_j dv_j$ is invariant under the dynamics [17]. This property,

together with the assumption of ergodicity of the dynamics, provides the usual starting point for the statistical mechanics of a Hamiltonian system [18].

We introduce now a macroscopic description of the system (2.1) in terms of a probability density $\rho^{(n)}(u_1, \dots, u_n, v_1, \dots, v_n)$ on the $2n$ -dimensional phase-space \mathbf{R}^{2n} . Thus, we seek a probability density that describes the statistical equilibrium state for the truncated dynamics. Following standard statistical mechanics and information theoretic practices [18, 19], we require this state to be the density $\rho^{(n)}$ on $2n$ -dimensional phase space which maximizes the Gibbs-Boltzmann entropy

$$S(\rho) = - \int_{\mathbf{R}^{2n}} \rho \log \rho \prod_{j=1}^n du_j dv_j, \quad (2.6)$$

subject to constraints on the density associated with the invariance of the Hamiltonian and the particle number under the dynamics (2.1). The key to constructing the statistical model lies in choosing the appropriate constraints. Our choice is motivated by the observation from numerical simulations that, for a large number of modes n , in the long-time limit, the field $(u^{(n)}, v^{(n)})$ decomposes into two essentially distinct components: a large-scale coherent structure, and small-scale radiation, or fluctuations. The simulations illustrate that, as time increases, the amplitude of the fluctuations decreases, until eventually the contribution of the

fluctuations to the particle number and the potential energy component of the Hamiltonian becomes negligible compared to the contribution from the coherent state. In the long-time limit, therefore, N_n and Θ_n are determined almost entirely by the coherent structure. We have checked that this effect becomes even more pronounced when the spatial resolution of the numerical simulations is improved. On the other hand, as the fluctuations exhibit rapid spatial variations, the amplitude of their gradient does not in general become negligible in the asymptotic time limit. In fact, the fluctuations typically make a significant contribution to the kinetic energy component K_n of the Hamiltonian. Figure 2 demonstrates this effect quite clearly.

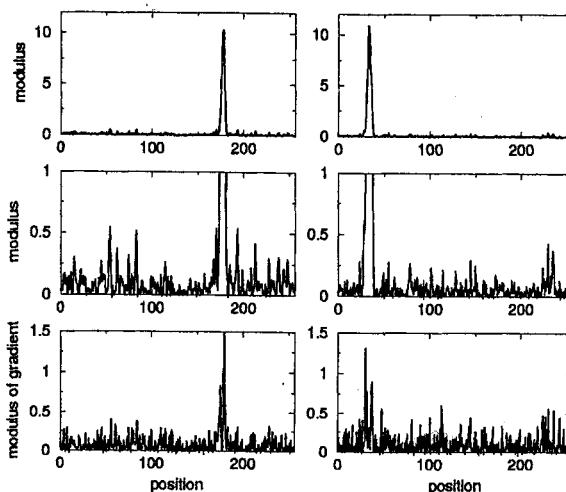


Figure 2. Numerical simulation for the saturated non-linearity $f(|\psi|^2) = |\psi|^2 / (1 + |\psi|^2)$ for periodic boundary conditions. The total number of modes is $n = 1024$ and the spatial grid size is $dx = 0.25$, so that the length of periodic interval is $L = 256$. Displayed are the squared modulus $|\psi|^2$ (first and second rows), and the squared modulus of the gradient of the field $|\psi_x|^2$ (third row) at unit times $t = 30,000$ (left) and $t = 220,000$ (right). The second row shows the same results as the first row, except that we have restricted the range on the vertical axis in order to focus in on the fluctuations of the field. The typical amplitude of the fluctuations of the field has decreased from $t = 30,000$ to $t = 220,000$ (second row), while the amplitude of the coherent structure has increased. On the other hand, the typical amplitude of the fluctuations of the gradient of the field has actually increased somewhat from $t = 30,000$ to $t = 220,000$.

We denote by $\langle u_j \rangle$ and $\langle v_j \rangle$ the means of the coefficients u_j and v_j with respect the admissible ensemble $\rho^{(n)}$. The coherent state is identified with the mean-field pair $(\langle u^{(n)}(x) \rangle, \langle v^{(n)}(x) \rangle) = \left(\sum_{j=1}^n \langle u_j \rangle e_j(x), \sum_{j=1}^n \langle v_j \rangle e_j(x) \right)$, and the fluctuations, or small-scale radiation inherent in the long-time state then correspond to the difference $(\delta u^{(n)}, \delta v^{(n)}) \equiv (u^{(n)} - \langle u^{(n)} \rangle, v^{(n)} - \langle v^{(n)} \rangle)$ between the state vector $(u^{(n)}, v^{(n)})$ and the mean-field vector. From the considerations of the preceding paragraph, it seems reasonable to require that the amplitude of the fluctuations of the field $\psi^{(n)}$ in the long-time state of the NLS system (2.1) vanish entirely (in some appropriate sense) in the continuum limit $n \rightarrow \infty$. Specifically, we shall make the following *vanishing of fluctuations hypothesis*:

$$\int_{\Omega} \left[\langle (\delta u^{(n)})^2 \rangle + \langle (\delta v^{(n)})^2 \rangle \right] dx \equiv \sum_{j=1}^n \left[\langle (\delta u_j)^2 \rangle + \langle (\delta v_j)^2 \rangle \right] \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (2.7)$$

Here, $\delta u_j = u_j - \langle u_j \rangle$ represents the fluctuations of the Fourier coefficient u_j about its mean value $\langle u_j \rangle$, and similarly for δv_j . It is important to emphasize that (2.7) is a hypothesis used to construct our statistical theory, and not a conclusion drawn from the theory itself.

The vanishing of fluctuations hypothesis immediately implies that, for n sufficiently large, the expectation $\langle N_n \rangle$ of the particle number is determined almost entirely by the mean $(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle)$. Furthermore, the hypothesis (2.7) implies that for n large, the expectation $\langle \Theta_n(u^{(n)}, v^{(n)}) \rangle$ of the potential energy is well approximated by $\Theta_n(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle)$, the potential energy contained in the mean. This may be verified by expanding the potential F about the mean $(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle)$ in equation (2.4), taking expectations, and noting that the vanishing of fluctuations hypothesis implies that $|\langle \Theta_n(u^{(n)}, v^{(n)}) \rangle - \Theta_n(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle)| = o(1)$ as $n \rightarrow \infty$ (see [15] for detailed calculations). It is important to recognize that the vanishing of fluctuations hypothesis does not imply that the contribution of the fluctuations to the expectation of the kinetic energy becomes negligible in the limit $n \rightarrow \infty$. This contribution is $(1/2) \sum_{j=1}^n k_j^2 \left[\langle (\delta u_j)^2 \rangle + \langle (\delta v_j)^2 \rangle \right]$, which need not tend to 0 as $n \rightarrow \infty$, even if (2.7) holds. Thus, from these arguments, we conclude that for n sufficiently large, $\langle H_n \rangle \approx \frac{1}{2} \sum_{j=1}^n k_j^2 \left(\langle u_j^2 \rangle + \langle v_j^2 \rangle \right) - \frac{1}{2} \int_{\Omega} F \left(\langle u^{(n)} \rangle^2 + \langle v^{(n)} \rangle^2 \right) dx$.

Based on these considerations, we impose the following *mean-field constraints* on the admissible probability densities $\rho^{(n)}$:

$$\begin{aligned}\tilde{N}_n(\rho^{(n)}) &\equiv \frac{1}{2} \sum_{j=1}^n (\langle u_j \rangle + \langle v_j \rangle^2) = N^0 \\ \tilde{H}_n(\rho^{(n)}) &\equiv \frac{1}{2} \sum_{j=1}^n k_j^2 (\langle u_j^2 \rangle + \langle v_j^2 \rangle) - \frac{1}{2} \int_{\Omega} F(\langle u^{(n)} \rangle^2 + \langle v^{(n)} \rangle^2) dx = H^0.\end{aligned}\quad (2.8)$$

Here, N^0 and H^0 are the conserved values of the particle number and the Hamiltonian, as determined from initial conditions. The statistical equilibrium states are then taken to be the probability densities $\rho^{(n)}$ on the phase-space \mathbf{R}^{2n} that maximize the entropy (2.6) subject to the constraints (2.8). We shall refer to the constrained maximum entropy principle that determines the statistical equilibria as (MEP).

It has been proved in [15] that the solutions $\rho^{(n)}$ of (MEP) concentrate on the phase-space manifold on which $H_n = H^0$ and $N_n = N^0$ in the continuum limit $n \rightarrow \infty$, in the sense that $\langle N_n \rangle \rightarrow N^0$, $\langle H_n \rangle \rightarrow H^0$, and $\text{var } N_n \rightarrow 0$, $\text{var } H_n \rightarrow 0$ in this limit. Here, $\text{var } W$ denotes the variance of the random variable W . This concentration property establishes a form of asymptotic equivalence between the mean-field ensembles $\rho^{(n)}$ and the microcanonical ensemble, which is the invariant measure concentrated on the phase-space manifold on which $H_n = H^0$ and $N_n = N^0$. This is a crucial result, because it is an accepted axiom of statistical mechanics that the microcanonical ensemble is the appropriate equilibrium ensemble for an isolated ergodic system [18]. Thus, this equivalence of ensembles property provides a strong theoretical justification for the mean-field statistical model, and it substantiates *a posteriori* the use of the vanishing of fluctuations hypothesis in constructing our statistical model. A detailed discussion of these issues can be found in [21].

3. CALCULATION AND ANALYSIS OF EQUILIBRIUM STATES

The solutions $\rho^{(n)}$ of (MEP) are calculated by an application of the Lagrange multiplier rule

$$S'(\rho^{(n)}) = \mu \tilde{N}'_n(\rho^{(n)}) + \beta \tilde{H}'_n(\rho^{(n)}),$$

where β and μ are the Lagrange multipliers to enforce that the probability density $\rho^{(n)}$ satisfy the constraints (2.8). Fairly straightforward, but somewhat tedious, calculations lead to the following expression for the maximum entropy distribution $\rho^{(n)}$ [15]:

$$\rho^{(n)}(u_1, \dots, u_n, v_1, \dots, v_n) = \prod_{j=1}^n \rho_j(u_j, v_j), \quad (3.1)$$

where, for $j = 1, \dots, n$,

$$\rho_j(u_j, v_j) = \frac{\beta k_j^2}{2\pi} \exp \left\{ -\frac{\beta k_j^2}{2} \left((u_j - \langle u_j \rangle)^2 + (v_j - \langle v_j \rangle)^2 \right) \right\}, \quad (3.2)$$

with

$$\begin{aligned} \langle u_j \rangle &= \frac{1}{k_j^2} \left(f \left(\langle u^{(n)} \rangle^2 + \langle v^{(n)} \rangle^2 \right) \langle u^{(n)} \rangle \right)_j - \frac{\mu}{\beta k_j^2} \langle u_j \rangle \\ \langle v_j \rangle &= \frac{1}{k_j^2} \left(f \left(\langle u^{(n)} \rangle^2 + \langle v^{(n)} \rangle^2 \right) \langle v^{(n)} \rangle \right)_j - \frac{\mu}{\beta k_j^2} \langle v_j \rangle. \end{aligned} \quad (3.3)$$

It follows that, for each j , u_j and v_j are independent Gaussian random variables with means given by the nonlinear equations (3.3) and with variances

$$\text{var } u_j = \text{var } v_j = \frac{1}{\beta k_j^2}. \quad (3.4)$$

Note that $\text{var } u_j = \langle (\delta u_j)^2 \rangle$ by definition, and similarly for v_j . Notice also that, since the probability density $\rho^{(n)}$ is Gaussian and factors according to (3.1), the Fourier modes $u_j, v_j, j = 1, \dots, n$, are statistically independent. Setting $\lambda = \mu/\beta$, the equations (3.3) imply that the complex mean-field $\langle \psi^{(n)} \rangle = \langle u^{(n)} \rangle + i \langle v^{(n)} \rangle$ is solution of

$$\langle \psi^{(n)} \rangle_{xx} + P^n \left(f \left(|\langle \psi^{(n)} \rangle|^2 \right) \langle \psi^{(n)} \rangle \right) - \lambda \langle \psi^{(n)} \rangle = 0. \quad (3.5)$$

This is obviously the spectral truncation of the eigenvalue equation (1.4) for the continuous NLS system (1.1). The important conclusion to be drawn from this is that the mean-field corresponds to a solitary wave solution of the NLS equation.

Since the maximum entropy distribution $\rho^{(n)}$ is required to satisfy the mean-field Hamiltonian constraint (2.8), we have from (3.1)-(3.5) that

$$H^0 = \frac{n}{\beta} + H_n \left(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle \right). \quad (3.6)$$

The term $H_n \left(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle \right)$ is the Hamiltonian of the mean, while the term n/β represents the contribution to the kinetic energy from the random fluctuations. We see that the contribution of the fluctuations to the kinetic energy is equipartitioned among the n Fourier modes. From (3.6), we obtain the following expression for β in terms of the number of modes n and the Hamiltonian of the mean:

$$\beta = \frac{n}{H^0 - H_n \left(\langle u^{(n)} \rangle, \langle v^{(n)} \rangle \right)}. \quad (3.7)$$

We are now ready to establish an essential result: *The mean field* $\langle\langle u^{(n)} \rangle\rangle, \langle\langle v^{(n)} \rangle\rangle$ *corresponding to the maximum entropy density* $\rho^{(n)}$ *is an absolute minimizer of the Hamiltonian* H_n *subject to the particle number constraint* $N_n = N^0$.

Indeed, using equations (3.1)-(3.7), we find after some algebraic manipulations that the entropy of any solution $\rho^{(n)}$ of (MEP) is

$$S(\rho^{(n)}) = C(n) + n \log \left(\frac{L^2 [H^0 - H_n(\langle\langle u^{(n)} \rangle\rangle, \langle\langle v^{(n)} \rangle\rangle)]}{n} \right),$$

where $C(n) = n - \sum_{j=1}^n \log(j^2 \pi/2)$ depends only on the number of Fourier modes n .

Clearly then, the entropy $S(\rho^{(n)})$ will be maximum if and only if the mean field pair $(\langle\langle u^{(n)} \rangle\rangle, \langle\langle v^{(n)} \rangle\rangle)$ realizes the minimum possible value of H_n over all fields $(u^{(n)}, v^{(n)})$ that satisfy the constraint $N_n(u^{(n)}, v^{(n)}) = N^0$. This is the desired conclusion.

The preceding argument reveals that, in statistical equilibrium the entropy is, up to additive and multiplicative constants, the logarithm of the kinetic energy contained in the turbulent fluctuations about the mean state. This result, therefore, provides a precise interpretation to the notions set forth by Zakharov et al. [10] and Pomeau [14] that the entropy of the NLS system is directly related to the amount of kinetic energy contained in the small-scale fluctuations, and that it is “thermodynamically advantageous” for the solution of NLS to approach a ground state which minimizes the Hamiltonian for the given number of particles.

We now know that $H_n(\langle\langle u^{(n)} \rangle\rangle, \langle\langle v^{(n)} \rangle\rangle) = H_n^*$, where H_n^* is the minimum value of H_n allowed by the particle number constraint $N_n = N^0$. Consequently, the Lagrange multiplier β is uniquely determined by (3.7):

$$\beta = \frac{n}{H^0 - H_n^*}. \quad (3.8)$$

The Lagrange multiplier λ (which also depends on n) is determined by the requirement that the mean $(\langle\langle u^{(n)} \rangle\rangle, \langle\langle v^{(n)} \rangle\rangle)$ minimizes the Hamiltonian H_n given the particle number constraint $N_n = N^0$.

Using equations (3.4) and (3.8), we may derive an exact expression for the contribution of the fluctuations to the expectation of the particle number. This is

$$\frac{1}{2} \sum_{j=1}^n \left[\langle (\delta u_j)^2 \rangle + \langle (\delta v_j)^2 \rangle \right] = \frac{H^0 - H_n^*}{n} \sum_{j=1}^n \frac{1}{k_j^2} = O(n^{-1}), \quad \text{as } n \rightarrow \infty. \quad (3.9)$$

In the derivation of the mean-field constraints (2.8), we assumed the vanishing of fluctuations condition (2.7). The calculation of (3.9) shows, therefore, that the

maximum entropy distribution $\rho^{(n)}$ indeed satisfy the hypothesis (2.7), and hence, that the mean-field statistical theory is consistent with the assumption that was made to derive it.

The statistical theory also provides predictions for the particle number and kinetic energy spectral densities, at least for the $2n$ -dimensional spectrally truncated NLS system (2.1) with n large. In fact, we have the following prediction for the particle number spectral density

$$\langle |\psi_j|^2 \rangle = \left| \langle \psi_j \rangle \right|^2 + \frac{H^0 - H_n^*}{nk_j^2}, \quad (3.10)$$

where we have used the identity $\psi_j = u_j + i v_j$, and equations (3.4) and (3.8). The first term on the right hand side of (3.10) is the contribution to the particle number spectrum from the mean, and the second term is the contribution from the fluctuations. As the mean field is a smooth solution of the ground-state equation, its spectrum decays rapidly, so that for $j \gg 1$, we have the approximation $\langle |\psi_j|^2 \rangle \approx (H^0 - H_n^*)/(nk_j^2)$. The kinetic energy spectral density is obtained simply by multiplying equation (3.10) by k_j^2 . In particular, we have the prediction that the kinetic energy arising from the fluctuations is equipartitioned among the n spectral modes, with each mode contributing the amount $(H^0 - H_n^*)/n$.

While we have focused on homogeneous Dirichlet boundary conditions so far, it is straightforward to modify the statistical theory to accommodate periodic boundary conditions on an interval of length L , say. In this case, it is convenient to write the spectrally truncated complex field $\psi^{(n)}$ as

$$\psi^{(n)} = \sum_{j=-n/2}^{n/2} \psi_j \exp(i k_j x),$$

for n an even positive integer, where $k_j = 2\pi j/L$. The predictions of the statistical theory remain the same as in the case of Dirichlet boundary conditions. Namely, the mean field $\langle \psi^{(n)} \rangle$ is a minimizer of the Hamiltonian H_n given the particle number constraint $N_n = N^0$, and the particle number spectrum satisfies (3.10) for $j \neq 0$. The Fourier coefficient ψ_0 may be consistently chosen to be deterministic (i.e., $\text{var } \psi_0 = 0$ and $\langle \psi_0 \rangle \equiv \psi_0$), to eliminate the ambiguity arising from the 0 mode.

4. COMPARISONS WITH NUMERICAL SIMULATIONS

Our primary purpose in this section is to compare the predictions of the mean-field statistical theory described above with the results of high resolution direct numerical simulations of NLS. In particular, we wish to examine how closely the coherent structure and the spectra predicted by the statistical theory agree with those observed in long-time numerical simulations. Here, we will present numerical results primarily for periodic boundary conditions and for the focusing power law nonlinearity $f(|\psi|^2) = |\psi|$. That is, we shall solve numerically the particular NLS equation

$$i\psi_t + \psi_{xx} + |\psi|\psi = 0, \quad (4.1)$$

on a periodic interval of length L . This nonlinearity offers a nice compromise between the focusing effect and nonlinear interactions. For weaker nonlinearities (such as the saturated ones), the interaction between modes is weak, and the time required to approach an asymptotic equilibrium state is quite long. For stronger nonlinearities, the solitary wave structures that emerge exhibit narrow peaks of large amplitude, and consequently, higher spatial resolution is required in the numerical simulations. We have performed elsewhere similar numerical experiments for different focusing nonlinearities and for Dirichlet boundary conditions, and we have observed that the general qualitative features of the long-time dynamics are unaltered by such changes (see, for example [15, 20]).

On the whole real line, the nonlinear Schrödinger equation (4.1) has solitary wave solutions of the form $\psi(x, t) = \phi(x) e^{i\lambda^2 t}$, with

$$\phi(x) = \frac{3\lambda^2}{2 \cosh^2 \left(\frac{\lambda(x - x_0)}{2} \right)}. \quad (4.2)$$

The particle number N and the Hamiltonian H of these solutions are determined by the parameter λ via the relationships $N = 6\lambda^3$ and $H = -\frac{18}{5}\lambda^5$. For a given value of the particle number N , the solitary wave (4.2) is the global minimizer of the Hamiltonian H . Of course, the solitary wave solutions for the equation (4.1) on a finite interval, as well as those for the spectrally-truncated version (2.1), differ from the solution (4.2) over the infinite interval. However, as the solitary waves (4.2) exhibit exponential decay, such differences can be neglected for all practical purposes if the spatial interval is large enough and if the number of modes is sufficiently large. We shall be comparing the coherent structures observed in the numerical simulations to the expression (4.2).

In our numerical simulations, the initial condition is taken to be the spatially homogeneous solution (condensate) $\psi(x) = A$ (A constant) coupled with a small spatially uncorrelated random perturbation. By choosing different realizations of the initial random perturbation, we may perform an ensemble average over different initial conditions for a given A (and therefore for fixed N^0 and H^0). The initial conditions we consider here may be thought of as being far away from the expected statistical attractor described by the maximum entropy probability density $\rho^{(n)}$, as the spectrum of the condensate differs considerably from the predicted statistical equilibrium spectrum (3.10). As we shall see, the numerical simulations provide convincing evidence that the solutions of the spectrally truncated NLS system converge in the long-time limit to a state that may be considered as statistically steady, and whose average features are very well described by the mean-field statistical ensemble.

The numerical scheme that we use for solving (4.1) is the well-known split-step Fourier method for a given number n of Fourier modes. Throughout the duration of the simulations, the relative error in the particle number is kept at less than 10^{-6} percent, and the relative error in the Hamiltonian is no greater than 10^{-2} percent. Note that these numerical simulations, which necessarily pertain to a finite number of Fourier modes, provide a perfect setting for comparisons with the statistical model discussed above.

Figure 1 demonstrates that the dynamics can be roughly decomposed into three stages: in the first stage, illustrated in Figure 1a, the modulational instability creates an array of soliton-like structures separated by a typical distance $l_i = 2\pi/k_i$ associated with most unstable wave number k_i . The second stage is characterized by the interaction and coalescence of these solitons. In this stage, the number of solitons decreases, while the amplitudes of the surviving solitons increase. Eventually a single soliton of large amplitude persists amongst a sea of small-amplitude background radiation (Figures 1b and c).

During the final stage of the dynamics, the surviving large-scale soliton interacts with the small-scale fluctuations. As time increases, the amplitude of the soliton increases, while the amplitude of the radiation decreases (note the changes from Figure 1c to Figure 1d). In this stage of the dynamics, the mass (or number of particles) is gradually transferred from the small-scale fluctuations to the large-scale coherent soliton. Eventually, the coherent structure attains a maximum amplitude, and subsequently, the coherent soliton and small-scale radiation appear to coexist in a statistically steady state.

Further evidence of the tendency of the solution of the NLS system (4.1) to approach a state of statistical equilibrium is furnished by the time evolution of the kinetic and potential energies (see Figure 3). The sum of these two quantities is the

Hamiltonian, and therefore, remains constant in time. We observe, however, that the kinetic energy increases monotonically, and consequently, the potential energy decreases monotonically as time goes on. The initial time period where these quantities evolve most rapidly (say $t < 20000$) corresponds to the first two stages of the dynamics described above, in which the modulational instability creates an array of soliton-like structures which then coalesce into a single coherent soliton. After the coalescence has ended, the kinetic (potential) energy increases (decreases) very slowly to its saturation value. In the process, fluctuations develop on increasingly finer spatial scales, which accounts for the gradual increase of kinetic energy. Simultaneously, the surviving soliton slowly absorbs mass from the background fluctuations, thereby increasing the magnitude of the contribution to the potential energy from the coherent structure. In the long-time limit, the soliton accounts for the vast majority of the potential energy, while the fluctuations make a nonnegligible contribution to the kinetic energy.

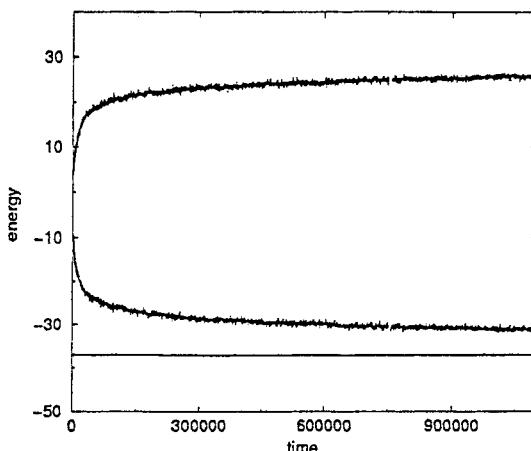


Figure 3. Time evolution of the kinetic (upper curve) and the potential (middle curve) energies. The kinetic energy is increasing and consequently the potential energy is decreasing, *in accord* with the statistical theory developed above. The lower line indicates the potential energy of the solitary wave that contains all the particles of the system. The curves are obtained from an ensemble average over 16 initial conditions for $n = 512$. The length of the system is $L = 128$, and the (conserved) values of the particle number and the Hamiltonian are, respectively, $N^0 = 20.48$ and $H^0 = -5.46$.

The statistical theory described above provides a prediction for the expected value of the kinetic energy K_n in statistical equilibrium for a given number of

modes n . This is $\langle K_n \rangle = K_n(\langle \psi^{(n)} \rangle) + H^0 - H_n^*$, which follows directly upon multiplying equation (3.10) by k_j^2 and summing over j . The first term in this expression for $\langle K_n \rangle$ is the contribution to the mean kinetic energy from the coherent soliton structure which minimizes the Hamiltonian H_n subject to the particle number constraint $N_n = N^0$. The second term in $\langle K_n \rangle$ is the contribution to the expectation of the kinetic energy from the fluctuations. H_n^* converges to $K(\psi^\infty) + H^0 - H^*$, where ψ^∞ is the minimizer of the Hamiltonian H given the particle number constraint $N = N^0$ for continuous NLS system on the interval $[0, L]$ and $H^* = H(\psi^\infty)$. Approximating $K(\psi^\infty)$ and $H(\psi^\infty)$ by $K(\phi)$ and $H(\phi)$, where ϕ is the solitary wave on the real line whose particle number is N^0 , we obtain for the setting considered in Figure 3 the large n estimates $K_n(\langle \psi^{(n)} \rangle) \approx 9.2$, $H^0 - H_n^* \approx 22.4$, and therefore, $\langle K_n \rangle \approx 31.6$. Also, according to the statistical theory, the expected value $\langle \Theta_n \rangle$ of the potential energy in statistical equilibrium should converge as $n \rightarrow \infty$ to $\Theta(\psi^\infty)$. Approximating this by $\Theta(\phi)$, with ϕ as above, we have the estimate $\langle \Theta_n \rangle \approx -37.1$, which we expect to be accurate for sufficiently large n . We see that the kinetic (potential) energy of the numerical solution is bounded above (below) by the estimate based on the statistical theory. As expected, the theoretically predicted value of the average kinetic energy for a finite number of modes is not attained. The reason for the difference is that a nonzero amount of the particle number and the potential energy are actually contained in the background radiation (according to the statistical theory, the expected contribution of the fluctuations to these quantities is $O(1/n)$, where n is the number of spectral modes – this follows from (3.10) [15]). It may be checked [20] that when the spatial resolution is improved, the contributions of the radiation to the particle number and the potential energy decrease, and the saturation values of the kinetic and potential energy attained in the numerical simulations approach more closely the predicted statistical equilibrium averages of these quantities.

Figures 1 and 3 offer evidence that, for a given (large) number of modes n , the dynamics converges in the long-time limit when to a state consisting of a large-scale coherent soliton, which accounts for all but a small fraction of the particle number and the potential energy integrals, coupled with small-scale radiation, or fluctuations, which account for the discrepancy between the total kinetic energy and the kinetic energy contained in the coherent structure. In fact, equation (3.10) suggests that, in the long-time limit, the coherent structure and the background radiation exist in balance (or in statistical equilibrium) with each other through the equipartition of the kinetic energy of the fluctuations. In Figure 4, we display the

particle number spectral density $|\psi_k|^2$ as a function of the wave number k for a long time run, where ψ_k is the Fourier transform of the field ψ . This spectrum is obtained through an ensemble average over 16 initial conditions, and a time average over the final 1000 time units for each run. For comparison, we display in this figure the spectrum of the solitary wave (4.2) whose particle number is equal to conserved value of the particle number for the simulation. There is both a qualitative and quantitative agreement between the spectrum of this solitary wave solution and the small wavenumber portion of the spectrum arising from the numerical simulations. This is in agreement with the statistical theory, which predicts that the coherent structure should coincide with this solitary wave (in the limit $n \rightarrow \infty$). For larger wavenumbers, the spectrum of the numerical solution is dominated by the small scale fluctuations. We have indicated on the graph the large wavenumber spectrum predicted by the statistical theory. This prediction comes from the second expression on the right hand side of equation (3.10), except that we have approximated the minimum value H_n^* of the Hamiltonian for the spectrally truncated system by the Hamiltonian H^* of the above-mentioned solitary wave solution for the continuum system. We observe a good qualitative agreement with the predicted $\propto k^{-2}$ spectrum, corresponding to the equipartition of kinetic energy amongst the small-scale fluctuations. Furthermore, there is an excellent quantitative agreement between the numerical results and the formula (3.10) for large k .

As mentioned above, the numerical spectrum shown in Figure 4 arises from an ensemble average over long time and over different initial conditions (with the same values of the particle number and the Hamiltonian). Under the assumption that the dynamics is ergodic, such an average should coincide with an average with respect to the microcanonical ensemble for the spectrally truncated NLS system [18]. Since it has been shown that the mean-field statistical ensembles $\rho^{(n)}$ constructed above concentrate on the microcanonical ensemble in the continuum limit $n \rightarrow \infty$ (see Theorem 3 of reference [15]), averages with respect to $\rho^{(n)}$ for large n should agree with the ensemble average of the numerical simulations over initial conditions and time, assuming ergodicity of the dynamics. While we have not shown that the dynamics is ergodic, we have, in fact, demonstrated a convincing agreement between the predictions of the mean-field ensembles $\rho^{(n)}$ and the results of direct numerical simulations. In [20], we have also compared the long-time saturation values of the quantities

$$S_m(\psi^{(n)}) = \sum k_j^{2m} |\psi_j|^2,$$

attained in numerical simulations with the predicted statistical equilibrium averages under the mean-field maximum entropy ensemble, where m is a positive integer.

Once again, a close agreement between the numerical and theoretically predicted values is found.

5. CONCLUSIONS

The primary purpose of the present work has been to test the predictions of a mean-field statistical model of self-organization in a generic class of nonintegrable focusing NLS equations defined by equation (1.1). This statistical theory, which has been summarized above, was originally developed and analyzed in [15]. In fact, we have demonstrated a remarkable agreement between the predictions of the statistical theory and the results of direct numerical simulations of the NLS system. There is a

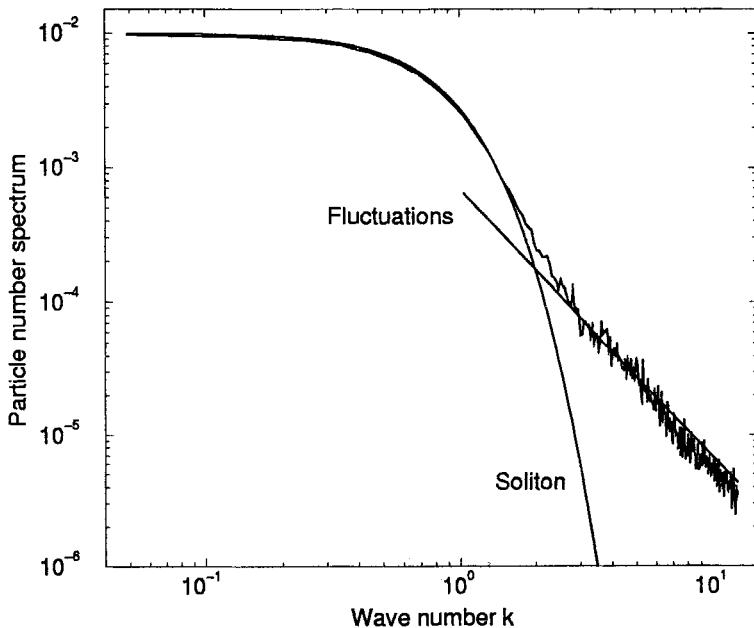


Figure 4. Particle number spectral density $|\psi_k|^2$ as a function of k for $t = 1.1 \times 10^6$ unit time (upper curve). The lower curve (smooth one) is the particle number spectral density for the solitary wave that contains all the particles of the system. The straight line drawn for large k corresponds to the statistical prediction (3.10) for the spectral density for large wavenumbers. The numerical simulation has been performed with $n = 512$, $dx = 0.25$, $N^0 = 20.48$ and $H^0 = -5.46$.

strong qualitative and quantitative agreement between the mean field predicted by the statistical theory and the large-scale coherent structure observed in the long-time numerical simulations. In addition, the statistical model accurately predicts the long-time spectrum of the numerical simulations. In addition, the statistical model accurately predicts the long-time spectrum of the numerical solution of the NLS system. The main conclusions we have reached are 1) The coherent structure that emerges in the asymptotic time limit is the solitary wave that minimizes the system Hamiltonian subject to the particle number constraint $N = N^0$, where N^0 is the given (conserved) value of N , and 2) The difference between the conserved Hamiltonian and the Hamiltonian of the coherent state resides in Gaussian fluctuations equipartitioned over wavenumbers. Further comparisons between the predictions of the statistical theory and the results of direct numerical simulations of NLS may be found in [20].

ACKNOWLEDGMENTS

It is a pleasure to thank Robert Almgren, Shiyi Chen, Richard Ellis, Leo Kadanoff, Yves Pomeau, Kim Rasmussen, and Scott Zoldi for valuable discussions and suggestions. R.J. also thanks Bruce Turkington and Craig Zirbel for collaborations on the mean-field statistical theory for NLS systems. R.J. acknowledges support from an NSF Mathematical Sciences Postdoctoral Research Fellowship. The research of C.J. has been supported by the ASCI Flash Center at the University of Chicago under DOE contract B341495.

REFERENCES

- [1] Rivera, M., Vorobieff, P., and Ecke, R., Turbulence in flowing soap films: Velocity, vorticity, and thickness, *Phys. Rev. Lett.*, 81, 1998, 1417-1420.
- [2] Segre, E. and Kida, S., Late states of incompressible 2D decaying vorticity fields, *Fluid Dyn. Res.*, 23, 1998, 89-112.
- [3] Montgomery, D., Matthaeus, W., Stribling, W., Martinez, D., and Oughton, S., Relaxation in two dimensions and the ‘Sinh-Poisson’ equation, *Phys. Fluids A*, 4, 1992, 3-6.
- [4] McWilliams, J.C., The emergence of isolated vortices in turbulent flow, *J. Fluid Mech.*, 146, 1984, 21-43.
- [5] Hasegawa, A., Self-organization processes in continuous media, *Adv. Phys.*, 34, 1985, 1-42.
- [6] Ablowitz, M. and Segur, H., On the evolution of packets of water waves, *J. Fluid Mech.*, 92, 1979, 691-715.

- [7] Pécseli, H.L., Solitons and weakly nonlinear waves in plasmas, IEEE Trans. Plasma Sci., 13, 1985, 53-86.
- [8] Hasegawa, A. and Kodama, Y. *Solitons in Optical Communications*, Oxford University Press, New York, 1995.
- [9] Zakharov, V.E. and Shabat, A.B., Exact theory in two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media, Soviet Phys. JETP, 34, 1972, 62-69.
- [10] Zakharov, V.E., Pushkarev, A.N., Shvets, V.F., and Yan'kov, V.V., Soliton turbulence, JETP Lett. 48, 1988, 83-87.
- [11] Bourgain, J., Periodic nonlinear Schrödinger equations and invariant measures, Commun. Math. Phys., 166, 1994, 1-26.
- [12] Rasmussen, J.J. and Rypdal, K., Blow-up in nonlinear Schrödinger equations, Physica Scripta, 33, 1986, 481-504.
- [13] Mihalache, D., Nazmitdinov, R.G., and Fedyanin, V.K., Nonlinear optical waves in layered structures, Soviet J. Part. Nucl., 20, 1989, 86-92.
- [14] Pomeau, Y., Asymptotic time behavior of solutions of nonlinear classical field equations, Nonlinearity, 5, 1992, 707-720.
- [15] Jordan, R., Turkington, B., Zirbel, C.L., A mean-field statistical theory for the nonlinear Schrödinger equation, Physica D, 137, 2000, 353-378.
- [16] Zhidkov, P.E., On an invariant measure for a nonlinear Schrödinger equation, Soviet Math. Dokl., 43, 1991, 431-434.
- [17] Bidégaray, B., Invariant measures for some partial differential equations, Physica D, 82, 1995, 340-364.
- [18] Balescu, R., *Equilibrium and Nonequilibrium Statistical Mechanics*, Wiley, New York, 1975.
- [19] Jaynes, E.T., Information theory and statistical mechanics, Phys. Rev., 106, 1957, 620-630.
- [20] Jordan, R. and Josserand, C., Self-organization in nonlinear wave turbulence, Phys. Rev. E., 61, 2000, 1527-1539.
- [21] Ellis, R.S., Jordan, R., and Turkington, B., Equilibrium ensembles for the nonlinear Schrödinger equation, in preparation.

14 FUZZY SETS AND FUZZY DIFFERENTIAL EQUATIONS

V. Lakshmikantham

Department of Applied Mathematics

Florida Institute of Technology

Melbourne, FL 32901

and

Ram N. Mohapatra

Department of Mathematics

University of Central Florida

Orlando, FL 32816

1. INTRODUCTION

Solution of real world problems often rely on solutions of mathematical models of empirical phenomena. It is well known that the precision and exactness necessary during the construction and solution of such models are not always true in real situations. The major difficulty encountered by a model builder is to express imprecise notions in a seemingly precise form. Conventional mathematics is not equipped to handle vagueness. As researchers and mathematical model builders continue their efforts to construct intelligent systems they are coming to grip with the issue of uncertainty in human knowledge and reasoning. As new fields of study like general system theory, robotics, artificial intelligence and language theory continue to grow, we are forced to specify imprecise notions and analyze them. In 1965, Zadeh [31] introduced a modification of set theory known as fuzzy set theory to study notions with prescribed vagueness.

Since the publication of Zadeh's paper [31], there has been a lot of research to extend algebraic and topological notions valid in crisp sets to fuzzy sets. Possibility theory has been introduced. Mathematicians and engineers are introducing new

notions to study geometrical properties and their applications to pattern recognition and image processing.

In spite of all the advancement in the theory of fuzzy sets, there are only a few results on fuzzy differential equations. It is our objective to show how basic properties of fuzzy differential equations can be studied with the help of appropriate comparison theorems. After stating our results on differential equations, we shall mention some results on stability for fuzzy differential equations and then outline the application of differential equations to study fixed points of fuzzy valued mappings.

2. FUZZY SETS

Hereafter, our ordinary notion of a set will be denoted as crisp set. One of the primary notions for a set is the notion of a membership from a universe of discourse. A sharp, unambiguous distinction exists between members and nonmembers of the collection in a crisp set. However, many of the collections we commonly employ do not exhibit this character. Classes of tall people, expensive cars, highly contagious diseases, numbers close to zero, sunny or cloudy days, present examples where membership does not follow from our two valued logic.

Fuzzy set introduces vagueness with the aim of reducing complexity of eliminating sharp boundaries dividing members of a collection from nonmembers. It can be defined by assigning to each possible individual in the universe of discourse a value representing its membership grade. This grade of membership represents the degree to which an individual member is similar or compatible with the concept represented by the fuzzy set.

A fuzzy subset A of a set X with membership function μ_A is a mapping $\mu_A : X \rightarrow [0,1]$. $\mu_A(x) = 0$ and $\mu_A(x) = 1$ represent the fact that the member $x \in X$ "does not" and "does" belong to A , respectively.

If A is the set of real numbers close to zero, then a possible membership function μ_A is given by

$$\mu_A(x) = (1 + 10x^2)^{-1}.$$

Some values of $\mu_A(x)$ are

x	3	1	.85	.25	.1	0
$\mu_A(x)$.01	.09	.12	.62	.91	1

It must be remarked that any function from X into $[0,1]$ will not be a membership function (see [11]).

Although the range of values between 0 and 1, inclusive, is the one most commonly used for $\mu_A(x)$, any arbitrary set with some natural full or partial ordering can be used. Thus

$$\mu_A : X \rightarrow L$$

where L is a lattice with respect to some partial ordering may be used. Fuzzy sets with such a membership function are called L -fuzzy sets (see [1], [2], [11], [21]).

The usual question that do fuzzy sets indicate some form of probability has a negative answer. It can be seen by using $\mu_A(x) = (1 + 10x^2)^{-1}$ that

$$\mu_A(3) + \mu_A(.85) + \mu_A(.25) + \mu_A(.1) = 1.66.$$

If the set $L = \{0,1\}$, then the fuzzy set reduces to the crisp set A . Thus, the membership function μ_A is a generalization of the characteristic function for a set A .

Given the isomorphism between usual set theory and the two-valued propositional logic where each proposition is either true or false, one wonders if there is an infinite valued logic which corresponds to fuzzy set theory. In order to understand that more than two valued logics have naturally struck the minds of logicians, we give a brief account of 2, 3, n valued and infinite valued logics.

Even during the formative years of the initial concept of propositional logic, Aristotle had questioned the universal appropriateness of two-valued logic. He maintained that there are matters whose status may be future contingent and hence, the truth status cannot be one of “true” or “false”, but can be potentially either of the two.

As a consequence of uncertainty principle in quantum mechanics truth values of certain propositions are inherently “indeterminate” due to fundamental limitations of measurement. In order to deal with such propositions the true, false dichotomy of the two-valued logic must be relaxed. Thus, a third truth-value called “indeterminate” can be assigned to propositions. Thus,

$$\{\text{false, indeterminate, true}\} \leftrightarrow \text{truth values are } \left\{0, \frac{1}{2}, 1\right\}.$$

With negation of a proposition with truth value “a” is “1-a” led the foundations of three-valued logics. Five best known three-valued logics are “Lukasiewicz”, “Bochvar”, “Kleene”, “Heyting”, and “Reichenbach”. In the 1930’s n -valued logics with truth values $\left\{0, \frac{1}{n-1}, \dots, \frac{n-2}{n-1}, 1\right\}$ were introduced. With these developments, it was clear that infinite-valued logics are possible where the truth

value is any member of the interval $[0, 1]$. It is known as standard Lukasiewicz logic L_{χ_1} or L_1 .

Now it is easy to see that the infinite-valued logic L_1 is isomorphic to fuzzy set theory.

An extension of nonfuzzy mathematical concepts to fuzzy environment is accomplished through the use of Zadeh's extension principle (see [32], [33], and [34]).

In 1968 Chang (see [3] and [25]) introduced the notion of fuzzy topological spaces. In 1976 Lowen (see [25]) modified Chang's definition and obtained results on fuzzy topological spaces and paracompactness. Using the concept of fuzzy space Dib [3] defined fuzzy topologies on the fuzzy space. Pascali and Ajmal [25] have discussed the trend of research which clearly had two prominent directions. In one, the research was carried out keeping as a model the concepts and results of general topological spaces and extending them to the framework of fuzzy setting. Although this type of study allows one to obtain the results of general topology as a particular case, it raises a pertinent question viz. What new fuzzy topology is doing to mathematics? Is there sufficient validity to seek such generalizations? In the other direction the research is more of a philosophical nature where a categorical framework leads the researchers to seek new categories. Attempts have been made to define categories of fuzzy topological spaces which behave nicely as the category of general topological spaces and also contain it as a proper subcategory. Pascali and Ajmal [25] have shown peculiarities and specialties of fuzzy settings by introducing certain closure and interior operators. They have shown that when these operators act on a given fuzzy topological space, they split the space into two fuzzy topological spaces. They call this an α -decomposition. They have obtained conditions under which a fuzzy topology admits an α -decomposition and have provided interesting examples.

The notion of convex and concave mappings play an important role on fuzzy sets and their applications. We begin with the concept of a fuzzy number.

Definition. Let R^1 be the set of reals. A fuzzy number is a fuzzy set u such that $u : R^1 \rightarrow [0, 1]$ satisfies the following:

$$u \text{ is upper semi-continuous;} \quad (2.1)$$

$$u \text{ is fuzzy convex, i.e.,} \quad (2.2)$$

$$u(\lambda t_1 + (1 - \lambda) t_2) \geq \min\{u(t_1), u(t_2)\}$$

for all $t_1, t_2 \in R^1$ and $\lambda \in [0, 1]$;

$$u \text{ is normal, i.e., there exists some } t_0 \in R^1 \text{ such that } u(t_0) = 1; \quad (2.3)$$

the closure of the support of u is compact. (2.4)

A concept that is very useful is the concept of α -level sets.

Definition. Let \mathcal{F}_0 be the set of all fuzzy numbers. The α -level set of a fuzzy number $u \in \mathcal{F}_0$ denoted by $[u]^\alpha$, $0 \leq \alpha \leq 1$, is defined by

$$[u]_\alpha = \begin{cases} \{t \in R^1 : u(t) \geq \alpha\}, & 0 < \alpha \leq 1; \\ \text{closure (support of } u\text{)}, & \alpha = 0. \end{cases}$$

It can be shown (see [24]) that α -level sets of a fuzzy number is a closed and bounded interval $[a^\alpha, b^\alpha]$ for some real numbers a and b .

Since each $r \in R^1$ can be considered as a fuzzy number \tilde{r} defined by

$$\tilde{r}(t) = \begin{cases} 1, & t = r; \\ 0, & t \neq r; \end{cases}$$

R^1 can be embedded in \mathcal{F}_0 (see [28]).

Convex and concave fuzzy mappings are defined as follows:

Definition. A fuzzy mapping $F : K \rightarrow \mathcal{F}_0$ defined on a convex set K is said to be convex if

$$F(\lambda x + (1 - \lambda) y) \leq \lambda F(x) + (1 - \lambda) F(y) \quad (2.5)$$

for $0 \leq \lambda \leq 1$ and $x, y \in K$. F is strictly convex if strict inequality holds for $x \neq y$ and $\lambda \in (0, 1)$.

One can define concavity of F when the relation " \leq " in (2.5) is replaced by \geq .

Characterization of convex and concave fuzzy mappings are given by the following two theorems.

Theorem A ([28]). A fuzzy mapping $F : K \rightarrow \mathcal{F}_0$ is convex if and only if for all $x, y \in K$, $\lambda \in R^1$, and all $u, v \in \mathcal{F}_0$ such that $F(x) < u$, $F(y) < v$, $0 < \lambda < 1$,

$$F(\lambda x + (1 - \lambda) y) < \lambda u + (1 - \lambda) v. \quad (2.6)$$

Theorem B ([28]). A fuzzy mapping $F : K \rightarrow \mathcal{F}_0$ is concave if and only if for all $x, y \in K$, $\lambda \in R^1$, and all $u, v \in \mathcal{F}_0$ such that $F(x) > u$, $F(y) > v$, $0 < \lambda < 1$,

$$F(\lambda x + (1 - \lambda) y) > \lambda u + (1 - \lambda) v. \quad (2.7)$$

Using Theorem A (or Theorem B), Syau [28] have obtained the following results.

Theorem C [28]. A fuzzy mapping $F : K \rightarrow \mathcal{F}_0$ is convex (or concave) if and only if the set

$$\{(x, u) : u \in K, u \in \mathcal{F}_0, f(x) < u\} \quad (2.8)$$

(or $\{(x, u) : u \in K, u \in \mathcal{F}_0, f(x) > u\}$) is convex (or concave).

Biggest success of fuzzy systems in industrial and commercial applications is achieved with fuzzy controllers (see [4], [29], and [35]). In classical control approaches a physical model of the process is developed. Fuzzy control tries to model a human expert. Fuzzy control is a knowledge based model which can be perceived as a way of defining a nonlinear table based system where the definition of the nonlinear transition function can be made without the need to specify each entry of the table individually. Examples of such controllers can be the following:

- (i) An electrical device moving an elevator.
- (ii) An electrical device used for heating a large shopping complex.

Tang et al. [29] have designed a globally stable adaptive indirect controller for a class of continuous time systems for which an explicit linear parameterization of the uncertainty in the dynamics is either unknown or impossible to determine with an external disturbance. The designing procedure comprises of a fuzzy logic system which is used to approximate nonlinear functions and an adaptive law is derived. Then a fuzzy sliding controller is constructed and is added to the adaptive controller for compensating for the uncertainties and for increasing the robustness.

Zhang and Knoll [35] have designed fuzzy controllers based on the B -spline model. They have used their system for prediction of a chaotic system generated by the Mackey-Glass equation.

3. FUZZY DIFFERENTIAL EQUATIONS

When a real world problem is modeled by a deterministic initial value problem (IVP) viz.

$$x' = f(t, x), \quad x(t_0) = x_0 \quad \left(x' = \frac{dx}{dt} \right), \quad (3.1)$$

we cannot always be sure that the model is perfect. The initial value may not be known exactly, and the function f may contain uncertain parameters. If the parameters are estimated through certain measurements, then they may be subject to errors. The analysis of these errors leads to the study of the qualitative behavior of the solutions of (3.1). If the underlying structure is not probabilistic because of the subjective choices, it would be natural to employ fuzzy differential equations.

Calculus of fuzzy functions have been investigated by Dubois and Prade (see [5], [6], and [7]). Puri and Ralescu [26] have studied the differentials of fuzzy

functions and structurally stable differential systems have been studied by Markus [19].

In 1987 Kaleva [9] investigated fuzzy-set valued mappings of a real variable whose values are normal, convex, upper semicontinuous, and compactly supported fuzzy sets in R^n . He investigated differentiability and integrability properties of such functions and proved an existence and uniqueness theorem for a solution to a fuzzy differential equation.

Existence and uniqueness of the solution for fuzzy initial value problem was considered under Lipschitz conditions by Kaleva [9] and [10]. The result corresponding to the classical Peano's theorem for local existence, uniqueness, and continuity has been studied under various conditions but has not reached a satisfactory level (see [8], [9], [10], and [12]). In 1997 Nieto [23] discussed all the existing results on fuzzy IVP and remarked that one can consider fuzzy differential equations as differential equations on a Banach space. He proved a version of the classical existence theorem of Peano for fuzzy differential equations. Using Ascoli's theorem he has shown that the initial value problem for the fuzzy differential equation (3.1) has a solution if f is continuous and bounded. Since continuity and boundedness of f does not imply Lipschitz continuity Nieto's result complements that of Kaleva [9, Theorem 6.1].

The remaining part of this paper will be devoted to the study of basic properties of solutions of IVP for fuzzy differential equations, their stability and applications. To precisely state our results, we need the following preliminary notations.

Let $P_k(R^n)$ denote the family of all nonempty compact, convex subsets of R^n . If $\alpha, \beta \in R^1$ and $A, B \in P_k(R^n)$, then $\alpha(A + B) = \alpha A + \alpha B$, $\alpha(\beta A) = (\alpha\beta)A$, $1A = A$ and if $\alpha, \beta \geq 0$, then $(\alpha + \beta)A = \alpha A + \beta A$. Let $I = [t_0, t_0 + a]$ and $a > 0$.

Let us denote

$E^n := u : R^n \rightarrow [0, 1]$ such that u satisfies (i) to (iv) mentioned below:

- (i) u is normal, that is, there exists an $x_0 \in R^n$ such that $u(x_0) = 1$;
- (ii) u is fuzzy convex, i.e., for $x, y \in R^n$ and $0 \leq \lambda \leq 1$,
 $u(\lambda x + (1 - \lambda)y) \geq \min[u(x), u(y)]$;
- (iii) u is upper semicontinuous;
 $[u]^\alpha = \{x \in R^n : u(x) \geq \alpha\}$ $0 < \alpha < 1$,
- (iv) $[u]^0 = \text{cl. } \{x \in R^n : u(x) > 0\}$ is compact.

Remark. It follows from (i) to (iv) that the α -level sets $[u]^\alpha \in P_k(R^n)$, for $0 \leq \alpha \leq 1$.

If C, D are bounded subsets of \mathbb{R}^n , then $d_H(C, D)$, the Hausdorff distance between the sets C and D is defined by

$$d_H(C, D) := \max \left\{ \sup_{a \in C} \inf_{b \in D} \|a - b\|, \sup_{b \in D} \inf_{a \in C} \|a - b\| \right\}$$

where $\|\cdot\|$ denotes the usual Euclidean norm in \mathbb{R}^n .

We define

$$d(u, v) := \sup_{0 \leq \alpha \leq 1} d_H([u]^\alpha, [v]^\alpha).$$

d defines a metric in E^n and (E^n, d) is a complete metric space. The following properties of d can be verified easily (see [9]):

$$d[u + \omega, v + \omega] = d[u, v], \quad (3.2)$$

$$d[\lambda u, \lambda v] = |\lambda| d[u, v], \quad (3.3)$$

$$d[u, v] \leq d[u, \omega] + d[\omega, v], \quad (3.4)$$

for all $u, v \in E^n$ and $\lambda \in R^1$.

For $x, y \in E^n$ if there exists $z \in E^n$ such that $x = y + z$, then z is called the H -difference of x and y and is denoted by $x - y$.

Definition. A mapping $F : I \rightarrow E^n$ is differentiable at $t \in I$, if there exists a $F'(t) \in E^n$ such that the limits

$$\lim_{h \rightarrow 0^+} \frac{F(t+h) - F(t)}{h} \text{ and } \lim_{h \rightarrow 0^+} \frac{F(t) - F(t-h)}{h}$$

exists and are equal to $F'(t)$ the limits being taken in the metric space (E^n, d) .

Moreover, if $F : I \rightarrow E^n$ is continuous, then it is integrable and for $a < c < b$

$$\int_a^b F = \int_A^C F + \int_c^b F.$$

The following properties of the integral are valid (see [6], [7], [8], and [9]).

If $F, G : I \rightarrow E^n$ are integrable, $\lambda \in R$, then the following hold:

$$\int (F + G) = \int f + \int G;$$

$$\int \lambda F = \lambda \int F, \quad \forall \lambda \in R;$$

$d[F, G]$ is integrable

$$d \left[\int F, \int G \right] \leq \int d[F, G].$$

The analogue of the fundamental theorem of differential calculus is also true. Precisely, if $F : I \rightarrow E^n$ is continuous, then the integral $G(t) = \int_{t_0}^t F$ is differentiable and $G'(t) = F(t)$. Furthermore,

$$F(t) - F(t_0) = \int_{t_0}^t F'(s) ds.$$

4. COMPARISON PRINCIPLES AND THEIR APPLICATIONS

Consider the fuzzy differential system

$$u' = f(t, u), \quad u(t_0) = u_0, \quad (4.1)$$

where $f \in C[I \times E^n, E^n]$ and $I = [t_0, t_0 + a]$, $t_0 \geq 0$, $a > 0$. $u : I \rightarrow E^n$ is a solution of the IVP (4.1) if and only if it is continuous and it satisfies the integral equation

$$u(t) = u_0 + \int_{t_0}^t f(s, u(s)) ds \quad \text{for } t \in I.$$

Using the properties of $d[u, v]$ and the integral listed above, and the theory of differential and integral inequalities Lakshmikantham and Mohapatra [13] established the following comparison principle.

Theorem 1. Assume that $f \in C[I \times E^n, E^n]$ and for $t \in I$, $u, v \in E^n$,

$$d[f(t, u), f(t, v)] \leq g(t, d[u, v]) \quad (4.2)$$

where $g \in C[I \times R_+, R_+]$ and $g(t, \omega)$ is nondecreasing in ω for each t . Suppose further that the maximal solution $r(t, t_0, \omega)$ of the scalar differential equation

$$\omega' = g(t, \omega), \quad \omega(t_0) = \omega_0 \geq 0, \quad (4.3)$$

exists on I . Then, if $u(t)$, $v(t)$ are any two solutions of (4.1) through (t_0, u_0) and (t_0, v_0) respectively on I , we have

$$d[u(t), v(t)] \leq r(t, t_0, \omega_0), \quad t \in I, \quad (4.4)$$

provided $d[u_0, v_0] \leq \omega_0$.

Remark. If one employs the theory of differential inequalities instead of integral inequalities, then it is possible to dispense with the monotone character of $g(t, \omega)$ assumed in Theorem 1.

Our next result is

Theorem 2 [13]. Let the assumptions of Theorem 1 hold except the nondecreasing property of $g(t, \omega)$ in ω . Then (4.4) holds.

The next comparison principle provides an estimate under weaker conditions:

Theorem 3. Let $f \in C[I \times E^n, E^n]$ and

$$\begin{aligned} & \limsup_{h \rightarrow 0^+} \frac{1}{h} \left[d[u + h f(t, u), v + h f(t, v)] \right] - d[u, v] \\ & \leq g(t, d[u, v]), \quad t \in I, \quad u, v \in E^n, \end{aligned} \tag{4.5}$$

where $g \in C[I \times R_+, R]$. The maximal solution $r(t, t_0, \omega_0)$ of (16) exists on I . Then the conclusion of Theorem 1 holds.

Let us define \hat{O} by

$$\hat{O}(t) = \begin{cases} 1, & t = t_0, \\ 0, & \text{elsewhere.} \end{cases}$$

As a special case of Theorems 1, 2, and 3, we have the following.

Corollary 1. Assume that $f \in C[I \times E^n, E^n]$ and either

$$(a) \quad d[f(t, u), \hat{O}] \leq g(t, d[u, \hat{O}])$$

or

$$\begin{aligned} (b) \quad & \limsup_{h \rightarrow 0^+} \frac{1}{h} \left[d[u + h f(t, u), \hat{O}] - d[u, \hat{O}] \right] \\ & \leq g(t, d[u, \hat{O}]) \end{aligned}$$

where $g \in C[I \times R_+, R]$. Then, if $d[u_0, \hat{O}] \leq \omega_0$, we have

$$d[u(t), \hat{O}] \leq r(t, t_0, \omega_0), \quad t \in I,$$

where $r(t, t_0, \omega_0)$ is the maximal solution of (16) on I .

Our next result gives the existence and uniqueness of the solution of the IVP (4.1) under an assumption more general than the Lipschitz type condition.

Theorem 4 (see [13]). Assume that

- (a) $f \in C[R_0, E^n]$ where $R_0 = [I \times B(u_0, b)]$, $B(u_0, b) = \{x \in E^n : d[u, u_0] \leq b\}$ and $d[f(t, x), \hat{O}] \leq M_0$ on R_0 ;
- (b) $g \in C[I \times [0, 2b], R_+]$, $0 \leq g(t, \omega) \leq M_1$ on $I \times C_0[0, 2b]$, $g(t, 0) = 0$, $g(t, \omega)$ is nondecreasing in ω for each $t \in I$ and $\omega(t) \equiv 0$ is the unique solution of (4.3) on I ;

(c) $d[f(t, u), f(t, v)] \leq g(t, d[u, v])$ on R_0 .

Then the successive approximations defined by

$$u_{n+1}(t) = u_0 + \int_{t_0}^t f(s, u_n(s)) ds, \quad n = 0, 1, \dots, \quad (4.6)$$

exist on $[t_0, t_0 + \eta]$ where $\eta = \min \left[a, \frac{b}{M} \right]$, $M = \max(M_0, M_1)$ as continuous functions and converge uniformly to the unique solution $u(t)$ of (4.1) on $[t_0, t_0 + \eta]$.

The next result discusses the continuous dependence of solutions with initial values.

Theorem 5 (see [13]). Let the assumptions of Theorem 4 hold. Also, further that the solutions $\omega(t, t_0, \omega_0)$ of (4.3) through every point (t_0, ω_0) are continuous with respect to (t_0, ω_0) . Then the solutions $u(t, t_0, u_0)$ of (4.1) are continuous relative to (t_0, u_0) .

Remark. For the system (4.1), local existence has been considered in [8], [9], [12], and [23]. Assuming local existence of the solution can one prove global existence?

Our next result answers this question in the affirmative.

Theorem 6 (see [13]). Assume $f \in C[R_+ \times E^n, E^n]$ and

$$d[f(t, u), \hat{O}] \leq g(t, d[u, \hat{O}]), \quad (t, u) \in R_+ \times E^n,$$

where $g \in C[R_+, R_+]$, $g(t, \omega)$ is nondecreasing in ω for each $t \in R_+$ and the maximal solution $r(t, t_0, \omega_0)$ of (4.3) exist on $[t_0, \infty)$. Suppose further that f is smooth enough to guarantee local existence of solutions of (4.1) for any $(t_0, u_0) \in R_+ \times E^n$. Then the largest interval of existence of any solution $u(t, t_0, u_0)$ of (4.1) such that $d[u_0, \hat{O}] \leq \omega_0$ is $[t_0, \infty)$.

5. STABILITY CRITERIA

Study of stability of solutions for ordinary differential equations are well known (see [14], [16], [19], and [20]). In this section we shall discuss stability of solutions of fuzzy differential equations. All the results presented are due to Lakshmikantham and Leela [15].

Consider the fuzzy differential equation

$$u' = f(t, u), \quad u(t_0) = u_0, \quad (5.1)$$

where $f \in C[R_+ \times S(\rho), E^n]$ and $S(\rho) = \{u \in E^n : d[u, \hat{O}] < \rho\}$. We assume that $f(t, \hat{O}) = \hat{O}$ so that we have the trivial solution for (5.1).

To investigate stability criteria, the following comparison result in terms of a Lyapunov function is very important which can be proved via the theory of differential inequalities. Here Lyapunov function serves as a vehicle to transform the fuzzy differential equation into a scalar comparison differential equation and therefore, it is enough to consider the stability properties of the simpler comparison equation.

Theorem 7 (see [15]). Assume that

$$(i) \quad V \in C[R_+ \times S(\rho), R_+], \quad |V(t, u_1) - V(t, u_2)| \leq L d[u_1, u_2], \quad L > 0 \text{ and}$$

$$\begin{aligned} D^+ V(t, u) := \limsup_{h \rightarrow 0^+} \frac{1}{h} [V(t+h, u+h f(t, u)) - V(t, u)] \\ \leq g(t, V(t, u)), \end{aligned}$$

where $g \in C[R_+, R]$.

Then, if $u(t)$ is any solution of (5.1) existing on $[t_0, \infty)$ such that $V(t_0, u_0) \leq \omega_0$, we have

$$V(t, u(t)) \leq r(t, t_0, \omega_0), \quad t \geq t_0,$$

where $r(t, t_0, \omega_0)$ is the maximal solution of the scalar differential equation

$$\omega' = g(t, \omega), \quad \omega(t_0) = \omega_0 \geq 0$$

existing on $[t_0, \infty)$.

We can deduce the following useful corollaries:

Corollary 2 (see [15]). The function $g(t, \omega) \equiv 0$ is admissible in Theorem 5 to yield the estimate

$$V(t, u(t)) \leq V(t_0, u_0), \quad t \geq t_0.$$

Corollary 3 (see [15]). If, in Theorem 5, we strengthen the assumption on $D^+ V(t, u)$ to

$$D^+ V(t, u) \leq -C[\omega(t, u)] + g(t, V(t, u)),$$

where $\omega \in C[R_+ \times S(\rho), R_+]$, $C \in K = \{a \in C[[0, \rho], R_+] : a(\omega) \text{ is increasing in } \omega \text{ and } a(0) = 0\}$, and $g(t, \omega)$ is nondecreasing in ω for each $t \in R_+$, then we get the estimate

$$V(t, u(t)) + \int_{t_0}^t C[\omega(s, u(s))] ds \leq r(t, t_0, \omega_0), \quad t \geq t_0,$$

whenever $V(t_0, u_0) \leq \omega_0$.

Let $u(t)$ be any solution of (5.1) existing on $[t_0, \infty)$.

Definition. We say that the trivial solution of (5.1) is equi-stable, if given $0 < \varepsilon < \rho$ and $t_0 \in R_+$, there exists a $\delta = \rho(t_0, \varepsilon) > 0$ such that

$$d[u_0, \hat{O}] < \delta \text{ implies } d[u(t), \hat{O}] < \varepsilon, \quad t \geq t_0.$$

If δ is independent of t_0 , then the stability is uniform.

Based on this definition, the other notions of stability can be formulated.

The following result provides nonuniform stability criteria under weaker assumptions.

Theorem 7 (see [15]). Assume that

$$(A1) \quad V_1 \in C[R_+ \times S(\rho), R_+], \quad |V_1(t, u_1) - V_1(t, u_2)| \leq L_1 d[u_1, u_2],$$

$$L_1 > 0, \quad V_1(t, u) \leq a_0(t, d[u, \delta]), \quad \text{where } a \in C[R_+ \times [0, p), R_+]$$

and $a_0(t, \cdot) \in k$ for each $t \in R_+$;

$$(A2) \quad D^+ V_1(t, u) \leq g_1(t, V_1(t, u)), \quad (t, u) \in R^+ \times S(\rho), \quad \text{where}$$

$$g_1 \in C[R_+^2, R] \text{ and } g_1(t, 0) \equiv 0;$$

$$(A3) \quad \text{for every } \eta > 0, \quad \text{there exists a } V_\eta \in C[R_+ \times S(\rho) \cap S^c(\eta), R_+]$$

$$|V_\eta(t, u_1) - V_\eta(t, u_2)| \leq L_\eta d[u_1, u_2],$$

$$b(d[u, \hat{O}]) \leq V(t, u) \leq a(d[u, \hat{O}]), \quad a, b \in k,$$

and

$$D^+ V_1(t, u) + D^+ V_\eta(t, u) \leq g_2(t, V_1(t, u) + V_\eta(t, u))$$

$$\text{for } (t, u) \in R_+ \times S(\rho) \cap S^c(\eta),$$

(A4) the trivial solution $\omega_1 \equiv 0$ of

$$\omega'_1 = g_1(t, \omega_1), \quad \omega_1(t_0) = \omega_{10} \geq 0,$$

is equi-stable.

(A5) The trivial solution $\omega_2 = 0$ of

$$\omega'_2 = g_2(t, \omega_2), \quad \omega_2(t_0) = \omega_{20} \geq 0,$$

is uniformly stable.

Then the trivial solution of (5.1) is equi-stable.

The next result offers conditions for equi-asymptotic stability.

Theorem 8. Let the assumptions of Theorem 7 hold except that the condition (A2) is strengthened to

$$(A2^*) \quad D^+ V_1(t, u) \leq -c(\omega(t, u)) + g_1(t, V_1(t, u)), \quad c \in K,$$

$$\omega \in C[R_+ \times S(\rho), R_+],$$

$$|\omega(t, u_1) - \omega(t, u_2)| \leq N d[u_1, u_2], \quad N > 0 \text{ and}$$

$D^+ \omega(t, u)$ is either bounded above or below.

Then the trivial solution of (5.1) is equi-asymptotically stable, if $g_1(t, \omega)$ is monotone nondecreasing in ω and $\omega(t, u) \geq b_0(d[u, \hat{O}])$, $b_0 \in K$. In the above results K is as in Corollary 3.

6. FIXED POINTS OF FUZZY OPERATORS

In this section we shall use the theory of fuzzy differential equations combined with the contraction mapping principle to obtain fixed points of fuzzy mappings. The results are from Lakshmikantham and Vatsala [17].

Consider the autonomous IVP

$$u' = f(u), \quad u(0) = u_0, \quad (6.1)$$

where $f \in C[E^n, E^n]$. Then the following global existence result is in [17].

Theorem 9. Assume that

$$(i) \quad \limsup_{h \rightarrow 0^+} \frac{1}{h} [d(u + h f(u)), v + h f(v)] - d[u, v] \leq -\beta d[u, v] \text{ for some}$$

$$\beta > 0;$$

$$(ii) \quad d[f(u(t)), \hat{O}] \leq M, \text{ whenever } d[u, \hat{O}] \leq L;$$

$$(iii) \quad \text{for each } x_0 \in E^n, \text{ there exists a solution locally on } [0, a].$$

Then for each $x_0 \in E^n$, there is a unique solution $u(t, x_0)$ existing on $[0, \infty)$.

Now consider the fuzzy operator $S \in C[E^n, E^n]$. We are interested to find a fuzzy fixed point of S . We define

$$S(u) = f(u) + u,$$

so that (6.1) possesses a fuzzy constant solution. Then that solution will be the desired fixed point.

To that effect, we have the following.

Theorem 10 (see [17]). Let the assumptions of Theorem 7 hold with $f(u) = S(u) - u$. Then there exists a $u^* \in E^n$ such that $Su^* = u^*$.

Remark. Research in the area of fuzzy differential equation is at its infancy. Fuzzy boundary value problems, solution of fuzzy differential equations with delay, impulsive fuzzy differential equations and solution of a system of fuzzy differential equations are still to be tried.

In addition to this, it is interesting to investigate the real world problems which can be modeled by fuzzy differential equations of one type or the other mentioned above.

REFERENCES

- [1] Chakrabarty, K., Biswas, R., and Nanda, S., Fuzzy L-structure, *Fuzzy Sets and Systems*, 103, 1999, 177-182.
- [2] Cruise, R., Gebhardt, J., and Klawonn, F., *Foundations of Fuzzy Systems*, John Wiley & Sons, 1994.
- [3] Dib, K.A., The fuzzy topological spaces on a fuzzy space, *Fuzzy Sets and Systems*, 108, 1999, 103-110.
- [4] Driankov, D., Hellendoorn, H., and Reinfrank, M., *An Introduction to Fuzzy Control*, Springer Verlag, Berlin, 1996.
- [5] Dubois, D. and Prade, H., Towards fuzzy differential calculus, part I, *Fuzzy Sets and Systems*, 8, 1982, 1-17.
- [6] Dubois, D. and Prade, H., Towards fuzzy differential calculus, part II, *Fuzzy Sets and Systems*, 8, 1982, 105-116.
- [7] Dubois, D. and Prade, H., Towards fuzzy differential calculus, part III, *Fuzzy Sets and Systems*, 8, 1982, 225-234.
- [8] Friedman, M., Ming, M., and Kandel, A., On the validity of the Peano's theorem for fuzzy differential equations, *Fuzzy Sets and Systems*, 86, 1997, 331-334.
- [9] Kaleva, O., Fuzzy differential equations, *Fuzzy Sets and Systems*, 24, 1987, 301-317.
- [10] Kaleva, O., The Cauchy problem for fuzzy differential equations, *Fuzzy Sets and Systems*, 35, 1990, 389-396.
- [11] Kendal, A., *Fuzzy Mathematical Techniques With Applications*, Addison-Wesley, 1986.

- [12] Kloeden, P.E., Remarks on Peano-like theorems for fuzzy differential equations, *Fuzzy Sets and Systems*, 44, 1991, 161-163.
- [13] Lakshmikantham, V. and Mohapatra, R.N., Basic properties of solutions of fuzzy differential equations, to appear.
- [14] Lakshmikantham, V. and Mohapatra, R.N., Strict stability of differential equations, to appear in *Nonlinear Analysis, Theory, Methods and Applications*.
- [15] Lakshmikantham, V. and Leela, S.G., Stability theory of fuzzy differential equations via differential inequalities, *Mathematical Inequalities and Applications*, 4, 1999, 551-559.
- [16] Lakshmikantham, V., Matrosov, V.M., and Sivasundaram, S., *Vector Lyapunov Functions and Stability Analysis of Nonlinear Systems*, Kluwer Academic Publishers, 1991.
- [17] Lakshmikantham V. and Vatsala, A.S., Existence of fixed points of fuzzy mappings via theory of fuzzy differential equations, *Journal of Inequalities and Applications*, 3, 1999, 233-244.
- [18] Lupiáñez, F.G., On fuzzy relative paracompactness, *Fuzzy Sets and Systems*, 101, 1999, 485-488.
- [19] Markus, L., Structurally stable differential systems, *Annals of Mathematics*, 73, 1961, 1-19.
- [20] Massera, J.L., The meaning of stability, *Bol. Fac. Ingen Agrimens. Montevideco*, 8, 1964, 405-429.
- [21] Negoita, C.V. and Ralescu, D.A., *Application of Fuzzy Sets to System Analysis*, Birkhauser Verlag, Basel, 1995.
- [22] Nguyen, H.T., A note on the extension principle for fuzzy sets, *Jour. Math. Anal. Appl.*, 64, 1978, 369-380.
- [23] Nicto, J.J., The Cauchy problem for fuzzy differential equations, *Fuzzy Sets and Systems*, to appear.
- [24] Novak, V., *Fuzzy sets and their applications*, Adam Hilger, Bristol, 1988.
- [25] Pascoli, E. and Ajmal, N., Fuzzy topologies and a type of their decomposition, *Rendiconti di Matematica*, 17, 1997, 305-328.
- [26] Puri, M.L. and Ralescu, D.A., Differentials of fuzzy functions, *Journal of Math. Anal. Appl.*, 91, 1983, 552-558.
- [27] Srivastava, R., Lal, S.N., and Srivastava, A.K., Fuzzy Hausdorff topological spaces, *Jour. Math. Anal. Appl.*, 81, 1981, 497-506.
- [28] Syau, Y.-R., On convex and concave fuzzy mappings, *Fuzzy Sets and Systems*, 103, 1999, 163-168.

- [29] Tang, S.C., Li, Q., and Chai, T., Fuzzy adaptive control for a class of nonlinear systems, *Fuzzy Sets and Systems*, 101, 1999, 31-39.
- [30] Yorke, J.A., Extending Liapunov's second method to non-Lipschitz Liapunov functions, Seminar in Differential Equations and Dynamical Systems (G.S. Jones ed.), Lecture notes in Math., Vol. 60, Springer-Verlag, 1968.
- [31] Zadeh, L.A., Fuzzy sets, *Information and Control*, 8, 1965, 338-353.
- [32] Zadeh, L.A., The concept of linguistic variable and its application to approximate reasoning, *Inform. Sci.*, 8, 1975, 199-249.
- [33] Zadeh, L.A., The concept of linguistic variable and its application to approximate reasoning II, *Inform. Sci.*, 8, 1975, 301-357.
- [34] Zadeh, L.A., The concept of linguistic variable and its application to approximate reasoning III, *Inform. Sci.*, 9, 1975, 43-80.
- [35] Zhang, J. and Knoll, A., Designing fuzzy controllers by rapid learning, *Fuzzy Sets and Systems*, 101, 1999, 287-301.

15 NUMERICAL SOLUTIONS OF COUPLED PARABOLIC SYSTEMS WITH TIME DELAYS

Xin Lu

Department of Mathematics and Statistics
University of North Carolina at Wilmington
Wilmington, NC 28403

ABSTRACT

A monotone iterative scheme is developed for the numerical solutions of coupled parabolic systems with time delays. The differential equation system is discretized by the finite difference method. Using upper and lower solutions as initial iterations, we construct two sequences that converge to a unique solution of the discretized system. The convergence and stability of this numerical scheme are also obtained.

1. INTRODUCTION

Through years of study on real-life problems, people have discovered that the future state of a system is not only determined by its present but also by its past. In order to better describe many physical, biological and ecological processes, the effects of time delays were introduced into differential equations. In recent years, various ordinary and partial differential equations with time delays have been studied rather extensively [1, 2, 3, 4, 6, 7, 11]. But there has not been much work concerning a numerical scheme for coupled systems with time delays. In this paper, we consider the following coupled parabolic differential equation system with time delays

$$\begin{aligned} \partial u_j / \partial t - D_j \Delta u_j &= f_j(x, t, \mathbf{u}(x, t), \mathbf{u}_\tau(x, t)) && \text{in } D_T, \\ B_j u_j &= h_j(x, t) && \text{on } S_T, \quad j = 1, 2, \dots, r, \\ u_j(x, t) &= \eta_j(x, t) && \text{in } \Omega \times J_j \end{aligned} \tag{1.1}$$

where $D_j = D_j(x, t)$ is continuous and positive. T, τ_1, \dots, τ_r , and τ are positive constants with $\tau = \max\{\tau_1, \dots, \tau_r\}$, $T > \tau > 0$. Ω is a bounded domain in R^p ($p = 1, 2, \dots$). $D_T \equiv \Omega \times (0, T]$, $S_T \equiv \partial\Omega \times (0, T]$, and $J_j = [-\tau_j, 0]$. Also $\mathbf{u}(x, t) \equiv (u_1, \dots, u_r)$, $\mathbf{u}_\tau(x, t) \equiv (u_1(x, t - \tau_1), \dots, u_r(x, t - \tau_r))$. The operators B_j for each j are given by

$$B_j u_j \equiv \alpha_j \partial v / \partial \nu + \beta_j(x) u_j,$$

where $\partial/\partial \nu$ denotes the outward normal derivative on $\partial\Omega$. B_j is of either Dirichlet type or Neumann-Robin type, and is allowed to be of different type for different j . We also assume that $\partial\Omega$ is of class $C^{1+\alpha}$, and for each j , h_j and η_j are Hölder continuous on S_T and $\Omega \times J_j$, respectively, and $f_j(x, t, \mathbf{u}, \mathbf{v})$ is Hölder continuous in (x, t) and locally Lipschitz continuous in \mathbf{u} and \mathbf{v} .

The existence and uniqueness of a continuous solution in system (1.1) had been studied by Pao (see [8, 11]) using the method of upper-lower solutions. Generally speaking, the computation for the numerical solution involves solving a nonlinear system of algebraic equation and therefore, requires some kind of iteration scheme. Our main goal in this paper is to derive an algorithm for the numerical solution of (1.1) and to study the convergence and stability of the algorithm. The techniques we use are the method of upper-lower solutions and its associated monotone iterations which have been applied several previous articles for numerical solutions of single equation and coupled systems (see [5, 6, 9, 10, 12]).

This paper is arranged as follows. In Section 2, we first write (1.1) in an alternative form and then discretize the system by the finite-difference method. The upper-lower solutions for the discretized finite-difference system are defined by the quasimonotone property. In Section 3, we develop an iterative algorithm which generates two sequences (namely, the upper and lower sequences), with the upper and lower solutions as initial iterations. We also prove the monotone and convergence properties of the upper and lower sequences. Finally, the convergence of approximation to the true solution of (1.1) and numerical stability of the iterative algorithm are discussed.

2. FINITE-DIFFERENCE SYSTEM AND UPPER-LOWER SOLUTIONS

In order to describe an iterative process for the monotone sequences, we write (1.1) in a special form based on the quasimonotone property of (f_1, f_2, \dots, f_r) [11]. Specifically, by writing \mathbf{u} and \mathbf{u}_τ in split form

$$\mathbf{u} \equiv \left(u_j, [\mathbf{u}]_{a_j}, [\mathbf{u}]_{b_j} \right), \quad \mathbf{u}_\tau \equiv \left([\mathbf{u}_\tau]_{c_j}, [\mathbf{u}_\tau]_{d_j} \right),$$

where a_j, b_j, c_j and d_j are nonnegative integers with

$$a_j + b_j = r - 1, \quad c_j + d_j = r, \quad j = 1, \dots, r, \quad (2.1)$$

we express system (1.1) in the form

$$\begin{aligned} \partial u_j / \partial t - D_j \Delta u_j &= f_j \left(x, t, u_j, [\mathbf{u}]_{a_j}, [\mathbf{u}]_{b_j}, [\mathbf{u}_\tau]_{c_j}, [\mathbf{u}_\tau]_{d_j} \right) \quad \text{in } D_T, \\ B_j u_j &= h_j(x, t) \quad \text{on } S_T, \quad j = 1, 2, \dots, r, \\ u_j(x, t) &= \eta_j(x, t) \quad \text{in } \Omega \times J_j. \end{aligned} \quad (2.2)$$

Here for each fixed $i, [\mathbf{u}]_{a_j}$ and $[\mathbf{u}]_{b_j}$ consist of a_j -components and b_j -components of \mathbf{u} , respectively, and similar interpretation holds for $[\mathbf{u}]_{c_j}$ and $[\mathbf{u}]_{d_j}$.

The exact values of a_j, b_j, c_j and d_j depend on the mixed quasimonotone property of $f \equiv (f_1, f_2, \dots, f_r)$ which is defined as following.

Definition 2.1. A vector function $f(\mathbf{u}, \mathbf{v}) \equiv (f_1(\mathbf{u}, \mathbf{v}), f_2(\mathbf{u}, \mathbf{v}), \dots, f_r(\mathbf{u}, \mathbf{v}))$ defined for \mathbf{u} and \mathbf{v} in a set Λ is said to have a mixed quasimonotone property in Λ if for each $j = 1, \dots, r$, there exist nonnegative integer a_j, b_j, c_j and d_j satisfying (2.1) such that for any $\mathbf{u} \equiv (u_j, [\mathbf{u}]_{a_j}, [\mathbf{u}]_{b_j})$ and $\mathbf{v} \equiv ([\mathbf{v}]_{c_j}, [\mathbf{v}]_{d_j})$ in Λ , $f_j(\mathbf{u}, \mathbf{v})$ is nondecreasing in $[\mathbf{u}]_{a_j}, [\mathbf{v}]_{c_j}$ and is nonincreasing $[\mathbf{u}]_{b_j}, [\mathbf{v}]_{d_j}$.

Now we discretize the system (2.2) by finite difference [13]. Let $i = (i_1, i_2, \dots, i_p)$ be a multiple index with $i_\mu = 0, 1, 2, \dots, M_\mu + 1$ and let $x_i = (x_{i_1}, x_{i_2}, \dots, x_{i_p})$ be an arbitrary mesh point in \overline{D}_T where p is the dimension of Ω and M_μ is the total number of interior points in the x_μ -coordinate direction. Denote by Ω_p, S_p, Λ_p and $\overline{\Lambda}_p$ the set of mesh points in Ω, S_T, D_T and \overline{D}_T , respectively, and by D_p the set of mesh points in $D_{-\tau}$. Also, denote by (i, n) an arbitrary mesh point in S_p, Λ_p and $\overline{\Lambda}_p$.

Let $k_n = t_n - t_{n-1}$ be the n -th time increment with $\sum_{l=1}^{n_j} k_l = \tau_j$ for some n_j , $j = 1, \dots, r$ and $\sum_{l=1}^{n_o} k_l = \tau$ with $n_o = \max\{n_1, \dots, n_r\}$, and let h_μ be the spatial increment in the x_{i_μ} -direction and $|h|^2 = h_1^2 + \dots + h_p^2$. We also choose the same time increments k_1, k_2, \dots, k_{n_o} in the time intervals $[l\tau, (l+1)\tau]$, $l \geq 1$. Define

$$\begin{aligned} \left(u_j\right)_{i,n} &\equiv u_j(x_i, t_n), \\ \left(f_j\right)_{i,n}\left(\left(u_j\right)_{i,n}, [\mathbf{u}]_{a_j}, [\mathbf{u}]_{b_j}, [\mathbf{u}_\tau]_{c_j}, [\mathbf{u}_\tau]_{d_j}\right) &\equiv \\ f_j\left(x_i, t_n, u(x_i, t_n)_j, [\mathbf{u}(x_i, t_n)]_{a_j}, [\mathbf{u}(x_i, t_n)]_{b_j}, [\mathbf{u}_\tau(x_i, t_n)]_{c_j}, [\mathbf{u}_\tau(x_i, t_n)]_{d_j}\right). \end{aligned}$$

Using the standard second order and first order finite-difference approximation, for each j ,

$$\Delta^{(\mu)}\left(u_j\right)_{i,n} \equiv h_\mu^{-2}\left[u_j\left(x_i + h_\mu e_\mu, t_n\right) - 2u_j(x_i, t_n) + u_j\left(x_i - h_\mu e_\mu, t_n\right)\right] \quad (2.3)$$

where $e_\mu = (0, \dots, 1, \dots, 0)$ is the unit vector in R^p , and

$$B\left[\left(u_j\right)_{i,n}\right] = \alpha(x_i, t_n)|x_i - \hat{x}_i|^{-1}\left[\left(u_j\right)(x_i, t_n) - \left(u_j\right)(\hat{x}_i, t_n)\right] + \beta(x_i, t_n)\left(u_j\right)(x_i, t_n),$$

where \hat{x}_i is a suitable point in Ω and $|x_i - \hat{x}_i|$ is the distance between x_i and \hat{x}_i .

Together with the implicit scheme for the time discretization, the finite-difference version of (1.1) is given by

$$\begin{aligned} L\left[\left(u_j\right)_{i,n}\right] &= k_n^{-1}\left(\left(u_j\right)_{i,n} - \left(u_j\right)_{i,n-1}\right) - \sum_{\mu=1}^p D_j \Delta^{(\mu)}\left(u_j\right)_{i,n} \\ &= \left(f_j\right)_{i,n}\left(\left(u_j\right)_{i,n}, [\mathbf{u}]_{a_j}, [\mathbf{u}]_{b_j}, [\mathbf{u}_\tau]_{c_j}, [\mathbf{u}_\tau]_{d_j}\right), \quad (i, n) \in \Lambda_p, \\ B\left[\left(u_j\right)_{i,n}\right] &= h_j(x_i, t_n), \quad (i, n) \in S_p, \\ \left(u_j\right)_{i,n} &= \eta_j(x_i, t_n), \quad (i, n) \in D_p. \end{aligned} \quad (2.4)$$

Using the same idea as for the continuous problem (see [6], [8], [12]), we have the following definition of upper and lower solutions of (2.4).

Definition 2.2. A pair of smooth functions $\tilde{\mathbf{u}}_{i,n} \equiv \left(\left(\tilde{u}_1\right)_{i,n}, \dots, \left(\tilde{u}_r\right)_{i,n}\right)$, $\hat{\mathbf{u}}_{i,n} \equiv \left(\left(\hat{u}_1\right)_{i,n}, \dots, \left(\hat{u}_r\right)_{i,n}\right)$ are called ordered upper and lower solutions of (2.4) if $\tilde{\mathbf{u}}_{i,n} \geq \hat{\mathbf{u}}_{i,n}$ and the following inequalities are satisfied:

$$\begin{aligned} L\left[\left(\tilde{u}_j\right)_{i,n}\right] &\geq \left(f_j\right)_{i,n}\left(\left(\tilde{u}_j\right)_{i,n}, [\tilde{\mathbf{u}}]_{a_j}, [\tilde{\mathbf{u}}]_{b_j}, [\tilde{\mathbf{u}}_\tau]_{c_j}, [\tilde{\mathbf{u}}_\tau]_{d_j}\right), \quad (i, n) \in \Lambda_p, \\ B\left[\left(\tilde{u}_j\right)_{i,n}\right] &\geq h_j(x_i, t_n), \quad (i, n) \in S_p, \\ \left(\tilde{u}_j\right)_{i,n} &\geq \eta_j(x_i, t_n), \quad (i, n) \in D_p \end{aligned} \quad , \quad (2.5)$$

and

$$\begin{aligned} L\left[\left(\hat{u}_j\right)_{i,n}\right] &\leq \left(f_j\right)_{i,n}\left(\left(\hat{u}_j\right)_{i,n}, [\hat{u}]_{a_j}, [\hat{u}]_{b_j}, [\hat{u}_\tau]_{c_j}, [\tilde{u}_\tau]_{d_j}\right), \quad (i, n) \in \Lambda_p, \\ B\left[\left(\hat{u}_j\right)_{i,n}\right] &\leq h_j(x_i, t_n), \quad (i, n) \in S_p, \\ \left(\hat{u}_j\right)_{i,n} &\leq \eta_j(x_i, t_n), \quad (i, n) \in D_p. \end{aligned} \quad (2.6)$$

Here the inequality $\tilde{u}_{i,n} \geq \hat{u}_{i,n}$ is in the usual componentwise sense, that is, for each $j = 1, \dots, r$, $(\tilde{u}_j)_{i,n} \geq (\hat{u}_j)_{i,n}$.

3. MONOTONE ITERATIVE SCHEME FOR THE FINITE DIFFERENCE SYSTEM

Assume that a pair of ordered upper and lower solutions $\tilde{\mathbf{u}}_{i,n}$ and $\hat{\mathbf{u}}_{i,n}$ exist. By the mixed quasimonotone property of $f(\mathbf{u}, \mathbf{u}_\tau) \equiv (f_1(\mathbf{u}, \mathbf{u}_\tau), f_2(\mathbf{u}, \mathbf{u}_\tau), \dots, f_r(\mathbf{u}, \mathbf{u}_\tau))$,

$$\begin{aligned} &\left(f_j\right)_{i,n}\left(\left(u_j\right)_{i,n}, [\mathbf{u}]_{a_j}, [\mathbf{w}]_{b_j}, [\mathbf{u}_\tau]_{c_j}, [\mathbf{w}_\tau]_{d_j}\right) \\ &\geq \left(f_j\right)_{i,n}\left(\left(w_j\right)_{i,n}, [\mathbf{w}]_{a_j}, [\mathbf{u}]_{b_j}, [\mathbf{w}_\tau]_{c_j}, [\mathbf{u}_\tau]_{d_j}\right) \end{aligned} \quad (3.1)$$

whenever $\tilde{\mathbf{u}}_{i,n} \geq \mathbf{u}_{i,n} \geq \mathbf{w}_{i,n} \geq \hat{\mathbf{u}}_{i,n}$ in the usual componentwise sense. Moreover, by the locally Lipschitz continuous property of f there exist positive constants K_j , $j = 1, \dots, r$, such that

$$\begin{aligned} &\left\| \left(f_j\right)_{i,n}\left(\left(u_j\right)_{i,n}, [\mathbf{u}]_{a_j}, [\mathbf{u}]_{b_j}, [\mathbf{u}_\tau]_{c_j}, [\mathbf{u}_\tau]_{d_j}\right) - \left(f_j\right)_{i,n}\left(\left(v_j\right)_{i,n}, [\mathbf{v}]_{a_j}, [\mathbf{v}]_{b_j}, [\mathbf{v}_\tau]_{c_j}, [\mathbf{v}_\tau]_{d_j}\right) \right\| \\ &\leq K_j \left(\|\mathbf{u}_{i,n} - \mathbf{v}_{i,n}\| + \|(\mathbf{u}_\tau)_{i,n} - (\mathbf{v}_\tau)_{i,n}\| \right) \end{aligned}$$

whenever $\tilde{\mathbf{u}}_{i,n} \geq \mathbf{u}_{i,n}$, $\mathbf{v}_{i,n} \geq \hat{\mathbf{u}}_{i,n}$ in the usual componentwise sense, with $\|\cdot\|$ denoting the maximum norm.

Now we will construct two sequences $\left\{(\tilde{u}_j)_{i,n}^{(m)}\right\}$ and $\left\{(\hat{u}_j)_{i,n}^{(m)}\right\}$, $j = 1, \dots, r$, (called upper and lower sequences, respectively), for $m = 1, 2, \dots$, by the following monotone iterative algorithm:

Step 1: Choose the initial iterations

$$(\bar{u}_j)_{i,n}^{(0)} = (\tilde{u}_j)_{i,n} \quad (\text{Upper solution})$$

$$(\underline{u}_j)_{i,n}^{(0)} = (\hat{u}_j)_{i,n} \quad (\text{Lower solution})$$

Step 2: For $m = 1, 2, \dots$, solve the following linear systems for $j = 1, \dots, r$, to construct $\{\bar{u}_j\}_{j,n}^{(m)}$ and $\{\underline{u}_j\}_{j,n}^{(m)}$:

$$\begin{aligned} & L \left[\left(\bar{u}_j \right)_{i,n}^{(m)} \right] + K_j \left(\bar{u}_j \right)_{i,n}^{(m)} \\ &= K_j \left(\bar{u}_j \right)_{i,n}^{(m-1)} + \left(f_j \right)_{i,n} \left(\left(\bar{u}_j \right)_{i,n}^{(m-1)}, [\bar{u}]_{a_j}^{(m-1)}, [\underline{u}]_{b_j}^{(m-1)}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right), \quad (i, n) \in \Lambda_p, \\ & B \left[\left(\bar{u}_j \right)_{i,n}^{(m)} \right] = h_j(x_i, t_n), \quad (i, n) \in S_p, \\ & \left(\bar{u}_j \right)_{i,0}^{(m)} = \eta_j(x_i, 0), \quad (i, 0) \in D_p \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} & L \left[\left(\underline{u}_j \right)_{i,n}^{(m)} \right] + K_j \left(\underline{u}_j \right)_{i,n}^{(m)} \\ &= K_j \left(\underline{u}_j \right)_{i,n}^{(m-1)} + \left(f_j \right)_{i,n} \left(\left(\underline{u}_j \right)_{i,n}^{(m-1)}, [\underline{u}]_{a_j}^{(m-1)}, [\bar{u}]_{b_j}^{(m-1)}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right), \quad (i, n) \in \Lambda_p, \\ & B \left[\left(\underline{u}_j \right)_{i,n}^{(m)} \right] = h_j(x_i, t_n), \quad (i, n) \in S_p, \\ & \left(\underline{u}_j \right)_{i,0}^{(m)} = \eta_j(x_i, 0), \quad (i, 0) \in D_p \end{aligned} \tag{3.3}$$

Step 3: Choose the tolerance δ and terminate iteration if

$$\left\| \left(u_j \right)_{*,n}^{(m)} - \left(u_j \right)_{*,n}^{(m-1)} \right\| \leq \delta, \quad \text{or}$$

$$\left\| \left(\bar{u}_j \right)_{*,n}^{(m)} - \left(\bar{u}_j \right)_{*,n}^{(m-1)} \right\| \leq \delta,$$

where $\left(u_j \right)_{*,n}^{(m)} = \left(\left(u_j \right)_{1,n}^{(m)}, \dots, \left(u_j \right)_{N,n}^{(m)} \right)$, $\left(\bar{u}_j \right)_{*,n}^{(m)} = \left(\left(\bar{u}_j \right)_{1,n}^{(m)}, \dots, \left(\bar{u}_j \right)_{N,n}^{(m)} \right)$ and

$\left(\underline{u}_j \right)_{*,n}^{(m)} = \left(\left(\underline{u}_j \right)_{1,n}^{(m)}, \dots, \left(\underline{u}_j \right)_{N,n}^{(m)} \right)$, N is determined by the number of mesh points and the way that the boundary conditions are discretized.

The following discrete version of the positivity lemma [10] is used for proving our theorem on the monotonicity of the upper and lower sequences.

Lemma 3.1 Let $k = \max \{k_n : n = 1, 2, \dots, N\}$ and $C_{i,n} \geq -k^{-1}$. If a function $w_{i,n}$ satisfies the inequalities:

$$\begin{aligned} L \left[w_{i,n} \right] + C_{i,n} w_{i,n} &\geq 0, \quad (i, n) \in \Lambda_p, \\ B \left[w_{i,n} \right] &\geq 0, \quad (i, n) \in S_p, \\ w_{i,0} &\geq 0, \quad (i, 0) \in D_p, \end{aligned} \tag{3.4}$$

then,

$$w_{i,n} \geq 0 \text{ on } \bar{\Lambda}_p.$$

Now we state and prove the following existence-comparison and uniqueness theorem for sequences $\{(\bar{u}_j)_{i,n}^{(m)}\}$ and $\{(\underline{u}_j)_{i,n}^{(m)}\}$, $j = 1, \dots, r$, constructed by the algorithm (3.2) and (3.3).

Theorem 3.1 Let $(\tilde{u}_j)_{i,n}$ and $(\hat{u}_j)_{i,n}$, $j = 1, \dots, r$, be a pair of ordered upper and lower solutions for (2.4) on $\bar{\Lambda}_p$. If $K_j \leq k^{-1}$ for $j = 1, \dots, r$, then both the upper sequence $\{(\bar{u}_j)_{i,n}^{(m)}\}$ and the lower sequence $\{(\underline{u}_j)_{i,n}^{(m)}\}$ converge to $(u_j)_{i,n}$ which is the unique solution of (2.4) between $(\tilde{u}_j)_{i,n}$ and $(\hat{u}_j)_{i,n}$, $j = 1, \dots, r$. Moreover, the following monotone property holds for each positive integer m ,

$$(\hat{u}_j)_{i,n} \leq (\underline{u}_j)_{i,n}^{(m)} \leq (\underline{u}_j)_{i,n}^{(m+1)} \leq (u_j)_{i,n} \leq (\bar{u}_j)_{i,n}^{(m+1)} \leq (\bar{u}_j)_{i,n}^{(m)} \leq (\tilde{u}_j)_{i,n}. \tag{3.5}$$

Proof: Let $(w_j)_{i,n}^{(0)} = (\bar{u}_j)_{i,n}^{(0)} - (\underline{u}_j)_{i,n}^{(0)} = (\tilde{u}_j)_{i,n} - (\bar{u}_j)_{i,n}^{(1)}$ for a fixed j . Then by the definition of the upper solution,

$$\begin{aligned} &L \left[(w_j)_{i,n}^{(0)} \right] + K_j (w_j)_{i,n}^{(0)} \\ &= L \left[(\tilde{u}_j)_{i,n} \right] - (f_j)_{i,n} \left((\tilde{u}_j)_{i,n}, [\tilde{u}]_{a_j}, [\hat{u}]_{b_j}, [\tilde{u}_\tau]_{c_j}, [\hat{u}_\tau]_{d_j} \right) \\ &\geq L \left[(\tilde{u}_j)_{i,n} \right] - (f_j)_{i,n} \left((\tilde{u}_j)_{i,n}, [\tilde{u}]_{a_j}, [\hat{u}]_{b_j}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right) \geq 0, \\ &B \left[(w_j)_{i,n}^{(0)} \right] = B \left[(\tilde{u}_j)_{i,n} \right] - h_j(x_i, t_n) \geq 0, \quad (i, n) \in S_p, \\ &(w_j)_{i,0}^{(0)} = (\tilde{u}_j)_{i,0} - \eta_j(x_i, 0) \geq 0, \quad (i, 0) \in D_p. \end{aligned}$$

Since $K_j \geq 0 \geq -k^{-1}$, by Lemma 3.1, and the definition of the upper solution, we have $(w_j)_{i,n}^{(0)} \geq 0$ which means that $(\bar{u}_j)_{i,n}^{(1)} \leq (\tilde{u}_j)_{i,n}$ for $j = 1, \dots, r$, in $\bar{\Lambda}_p$. A similar argument show that $(\underline{u}_j)_{i,n}^{(1)} \geq (\hat{u}_j)_{i,n}$ in $\bar{\Lambda}_p$.

Again let $(w_j)_{i,n}^{(1)} = (\bar{u}_j)_{i,n}^{(1)} - (\underline{u}_j)_{i,n}^{(1)}$ for a fixed j . Then by (3.1) and the Lipschitz continuous property of function f ,

$$\begin{aligned}
& L \left[\left(w_j \right)_{i,n}^{(1)} \right] + K_j \left(w_j \right)_{i,n}^{(1)} = K_j \left(\left(\tilde{u}_j \right)_{i,n} - \left(\hat{u}_j \right)_{i,n} \right) \\
& + \left(f_j \right)_{i,n} \left(\left(\tilde{u}_j \right)_{i,n}, [\tilde{\mathbf{u}}]_{a_j}, [\hat{\mathbf{u}}]_{b_j}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right) \\
& - \left(f_j \right)_{i,n} \left(\left(\hat{u}_j \right)_{i,n}, [\hat{\mathbf{u}}]_{a_j}, [\tilde{\mathbf{u}}]_{b_j}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right) \geq 0, \quad (i, n) \in \Lambda_p, \\
& B \left[\left(w_j \right)_{i,n}^{(1)} \right] = h_j(x_i, t_n) - h_j(x_i, 0) = 0, \quad (i, n) \in S_p, \\
& \left(w_j \right)_{i,0}^{(1)} = \eta_j(x_i, 0) - \eta_j(x_i, 0) = 0, \quad (i, 0) \in D_p.
\end{aligned}$$

By Lemma 3.1 we also have $\left(w_j \right)_{i,n}^{(1)} \geq 0$ which means that $\left(\bar{u}_j \right)_{i,n}^{(1)} \geq \left(u_j \right)_{i,n}^{(1)}$ in $\bar{\Lambda}_p$ for $j = 1, \dots, r$. This conclusion leads to the relation

$$\left(\hat{u}_j \right)_{i,n} = \left(u_j \right)_{i,n}^{(0)} \leq \left(u_j \right)_{i,n}^{(1)} \leq \left(\bar{u}_j \right)_{i,n}^{(1)} \leq \left(\bar{u}_j \right)_{i,n}^{(0)} = \left(\tilde{u}_j \right)_{i,n} \in \bar{\Lambda}_p. \quad (3.6)$$

Assume for each $j = 1, \dots, r$, by induction that

$$\left(u_j \right)_{i,n}^{(m-1)} \leq \left(u_j \right)_{i,n}^{(m)} \leq \left(\bar{u}_j \right)_{i,n}^{(m)} \leq \left(\bar{u}_j \right)_{i,n}^{(m-1)} \quad \text{in } \bar{\Lambda}_p \quad (3.7)$$

and let $\left(w_j \right)_{i,n}^{(m)} = \left(\bar{u}_j \right)_{i,n}^{(m)} - \left(\bar{u}_j \right)_{i,n}^{(m+1)}$. Then by (3.1) and the Lipschitz continuous property of function f ,

$$\begin{aligned}
& L \left[\left(w_j \right)_{i,n}^{(m)} \right] + K_j \left(w_j \right)_{i,n}^{(m)} = K_j \left(\left(\bar{u}_j \right)_{i,n}^{(m-1)} - \left(\bar{u}_j \right)_{i,n}^{(m)} \right) \\
& + \left(f_j \right)_{i,n} \left(\left(\bar{u}_j \right)_{i,n}^{(m-1)}, [\bar{\mathbf{u}}]_{a_j}^{(m-1)}, [\bar{\mathbf{u}}]_{b_j}^{(m-1)}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right) \\
& - \left(f_j \right)_{i,n} \left(\left(\bar{u}_j \right)_{i,n}^{(m-1)}, [\bar{\mathbf{u}}]_{a_j}^{(m)}, [\bar{\mathbf{u}}]_{b_j}^{(m)}, (u_1)_{i,n-n_1}, \dots, (u_r)_{i,n-n_r} \right) \geq 0, \quad (i, n) \in \Lambda_p, \\
& B \left[\left(w_j \right)_{i,n}^{(m)} \right] = h_j(x_i, t_n) - h_j(x_i, 0) = 0, \quad (i, n) \in S_p, \\
& \left(w_j \right)_{i,0}^{(m)} = \eta_j(x_i, 0) - \eta_j(x_i, 0) = 0, \quad (i, 0) \in D_p.
\end{aligned}$$

It follows again from Lemma 3.1 that $\left(w_j \right)_{i,n}^{(m)} \geq 0$ which gives $\left(\bar{u}_j \right)_{i,n}^{(m)} \geq \left(\bar{u}_j \right)_{i,n}^{(m+1)}$ in $\bar{\Lambda}_p$ for $j = 1, \dots, r$. Similarly, we can show that $\left(u_j \right)_{i,n}^{(m+1)} \geq \left(u_j \right)_{i,n}^{(m)}$ and $\left(\bar{u}_j \right)_{i,n}^{(m+1)} \geq \left(\bar{u}_j \right)_{i,n}^{(m+1)}$ in $\bar{\Lambda}_p$ for $j = 1, \dots, r$. An application of the induction argument yields

$$\left(\bar{u}_j \right)_{i,n}^{(m)} \leq \left(u_j \right)_{i,n}^{(m+1)} \leq \left(\bar{u}_j \right)_{i,n}^{(m+1)} \leq \left(\bar{u}_j \right)_{i,n}^{(m)} \quad \text{in } \bar{\Lambda}_p. \quad (3.8)$$

This proves the monotone property of the upper and lower sequences. In view of the monotone property (3.8), the limits

$$\left(\bar{u}_j \right)_{i,n} = \lim_{m \rightarrow \infty} \left(\bar{u}_j \right)_{i,n}^{(m)} \quad \text{and} \quad \left(u_j \right)_{i,n} = \lim_{m \rightarrow \infty} \left(u_j \right)_{i,n}^{(m)}, \quad j = 1, \dots, r$$

exist and satisfy the relation

$$(\hat{u}_j)_{i,n} \leq (\underline{u}_j)_{i,n}^{(m)} \leq (\underline{u}_j)_{i,n}^{(m+1)} \leq (\underline{u}_j)_{i,n} \leq (\bar{u}_j)_{i,n} \leq (\bar{u}_j)_{i,n}^{(m+1)} \leq (\bar{u}_j)_{i,n}^{(m)} \leq (\tilde{u}_j)_{i,n}.$$

To show that $(\bar{u}_j)_{i,n}$ and $(\underline{u}_j)_{i,n}$ are solutions of (2.4), it suffices to show that $(\bar{u}_j)_{i,n} = (\underline{u}_j)_{i,n}$. Since in (3.2) and (3.3) all terms with time delays are known by initial conditions and previous calculation the system (3.2) and (3.3) may be considered as a system of coupled system without time delays. By Theorem 3.1 on page 400 in [12], $(\bar{u}_j)_{i,n} = (\underline{u}_j)_{i,n}$. Now letting $m \rightarrow \infty$ in (3.2) and (3.3) shows that $(\bar{u}_j)_{i,n}$ and $(\underline{u}_j)_{i,n}$ are solutions of (3.2) and (3.3). Therefore, $(\bar{u}_j)_{i,n} = (\underline{u}_j)_{i,n}$, $j = 1, \dots, r$ is the unique solution of (2.4).

The convergence of the numerical solution to the continuous solution and the stability of numerical algorithm are given in the following theorem. This theorem can be proven through similar arguments as in [10, 12] for single-equation and coupled systems. We hereby omit the proof.

Theorem 3.2 If all conditions in Theorem 3.1 are satisfied, then for each $j = 1, \dots, r$, the discrete solution $(u_j)_{i,n}$ of (2.4) converges to the continuous solution $u_j(x_i, t_n)$ of (1.1) as $(k_n + |h|^2) \rightarrow 0$. Moreover the monotone iterative scheme defined in (3.2) and (3.3) is numerically stable.

Let $\tau_{\min} = \min \{\tau_1, \dots, \tau_r\}$. For any time $t = t_0$, we define a pair of ordered local upper and lower solutions $\tilde{u}_{i,n}$ and $\hat{u}_{i,n}$ of (2.4), over the time interval $[t_0, t_0 + \tau_{\min}]$, by replacing $(f_j)_{i,n}((\tilde{u}_j)_{i,n}, [\tilde{u}]_{a_j}, [\hat{u}]_{b_j}, [\tilde{u}_\tau]_{c_j}, [\hat{u}_\tau]_{d_j})$ with $(f_j)_{i,n}((\tilde{u}_j)_{i,n}, [\tilde{u}]_{a_j}, [\hat{u}]_{b_j}, \mathbf{u}_\tau)$ in (2.5) and $(f_j)_{i,n}((\hat{u}_j)_{i,n}, [\hat{u}]_{a_j}, [\tilde{u}]_{b_j}, [\hat{u}_\tau]_{c_j}, [\tilde{u}_\tau]_{d_j})$ with $(f_j)_{i,n}((\hat{u}_j)_{i,n}, [\hat{u}]_{a_j}, [\tilde{u}]_{b_j}, \mathbf{u}_\tau)$ in (2.6), where \mathbf{u}_τ is given by initial conditions or previous computation. By induction argument over time intervals $[0, \tau_{\min}], \dots, [(n-1) \times \tau_{\min}, n \times \tau_{\min}]$, we have:

Theorem 3.3 If all conditions in Theorem 3.1 and 3.2 are satisfied except that $f(\mathbf{u}, \mathbf{u}_\tau)$ is only quasimonotone with respect to \mathbf{u} and the local upper and lower solutions are used as initial iterations, then all conclusions in Theorem 3.1 and 3.2 hold.

Remark.

1. The upper solution $\tilde{u}_{i,n}$ and the lower solution $\hat{u}_{i,n}$ are defined by (2.5) and (2.6) globally (over the entire domain). However, in practice the upper solution $\tilde{u}_{i,n}$ and the lower solution $\hat{u}_{i,n}$ may be found locally at each time step. Please refer to the book [8] by Pao for methods about finding upper solutions and lower solutions.
2. The monotone coefficients K_j , used in the iterative algorithm (3.2) and (3.3) may also be determined by the following:

$$K_j(x,t) \geq \max \left\{ -\frac{\partial f_j}{\partial u_j}(x,t,u,v; \hat{u} \leq u, v \leq \hat{u}) \right\}, \quad j = 1, \dots, r.$$

REFERENCES

- [1] Cushing, J.M., *Integrodifferential Equations and Delay Models in Population Dynamics*, Lecture Notes in Biomathematics, 20, Springer-Verlag, Berlin-Heidleberg-New York, 1977.
- [2] Erbe, L. and Liu, X., Monotone iterative methods for differential systems with finite delays, *Appl. Math. Compu.*, 43, 1991, 43-64.
- [3] Friesenecker, G., Exponentially growing solutions for a delay-diffusion equation with negative feedback, *J. Diff. Eq.*, 98, 1992, 1-18.
- [4] Kuang, Y., *Delay Differential Equations with Applications in Population Dynamics*, Mathematics in Science and Engineering, Volume 191, Academic Press, Inc., 1993.
- [5] Ladde, G.S., Lakshmikantham, V., and Vatsala, A., *Monotone Iterative Techniques for Nonlinear Differential Equations*, Pitman, Boston, 1985.
- [6] Lu, X., Monotone method and convergence acceleration for finite-difference solutions of parabolic problems with time delays, *Numerical Methods for Partial Differential Equations*, 11, 1995, 591-602.
- [7] Martin, R.H. and Smith, H.L., Convergence in Lotka-Volterra systems with diffusion and delay, *Lecture Notes in Pure and Applied Math.*, 133, 1991, 259-266.
- [8] Pao, C.V., *Nonlinear Parabolic and Elliptic Equations*, Plenum Press, New York, 1992.
- [9] Pao, C.V., Numerical methods for semilinear parabolic equations, *SIAM J. Numer. Anal.*, 24, 1987, 24-35.
- [10] Pao, C.V., Numerical method for coupled systems of nonlinear boundary value problem, *J. Math. Anal. Appl.*, 151, 1990, 581-607.

- [11] Pao, C.V., Coupled nonlinear parabolic systems with time-delays, *J. Math. Anal. Appl.*, 196, No. 1, 1995, 237-265.
- [12] Pao, C.V., Numerical analysis of coupled system of nonlinear parabolic equations, *SIAM J. Numer. Anal.*, 36, 1999, 393-416.
- [13] Smith, G.D., *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Clarendon Press, Oxford, 1984.

16 GLOBAL ANALYSIS FOR THE FLUIDS OF A POWER-LAW TYPE

Josef Málek

Mathematical Institute of Charles University
Sokolovská 83, 186 75 Prague 8, Czech Republic

ABSTRACT

We present a survey of the results regarding the analysis of a system of partial differential equations describing the unsteady motions of incompressible fluids with shear dependent viscosity. The analysis is focused on the questions of existence, uniqueness, regularity, and large time behavior of solutions with no restriction on the size of data.

1. THE FLUIDS OF A POWER-LAW TYPE

This article surveys results concerning *global* properties of solutions to nonlinear systems of partial differential equations. By ‘properties of solutions’, we mean its existence, uniqueness, and continuous dependence on data (these are initial conditions and the right-hand side in our case), higher smoothness (regularity), large-time behavior, and the stability of steady solutions.

The notion *global* means that we want the mentioned properties without restricting to the small size of data or to a short length of the time interval.

We consider a system of equations describing unsteady flows of an incompressible fluid in an open subset $\Omega \subset \mathbb{R}^d$, $d > 1$ written in the form

$$\operatorname{div} v = 0,$$

$$\rho \frac{\partial v}{\partial t} + \rho \operatorname{div}(v \otimes v) - \operatorname{div} T^E = -\nabla \pi + \rho f, \quad (1.1)$$

where $v = (v_1, v_2, \dots, v_d)$ denotes the velocity field, ρ is a given positive constant determining the density of the material and $f = (f_1, f_2, \dots, f_d)$ captures the given

density of the external body forces; the scalar quantity π , sometimes called the pressure and the extra stress $\mathbf{T}^E = (T_{ij}^E)_{i,j=1}^d$ represent (up to a multiplicative constant) the trace and the deviatoric (traceless) part of the (total) stress tensor, respectively. The constitutive equation for \mathbf{T}^E (see equation (1.2) below) specifies the type of materials we are interested in dealing with.

Let $T > 0$ be given. If not specified otherwise, the functions $\mathbf{v}, \pi, \mathbf{f}$, and \mathbf{T}^E are supposed to depend on (t, x) , where $t \in [0, T]$ and $x \in \Omega$.

Let further \mathbf{D} denote the symmetric part of the gradient of the velocity, i.e.,

$$\mathbf{D} := \mathbf{D}(\mathbf{v}) \equiv \frac{1}{2} \left[(\nabla \mathbf{v}) + (\nabla \mathbf{v})^T \right].$$

In this paper, we deal with fluids of a power-law type that fall into the class of materials given by the constitutive equation

$$\mathbf{T}^E = \mathbf{T}(\mathbf{D}), \quad \text{tr } \mathbf{T} = 0, \quad (1.2)$$

where $\mathbf{T} : \mathbb{R}_{\text{sym}}^{d \times d} \rightarrow \mathbb{R}_{\text{sym}}^{d \times d}$; here $\mathbb{R}_{\text{sym}}^{d \times d}$ consists of all symmetric $(d \times d)$ -matrices.

Relation (1.2) makes the system (1.1) closed: we obtain $(d+1)$ equations for $(d+1)$ unknowns $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d)$ and π dealing generally with two nonlinearities. The first one is due to the convective term $\text{div}(\mathbf{v} \otimes \mathbf{v})$, the second one comes out from the equation for \mathbf{T}^E . The number of nonlinearities reduces when we consider the simplest example of (1.2) the so-called Stokes law. Then \mathbf{T} is a linear function of \mathbf{D} , i.e.,

$$\mathbf{T}^E = 2\mu\mathbf{D}, \quad \mu > 0, \quad (1.3)$$

and (1.3) together with (1.1) gives the Navier-Stokes equations (*NSEs* for short)
 $\text{div } \mathbf{v} = 0$,

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho \text{div}(\mathbf{v} \otimes \mathbf{v}) - \mu \Delta \mathbf{v} = -\nabla \pi + \rho \mathbf{f}. \quad (1.4)$$

From the mathematical point of view *NSEs* in *two dimensions* ($2d$ *NSEs* for short) represent the example of a nonlinear system where the solution enjoys all the above discussed properties: for an arbitrarily long time interval and for arbitrarily large (convenient) norms of initial values and \mathbf{f} 's there exist (weak) solutions which are unique and smooth as data permits. Moreover, all solutions (considered as trajectories in the phase space) converge uniformly (as time goes to infinity) to a small (i.e., compact) set with finite fractal dimension. In addition, the rate of convergence can be made exponential. The reader can find these results in [61], [24], [5], [35], for example.

On the other hand, the *three-dimensional* Navier-Stokes equations ($3d$ *NSEs* for short) are the type of the system with clear physical meaning for which the fundamental mathematical question of the global existence of a uniquely defined

solution is up-to-now unresolved. Let us recall that the only type of solution for which the existence is known (for the Cauchy problem almost 70 years, see Leray's paper [32]) is weak solution (see Definition 4.1). It is not at all clear whether this solution is unique or it can bifurcate in time. The question of uniqueness is closely related to the question of regularity: there are known minimal (Prodi-Serrin's) conditions on smoothness of weak solutions that would imply uniqueness of such 'smoother' solutions in the class of all weak solutions. More details are given in Section 4.

As 3d NSEs provide from our global point of view an unsatisfactory answer, it is natural to ask if it is possible to modify the Stokes law (1.3) in such a way that solutions of (1.1) with this modified constitutive law will have the desired (prospective) global properties.

In order to answer this question, O.A. Ladyzhenskaya, in the mid sixties, put before the mathematical audience the model ([22], [23])

$$\mathbf{T}^E = 2\mu \left(1 + \varepsilon |\mathbf{D}|^{p-2}\right) \mathbf{D}, \quad \varepsilon > 0, \quad (1.5)$$

where p is a real parameter; generally $p \in (1, \infty)$. As the Navier-Stokes equations are a sub-case when $p = 2$, we expect better global properties when $p > 2$. Other typical variants of the model (1.5) are

$$\mathbf{T}^E = 2\mu |\mathbf{D}|^{p-2} \mathbf{D},$$

$$\mathbf{T}^E = 2\mu \left(1 + \varepsilon |\mathbf{D}|^2\right)^{\frac{p-2}{2}} \mathbf{D}.$$

Notice that all of them can be expressed in the form

$$\mathbf{T}^E = 2\mu v(|\mathbf{D}|^2) \mathbf{D}, \quad (1.6)$$

where $v : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ is the so-called generalized viscosity, $\mu > 0$. Because system (1.1) with (1.6) remain of the second order, we can understand it as the simplest generalization¹ of the Stokes law. This is why the system (1.1) with (1.6) is oftentimes called the generalized Navier-Stokes equations (shortly *GNSEs*).

¹ Consequently, many of the existing numerical codes can be easily extended to these models. Let us also remark that other generalizations exhibiting nice mathematical properties are multi-polar fluids (see [50] for a detailed exposition and further references (mainly for compressible model); and [37] where the global existence of regular unique solutions and their large time behavior for these models are successfully studied in the context of incompressible fluids). Then, however, we obtain the equations of higher order requiring different kind of numerical approximations. Another disadvantage of bipolar (or multi-polar) fluids is the lack of any experimental work supporting the existence of such fluids. On the other hand, there are materials that exhibit experimentally measured responses that coincide with analytical solutions of *GNSEs* for special kinds of flows, and particular forms of the

Observe also that if \mathbf{T}^E is of the form (1.6) then there is always a scalar potential $\Phi : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ to \mathbf{T}^E given by

$$\Phi(|\mathbf{D}|^2) \equiv \mu \int_0^{|\mathbf{D}|^2} v(s) ds.$$

Thus, for $r, s = 1, 2, \dots, d$ we have

$$\begin{aligned} T_{rs}(\cdot) &= \partial_{rs} \Phi(\cdot) \equiv \frac{\partial \Phi(\cdot)}{\partial D_{rs}}, \\ \Phi(0) &= \partial_{rs} \Phi(0) = 0. \end{aligned} \quad (1.7)$$

It might be worth mentioning that models (1.6) are frequently used at various scientific areas and engineering applications as chemistry, geology, glaciology, biology. From this perspective (1.6) represents a significant class of models in continuum mechanics. An interested reader can find a representable list of applications of models of the type (1.6) in [44].

Using the terminology of the non-Newtonian fluid mechanics, the models (1.6) have the ability to capture the phenomena called shear-thickening (if $p > 2$) and shear-thinning (when $p < 2$). Interestingly, the most of chemical solutions, dyes, polymeric liquids, glacier, etc. are modeled by (1.6) when parameter p is close to 1 (typical values are $4/3, 6/5$, etc.). On the other hand, these models cannot exhibit other non-Newtonian effects as non-zero normal stress differences, stress relaxation, nonlinear creep, yield or pre-stress – see [54] for the more details and definitions of these notions.

We have seen that the parameter p splits our attention into three cases:

- (1) *NSEs* – understood as a special subclass when $p = 2$ and the elliptic operator is linear (see Section 4);
- (2) *GNSEs* with $p \geq 2$ – we expect better global properties of the system (1.1).

Our expectations will be confirmed by the results presented in Sections 2 and 3;

- (3) *GNSEs* with $p < 2$ – the theoretical results up to the nineties are more or less missing (both in two and three dimensions). As the energy inequality in this case gives worse starting information than in the case of *NSEs*, the results are valuable and nontrivial even in two dimensions and for steady flows (see Section 2).

viscosity function v in (1.6). It may happen that while the modified Stokes law is not experimentally observable for classical fluids (as water) in the range of shear rates the viscosity is measured it can occur if the shear rates would increase. In such a case the model (1.5) should be rather used instead of (1.3).

Note that the model (1.6) where v is non-constant but bounded function of $|D|^2$ falls into the category (2) above.

We complete the equations by initial and boundary conditions, and the assumption on potential Φ introduced in (1.7).

We suppose that for $p \geq 1$ there are positive constants C_1, C_2 so that for all matrices $\xi \in \mathbb{R}^{d \times d}$ and for all symmetric matrices $A \in \mathbb{R}^{d \times d}$ it holds

$$C_1 (1 + |A|)^{p-2} |\xi|^2 \leq \xi^T \cdot \partial^2 \Phi(|A|^2) \xi \leq C_2 (1 + |A|)^{p-2} |\xi|^2. \quad (1.8)$$

Let $v_0 : \Omega \rightarrow \mathbb{R}^d$ satisfy $\operatorname{div} v_0 = 0$. We say that v satisfies the initial condition if we have (in some sense)

$$v(0, x) = v_0(x), \quad x \in \Omega. \quad (1.9)$$

In this exposition, we restrict ourselves to three kind of boundary-value problems:

(i) *Dirichlet problem* – Ω is a bounded domain with (at least) Lipschitz boundary and

$$v(t, x) = \mathbf{0} \quad \text{for } (t, x) \in [0, T] \times \partial\Omega; \quad (1.10)$$

(ii) *Space periodic problem* – Ω is a d -dimensional cube with the side of the length $L > 0$, $\Omega \equiv (0, L)^d$ and

$$v, \pi \text{ are periodic with period } L \text{ at each variable } x_i, i = 1, 2, \dots, d; \quad (1.11)$$

(iii) *Cauchy problem* – $\Omega = \mathbb{R}^d$ and

$$v \text{ is vanishing as } |x| \rightarrow \infty. \quad (1.12)$$

We abbreviate the problem (1.1), (1.2), (1.8), (1.9) with boundary conditions (1.10), (1.11), and (1.12), respectively, by $(Dir)_p$, $(Per)_p$, and $(Cau)_p$, respectively. If we speak about steady flows, we call them analogously $(Dir-st)_p$ and $(Per-st)_p$, respectively.

Note that $(Per)_p$ and $(Cau)_p$ are mathematical simplifications of the problem when we eliminate the influence of the boundary; the advantage of $(Per)_p$ is that the domain of interest still remains bounded.

The reader who is not familiar with certain notation can find worth to look at the appendix where also definitions of function spaces used in the next chapters are provided.

2. EXISTENCE, UNIQUENESS, AND REGULARITY

2.1 Definition. We say that

$$v \in L^\infty(I; L_{\operatorname{DIV}}^2(\Omega)^d) \cap L^p(I; W_{\operatorname{DIV}}^{1,p}(\Omega)^d) \quad (2.1)$$

is a weak solution to $(Dir)_p$ if for all $\varphi \in \mathcal{D}(-\infty, T; \mathcal{D}_{\text{DIV}}(\Omega)^d)$

$$\begin{aligned} & \int_{Q_T} \left[-\nu \cdot \frac{\partial \varphi}{\partial t} + T_{ij}(\mathbf{D}(\nu)) D_{ij}(\varphi) - \nu_k \nu \cdot \frac{\partial \varphi}{\partial x_k} \right] dx dt \\ &= \int_{Q_T} f \cdot \varphi dx dt + \int_{\Omega} \nu_0 \cdot \varphi(0_1 \cdot) dx. \end{aligned} \quad (2.2)$$

The choice of functions spaces follows from the energy inequality which reads

$$\sup_{t \in I} \|\nu(t)\|_2^2 + \int_0^T \|\mathbf{D}(\nu(t))\|_p^p dt \leq \|\nu_0\|_2^2 + c \int_0^T \|f(t)\|_{(W_{\text{DIV}}^{1,p}(\Omega))^*}^{p'} dt. \quad (2.3)$$

Using the Hölder and the Korn inequalities, we obtain from (2.1) with $p' = p/(p-1)$

$$\text{div}(\nu \otimes \nu) \in L^{p'} \left(I; (W_{\text{DIV}}^{1,p}(\Omega))^* \right) \quad \text{if } p \geq \frac{3d+2}{d+2}; \quad (2.4)$$

$$\nu_k \frac{\partial \nu}{\partial x_k} \in L^1 \left(I; L^1(\Omega)^d \right) \quad \text{if } p \geq \frac{2(d+1)}{d+2}; \quad (2.5)$$

$$\nu \otimes \nu \in L^1 \left(I; L^1(\Omega)^{d \times d} \right) \quad \text{if } p \geq \frac{2d}{d+2}. \quad (2.6)$$

Monotone operator theory for $(Per)_p$, $(Dir)_p$ and $(Cau)_p$.

If (2.4) holds, then (2.2) can be rewritten as

$$\int_0^T \left\langle \frac{d\nu}{dt} + \mathbf{T}(\mathbf{D}(\nu)) - \text{div}(\nu \otimes \nu) - f, \varphi \right\rangle dt = 0 \quad (2.7)$$

valid for all $\varphi \in L^{p'} \left(I; W_{\text{DIV}}^{1,p}(\Omega)^d \right)^*$. In particular, due to (2.1), one can set $\varphi = \nu$ in (2.7).

Due to the assumption (1.8), the main operator is monotone (in fact it is even strictly monotone) and we can successfully combine the theory of monotone operators together with the compactness arguments needed to justify the passage to the limit in the convective term. It results in the following theorem (proved in [24] (see also [33])).

2.1 Theorem. Let $\nu_0 \in L^2_{\text{DIV}}(\Omega)^d$, $f \in L^{p'} \left(I; W_{\text{DIV}}^{1,p}(\Omega)^d \right)^*$ be given and let

$$p \geq \frac{3d+2}{d+2}. \quad (2.8)$$

Then there is at least one weak solution to $(Dir)_p$. Moreover, if

$$p \geq \frac{d+2}{2} \quad (2.9)$$

this solution is unique in the class of weak solutions and depends continuously on data.

If $d = 2$, then the bounds in (2.8) and (2.9) coincide. This means that for $p \geq 2$, we can ensure both the existence and uniqueness, and $2d$ NSEs are nicely included into this theory. If $d = 3$, then the monotone operator theory gives the existence of solution for $p \geq \frac{11}{5}$ and its uniqueness for $p \geq \frac{5}{2}$. We see that $3d$ NSEs are not “covered” by Theorem 2.1 although the existence of a weak solution is well-known. This has been a motivation to attempt to extend the results given by the above theorem.

For this purpose, three new approaches have been developed:

- (1) the regularity method,
- (2) the truncation method,
- (3) the Lipschitz “truncation” method.

Regularity method for $(Per)_p$ and $(Cau)_p$.

The *regularity method* was applied firstly in [3] and [38] for different variants of approximations v^n . Demanded compactness of the velocity gradient is obtained using the uniform estimates of the time integral of the fraction of a convenient L^q -norm of the second derivatives of v^n versus L^2 -norm of the first derivatives of v^n . More precisely, setting

$$\lambda = 2 \frac{3-p}{3p-5}$$

one can obtain the second a priori estimates²

$$\frac{\left(1 + \|\nabla v^n(t)\|_2^2\right)^{1-\lambda}}{1-\lambda} + \int_0^t \frac{\int_{\Omega} \left(1 + |\mathcal{D}(v^n(\tau))|^2\right)^{(p-2)/2} |\mathcal{D}(\nabla v^n(\tau))|^2 dx}{\left(1 + \|\nabla v^n(\tau)\|_2^2\right)^{\lambda}} d\tau \leq K.$$

In three dimensions, this method gives the existence of a weak solution for $p > 9/5$. Consequently, the gap in the existence theory between the result of Leray for $3d$ NSEs and Ladyzhenskaya for $3d$ GNSEs when $p \geq 11/5$ has been removed. Despite this fact, the main importance of the regularity method lies, in the author’s opinion, in new regularity and uniqueness results. In three dimensions, we thus obtain the existence of regular (strong) solution if $p \geq 11/5$; in addition, these solutions are unique in the class of weak solutions provided that initial values belong to $W_{\text{div}}^{1,2}(\Omega)^d$.

In two dimensions, this result can be strengthened due to the identity

² If $\lambda = 1$, the first term is replaced by $\log \left(1 + \|\nabla v^n\|_2^2\right)$.

$$\int_{\Omega} \frac{\partial v_k}{\partial x_s} \frac{\partial v_i}{\partial x_k} \frac{\partial v_i}{\partial x_s} dx = 0 \quad (2.10)$$

valid if $\operatorname{div} v = 0$. As shown in monograph [41], this gives the existence of regular (strong) solutions for all $p > 1$; if $p \in (1, 2)$ these solutions are unique in the class of strong solutions.

Regularity techniques, as the difference-quotient method, can be easily applied also to $(Cau)_p$, thanks to missing boundary. However, one has to overcome difficulties due to unboundedness of the domain. Yet, the results for $(Per)_p$ hold also for $(Cau)_p$ with the same range of parameters as shown in [53].

We can summarize the results discovered by the regularity method in the following theorem.

2.2 Theorem. Let $v_0 \in W_{\text{DIV}}^{1,2}(\Omega)^d \cap W_{\text{DIV}}^{1,p}(\Omega)^d$ and $f \in L^2(Q_T)^d$ for $p \geq 2$, and $f \in L^{p'}(Q_T)^d$ if $p \leq 2$. Let

$$p \geq \frac{3d}{d+2}. \quad (2.11)$$

Then there are weak solutions to $(Per)_p$ and $(Cau)_p$. Moreover, if (2.8) holds, then such solutions are smoother, i.e.,

$$\begin{aligned} v &\in C\left(I; L_{\text{DIV}}^2(\Omega)^d\right) \cap L^\infty\left(I; W_{\text{DIV}}^{1,p}(\Omega)^d\right) \cap L^2\left(I; W^{2,2}(\Omega)^d\right), \\ \frac{\partial v}{\partial t} &\in L^2\left(I; L^2(\Omega)^d\right), \end{aligned} \quad (2.12)$$

and unique (in the class of weak solutions). If $d = 2$, $v_0 \in W_{\text{DIV}}^{1,2}(\Omega)^d$ and $p \in (1, 2)$, then there are solutions v such that

$$\begin{aligned} v &\in C\left(I; L_{\text{DIV}}^2(\Omega)^d\right) \cap L^\infty\left(I; W_{\text{DIV}}^{1,2}(\Omega)^d\right) \cap L^2\left(I; W^{2,p}(\Omega)^d\right), \\ \frac{\partial v}{\partial t} &\in L^2\left(I; L^2(\Omega)^d\right). \end{aligned} \quad (2.13)$$

The complete proof of the theorem can be found in [41], Section 5. The Cauchy problem is treated in [53].

Truncation method for $(Per)_p$.

The reader will immediately ask whether it is possible to obtain the existence of weak solution even for $p \leq 9/5$. The second *truncation method* gives the positive answer for the $p > 8/5$. More precisely we have:

2.1 Theorem. Let $d \geq 2$, $v_0 \in L_{\text{DIV}}^2(\Omega)^d$ and $f \in L^{p'}\left(I; \left(W_{\text{DIV}}^{1,p}(\Omega)^d\right)^*\right)$. If

$$p > \frac{2(d+1)}{d+2}, \quad (2.14)$$

then there is at least one weak solution to $(Per)_p$.

The method of the proof is based on the strict monotonicity of the elliptic operator and on the construction of a special test function belonging to $L^\infty(Q_T)^d \cap L^p(I; W_{\text{DIV}}^{1,p}(\Omega)^d)$ – we call it the truncation function. The bound (2.14) is due to integrability of the convective term, see (2.5). The detailed proof of Theorem 2.3 can be found in [13].

The origin of the method goes back to the works [11], [2], and [8] aimed to the analysis of nonlinear elliptic equations, or even earlier to the results of Leray and Lions [32]. The method is also applicable to the steady problem $(Dir - St)_p$ and $(Per - St)_p$, see [12] and [55].

Lipschitz “truncation” method for $(Per)_p$.

It seems that we can obtain the existence of weak solutions even for $p > 6/5$ using the so-called *Lipschitz “truncation” method*. Because of complicated analysis, steady problems are treated first. The following result is proved in [14].

2.2 Theorem. Let $d \geq 2$, $f \in L^{p'}(\Omega)^d$ and

$$p > \frac{2d}{d+2}, \quad (2.15)$$

then there are weak solutions to $(Dir - St)_p$.

Regularity method for $(Dir)_p$.

The extension of the results to $(Dir)_p$ is not simple. Up to now we can present only results that are based on the regularity method. In this method, relying on the local techniques of the regularity theory (estimates in interior, then in tangential directions near the boundary and finally in the normal direction at the boundary), one has to overcome several technical difficulties stemming from: (i) missing uniform estimates of the L^2 -norm of the time derivative and pressure if $p < 12/5$; (ii) the dependence of the elliptic operator on the symmetric velocity gradient; in order to obtain information on the full gradient, one uses the Korn inequality which is a type of integral inequality requiring the information on the symmetric velocity gradient in the whole domain Ω ; (iii) missing information on the gradient of the pressure at the boundary does not allow to estimate the second derivatives of v at the normal direction at the boundary. This is eliminated by using the curl-operator

on equation (1.1). The information on the second derivatives is then obtained using the relation between L^q -norm of a function and the estimates of its gradient in the norm of the ‘negative’ Sobolev spaces. Due to these difficulties, the results for $(Dir)_p$ are up-to-now weaker than those for $(Per)_p$.

The following result is proved in [40].

2.3 Theorem. Let $d = 3$ and $v_0 \in W_{\text{DIV}}^{1,p}(\Omega)^d$ and $f \in L^2(Q_r)^d$. If $p \in (2, 3)$, then there are weak solutions to $(Dir)_p$. Moreover, if

$$p > \frac{9}{4}, \quad (2.16)$$

then this solution is unique (in the class of weak solutions) and regular, i.e.,

$$\begin{aligned} v &\in C(I; L^2(\Omega)^d) \cap L^\infty(I; W_{\text{DIV}}^{1,p}(\Omega)^d) \cap L^{p'}(I; W^{2,p'}(\Omega)^d), \\ \frac{\partial v}{\partial t} &\in L^2(I; L^2(\Omega)^d). \end{aligned} \quad (2.17)$$

Let us remark that only this result removes the gap between the existence of weak solution to 3d NSEs in case of the Dirichlet problem (see [18] and [24]) and the existence results of Ladyzhenskaya for a weak solution to $(Dir)_p$ in three dimensions, when $p \geq \frac{11}{5}$.

$C^{1,\alpha}$ -regularity.

Suppose that $p > 9/4$ for $(Dir)_p$ (or $p \geq 11/5$ for $(Per)_p$). One can ask the following question: Does then the solution to $(Dir)_p$ (or $(Per)_p$) have the same properties as the solution of 2d NSEs? Otherwise: Can one consider system (1.1), (1.2) with $p > 9/4$ (or $p \geq 11/5$) as the well-posed alternative to 3d NSEs?

The answer is not simple, and depends on properties we want to analyze.

If we are interested in studying the large time behavior of all solutions, we will see in Section 4 that we can indeed prove the same results as for 2d NSEs. However, new approaches had to be developed in order to achieve them.

If we are interested in proving higher regularity, for example, one could ask whether solution belongs to

$$L^2(I; W^{3,2}(\Omega)^d), \quad (2.18)$$

we immediately find out that this is very difficult task, which is not generally known even for elliptic systems (then (2.18) is replaced by $W^{3,2}(\Omega)$ regularity) with $p = 2$ where the relation between \mathbf{T}^E and \mathbf{D} (or even $\nabla \mathbf{v}$) is nonlinear³.

It is not difficult to see that (2.18) can be obtained provided that we can show that $\nabla \mathbf{v} \in L^\infty(Q_T)^{d \times d}$. This key step follows from slightly better information, namely

$$\mathbf{v} \in C^{1,\alpha}(Q_T)^d.$$

Thus the question of the Hölder continuity of the velocity gradients is interesting and nontrivial even for GNSEs in two dimensions. The rest of this section is devoted to this problem.

In the sequel, let $d = 2$. Regarding the elliptic systems the existence and uniqueness of $C^{1,\alpha}$ -solutions are proved firstly in [59] generalizing the result [48], [49], where the scalar elliptic equations were studied.

Regarding parabolic systems, interior $C^{1,\alpha}$ -regularity was proved in [51], but only for $p = 2$. The same result, i.e., the interior $C^{1,\alpha}$ -regularity to $(Dir)_p$ was shown in [56], but also for $p = 2$; see also [29].

In order to understand better behavior of the models with $p \neq 2$ and also behavior of the solutions near the boundary it is useful to discuss first the stationary problems $(Dir - St)_p$ and $(Per - St)_p$. Since the identity (2.10) can be used only at the interior of the domain, the results differ significantly if we deal with $(Dir - St)_p$ or $(Per - St)_p$. We can summarize them into the following theorem.

2.4 Theorem. Let $d = 2$ and $\Omega \in C^2$. For $\varepsilon_0 > 0$, let f belong to $L^{2+\varepsilon_0}(\Omega)^d$.

- (i) If $p > \frac{6}{5}$, then there is a solution to $(Dir - St)_p$ belonging to $C_{loc}^{1,\alpha}(\Omega)$ (interior regularity).
- (ii) If $p > \frac{3}{2}$, then there is a solution to $(Dir - St)_p$ belonging to $C^{1,\alpha}(\overline{\Omega})$ (global regularity).
- (iii) If $p > 1$, then there is a solution to $(Per - st)_p$ belonging to $C^{1,\alpha}(\mathbb{R}^2)$ (global regularity for space period problem).

The proof of (iii) can be found in [18], while (i) and (ii) is presented in [19].

³ Special problems including the p -Laplace operators can be solved, see [9].

It reveals that techniques used to prove the above mentioned theorem is a good strategy to treat the evolutionary models, as shows the following result, see [20] for the proof.

2.5 Theorem. Let $d = 2$. For $\varepsilon_0 > 0$, let f belong to $L^{2+\varepsilon_0}(Q_T)^d$. If $p > 4/3$, then there is a solution to $(Per)_p$ belonging to $C_{loc}^{1,\alpha}(I; C^{1,\alpha}(\Omega)^d)$. In addition, this solution is unique in the class of weak solutions.

In three dimensions, there are known sufficient conditions (for a certain range of parameters p) that guarantee $C^{1,\alpha}$ -regularity of solutions for GNSEs, see [57].

3. LARGE TIME BEHAVIOR – FINITE-DIMENSIONAL GLOBAL AND EXPONENTIAL ATTRACTORS

Consider first a general nonlinear system of partial differential equations written as an abstract evolutionary problem

$$\begin{aligned} u'(t) &= F(u(t)) \quad \text{in } X, \quad (t > 0) \\ u(0) &= u_0, \end{aligned} \tag{3.1}$$

where X is an (infinite-dimensional) Banach space, $F : X \rightarrow X$ is a nonlinear operator and $u_0 \in X$.

The objective is to present a method of proving both the existence of a global attractor with finite fractal dimension and the existence of an exponential attractor under assumptions on F and on the solution of (3.1) that are more general than those usually traced in the literature (see [1], [10], [16], [27], [60] for example) and consequently, they are well-suited for studying large time behavior of our problems $(Per)_p$ and $(Dir)_p$. We call this new approach the *methods of ℓ -trajectories*.

To be more specific, we are going to address one of the advantages of this method. In classical theory, the nonlinearity $F(u)$ is usually of the type $L(u) + Q(u)$, where $L : X \rightarrow X$ is a linear elliptic operator and Q is a nonlinearity of lower order. Due to the presence of *linear* operator L , one uses the procedures that require regularity of solution for both the original and the linearized problem. A good example is the method of Lyapunov exponents. (The key step in the method is to show that $\|S_t u - S_t v - S'_t(u)(u - v)\|_X = o(\|u - v\|_X)$ as $\|u - v\|_X \rightarrow 0$, where S_t are solution operators and $S'_t(u)$ is the Fréchet derivative of attractors (see [6] or [61]), or the schemes where L commutes with projector operators (see for example the Ladyzenskaya's criterion of finiteness of fractal dimension of compact sets, cf. [28]).

However, if F is of the form $F(u) = N(u) + Q(u)$, where $N(u)$ is a *nonlinear* elliptic operator, then these procedures and schemes can hardly be used. As mentioned in the previous section, the question of $C^{1,\alpha}$ -regularity of weak solutions is open even for parabolic systems in three (but generally even in two) dimensions. Then the *method of ℓ -trajectories* that has been developed in [36] and [42] and described in general setting with consequences to other evolutionary dissipative systems in [43] reveals itself to be a powerful tool.

The characteristic features of the method are the following:

- (i) Instead of working in the original phase space X , we deal with the set of all ℓ -trajectories (for all possible initial values in X), considered as a subset of $X_\ell = L^2(0, \ell; X)$.
- (ii) It reveals that the proofs of the existence of a finite-dimensional global attractor \mathcal{A}_ℓ and the existence of an exponential attractor \mathcal{E}_ℓ in the phase space of ℓ -trajectories are much easier. Furthermore, their existence in X_ℓ is proved even in situations where the known direct methods to show the existence of attractors in X fail.
- (iii) Finally, defining $\mathcal{A} \subset X$ as the set of end-points of all ℓ -trajectories from \mathcal{A}_ℓ , we observe that \mathcal{A} enjoys all properties of global attractor with finite fractal dimension in the original phase space. Analogously, we observe that the set $\mathcal{E} \subset X$ of end-points of all ℓ -trajectories from \mathcal{E}_ℓ enjoys all properties of exponential attractor in X .

Despite the fact that the uniqueness results require different smoothness of initial values depending if p satisfies (2.9) or (2.8) (compare also with Theorems 2.1 and 2.2), we are able to prove the following theorem.

3.1 Theorem. Assume that $f \in L^2(\Omega)^d$ be arbitrary, and

$$p \geq \frac{3d+2}{d+2} \quad (d = 2, 3).$$

Then the problem $(Per)_p$ possesses both a global attractor \mathcal{A} with finite fractal dimension and an exponential attractor \mathcal{E} .

In two dimensions we have:

3.2 Theorem. Let $f \in L^{p'}(\Omega)^d$ and $p \in (1, 2)$, $d = 2$. Then dynamical system (3.1) possesses both a global attractor with finite fractal dimension and an exponential attractor.

The proofs can be found in [43]. We believe this method can be applied to various problems of physics and mechanics (as examples, we can put Boussinesq

approximation, wave equations with various kind of dissipations, evolutionary dissipative problems with delay and memory terms, etc.).

The question of large time behavior for *GNSEs* is treated systematically also by Ladyzhenskaya and Serëgin, see [30], [31], [58] for example.

4. CLASSICAL NAVIER-STOKES EQUATIONS

Navier-Stokes equations are the source of many mathematical developments in analysis and without exaggeration they represent the most studied system of nonlinear partial differential equations in the theory of PDEs. *NSEs* have occurred in the titles of more than 2800 mathematical publications published since 1940. Need to say that the most fundamental papers (from the point of view of modern applied functional analysis and theory of PDEs) were published even earlier in 1933-34 by Jean Leray (see [32]).

For simplicity we restrict ourselves to the Cauchy problem with $f \equiv 0$, and we study (by a nondimensionalization and a convenient scaling one can arrange the viscosity equals to 1),

$$\begin{aligned} \operatorname{div} v &= 0, \\ \frac{\partial v}{\partial t} + v_k \frac{\partial v}{\partial x_k} - \Delta v &= -\nabla \pi, \end{aligned} \quad (4.1)$$

with the initial value

$$v(0, x) = v_0(x) \in \mathbb{R}^3, \quad \operatorname{div} v_0 = 0. \quad (4.2)$$

4.1 Definition. Let $v_0 \in L^2_{\operatorname{DIV}}(\mathbb{R}^3)^3$. We say that

$$v \in C\left(I; L^2_{\operatorname{DIV}, w}(\mathbb{R}^3)^3\right) \cap L^2\left(I; W^{1,2}_{\operatorname{DIV}}(\mathbb{R}^3)^3\right) \quad (4.3)$$

is a weak solution to (4.1) if the identity

$$\left\langle \frac{\partial v(t)}{\partial t}, \varphi \right\rangle_{W^{1,2}_{\operatorname{DIV}}(\Omega)^d} + \left\langle v_k(t) \frac{\partial v(t)}{\partial x_k}, \varphi \right\rangle + v(\nabla v(t), \nabla \varphi) = 0 \quad (4.4)$$

holds for a.a. $t \in I$ and for all $\varphi \in W^{1,2}_{\operatorname{DIV}}(\mathbb{R}^3)^3$.

The fact that v satisfies (4.3) implies that

$$\frac{\partial v}{\partial t} \in L^{\frac{4}{3}}\left(I; \left(W^{1,2}_{\operatorname{DIV}}(\mathbb{R}^3)^3\right)^*\right).$$

We immediately see that we cannot set $\varphi = v$ as a test function in (4.4) to obtain energy identity. Nevertheless, there are weak solutions satisfying the energy inequality having the form

$$\sup_{t \in I} \|\boldsymbol{v}(t)\|_2^2 + \int_0^T \|\nabla \boldsymbol{v}(t)\|_2^2 dt \leq \|\boldsymbol{v}_0\|_2^2. \quad (4.5)$$

Existence of a weak solution satisfying (4.5) is proved in [32]. Leray investigates also the two-dimensional Cauchy problem and shows that then this solution is unique and smooth as data. Leray (and up-to-now anybody else) does not succeed in solvability of the question of uniqueness in three space dimensions. It is known, however, that if weak solution \boldsymbol{v} satisfies besides (4.2) the condition

$$\boldsymbol{v} \in L^r(I; L^s(\mathbb{R}^3)^3); \quad \frac{2}{r} + \frac{3}{s} \leq 1, \quad s \in (3, \infty), \quad r \in (2, \infty), \quad (4.6)$$

then such a solution is unique even in the class of weak solutions.

Recall that (4.3) yields $\boldsymbol{v} \in L^{\frac{10}{3}}(I; L^{\frac{10}{3}}(\mathbb{R}^3)^3)$, while uniqueness is guaranteed if $\boldsymbol{v} \in L^5(I; L^5(\mathbb{R}^3)^3)$. Condition (4.6) is also the condition implying the higher space regularity of weak solution.

It is natural to ask whether one can improve the information following from (4.5). Many attempts have been made in this direction. None of them implies (4.6), even more, none of them improves the so-called scaling number. The result related closely to the regularity method in this exposition is due to [64], see also [61]. It is shown that there is a weak solution satisfying (in addition to (4.3)-(4.5))

$$\int_0^T \|\nabla^{(2)} \boldsymbol{v}(t)\|_2^{2/3} dt < \infty,$$

which implies $\boldsymbol{v} \in L^1(I; L^\infty(\mathbb{R}^3)^3)$. A slight improvement of this result can be found in [46].

Coming back to (4.6), we concentrate on the limiting case when

$$\boldsymbol{v} \in L^\infty(I; L^3(\mathbb{R}^3)^3), \quad (4.7)$$

which is interesting from the following two reasons:

- (1) The space $L^3(\mathbb{R}^3)^3$ is the largest Lebesgue space for which the global existence of unique smooth solution is known for small data. This result should be addressed to Kato [21], see also recent works by Cannone, Planchon [7].
- (2) J. Leray in 1933 expected weak solutions to be irregular and he also suggested a construction of the singular solution in (backward) self-similar form. These possible solutions are of the form

$$\boldsymbol{v}(t, x) = \lambda(t) \boldsymbol{U}(\lambda(t)x),$$

$$\text{where } \lambda(t) = \frac{1}{\sqrt{a(T-t)}}. \quad (4.8)$$

$$\pi(t, x) = \lambda^2(t) P(\lambda(t)x),$$

Putting (4.8) into (4.1), one obtains Leray's system

$$\operatorname{div} \mathbf{U} = 0,$$

$$a\mathbf{U}(y) + ay_k \frac{\partial \mathbf{U}}{\partial y_k} + U_k \frac{\partial \mathbf{U}}{\partial y_k} - \Delta \mathbf{U} = -\nabla P. \quad (4.9)$$

If one shows that there are nontrivial solutions, \mathbf{U} of (4.9) such that $\mathbf{U} \in L^\infty(\mathbb{R}^3)^3 \cap L^2(\mathbb{R}^3)^3$ (these function spaces were originally proposed by Leray), then solution of the type (4.8) is a weak solution such that

$$\lim_{t \rightarrow T^-} \|\nabla v(t)\|_2 = \infty. \quad (4.10)$$

Let $\mathbf{U} \in W^{1,2}(\mathbb{R}^3)^3$. Because of the particular form of self-similar solution (4.8), we see that v satisfies not only (4.3), but also

$$v \in L^\infty(I; L^3(\mathbb{R}^3)^3). \quad (4.11)$$

If one knows that $L^\infty(I; L^3(\mathbb{R}^3)^3)$ is a class of functions that implies regularity then Leray's construction is immediately excluded. Although specialists believed that this construction is not possible (see [61], p. 27), it took more than sixty years when the problem was completely resolved in case of the whole space \mathbb{R}^3 . There are three works devoted to solving this problem. First of all, Nečas, Růžička, and Šverák [52] show that $\mathbf{U} \equiv 0$ if $\mathbf{U} \in L^3(\mathbb{R}^3)^3$. The proof uses results of [4]. Later on, a simple proof was given in [39] under the assumption that $\mathbf{U} \in W^{1,2}(\mathbb{R}^3)^3$.

Finally, Tsai in [62] proves the nonexistence of nontrivial solutions provided very general assumptions on v respective \mathbf{U} . It is enough to assume that the energy inequality holds only locally, i.e., that for all bounded $B \subset \mathbb{R}^3$

$$\sup_{t \in I} \int_B |v(t, x)|^2 dx + \int_0^T \int_B |\nabla v(t, x)|^2 dx \leq \|v_0\|_{2,B}^2,$$

or $\mathbf{U} \in L^q(\mathbb{R}^3)$ for q arbitrarily large. The feature that is common to all three papers is the maximum principle for the quantity

$$X(y) = \frac{|\mathbf{U}(y)|^2}{2} + P(y) + ay \cdot \mathbf{U}(y),$$

which satisfies

$$-\nu \Delta X(y) + ay_k \frac{\partial X(y)}{\partial y_k} + U_k(y) \frac{\partial X(y)}{\partial y_k} \leq -|\operatorname{curl} \mathbf{U}|^2 \leq 0. \quad (4.12)$$

Equation (4.12) can be then used if one has certain control of $U(y)$ and $P(y)$ for $|y|$ sufficiently large. This is the point where the papers differ.

The aim of this section was to mention results for *NSEs* connected with previous sections and to summarize the results concerning self-similar solutions. This section was not meant as to survey all significant recent results concerning *NSEs*. For this purpose, we suggest the reader to look into papers by Heywood and Wiegner, see [17], [63]; or the monographs [61], [5] and [34].

5. APPENDIX

1. Sets in \mathbb{R}^m , $m \in \mathbb{N}$: $\Omega \subset \mathbb{R}^d$ is an open set (mostly $\Omega = \mathbb{R}^d$ or $\Omega = (0, L)^d$) or Ω is a bounded open set in \mathbb{R}^d ; $I \equiv (0, T)$ is a time interval, $T \in (0, \infty)$; $Q_T \equiv (0, T) \times \Omega = I \times \Omega$.
2. Functions

Scalar functions are written in italics, vector-valued functions are boldfaced, and tensor-valued functions are typeset in boldfaced capitals. For steady problems, all functions depend on $x \in \Omega$, while in evolutionary case they usually depend on $(t, x) \in Q_T$. Examples (with physical interpretation used in the text):

$\pi : Q_T \rightarrow \mathbb{R}$ (scalar) pressure, $v \equiv (v_1, v_2, \dots, v_d) : Q_T \rightarrow \mathbb{R}^d$ the velocity,

$T^E \equiv (T_{ij}^E)_{i,j=1}^d : Q_T \rightarrow \mathbb{R}^{d \times d} \equiv \mathbb{R}^d \times \mathbb{R}^d$ the stress tensor.

$v \otimes v \equiv (v_i v_j)_{i,j=1}^d$.

3. Operations

Einstein summation convention is used throughout the whole text, for example:

$$\operatorname{div} v = \sum_{i=1}^d \frac{\partial v_i}{\partial x_i} = \frac{\partial v_i}{\partial x_i};$$

$$\operatorname{div} T^E = \left((\operatorname{div} T^E)_1, \dots, (\operatorname{div} T^E)_d \right) \text{ with } (\operatorname{div} T^E)_i = \sum_{j=1}^d \frac{\partial T_{ij}^E}{\partial x_j} = \frac{\partial T_{ij}^E}{\partial x_j}, i = 1, \dots, d;$$

Similarly the i -th component of $\operatorname{div}(v \otimes v)$ is given by $\frac{\partial}{\partial x_j} v_i v_j$.

$$\nabla \pi = \left(\frac{\partial \pi}{\partial x_1}, \dots, \frac{\partial \pi}{\partial x_d} \right).$$

$$\text{If } T = T(A) : \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d} \text{ then } \partial T = (\partial_{rs} T)_{r,s=1}^d \text{ and } \partial_{rs} T \equiv \frac{\partial T}{\partial A_{rs}}.$$

4. Function spaces

Let $(X(\Omega), \|\cdot\|_{X(\Omega)})$ be a Banach space of scalar functions defined on Ω . Then $X(\Omega)^d$ (resp. $X(\Omega)^{d \times d}$) represents the space of vector-valued (tensor-valued) functions with components belonging to $X(\Omega)$. The symbol $X(\Omega)^*$ denotes the dual space to $X(\Omega)$.

Let $p, q > 1$ and $k > 0$. Then $(L^p(\Omega), \|\cdot\|^p)$ denotes usual Lebesgue spaces and $(W^{k,p}(\Omega), \|\cdot\|_{k,p})$ denotes standard Sobolev spaces.

Bochner spaces are designated by symbols

$$\left(L^q((0, T); X(\Omega)), \left(\int_0^T \|\cdot\|_{X(\Omega)}^q dt \right)^{1/q} \right).$$

$\mathcal{D}(\Omega)$ denotes the space of smooth functions with compact support in Ω .

We further introduce

$$\mathcal{D}_{\text{DIV}}(\Omega)^d \equiv \{\boldsymbol{\psi} \in \mathcal{D}(\Omega)^d; \operatorname{div} \boldsymbol{\psi} = 0\},$$

$L_{\text{DIV}}^2(\Omega)^d \equiv$ the closure of $\mathcal{D}_{\text{DIV}}(\Omega)$ w.r.t. $\|\cdot\|_2$ -norm;

and we define the space $W_{\text{DIV}}^{1,p}(\Omega)^d$ in dependence on the type of boundary conditions. In case of the Dirichlet or the Cauchy problems, we set

$$W_{\text{DIV}}^{1,p}(\Omega)^d \equiv \text{the closure of } \mathcal{D}_{\text{DIV}}(\Omega)^d \text{ w.r.t. } \|\cdot\|_p \text{-norm,}$$

while for the space periodic problem, we set for $i \in \{1, \dots, d\}$

$$\Gamma_i \equiv \{x \in \mathbb{R}^d; x_i = 0, x_j \in (0, L), j \neq i, j = 1, \dots, d\}$$

$$\Gamma_{i+d} \equiv \{x \in \mathbb{R}^d; x_i = L, x_j \in (0, L), j \neq i, j = 1, \dots, d\}$$

and put

$$W_{\text{DIV}}^{1,p}(\Omega)^d \equiv \left\{ \boldsymbol{v} \in W^{1,p}(\Omega)^d; \forall i \int_{\Omega} v_i dx = 0, \boldsymbol{v}|_{\Gamma_i} = \boldsymbol{v}|_{\Gamma_{i+d}}, \operatorname{div} \boldsymbol{v} = 0 \right\}.$$

The brackets (\mathbf{h}, \mathbf{g}) denote $\int_{\Omega} \mathbf{h} \cdot \mathbf{g} dx$, where $\mathbf{h} \cdot \mathbf{g} = h_i g_i \in L^1(\Omega)$, while

$\langle \cdot, \cdot \rangle_{X(\Omega)}$ represents the duality between $X(\Omega)$ a $X(\Omega)^*$.

Acknowledgement: The research was supported by CEZ: J13/98113200007, and by the grant 2996K1020 supported by NSF and the Ministry of Education of the Czech Republic.

REFERENCES

- [1] Babin, A.V. and Vishik, V.B., *Attractors of Evolution Equations*, North-Holland, Amsterdam, London, New York, Tokyo, 1992.
- [2] Boccardo, L. and Murat, F., Almost everywhere convergence of the gradients of solutions to elliptic and parabolic equations, *Nonlinear Anal.*, Vol. 19, 1992, 581-597.
- [3] Bellout, H., Bloom, F., and Nečas, J., Young measure-valued solutions for non-Newtonian incompressible fluids, *Comm. in PDE*, Vol. 19 (11 & 12), 1994, 1763-1803.
- [4] Caffarelli, L., Kohn, R.V., and Nirenberg, L., Partial regularity of suitable weak solution of the Navier-Stokes equations, *Comm. Pure Appl. Math.*, 35, 1982, 771-831.
- [5] Constantin, P. and Foias, C., *The Navier-Stokes Equations*, Univ. of Chicago Press, Chicago, 1988.
- [6] Constantin, P. and Foias, C., Global Lyapunov exponents, Kaplan-Yorke formulas and the dimension for the attractors for two-dimensional Navier-Stokes equations, *Comm. Pure Appl. Math.*, 38, 1985, 1-27.
- [7] Cannone, M. and Planchon, F., On the nonstationary Navier-Stokes equations with an external force, *Adv. in Diff. Eq.*, 4, 1999, 697-730.
- [8] Dal Maso, G. and Murat, F., Almost everywhere convergence of gradients of solutions to nonlinear elliptic systems, *Nonlinear Anal.*, 31, 1998, 405-412.
- [9] DiBenedetto, E., *Degenerate Parabolic Equations*, Springer-Verlag, New York, 1993.
- [10] Eden, A., Foias, C., Nicolaenko, B., and Temam, R., *Exponential Attractors for Dissipative Evolution Equations*, Wiley, Masson, Chichester, Paris, 1994.
- [11] Frehse, J., A refinement of Rellich's theorem, *Ren. Mat.*, 5, 1985, 229-242.
- [12] Frehse, J., Málek, J., and Steinhauer, M., An existence result for fluids with shear dependent viscosity-steady flows, *Nonlinear Analysis-Theory, Methods, Applications*, 30, 1997, 3041-3049.
- [13] Frehse, J., Málek, J., and Steinhauer, M., On existence results for fluids with shear dependent viscosity-unsteady flows, in *Partial Differential Equations, Theory and Numerical Solution*, Pitman Research Notes in Mathematics Series (Eds. W. Jäger, O. John, K. Najzar, J. Nečas, J. Stará), Chapman & Hall/CRC, 1999, 121-129.
- [14] Frehse, J., Málek, J., and Steinhauer, M., An existence result for steady motions of fluids with shear dependent viscosity via the Lipschitz "truncation" method, In preparation.
- [15] Foias, C. and Olson, E., Finite fractal dimension and Hölder-Lipschitz parametrization, *Indiana Univ. Math. J.*, 45, 1996, 603-616.

- [16] Hale, J.K., *Asymptotic Behavior of Dissipative Systems*, Vol. 25, AMS, Providence, 1988.
- [17] Heywood, J.G., Remarks on the possible global regularity of solutions of the three-dimensional Navier-Stokes equations, *Progress in Theoretical and Computational Fluid Mechanics* (Eds. G.P. Galdi, J. Málek, J. Nečas), Pitman Research Notes in Mathematics Series 308, Longman Scientific & Technical, Essex, 1994, 1-32.
- [18] Kaplický P., Málek, J., and Stará, J., Full regularity of weak solutions to a class of nonlinear fluids in two dimensions-stationary periodic problem, *Comment. Mat. Univ. Carol.*, 38 (No. 4), 1997, 681-695.
- [19] Kaplický', P., Málek, J., and Stará, J., $C^{1,\alpha}$ regularity of weak solutions to a class of nonlinear fluids in two dimensions-stationary Dirichlet problem, *Zap. Nauchn. Sem. POMI*, 259 (No. 29), 1999, 89-121.
- [20] Kaplický', P., Málek, J., and Stará, J., $C^{1,\alpha}$ -solutions to a class of nonlinear fluids with shear dependent viscosity in two dimensions-evolutionary space periodic problem, submitted to *Nonlinear Differential Equations and Applications*.
- [21] Kato, T., Strong L^p solutions of the Navier-Stokes equations in \mathbb{R}^3 with applications to weak solutions, *Math. Zeit.*, 187, 1984, 471-480.
- [22] Ladyzhenskaya, O.A., On some new equations describing dynamics of incompressible fluids and on global solvability of boundary value problems to these equations, *Trudy Steklov's Math. Institute*, 102, 1967, 85-104.
- [23] Ladyzhenskaya, O.A., On some modifications of the Navier-Stokes equations for large gradients of velocity, *Zapiski Naukhnych Seminarov LOMI*, 7, 1968, 126-154.
- [24] Ladyzhenskaya, O.A., *The Mathematical Theory of Viscous Incompressible Flow*, Gordon and Beach, New York, 1969.
- [25] Ladyzhenskaya, O.A., On the dynamical system generated by the Navier-Stokes equations, *Zap. Nauchn. Sem. LOMI*, 27, 1972, 91-114.
- [26] Ladyzhenskaya, O.A., On finite dimensionality of bounded invariant sets for the Navier-Stokes equations and some other dissipative systems, *Zap. Nauchn. Sem. LOMI*, 115, 1982, 137-155.
- [27] Ladyzhenskaya, O.A., *Attractors for Semigroups and Evolution Equations*, Cambridge University Press, Cambridge, 1991.
- [28] Ladyzhenskaya, O.A., On finite dimensionality of bounded invariant sets for the Navier-Stokes equations and some other dissipative systems, *Zap. Nauchn. Sem. LOMI*, 115, 1982, 137-155.
- [29] Ladyzhenskaya, O.A. and Serëgin, G., Interior regularity for solutions to two-dimensional equations of the dynamics of fluids with nonlinear viscosity, *Zapiski, Nauchn. Seminar. POMI*, 259, 1998, 1-22.

- [30] Ladyzhenskaya, O.A. and Serëgin, G., On semigroups generated by initial-boundary value problems describing two-dimensional visco-plastic flows, Amer. Math. Soc. Transl. (2), 164, 1995, 99-123.
- [31] Ladyzhenskaya, O.A. and Serëgin, G., On smoothness of solutions to system describing the flow of generalized Newtonian fluids, and on evaluation of dimension of their attractors, Izvestia RAN, Ser. Math., 62, 1993, 59-122.

17 NONLINEAR HYPERBOLIC PARTIAL DIFFERENTIAL AND VOLTERRA INTEGRAL EQUATIONS: ANALYTICAL AND NUMERICAL APPROACHES

Roger G. Marshall

Department of Computer Science
Winston-Salem State University
Winston, Salem, NC 27110
and

Sudhakar G. Pandit
Department of Mathematics
Winston-Salem State University
Winston-Salem, NC 27110

ABSTRACT

Linear hyperbolic partial differential and Volterra integral inequalities are employed to solve nonlinear characteristic hyperbolic initial-boundary value problems and Volterra integral equations. Three types of monotone iterative schemes are presented: (i) The Alternating Sequence Scheme; (ii) The Monotone Iterative Scheme; and (iii) The Generalized Quasilinear Scheme. The iterates in all the three shcemes are linear and hence, can be easily computable by using the Variation of Parameters Formulas and the Resolvent Kernel techniques. Whereas the mode of convergence of the iterates in the first two schemes is linear, that in the third scheme is quadratic and hence, more rapid. Numerical Examples are given in support of the analytical methods.

INTRODUCTION

Consider the nonlinear hyperbolic partial differential equation

$$u_{xy} = f(x, y, u, u_x, u_y), \quad (1.1)$$

and the nonlinear Volterra integral equation

$$u(t) = h(t) + \int_0^t K(t, s, u(s)) ds. \quad (1.2)$$

Utilizing the theory of hyperbolic partial differential and Volterra integral inequalities [2, 4-7, 10], we develop the following three types of iterative schemes for (1.1) and (1.2):

- (i) The Alternating Sequence Scheme;
- (ii) The Monotone Iterative Scheme; and
- (iii) The Generalized Quasilinear Scheme.

Whereas the iterates in (i) and (ii) converge linearly, the convergence of the iterates in (iii) is quadratic and hence, more rapid. Of special interest are the cases when f in (1.1) is nonincreasing in the last three variables, and K in (1.2) is nonincreasing in the last variable. The boundary-layer theory problem, considered in [11], for example, can be transformed to the Volterra integral equation (1.2) in which the kernel K is nonincreasing in u . In this paper, we provide numerical computations for the Alternating Sequence Schemes only. For the other two schemes, see [8, 9].

NOTATIONS AND DEFINITIONS

We use the following notations and definitions concerning equations (1.1) and (1.2). For real numbers $a > 0$, $b > 0$, let I , J , and R denote the intervals $[0, a]$, $[0, b]$, and the rectangle $I \times J$ respectively. The interval $[0, T]$, $T > 0$ is denoted by \mathbb{T} . For $z \in C^2[R, \mathbb{R}]$, by $\langle z \rangle$ we mean the triple (z, z_x, z_y) . For $f \in C^2[R \times \mathbb{R}^3, \mathbb{R}]$, $f_i(x, y, \langle z \rangle)$, $3 \leq i \leq 5$, denote the second order partial derivatives of f with respect to the last three variables. When there is no danger of ambiguity, we shall suppress the independent variables x and y in the function(s). For example, the expression $\frac{\partial^2 f(x, y, z(x, y), z_x(x, y), z_y(x, y))}{\partial z \partial z_y}$ is represented by the abbreviation $f_{3,5}(\langle z \rangle)$. The

triple (f_3, f_4, f_5) is denoted by $[f]$. The expression $[f] \cdot \langle z \rangle$ is the inner product $f_3 z + f_4 z_x + f_5 z_y$. For $v, \omega \in C^2[R, \mathbb{R}]$, the inequality $\langle v \rangle \leq \langle \omega \rangle$ on R means that $v(x, y) \leq \omega(x, y)$, $v_z(x, y) \leq \omega_x(x, y)$, and $v_y(x, y) \leq \omega_y(x, y)$, for $(x, y) \in R$.

Under these notations, consider the nonlinear Characteristic Initial-Boundary Value Problem

$$u_{xy} = f(x, y, u, u_x, u_y), \quad (x, y) \in R; \quad (1.3)$$

$$u(x, 0) = \sigma(x), \quad x \in I; \quad u(0, y) = \tau(y), \quad y \in J; \quad \sigma(0) = \tau(0) = u_0,$$

where $f \in C[R \times \mathbb{R}^3, \mathbb{R}]$, $\sigma \in C^1[I, \mathbb{R}]$, $\tau \in C^1[J, \mathbb{R}]$, and the nonlinear Volterra integral equation

$$u(t) = h(t) + \int_0^t K(t, s, u(s)) ds, \quad t \in \mathbb{T}, \quad (1.4)$$

where $h \in C[\mathbb{T}, \mathbb{R}]$, and $K \in C[\mathbb{T} \times \mathbb{T} \times \mathbb{R}, \mathbb{R}]$.

Definition 1

A function $v \in C^2[R, \mathbb{R}]$ is called a lower solution of (1.3) on R , if

$$v_{xy} \leq f(x, y, \langle v \rangle), \quad (x, y) \in R;$$

$$v_x(x, 0) \leq \sigma'(x), \quad x \in I; \quad v_y(0, y) \leq \tau'(y), \quad y \in J; \quad v(0, 0) \leq u_0,$$

an upper solution of (1.3) on R , if the reversed inequalities hold.

Definition 2

Functions $v, \omega \in C^2[R, \mathbb{R}]$ are called coupled lower-upper solutions of (1.3) on R , if

$$v_{xy} \leq f(x, y, \langle \omega \rangle), \quad \omega_{xy} \geq f(x, y, \langle v \rangle), \quad (x, y) \in R;$$

$$v_x(x, 0) \leq \sigma'(x) \leq \omega_x(x, 0), \quad x \in I; \quad v_y(0, y) \leq \tau'(y) \leq \omega_y(0, y), \quad y \in J; \quad \text{and}$$

$$v(0, 0) \leq u_0 \leq \omega(0, 0).$$

Definition 3

A function $v \in C[\mathbb{T}, \mathbb{R}]$ is called a lower solution of (1.4) on \mathbb{T} , if

$$v(t) \leq h(t) + \int_0^t K(t, s, v(s)) ds, \quad t \in \mathbb{T},$$

and, an upper solution of (1.4) on \mathbb{T} , if the reversed inequality holds.

Definition 4

Functions $v, \omega \in C[\mathbb{T}, \mathbb{R}]$ are called coupled lower-upper solutions of (1.4) on \mathbb{T} , if

$$v(t) \leq h(t) + \int_0^t K(t, s, \omega(s)) ds, \quad \omega(t) \geq h(t) + \int_0^t K(t, s, v(s)) ds, \quad t \in \mathbb{T}.$$

When v^0, ω^0 , with $\langle v^0 \rangle \leq \langle \omega^0 \rangle$ on R , are lower-upper (coupled lower-upper) solutions of (1.3), we denote by Ω_H , the closed set

$$\Omega_H = \{(x, y, z, p, q) : (x, y) \in R, \text{ and } \langle v^0 \rangle \leq (z, p, q) \leq \langle \omega^0 \rangle \text{ on } R\}.$$

Similarly, when v_0, ω_0 , with $v_0 \leq \omega_0$ on \mathbb{T} , are lower-upper (coupled lower-upper solutions of (1.4) on \mathbb{T} , we denote by Ω_H , the closed set

$$\Omega_V = \{(t, s, u) : (t, s) \in \mathbb{T} \times \mathbb{T}, s \leq t, \text{ and } v_0(t) \leq u \leq \omega_0(t) \text{ on } \mathbb{T}\}.$$

THEORY OF HYPERBOLIC PARTIAL DIFFERENTIAL INEQUALITIES AND VOLTERRA INTEGRAL INEQUALITIES

Theorem 1 below is a fundamental result on hyperbolic partial differential inequalities.

Theorem 1

Suppose that

- (i) $f, g \in C[R \times \mathbb{R}^3, \mathbb{R}]$, f is nondecreasing and g is nonincreasing in the last three variables;
- (ii) $v^0, \omega^0 \in C^2[R, \mathbb{R}]$, $\langle v^0 \rangle \leq \langle \omega^0 \rangle$ on R , and satisfy
 $v_{xy}^0 \leq f(\langle v^0 \rangle) + g(\langle w^0 \rangle)$, $w_{xy}^0 \geq f(\langle w^0 \rangle) + g(\langle v^0 \rangle)$ on R ;
 $v_x^0(x, 0) \leq w_x^0(x, 0)$, $x \in I$; $v_y^0(0, y) \leq w_y^0(0, y)$, $y \in J$; and $v^0(0, 0) \leq w^0(0, 0)$;
- (iii) f and g satisfy the one-sided Lipschitz conditions
 $f(x, y, z, p, q) - f(x, y, \bar{z}, \bar{p}, \bar{q}) \leq L[(z - \bar{z}) + (p - \bar{p}) + (q - \bar{q})]$, and
 $g(x, y, z, p, q) - g(x, y, \bar{z}, \bar{p}, \bar{q}) \geq -M[(z - \bar{z}) + (p - \bar{p}) + (q - \bar{q})]$.
whenever $(z, p, q) \geq (\bar{z}, \bar{p}, \bar{q})$, $(x, y) \in R$, and $L > 0$, $M > 0$ are constants. Then, we have

$$\langle v^0 \rangle \leq \langle w^0 \rangle \text{ on } R. \quad (1.5)$$

Remark 1

The one-sided Lipschitz conditions are required for the validity of Theorem 1, as can be seen from the following.

Example 1

Consider $u_{xy} = f(u) = \sqrt{u}$, $u \geq 0$, $(x, y) \in [0, 1] \times [0, 1]$; $u(x, 0) \equiv u(0, y) \equiv 0$, $x \in [0, 1]$, $y \in [0, 1]$. Then, $v^0(x, y) = \frac{x^2 y^2}{16}$, $w_0(x, y) \equiv 0$, satisfy $v_{xy}^0 = f(\langle v^0 \rangle)$, $w_{xy}^0 = f(\langle w^0 \rangle)$, $v^0(x, 0) \equiv w^0(x, 0)$, $x \in [0, 1]$ and $v^0(0, y) \equiv w^0(0, y)$, $y \in [0, 1]$.

However, the conclusion (1.5) is false.

The nonlinear telegraph equation

$$u_{xy} = f(x, y, u), \quad u = u(x, y), \quad (1.6)$$

is the two dimensional analog of the ordinary differential equation

$$u' = f(t, u), \quad u = u(t). \quad (1.7)$$

The close relationship between the two, which can be traced in almost the entire theory, however, breaks down while dealing with the monotonicity results. It is therefore not surprising that (1.6), in contrast to (1.7), admits extremal (minimal and maximal) solutions only if f is monotonic in the last variable. A major difference between the behaviors of (1.6) and (1.7) is exhibited by the following example.

Example 2

It is easy to see that if $u'(t) + M u(t) \geq 0$, for $t \in \mathbb{T}$, and M is any constant, then, $u(0) \geq 0$ implies that $u(t) \geq 0$ for all $t \in \mathbb{T}$. In contrast, the function $u(x, y) = \cos(\pi(2x + y))$ satisfies the hyperbolic partial differential inequality $u_{xy} + 2\pi^2 u(x, y) \geq 0$ on the rectangle $\left[0, \frac{1}{4}\right] \times \left[0, \frac{1}{2}\right]$. Also “initially”, we have $u(x, 0) \geq 0$ for $x \in \left[0, \frac{1}{4}\right]$, and $u(0, y) \geq 0$ for $y \in \left[0, \frac{1}{2}\right]$. However, $u\left(\frac{1}{6}, \frac{1}{3}\right) = -\frac{1}{2}$.

Repeated application of the following (linear) comparison result (minimum principle) is made in the development of the Monotone Iterative Technique for the Problem (1.3).

Theorem 2

Suppose that

$$\begin{aligned} u_{xy} &\geq -M_1 u - M_2 u_x - M_3 u_y, \quad (x, y) \in R; \\ u(x, 0) &\geq 0, \quad x \in I; \quad u(0, y) \geq 0, \quad y \in J; \quad u(0, 0) = 0, \end{aligned}$$

and, the constants M_i , $1 \leq i \leq 3$, satisfy

$$M_2 M_3 \geq M_1. \quad (1.8)$$

Then, $u(x, y) \geq 0$, for all $(x, y) \in R$.

The expression $M_2 M_3 - M_1$ appearing in (1.8) is the Laplace Invariant [1, 3]. It is required to be nonnegative for the validity of Theorem 2, as can be seen from the following.

Example 3

The function $u(x, y) = \sin(x + y)$ satisfies the inequality $u_{xy} \geq -u + u_x - u_y$ on the rectangle $I \times J = [0, \pi] \times [0, \pi/2]$. Even though $u(x, 0) \geq 0$, for $x \in I$, $u(0, y) \geq 0$, for $y \in J$, and $u(0, 0) = 0$, the conclusion in Theorem 2 is false.

A fundamental result on the Volterra integral inequalities is the following.

Theorem 3

Suppose that

- (i) $K_1, K_2 \in C[\mathbb{T} \times \mathbb{T}, \mathbb{R}]$, $h, v, w \in C[\mathbb{T}, \mathbb{R}]$, and satisfy the inequalities

$$v(t) \leq h(t) + \int_0^t K_1(t, s, v(s)) ds + \int_0^t K_2(t, s, w(s)) ds, \quad t \in \mathbb{T},$$

$$w(t) \geq h(t) + \int_0^t K_1(t, s, w(s)) ds + \int_0^t K_2(t, s, v(s)) ds, \quad t \in \mathbb{T};$$

- (ii) K_1 is nondecreasing and K_2 is nonincreasing in u , for each fixed $(t, s) \in \mathbb{T} \times \mathbb{T}$;

- (iii) K_1 and K_2 satisfy one-sided Lipschitz conditions

$$K_1(t, s, u) - K_2(t, s, \bar{u}) \leq L(u - \bar{u}), \quad K_2(t, s, u) - K_1(t, s, \bar{u}) \geq -M(u - \bar{u}),$$

whenever $u \geq \bar{u}$, for some constants $L > 0$ and $M > 0$. Then,

$$v(0) \leq 0 \text{ implies } v(t) \leq w(t) \text{ for all } t \in \mathbb{T}.$$

When f is nonincreasing in the last three variables, and satisfies the Lipschitz condition

$$|f(x, y, z, p, q) - f(x, y, \bar{z}, \bar{p}, \bar{q})| \leq L(|z - \bar{z}| + |p - \bar{p}| + |q - \bar{q}|), \quad (1.9)$$

for $(x, y) \in R$, $z, p, q, \bar{z}, \bar{p}, \bar{q} \in \mathbb{R}$, and for some constant $L > 0$, we have

Theorem 4

Suppose that

- (i) $v^0, w^0 \in C^2[R, \mathbb{R}]$, $\langle v^0 \rangle \leq \langle w^0 \rangle$ on R , and v^0, w^0 are lower and upper solutions respectively of (1.3) on R , such that $v^0(0, 0) = w^0(0, 0)$;

- (ii) f is nonincreasing in the last three variables, and satisfies Lipschitz condition (1.9);

For $n = 1, 2, 3, \dots$, define the iterative scheme

$$v_{xy}^n = f\left(x, y, \langle v^{n-1} \rangle\right); \quad v^n(x, 0) = \sigma(x), \quad x \in I; \quad v^n(0, y) = \tau(y), \quad y \in J.$$

If $\langle v^2 \rangle \geq \langle v^0 \rangle$ on R , then the sequence $\{v^n\}$ satisfies the alternating inequalities

$\langle v^0 \rangle \leq \langle v^2 \rangle \leq \langle v^4 \rangle \leq \cdots \leq u \leq \cdots \leq \langle v^5 \rangle \leq \langle v^3 \rangle \leq \langle v^1 \rangle \leq w^0$ on R ,
and converges uniformly to the unique solution u of (1.3) on R .

Example 4

Consider the nonlinear telegraph equation

$$u_{xy} = f(x, y, u) = \begin{cases} 1, & \text{if } u < 0 \\ \exp(-u/(x+1)), & \text{if } u \geq 0 \end{cases} \quad (1.10)$$

$$u(x, 0) \equiv 0, \quad u(0, y) = \ln(y+1),$$

on the square $[0, 1] \times [0, 1]$. All the conditions of Theorem 4 are satisfied with $v^0(x, y) \equiv 0$, $w^0(x, y) = xy + \ln(y+1)$ as lower and upper solutions respectively of (1.10), and $L = 1$ as Lipschitz constant. We note in passing that, the unique solution of (1.10) is $u(x, y) = (x+1) \ln(y+1)$. Numerical results are given below in Table 1.

Table 1

Iteration	X value	Y value	V (X, Y)	Iteration	X value	Y value	V (X, Y)
0	0.20	0.20	0.00	6	0.20	0.20	0.247244
1	0.20	0.20	0.222322	7	0.20	0.20	0.247244
2	0.20	0.20	0.248389	8	0.20	0.20	0.247244
3	0.20	0.20	0.247211	9	0.20	0.20	0.247244
4	0.20	0.20	0.247244	10	0.20	0.20	0.247244
5	0.20	0.20	0.247244	Exact			0.218786

Iteration	X value	Y value	V (X, Y)	Iteration	X value	Y value	V (X, Y)
0	0.20	0.40	0.00	6	0.20	0.40	0.454557
1	0.20	0.40	0.416472	7	0.20	0.40	0.454557
2	0.20	0.40	0.457328	8	0.20	0.40	0.454557
3	0.20	0.40	0.454452	9	0.20	0.40	0.454557
4	0.20	0.40	0.454560	10	0.20	0.40	0.454557
5	0.20	0.40	0.454557	Exact			0.403766

Iteration	X value	Y value	V (X, Y)	Iteration	X value	Y value	V (X, Y)
0	0.40	0.80	0.00	6	0.40	0.80	0.946172
1	0.40	0.80	0.907787	7	0.40	0.80	0.946172
2	0.40	0.80	0.959210	8	0.40	0.80	0.946172
3	0.40	0.80	0.945150	9	0.40	0.80	0.946172
4	0.40	0.80	0.946218	10	0.40	0.80	0.946172
5	0.40	0.80	0.946170	Exact			0.822901

Iteration	X value	Y value	V (X, Y)	Iteration	X value	Y value	V (X, Y)
0	1.0	1.0	0.00	6	1.0	1.0	1.473578
1	1.0	1.0	1.693147	7	1.0	1.0	1.473577
2	1.0	1.0	1.500038	8	1.0	1.0	1.473577
3	1.0	1.0	1.468218	9	1.0	1.0	1.473577
4	1.0	1.0	1.474005	10	1.0	1.0	1.473577
5	1.0	1.0	1.473556	Exact			1.386329

Theorem 5

Suppose that

- (i) $v_0, w_0 \in C[\mathbb{T}, \mathbb{R}]$, $v_0 \leq w_0$ on \mathbb{T} , and v_0, w_0 are lower and upper solutions respectively of (1.4) on \mathbb{T} ;

(ii) K is nonincreasing in the last variable, and satisfies Lipschitz condition

$$|K(t, s, u) - K(t, s, \bar{u})| \leq L |u - \bar{u}|,$$

whenever $t, s \in \mathbb{T}$, $u, \bar{u} \in \mathbb{R}$, $L > 0$ being a constant. For $n = 1, 2, 3, \dots$, define the iterative scheme

$$v_n(t) = h(t) + \int_0^t K(t, s, v_{n-1}(s)) ds.$$

If $v_2(t) \geq v_0(t)$ on \mathbb{T} , then the sequence $\{v_n\}$ satisfies the alternating inequalities

$v_0 \leq v_2 \leq v_4 \leq \dots \leq u \leq \dots \leq v_5 \leq v_3 \leq v_1 \leq w_0$ on \mathbb{T} ,
and converges uniformly to the unique solution u of (1.4) on \mathbb{T} .

Example 5

The third order nonlinear ordinary differential equation

$$y''' + 2y'' = 0, \quad 0 \leq t < \infty$$

$$y(0) = 0, \quad y'(0) = 0, \quad y''(0) = 1,$$

studied by H. Weil [11], arises in connection with the boundary-layer theory of fluid flow. The transformation $u(t) = -\ln(y''(t))$ transforms it into the nonlinear Volterra integral equation

$$u(t) = \int_0^t K(t, s, u(s)) ds, \quad (1.11)$$

wherein the kernel K , given by

$$K(t, s, u) = \begin{cases} (t-s)^2, & \text{if } u < 0 \\ (t-s)^2 \exp(-u), & \text{if } u \geq 0, \end{cases}$$

is nonincreasing (and convex) in u . Moreover, K is Lipschitzian in u , with Lipschitz constant $L = 1$. The functions $v_0(t) \equiv 0$, and $w_0(t) = \frac{t^3}{3}$ form a pair of lower and upper solutions respectively of (1.11). Table 2 shows the numerical results.

MONOTONE ITERATIVE SCHEMES

Let v^0, w^0 be lower and upper solutions respectively of (1.3) on R , such that $v^0 \leq w^0$ on R , and $v^0(0, 0) = w^0(0, 0)$. For constants M_i , $1 \leq i \leq 3$, with $M_2 M_3 = M_1$, define the functions F and G as follows:

$$F(x, y, z, p, q) = -M_1 z - M_2 p - M_3 q; \quad \text{and}$$

$$G(x, y, z, p, q) = f(x, y, z, p, q) - F(x, y, z, p, q).$$

Table 2

Iteration	t value	U (t)	Iteration	t value	U (t)
0	0.240	0.0	6	0.240	0.004671
1	0.240	0.004672	7	0.240	0.004671
2	0.240	0.004671	8	0.240	0.004671
3	0.240	0.04671	9	0.240	0.004671
4	0.240	0.004671	10	0.240	0.004671
5	0.240	0.004671			

Iteration	t value	U (t)	Iteration	t value	U (t)
0	0.720	0.0	6	0.720	0.123841
1	0.720	0.124608	7	0.720	0.123841
2	0.720	0.123839	8	0.720	0.123841
3	0.720	0.123841	9	0.720	0.123841
4	0.720	0.123841	10	0.720	0.123841
5	0.720	0.123841			

Iteration	t value	U (t)	Iteration	t value	U (t)
0	0.960	0.0	6	0.960	0.290962
1	0.960	0.295168	7	0.960	0.290962
2	0.960	0.290948	8	0.960	0.290962
3	0.960	0.290962	9	0.960	0.290962
4	0.960	0.290962	10	0.960	0.290962
5	0.960	0.290962			

Iteration	t value	U (t)	Iteration	t value	U (t)
0	1.00	0.0	6	1.00	0.328254
1	1.00	0.333600	7	1.00	0.328254
2	1.00	0.328235	8	1.00	0.328254
3	1.00	0.328254	9	1.00	0.328254
4	1.00	0.328254	10	1.00	0.328254
5	1.00	0.328254			

For $n = 1, 2, 3, \dots$, define the iterative schemes:

$$v_{xy}^n = F(v^n) + G(v^{n-1}); \quad v^n(x, 0) = \sigma(x), \quad x \in I; \quad v^n(0, y) = \tau(y), \quad y \in J$$

and

$$w_{xy}^n = F(w^n) + G(w^{n-1}); \quad w^n(x, 0) = \sigma(x), \quad x \in I; \quad w^n(0, y) = \tau(y), \quad y \in J.$$

Theorem 6

Suppose that $f \in C[\Omega_H, \mathbb{R}]$, and satisfies the monotonicity condition

$$f(x, y, z, p, q) - f(x, y, \bar{z}, \bar{p}, \bar{q}) \geq -M_1(z - \bar{z}) - M_2(p - \bar{p}) - M_3(q - \bar{q}),$$

whenever $(x, y, z, p, q), (x, y, \bar{z}, \bar{p}, \bar{q}) \in \Omega_H$. Then, $\{v^n\}$ is nondecreasing, $\{w^n\}$ is nonincreasing, and $\lim_{n \rightarrow \infty} v^n = \alpha$, $\lim_{n \rightarrow \infty} w^n = \beta$ uniformly, where α and β are the minimal and the maximal solutions of (1.3) respectively, and satisfy $v^0 \leq \alpha \leq u \leq \beta \leq w^0$ on R , where u is any solution of (1.3) on R . If f also satisfies Lipschitz condition (1.9), then $\alpha \equiv \beta$ on R , and consequently, there exists a unique solution of u of (1.3) satisfying $v^0 \leq u \leq w^0$ on R . Moreover, the convergence of the sequences $\{v^n\}$ and $\{w^n\}$ is linear, in the sense that, there exists a constant $K > 0$ such that

$$|u - v^n| \leq K |u - v_{n-1}|, \quad \text{and} \quad |w^n - u| \leq K |w^{n-1} - u|,$$

for $n = 1, 2, 3, \dots$.

Example 6

For (x, y) in the rectangle $[0, 2] \times [0, 1]$, consider

$$u_{xy} = 2xy(e^u + 1)^{-1} - y \sin u_x + x \cos u_y;$$

$$u(x, 0) = x + 1, \quad x \in [0, 2]; \quad u(0, y) = 1 - y, \quad y \in [0, 1].$$

The constants $M_1 = 4$, $M_2 = M_3 = 2$, and the initializations (lower-upper solutions) $v^0 = x - y - 4xy + 1$, $w^0 = x - y + 8xy + 1$ satisfy all the conditions of Theorem 6.

Remark 2

The special case $M_1 = M_2 = M_3 = 0$, which implies that f is monotonic nondecreasing in the last three variables is covered by Theorem 6. We note, in this case, that the restriction $v^0(0, 0) = w^0(0, 0)$ is not required. The other special case, when f is monotonic nonincreasing in the last three variables is, however, not covered by Theorem 6.

GENERALIZED QUASILINEAR SCHEMES

For $v^0, w^0 \in C^2[R, \mathbb{R}]$, $\langle v^0 \rangle \leq \langle w^0 \rangle$ on R , and the closed set Ω_H , assume the following hypotheses:

- (H₁) f admits a decomposition $f = g + h$, where $g, h \in C^2[\Omega_H, \mathbb{R}]$, $g_i(\langle u \rangle) \leq 0$, $h_i(\langle u \rangle) \geq 0$ on R , for $3 \leq i \leq 5$, $g_{i,j}(\langle u \rangle) \geq 0$, $h_{i,j}(\langle u \rangle) \geq 0$ on R , for $3 \leq i, j \leq 5$.
- (H₂) $v_{xy}^0 \leq g(\langle w^0 \rangle) + h(\langle v^0 \rangle)$, $w_{xy}^0 \geq g(\langle v^0 \rangle) + h(\langle w^0 \rangle)$ on R ;
 $v_x^0(x, 0) \leq \sigma'(x) \leq w_x^0(x, 0)$, for $x \in I$; $v_y^0(0, y) \leq \tau'(y) \leq w_y^0(0, y)$, for $y \in J$; and $v^0(0, 0) \leq u_0 \leq w^0(0, 0)$.

Theorem 7

Under the hypotheses (H₁) and (H₂), there exists a monotone nondecreasing sequence $\{v^n\}$ and a monotone nonincreasing sequence $\{w^n\}$ in Ω_H such that $\lim_{n \rightarrow \infty} \langle v^n \rangle = \langle u \rangle = \lim_{n \rightarrow \infty} \langle w^n \rangle$ uniformly on R , where u is the unique solution of (1.3).

Moreover, the sequences satisfy the quadratic convergence estimates

$$|u - v^n| \leq K_1 |u - v^{n-1}|^2 + K_2 |w^{n-1} - u|^2,$$

and

$$|w^n - u| \leq K_3 |w^{n-1} - u|^2 + K_4 |u - v^{n-1}|^2,$$

for $n = 1, 2, 3, \dots$, where $K_i > 0$, $1 \leq i \leq 4$, are constants. Similar estimates hold also for the gradient terms $v_x^n, v_x^{n-1}, v_y^{n-1}, u_x, u_y, w_x^n, w_x^{n-1}, w_y^n$, and w_y^{n-1} .

Example 7

Consider once again the telegraph equation (1.10) in Example 4. Clearly, g is nonincreasing and convex in u . The same functions $v^0(x, y) \equiv 0$, and $w^0(x, y) = xy + \ln(y+1)$ satisfy the conditions of Theorem 7. For $n = 1, 2, 3, \dots$, define the iterative schemes

$$\begin{aligned} v_{xy}^n &= f(\langle w^{n-1} \rangle) + [f(w^{n-1})] \cdot \langle w^n - w^{n-1} \rangle, \quad (x, y) \in R; \\ v^n(x, 0) &\equiv 0, \quad x \in I; \quad v^n(0, y) = \ln(y+1), \quad y \in J, \end{aligned}$$

and

$$\begin{aligned} w_{xy}^n &= f(\langle v^{n-1} \rangle) + [f(v^{n-1})] \cdot \langle v^n - v^{n-1} \rangle, \quad (x, y) \in R; \\ w^n(x, 0) &\equiv 0, \quad x \in I; \quad w^n(0, y) = \ln(y+1), \quad y \in J. \end{aligned}$$

Next, consider the Volterra integral equation (1.4), under the following hypotheses:

(H₃) K admits a decomposition $K = K_1 + K_2$, where $K_1, K_2 \in C^2[\Omega_V, \mathbb{R}]$,

$$K_{1u} \leq 0, \quad K_{2u} \geq 0, \quad K_{1uu} \geq 0, \quad \text{and} \quad K_{2uu} \leq 0 \quad \text{on } \Omega_V.$$

(H₄) $v_0, w_0 \in C[\mathbb{T}, \mathbb{R}]$, $v_0 \leq w_0$ on \mathbb{T} , and satisfy the inequalities

$$v_0(t) \leq h(t) + \int_0^t K_1(t, s, w_0(s)) ds + \int_0^t K_2(t, s, v_0(s)) ds, \quad t \in \mathbb{T},$$

and

$$w_0(t) \geq h(t) + \int_0^t K_1(t, s, v_0(s)) ds + \int_0^t K_2(t, s, w_0(s)) ds, \quad t \in \mathbb{T}.$$

Theorem 8

Under the hypothesis (H₃) and (H₄), there exist monotone sequences $\{v_n\}$ and $\{w_n\}$ in Ω_V converging uniformly and quadratically to the unique solution u of (1.4) on \mathbb{T} .

Example 8

Consider once again the Volterra integral equation (1.11) in Example 5. For $n = 1, 2, 3, \dots$, and the coupled lower solutions $v_0(t) \equiv 0$, $w_0(t) = \frac{t^3}{3}$, define the iterative schemes

$$v_n(t) = \int_0^t R_1(t, s, w_{n-1}(s); w_n(s)) ds, \quad t \in \mathbb{T},$$

and

$$w_n(t) = \int_0^t R_2(t, s, w_{n-1}(s), v_{n-1}(s); v_n(s)) ds, \quad t \in \mathbb{T},$$

where

$$\begin{aligned} R_1(t, s, w_{n-1}(s); w_n(s)) &= K(t, s, w_{n-1}(s)) + \\ &\quad K_u(t, s, w_{n-1}(s))(w_n(s) - w_{n-1}(s)), \end{aligned}$$

and

$$\begin{aligned} R_2(t, s, w_{n-1}(s), v_{n-1}(s); v_n(s)) &= K(t, s, v_{n-1}(s)) + \\ &\quad K_u(t, s, w_{n-1}(s))(v_n(s) - v_{n-1}(s)). \end{aligned}$$

REFERENCES

- [1] Agmon, S., Nirenberg, L., and Protter, M.H., A maximum principle for a class of hyperbolic equations and applications to equations of mixed elliptic-hyperbolic type, Comm. Pure Appl. Math., 6, 1953, 455-470.
- [2] Blakley, R.D. and Pandit, S.G., On a sharp linear comparison result and an application to nonlinear Cauchy problem, Dynamic. Syst. Appl., 3, 1994, 135-140.
- [3] Darboux, G., *Lecons Sur la Théorie Générale des Surfaces*, Vol. 2, Gauthier Villars, Paris, 1912.
- [4] Deo, S.G. and Pandit, S.G., Method of generalized quasilinearization for hyperbolic initial-boundary value problems, Nonl. World, 3, 1996, 267-275.
- [5] Ladde, G.S., Lakshmikantham, V., and Pachpatte, B.G., The method of upper, lower solutions and Volterra integral equations, J. Integral. Eqs., 4, 1982, 353-360.
- [6] Ladde, G.S., Lakshmikantham, V., and Vatsala, A.S., *Monotone Iterative Techniques for Nonlinear Differential Equations*, Pitman, Boston, 1985.
- [7] Lakshmikantham, V. and Pandit, S.G., The method of upper, lower solutions and hyperbolic partial differential equations, J. Math. Anal. Appl., 105, 1985, 466-477.
- [8] Marshall, R.G. and Pandit, S.G., Numerical analysis of monotone methods for nonlinear hyperbolic partial differential equations, in progress.
- [9] Marshall, R.G. and Pandit, S.G., Numerical analysis of monotone methods for nonlinear Volterra integral equations, in progress.
- [10] Pandit, S.G., Quadratically converging iterative schemes for nonlinear Volterra integral equations and an application, Jour. Appl. Math. Stoch. Anal., 10:2, 1997, 169-178.
- [11] Weil, H., On differential equations of the simplest boundary-layer theory problems, Ann. of Math., 43, 1942, 381-407.

18 KRONECKER PRODUCT OPERATIONS ON TENSORS

David W. Nicholson
University of Central Florida
Orlando, FL 32816

1. INTRODUCTION

Kronecker product algebra (KPA) is widely applied in control theory, signal processing, image processing and statistics. However, it does not appear to commonly applied to continuum mechanics. The reason may be that, in its current state of development, KPA applies to matrices but not to third and fourth order tensors which are pervasive in continuum mechanics. In broad terms the goal of the current investigation is to extend Kronecker product algebra to tensors. In particular, Kronecker counterparts of quadratic products and of tensor outer products are presented. Kronecker product operators on third and fourth order tensors are introduced. The tensorial nature of Kronecker products of tensors is established. Finally, conditions for symmetry classes in fourth order tensors are stated in terms of Kronecker products. In several related studies on continuum mechanics, the KPA extensions have been shown to furnish compact expressions for elaborate quantities such as the tangent modulus tensor in thermohyperelasticity.

In its current state of development, Kronecker Product Algebra (KPA) applies to matrices but not to third and fourth order tensors, which may explain its minimal use in continuum mechanics. The goal of several recent investigations by the author has to extend Kronecker Product Algebra to furnish additional classes of relations applicable to tensors and to apply the extensions to continuum mechanics. In particular, in several related studies on continuum mechanics [1-10] the KPA extensions were shown to furnish compact expressions for elaborate quantities such as the tangent modulus tensor in thermohyperelasticity. The intent here is to provide a formal, comprehensive and unified presentation of the KPA extensions, with

concise statements and proofs of the major results. Kronecker product operators for third and fourth order tensors are introduced. The tensorial nature of Kronecker products of tensors is identified. A Kronecker product counterpart of the quadratic product of a second order tensor is introduced. Kronecker counterparts of tensor outer products are derived. Finally, Kronecker product expressions for symmetry conditions in fourth order tensors are presented. All matrices and tensors introduced below are assumed to be real.

2. KRONECKER PRODUCT ALGEBRA

Definition 1. [1] The *VEC* Operator

Let A be an $n \times n$ (second order) tensor. Kronecker product algebra [1] reduces A to a first order $n^2 \times 1$ tensor (vector) as follows.

$$VEC(A) = \{a_{11} \ a_{21} \ a_{31} \dots a_{nn}\}^T. \quad (2.1)$$

The vectorial properties of $VEC(A)$ are explained in Section (3.3). The I th entry of $VEC(A)$ corresponds to the $i \ j^{th}$ entry of A , in which $I = (j-1)n + i$.

Definition 2. [12] The inverse *VEC* operator, *IVEC(.)* is introduced by the relation

$$IVEC(VEC(A)) = A. \quad (2.2)$$

Definition 3. [1] For an $n \times m$ real matrix A and an $r \times s$ real matrix B , the Right Kronecker product denoted by \times_R and the left Kronecker product denoted by \times_L generate $nr \times ms$ matrices as follows:

$$A \times_R B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & & & \\ \vdots & & & \\ a_{n1}B & & & a_{nm}B \end{bmatrix} \quad A \times_L B = \begin{bmatrix} Ab_{11} & Ab_{12} & \dots & Ab_{1s} \\ Ab_{21} & & & \\ \vdots & & & \\ Ab_{r1} & & & Ab_{rs} \end{bmatrix} \quad (2.3)$$

Proposition 1. The Right and Left Kronecker products satisfy

$$B \times_L A = A \times_R B. \quad (2.4)$$

Proof: Let $I = (i-1)n + j$ and $J = (k-1)r + l$. From the foregoing definitions the product $a_{ij}b_{kl}$ occupies the IJ^{th} position of $A \times_R B$. But simple manipulation serves to verify that the product $b_{kl}a_{ij}$ occupies the IJ^{th} position of $B \times_L A$.

By virtue of Proposition 1, only the Right Kronecker product will be used in the following sections, now denoted by \times_R . We will later show that, if m, n, r and s

are equal to n and if \mathbf{A} and \mathbf{B} are tensors, $\mathbf{A} \times_K \mathbf{B}$ transforms as a second order $n \times n$ tensor in a restricted sense to be explained in Section (3.3).

Equation (2.3) implies that the $n^2 \times 1$ Kronecker product of two $n \times 1$ vectors \mathbf{a} and \mathbf{b} is written as

$$\mathbf{a} \times_K \mathbf{b} = \begin{pmatrix} \mathbf{a}_1 \mathbf{b} \\ \mathbf{a}_2 \mathbf{b} \\ \vdots \\ \mathbf{a}_n \mathbf{b} \end{pmatrix}. \quad (2.5)$$

Propositions (2-6) are introduced, together with a number of corollaries. The propositions are so named because they are proved by entry-wise arguments following the effect of Kronecker Product operators on the positions of matrix entries. However, the corollaries for the most part are proved using the propositions. Propositions (2-6) provide the fundamental relations of Kronecker Product Algebra in that their use suffices to derive the currently established properties of Kronecker products. The proofs of the first five propositions are presented in Ref [11].

Proposition 2. [11] Let \mathbf{A} denote an $n \times m$ real matrix with entry a_{ij} in the i -th row and j -th column. With I and J as defined above, let \mathbf{U}_v denote the $nm \times nm$ matrix, independent of \mathbf{A} , satisfying

$$\mathbf{u}_{JK} = \begin{cases} 1, & K = I \\ 0, & K \neq I \end{cases}$$

Then

$$\text{VEC}(\mathbf{A}^T) = \mathbf{U}_v \text{VEC}(\mathbf{A}). \quad (2.6)$$

Corollary 1. \mathbf{U}_v is symmetric if \mathbf{A} is square ($m = n$).

Proof: Note that $u_{JK} = u_{JI} = 1$ and $u_{IK} = u_{IJ} = 1$, with all other entries vanishing. If $m = n$, it follows that $u_{JI} = u_{IJ}$.

Proposition 3. [11] If \mathbf{A} and \mathbf{B} are second order $n \times n$ tensors, then

$$\text{TRACE}(\mathbf{AB}) = \text{VEC}^T(\mathbf{A}^T) \text{VEC}(\mathbf{B}). \quad (2.7)$$

Proposition 4. [11] If \mathbf{I}_n denotes the $n \times n$ identity matrix and if \mathbf{B} denotes an $n \times n$ tensor, then

$$\mathbf{I}_n \times_K \mathbf{B}^T = (\mathbf{I}_n \times_K \mathbf{B})^T. \quad (2.8)$$

Proposition 5. [11] Let A, B, C and D respectively denote $m \times n, r \times s, n \times p$, and $s \times q$ matrices. Then

$$(A \times_K B)(C \times_K D) = AC \times_K BD. \quad (2.9)$$

Proposition 6. [12] If A, B and C are $n \times m, m \times r$ and $r \times s$ matrices, then

$$VEC(ACB) = (B^T \times_K A) VEC(C). \quad (2.10a)$$

Clearly,

$$\text{if } A = I_n \text{ then } VEC(CB) = B^T \times_K I_n VEC(C) \quad (2.10b)$$

$$\text{if } B = I_n \text{ then } VEC(AC) = I_n \times_K A VEC(C). \quad (2.10c)$$

Corollary 2. The permutation matrix U_v satisfies

$$U_v = U_v^T = U_v^{-1}; \quad U_v^2 = I_v. \quad (2.11)$$

Proof Using KPA: Symmetry was established in Corollary 1. Note that

$$\begin{aligned} VEC(A) &= U_v VEC(A^T) \\ &= U_v^2 VEC(A) \end{aligned}$$

and hence,

$$U_v^2 = I_v \quad U_v = U_v^T = U_v^{-1}.$$

U_v is hereafter called the permutation tensor for $n \times n$ matrices. If A is symmetric, $(U_v - I_v) VEC(A) = \mathbf{0}$. If A is antisymmetric, $(U_v + I_v) VEC(A) = \mathbf{0}$.

Corollary 3: If A and B are second order $n \times n$ tensors, then

$$TRACE(AB) = TRACE(BA). \quad (2.12)$$

Proof using KPA:

$$\begin{aligned} TRACE(AB) &= VEC^T(B^T) VEC(A) \\ &= VEC^T(B) U_v VEC(A) \\ &= [U_v^T VEC(B)]^T VEC(A) \\ &= [U_v^T VEC(B)]^T VEC(A) \\ &= VEC^T(B^T) VEC(B) \\ &= TRACE(BA). \end{aligned}$$

thereby recovering a well-known relation.

Corollary 4: If I_n is the $n \times n$ identity tensor and $i_n = VEC(I_n)$ then

$$VEC(A) = I_n \times_K A i_n. \quad (2.13)$$

Proof: $\text{VEC}(A) = \text{VEC}(AI_n)$ and from Proposition (6) $\text{VEC}(AI_n) = (I_n \times_K A) i_n$.

Corollary 5: If I_v is the identity tensor in n^2 -dimensional space then

$$I_n \times_K I_n = I_v. \quad (2.14)$$

Proof using KPA: From Proposition (6) and Corollary (4),

$$\text{VEC}(I_n) = \text{VEC}(I_n I_n) = (I_n \times_K I_n) \text{VEC}(I_n),$$

from which $i_n = I_n \times_K I_n i_n$. But $i_n = I_v i_n$ and hence, $I_n \times_K I_n = I_v$.

Corollary 6: If A and B denote $n \times n$ tensors, then

$$B \times_K A = U_v (A \times_K B) U_v. \quad (2.15)$$

Proof using KPA: From Propositions (5 and 6)

$$\begin{aligned} \text{VEC}(ACB^T) &= (I_n \times_K A) \text{VEC}(CB^T) \\ &= (I_n \times_K A)(B \times_K I_n) \text{VEC}(C) \\ &= (B \times_K A) \text{VEC}(C). \end{aligned}$$

But by a parallel argument using Proposition (2)

$$\begin{aligned} \text{VEC}((ACB^T)^T) &= \text{VEC}(BC^T A^T) \\ &= (A \times_K B) \text{VEC}(C^T) \\ &= (A \times_K B) U_v \text{VEC}(C). \end{aligned}$$

But Proposition (2) also implies that

$$\text{VEC}((ACB^T)^T) = U_n \text{VEC}(ACB^T).$$

Consequently, if C is arbitrary

$$U_v (B \times_K A) \text{VEC}(C) = (A \times_K B) U_v \text{VEC}(C)$$

and hence, upon using the relation, $U_v = U_v^{-1}$, $B \times_K A = U_v (A \times_K B) U_v$.

Corollary 7: If A and B are nonsingular $n \times n$ tensors, then

$$[A \times_K B]^{-1} = A^{-1} \times_K B^{-1}. \quad (2.16)$$

Proof using KPA: From Proposition (2),

$$\begin{aligned} (A \times_K B)(A^{-1} \times_K B^{-1}) &= AA^{-1} \times_K BB^{-1} \\ &= I_n \times_K I_n \\ &= I_v. \end{aligned}$$

Definition 3: Kronecker sum and Kronecker difference.

The Kronecker sum denoted by the operator $+_K$ and the Kronecker difference denoted by the operator $-_K$ are defined as follows:

$$\mathbf{A} +_K \mathbf{B} = \mathbf{A} \times_K \mathbf{I}_n + \mathbf{I}_n \times_K \mathbf{B} \quad \mathbf{A} -_K \mathbf{B} = \mathbf{A} \times_K \mathbf{I}_n - \mathbf{I}_n \times_K \mathbf{B}. \quad (2.17)$$

Corollary 8: Let α_j and β_k denote the eigenvalues of \mathbf{A} and \mathbf{B} , and let \mathbf{y}_j and \mathbf{z}_k denote the corresponding eigenvectors. The Kronecker product, sum and difference have the following eigenstructures.

expression	jk^{th} eigenvalue	jk^{th} eigenvector	
$\mathbf{A} \times_K \mathbf{B}$	$\alpha_j \beta_k$	$\mathbf{y}_j \times_K \mathbf{z}_k$	
$\mathbf{A} +_K \mathbf{B}$	$\alpha_j + \beta_k$	$\mathbf{y}_j \times_K \mathbf{z}_k$	(2.18)
$\mathbf{A} -_K \mathbf{B}$	$\alpha_j - \beta_k$	$\mathbf{y}_j \times_K \mathbf{z}_k$.	

Proof using KPA: From Proposition (6),

$$\begin{aligned} \alpha_j \beta_k \mathbf{y}_j \times_K \mathbf{z}_k &= (\alpha_j \mathbf{y}_j) \times_K (\beta_k \mathbf{z}_k) \\ &= \mathbf{A}\mathbf{y}_j \times_K \mathbf{B}\mathbf{z}_k \\ &= (\mathbf{A} \times_K \mathbf{B})(\mathbf{y}_j \times_K \mathbf{z}_k). \end{aligned}$$

Hence the jk^{th} eigenvalue of $\mathbf{A}\mathbf{I}_n$ is $\alpha_j \times 1$ while the jk^{th} eigenvector is $\mathbf{y}_j \times_K \mathbf{w}_k$ in which \mathbf{w}_k is an arbitrary unit vector (eigenvector of \mathbf{I}_n). The counterparts for $\mathbf{I}_n \times_K \mathbf{B}$ are $1 \times \beta_k$ and $\mathbf{v}_j \times_K \mathbf{z}_k$ in which \mathbf{v}_j is an arbitrary eigenvector of \mathbf{I}_n . Upon selecting $\mathbf{w}_k = \mathbf{z}_k$ and $\mathbf{v}_j = \mathbf{y}_j$ the Kronecker sum is seen to possess eigenvalues $\alpha_j + \beta_k = \alpha_j \times 1 + 1 \times \beta_k$ and the eigenvectors $\mathbf{y}_j \times_K \mathbf{z}_k$.

Further, the Kronecker sum and difference of two $n \times n$ tensors are $n^2 \times n^2$ tensors in a restricted sense explained in Section (3.3) below.

3. EXTENSIONS OF KRONECKER PRODUCT ALGEBRA

3.1 Kronecker Form of Quadratic Products

Proposition 7: [7] If \mathbf{a} and \mathbf{b} are $n \times 1$ vectors, then

$$\mathbf{a} \times_K \mathbf{b} = VEC\left(\left[\mathbf{ab}^T\right]^T\right). \quad (3.1)$$

Proof: If $I = (j-1)n + i$, the I^{th} entry of $VEC(\mathbf{ba}^T)$ is $b_i a_j$ which is also the I^{th} entry of $\mathbf{a} \times_K \mathbf{b}$. Hence $\mathbf{a} \times_K \mathbf{b} = VEC(\mathbf{ba}^T) = VEC\left(\left[\mathbf{ab}^T\right]^T\right)$.

Corollary 9: [3] Let \mathbf{R} be a second order $n \times n$ tensor. Then, letting $\mathbf{r} = VEC(\mathbf{R})$,

$$\mathbf{a}^T \mathbf{R} \mathbf{b} = \mathbf{b}^T \times_K \mathbf{a}^T \mathbf{r}. \quad (3.2)$$

Proof using KPA: From Propositions (3 and 7)

$$\begin{aligned}\mathbf{a}^T \mathbf{R} \mathbf{b} &= \text{TRACE}(\mathbf{R} \mathbf{b} \mathbf{a}^T) \\ &= \text{TRACE}(\mathbf{b} \mathbf{a}^T \mathbf{R}) \\ &= \text{VEC}^T(\mathbf{b} \mathbf{a}^T)^T \text{VEC}(\mathbf{R}) \\ &= \text{VEC}^T(\mathbf{a} \mathbf{b}^T) \text{VEC}(\mathbf{R}) \\ &= (\mathbf{b}^T \times_K \mathbf{a}^T) \mathbf{r}.\end{aligned}$$

Alternatively, note that, if $\mathbf{d} = \mathbf{R} \mathbf{b}$ and \mathbf{R} is an $n \times n$ tensor, then from Equations (10.2 and 10.3)

$$\begin{aligned}\mathbf{d} &= \text{VEC}(\mathbf{d}) \\ &= \mathbf{b}^T \times_K \mathbf{I}_n \text{VEC}(\mathbf{R}),\end{aligned}$$

and

$$\begin{aligned}\mathbf{a}^T \mathbf{R} \mathbf{b} &= \mathbf{a}^T \mathbf{d} \\ &= (\mathbf{I}_n \times_K \mathbf{a}^T) \mathbf{d} \\ &= (\mathbf{I}_n \times_K \mathbf{a}^T) (\mathbf{b}^T \times_K \mathbf{I}_n) \text{VEC}(\mathbf{R}) \\ &= (\mathbf{b}^T \times_K \mathbf{a}^T) \mathbf{r}.\end{aligned}$$

3.2 Kronecker Operations for Third and Fourth Order Tensors

Kronecker product algebra appears to have been developed primarily for matrices (arrays of numbers in rows and columns) [11]. To extend it to third and fourth order tensors we have introduced the $\text{TEN22}(\cdot)$ operator [9], and the $\text{TEN21}(\cdot)$ and $\text{TEN12}(\cdot)$ operators [10], as follows. Let \mathbf{A} and \mathbf{B} be second order $n \times n$ tensors and let \mathbf{C} be a fourth order $n \times n \times n \times n$ tensor with entries c_{ijkl} . Suppose that

$$a_{ij} = c_{ijkl} b_{kl} \quad (3.3)$$

in which the range of i, j, k and l is (l, n) .

Definition 4: [9] The TEN22 Operator.

Under the stated conditions on \mathbf{A} , \mathbf{B} and \mathbf{C} , the unique $n^2 \times n^2$ matrix $\text{TEN22}(\mathbf{C})$ is introduced implicitly using

$$\text{VEC}(\mathbf{A}) = \text{TEN22}(\mathbf{C}) \text{VEC}(\mathbf{B}). \quad (3.4)$$

The KL^{th} entry of $\text{TEN22}(\mathbf{C})$ corresponds to c_{ijkl} , in which $K = (j-1)n + i$ and $L = (l-1)n + k$.

Proposition 8: [12] Under the stated conditions on \mathbf{A} , \mathbf{B} and \mathbf{C} ,

$$\text{TEN22}(\mathbf{ACB}) = \mathbf{I}_n \times_K \mathbf{A} \text{TEN22}(\mathbf{C}) \mathbf{I}_n \times_K \mathbf{B}. \quad (3.5)$$

Proof using KPA: Let \mathbf{D} denote an $n \times n$ tensor. Then, using Equation (3.4) and Proposition (6),

$$\begin{aligned} \text{TEN22}(\mathbf{ACB}) \text{VEC}(\mathbf{D}) &= \text{VEC}(\mathbf{ACBD}) \\ &= (\mathbf{I}_n \times_K \mathbf{A}) \text{VEC}(\mathbf{CBD}) \\ &= (\mathbf{I}_n \times_K \mathbf{A}) \text{TEN22}(\mathbf{C}) \text{VEC}(\mathbf{BD}) \\ &= (\mathbf{I}_n \times_K \mathbf{A}) \text{TEN22}(\mathbf{C}) (\mathbf{I}_n \times_K \mathbf{B}) \text{VEC}(\mathbf{D}). \end{aligned}$$

Corollary 10: [12] If \mathbf{C} is a fourth order nonsingular $n \times n \times n \times n$ tensor, then

$$\text{TEN22}(\mathbf{C}^{-1}) = \text{TEN22}^{-1}(\mathbf{C}). \quad (3.6)$$

Proof using KPA: Writing $\mathbf{B} = \mathbf{CA}$, it is immediate that

$$\text{VEC}(\mathbf{B}) = \text{TEN22}(\mathbf{C}^{-1}) \text{VEC}(\mathbf{A}).$$

But $\text{TEN22}(\mathbf{C}) \text{VEC}(\mathbf{B}) = \text{VEC}(\mathbf{A})$ and hence, $\text{VEC}(\mathbf{B}) = [\text{TEN22}(\mathbf{C})]^{-1} \text{VEC}(\mathbf{A})$, establishing Equation (3.6).

Corollary 11: [12] Let $\hat{\mathbf{C}}$ denote the fourth order $n \times n \times n \times n$ tensor whose $ijkl^{\text{th}}$ entry is c_{jilk} . Then

$$\text{TEN22}(\hat{\mathbf{C}}) = \mathbf{U}_v \text{TEN22}(\mathbf{C}) \mathbf{U}_v. \quad (3.7)$$

Proof using KPA: Writing $\mathbf{A}^T = \hat{\mathbf{C}}\mathbf{B}$, it is immediate from Proposition 1 that

$$\mathbf{U}_v \mathbf{a} = \text{TEN22}(\hat{\mathbf{C}}) \mathbf{U}_v \mathbf{b}.$$

Equation (3.7) follows from application of Corollary (2).

Definition 4: [12] *ITEN22* operator.

The inverse of the TEN22 operator is introduced using the relation

$$\text{ITEN22}(\text{TEN22}(\mathbf{C})) = \mathbf{C}. \quad (3.8)$$

Let \mathbf{A} be a second order $n \times n$ tensor and let \mathbf{b} be an $n \times 1$ vector. The $n \times n \times n$ third order tensor \mathbf{C}_{21} is introduced by the relation

$$a_{ij} = c_{ijk} b_k \quad \mathbf{A} = \mathbf{C}_{21} \cdot \mathbf{b}. \quad (3.9)$$

Definition 5: [10] $TEN21$ operator.

For A , b and C as defined above, the $TEN21$ operator on the third order tensor C_{21} is defined by the relation

$$VEC(A) = TEN21(C)b. \quad (3.10)$$

An example of an application of $TEN21(\cdot)$ is the tensor of Christoffel symbols of the second kind.

Now suppose that a is an $n \times 1$ vector and that B is an $n \times n$ second order tensor. The third order tensor C_{12} is introduced using the relation

$$a_i = c_{ijk} b_j k \quad a = C_{12} \cdot B \quad (3.11)$$

in which B is a second order $n \times n$ tensor.

Definition 6: [10] $TEN12$ operator.

The operator $TEN12(\cdot)$ on the third order tensor C_{12} is defined implicitly by

$$a = TEN12(C_{12}) VEC(B). \quad (3.12)$$

An example of the application of $TEN12(\cdot)$ is the permutation tensor ϵ_{ijk} . Properties of $TEN21$ and $TEN12$ which are counterparts of Proposition (8) and Corollaries (9) and (10) are readily derived.

3.3 Transformation Properties of VEC , $TEN22$, $TEN12$, and $TEN21$

Suppose that A and B are real second order $n \times n$ tensors and C is a fourth order $n \times n \times n \times n$ tensor such that $A = CB$. All are referred to a coordinate system denoted as Y . Let the unitary matrix (tensor) Q_n represent a rotation which gives rise to a coordinate system Y' . Let A' , B' , and C' denote the counterparts of A , B and C . Now, since $A' = Q_n A Q_n$ from Proposition (6)

$$VEC(A) = (Q_n \times_K Q_n) VEC(A). \quad (3.13)$$

Proposition 9: [10] For Q defined in the foregoing paragraph, the $n^2 \times n^2$ matrix $Q_n \times_K Q_n$ satisfies

$$(Q_n \times_K Q_n)^T = (Q_n \times_K Q_n)^{-1}.$$

Proof using KPA: Note that, since $Q_n^T = Q_n^{-1}$,

$$\begin{aligned} (Q_n \times_K Q_n)^T &= Q_n^T \times_K Q_n^T \\ &= Q_n^{-1} \times_K Q_n^{-1} \\ &= (Q_n \times_K Q_n)^{-1}. \end{aligned} \quad (3.14)$$

$Q_n \times_K Q_n$ may be shown to be a tensor in n^2 -dimensional space. In view of Equation (3.14) it is orthogonal. However, $Q_n \times_K Q_n$ represents a subset of the set

of all orthogonal tensors in n -dimensional space. It follows that $\text{VEC}(A)$ transforms as an $n^2 \times 1$ vector under orthogonal transformations of the form $\mathcal{Q}_n \times_K \mathcal{Q}_n$, but not under the full class of orthogonal transformations in n^2 -dimensional space.

Corollary 11: [10] The $\text{TEN}22$ operator transforms as

$$\text{TEN}22(\mathbf{C}) = (\mathcal{Q}_n \times_K \mathcal{Q}_n) \text{TEN}22(\mathbf{C})(\mathcal{Q}_n \times_K \mathcal{Q}_n) \quad (3.15)$$

Proof using KPA: Now $A' = C'B'$. Invoking Propositions (8 and 9) furnishes

$$(\mathcal{Q}_n \times_K \mathcal{Q}_n) \text{VEC}(A) = \text{TEN}22(\mathbf{C})(\mathcal{Q}_n \times_K \mathcal{Q}_n) \text{VEC}(\mathbf{B}).$$

Upon multiplication using $\mathcal{Q} \times_K \mathcal{Q}$ and recalling the definition of $\text{TEN}22$, we conclude that

$$\text{TEN}22(\mathbf{C}) = (\mathcal{Q}_n \times_K \mathcal{Q}_n) \text{TEN}22(\mathbf{C})(\mathcal{Q}_n \times_K \mathcal{Q}_n)^T.$$

Consequently,

$$\begin{aligned} (\mathcal{Q}_n \times \mathcal{Q}_n) \text{TEN}22(\mathbf{C}) \mathcal{Q}_n \times \mathcal{Q}_n &= (\mathcal{Q}_n \times \mathcal{Q}_n)^T (\mathcal{Q}_n \times \mathcal{Q}_n) \text{TEN}22(\mathbf{C}') (\mathcal{Q}_n \times \mathcal{Q}_n) \\ &= \text{TEN}22(\mathbf{C}'). \end{aligned}$$

Clearly $\text{TEN}22(\mathbf{C})$ transforms as a second order $n^2 \times n^2$ tensor under the restricted class of orthogonal transformations represented by $\mathcal{Q}_n \times_K \mathcal{Q}_n$. Finally, we state the following corollaries without proof [10]:

Corollary 12:

$$\text{TEN}21(\mathbf{C}'_{21}) = (\mathcal{Q}_n \times_K \mathcal{Q}_n) \text{TEN}21(\mathbf{C}_{21}) \mathcal{Q}_n. \quad (3.16)$$

Corollary 13:

$$\text{TEN}12(\mathbf{C}'_{12}) = \mathcal{Q}_n \text{TEN}12(\mathbf{C}_{12})(\mathcal{Q}_n \times_K \mathcal{Q}_n) \quad (3.17)$$

and we say that $\text{TEN}21$ and $\text{TEN}12$ transform as tensors of order $1+1/2$ and $1/2+1$, respectively, under orthogonal transformations of the form \mathcal{Q} in n -space and $(\mathcal{Q} \times_K \mathcal{Q})$ in n^2 -space.

3.4 KPA Expressions for Tensor Outer Products

Let \mathbf{A} and \mathbf{B} be two nonsingular $n \times n$ real second order tensors with entries a_{ij} and b_{ij} ; let $\mathbf{a} = \text{VEC}(\mathbf{A})$ and $\mathbf{b} = \text{VEC}(\mathbf{B})$. There are twenty four permutations of the indices $ijkl$ corresponding to outer products of \mathbf{A} and \mathbf{B} . Recalling the definitions of the (right) Kronecker product, we introduce three basic Kronecker product functions:

Definition 7: [10] The Kronecker product functions C_1, C_2 and C_3 of A and B are defined as follows:

$$C_1(A, B) = ab^T \quad C_2(A, B) = A \times_K B \quad C_2(A, B) = (A \times_K B) U_v. \quad (3.18)$$

Proposition 10: [10] The twenty four tensor outer products can be expressed in terms of the Kronecker product functions C_1, C_2 and C_3 as follows:

$$\begin{aligned} TEN22(a_{ij}b_{kl}) &= C_1(A, B) & TEN22(b_{ij}a_{kl}) &= C_1(B, A) \\ TEN22(a_{ji}b_{kl}) &= C_1(A^T, B) & TEN22(b_{ji}a_{kl}) &= C_1(B^T, A) \\ TEN22(a_{ij}b_{lk}) &= C_1(A, B^T) & TEN22(b_{ij}a_{lk}) &= C_1(B, A^T) \\ TEN22(a_{ji}b_{lk}) &= C_1(A^T, B^T) & TEN22(b_{ji}a_{lk}) &= C_1(B^T, A^T) \\ \\ TEN22(a_{ik}b_{jl}) &= C_2(B, A) & TEN22(b_{ki}a_{jl}) &= C_2(A, B) \\ TEN22(a_{kl}b_{jl}) &= C_2(B, A^T) & TEN22(b_{ik}a_{lj}) &= C_2(A, B^T) \\ TEN22(a_{ik}b_{lj}) &= C_2(B^T, A) & TEN22(b_{ki}a_{lj}) &= C_2(A^T, B) \\ TEN22(a_{ki}b_{lj}) &= C_2(B^T, A^T) & TEN22(b_{ki}a_{lj}) &= C_2(A^T, B^T) \\ \\ TEN22(a_{il}b_{jk}) &= C_3(B, A) & TEN22(b_{il}a_{jk}) &= C_3(A, B) \\ TEN22(a_{li}b_{jk}) &= C_3(B, A^T) & TEN22(b_{li}a_{jk}) &= C_3(A, B^T) \\ TEN22(a_{il}b_{kj}) &= C_3(B^T, A) & TEN22(b_{il}a_{kj}) &= C_3(A^T, B) \\ TEN22(a_{li}b_{kj}) &= C_3(B^T, A^T) & TEN22(b_{li}a_{kj}) &= C_3(A^T, B^T) \end{aligned} \quad (3.19)$$

Proof: The proof is now presented for three of the relations in Equation (3.19). The remainder of the relations may be derived by appropriately modifying the arguments of $TEN22$. We introduce the $n \times n$ tensors R and S with entries r_{ij} and s_{ij} . Also let $s = VEC(S)$ and $r = VEC(R)$.

(a). Suppose that $s_{ij} = a_{ij}b_{kl}r_{kl}$. But $a_{ij}b_{kl}r_{kl} = a_{ij}b_{lk}^Tr_{kl}$ in which b_{lk}^T is the $k l^{\text{th}}$ entry of B^T . It follows that $S = TRACE(B^TR)A$. Hence, letting $s = VEC(S)$,

$$\begin{aligned} s &= VEC^T((B^T)^T)VEC(R)a \\ &= VEC^T(B)VEC(R)a \\ &= ab^Tr \\ &= C_1(A, B). \end{aligned}$$

Since $s = TEN 22(\mathbf{a}_{ij}\mathbf{b}_{kl})\mathbf{r}$, it follows that $TEN 22(\mathbf{a}_{ij}\mathbf{b}_{kl}) = \mathbf{C}_1(\mathbf{A}, \mathbf{B})$.

b. Suppose that $s_{ij} = \mathbf{a}_{ik}\mathbf{b}_{jl}\mathbf{r}_{kl}$. But $\mathbf{a}_{ik}\mathbf{b}_{jl}\mathbf{r}_{kl} = \mathbf{a}_{ik}\mathbf{r}_k\mathbf{b}_{lj}^T$, and hence, $\mathbf{S} = \mathbf{A}\mathbf{R}^T\mathbf{B}^T$.

Now

$$\begin{aligned}s &= VEC^T(\mathbf{ARB}^T) \\&= (\mathbf{I}_n \times_K \mathbf{A}) VEC(\mathbf{RB}^T) \\&= (\mathbf{I}_n \times_K \mathbf{A})(\mathbf{B} \times_K \mathbf{I}_n)\mathbf{r} \\&= (\mathbf{B} \times_K \mathbf{A})\mathbf{r} \\&= \mathbf{C}_2(\mathbf{B}, \mathbf{A}).\end{aligned}$$

c. Suppose $s_{ij} = \mathbf{a}_{il}\mathbf{b}_{jk}\mathbf{r}_{kl}$. But $\mathbf{a}_{il}\mathbf{b}_{jk}\mathbf{r}_{kl} = \mathbf{a}_{il}\mathbf{r}_{lk}^T\mathbf{b}_{kj}$, and hence, $\mathbf{S} = \mathbf{A}\mathbf{R}^T\mathbf{B}^T$.

Now

$$\begin{aligned}s &= VEC(\mathbf{AR}^T\mathbf{B}^T) \\&= (\mathbf{I}_n \times_K \mathbf{A})(\mathbf{B} \times_K \mathbf{I}_n) VEC(\mathbf{R}^T) \\&= (\mathbf{B} \times_K \mathbf{A}) \mathbf{U}_v VEC(\mathbf{R}) \\&= \mathbf{C}_3(\mathbf{B}, \mathbf{A}).\end{aligned}$$

3.5 Symmetry Classes for Fourth Order Tensors

Again \mathbf{C} is a fourth order $n \times n \times n \times n$ tensor with entries c_{ijkl} , assumed for convenience to be nonsingular.

Definition 8: [12] Symmetry and total symmetry of fourth order tensors.

The 4th order $n \times n \times n \times n$ tensor \mathbf{C} is symmetric if its entries satisfy

$$c_{ijkl} = c_{klji}. \quad (3.20a)$$

\mathbf{C} is totally symmetric if it is symmetric and its entries also satisfy

$$c_{ijkl} = c_{jikl} \quad (3.20b)$$

$$c_{ijkl} = c_{ijlk}. \quad (3.20c)$$

Equation (3.20a) clearly expresses symmetry with respect to exchange of ij and kl in \mathbf{C} .

Proposition 11: [12] The fourth order $n \times n \times n \times n$ tensor \mathbf{C} is totally symmetric if and only if

$$TEN 22(\mathbf{C}) = TEN 22^T(\mathbf{C}) \quad (3.21a)$$

$$\mathbf{U}_v TEN 22(\mathbf{C}) = TEN 22(\mathbf{C}) \quad (3.21b)$$

$$TEN\ 22(\mathbf{C})\mathbf{U}_v = TEN\ 22(\mathbf{C}). \quad (3.21c)$$

Proof: (Using KPA in (b) and (c)):

(a) The $i\ j\ k\ l^{\text{th}}$ entry of \mathbf{C} , c_{ijkl} occupies the IJ^{th} entry of $TEN\ 22(\mathbf{C})$ in which $I = (i - l)n + j$ and $J = (k - 1)n + l$. In addition, the $klij^{\text{th}}$ entry of \mathbf{C} , c_{klij} , occupies the JI^{th} entry of $TEN\ 22(\mathbf{C})$. But, by virtue of symmetry, $c_{ijkl} = c_{klij}$ and hence, $TEN\ 22(\mathbf{C})|_{IJ} = TEN\ 22(\mathbf{C})|_{JI}$. Consequently Equation (3.21a) is equivalent to Equation (3.20a).

(b) Total symmetry (in particular Equation (3.20a)) implies that, for *any* second order $n \times n$ tensor \mathbf{B} , the corresponding $n \times n$ tensor $\mathbf{A} = \mathbf{CB}$ is symmetric. Hence, $VEC(\mathbf{A}) = U_v VEC(\mathbf{A})$. If $a = VEC(\mathbf{A})$ and $b = VEC(\mathbf{B})$,

$$\mathbf{a} = TEN\ 22(\mathbf{C})\mathbf{b} \quad (3.22a)$$

and also

$$\mathbf{U}_v \mathbf{a} = TEN\ 22(\mathbf{C})\mathbf{b}. \quad (3.22b)$$

Multiplying through the latter expression with \mathbf{U}_v implies that

$$\mathbf{a} = \mathbf{U}_v \ T EN\ 22(\mathbf{C})\mathbf{b}. \quad (3.22c)$$

Eliminating \mathbf{a} in Equations (3.22a) and (3.22c) furnishes Equation (3.21b), establishing its equivalence to Equation (3.12b).

From Equation (3.20c), for any $n \times n$ tensor \mathbf{A} , the tensor $\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}$ is symmetric. Using Corollary 9, it follows that

$$\mathbf{b} = TEN\ 22(\mathbf{C}^{-1})\mathbf{a} \quad (3.23a)$$

$$= (TEN\ 22(\mathbf{C}))^{-1} \mathbf{a} \quad (3.23b)$$

and also

$$\mathbf{U}_v \mathbf{b} = TEN\ 22(\mathbf{C}^{-1})\mathbf{a}. \quad (3.23c)$$

Thus,

$$TEN\ 22^{-1}(\mathbf{C}) = \mathbf{U}_v \ T EN\ 22^{-1}(\mathbf{C}).$$

Also $TEN\ 22(\mathbf{C}) = [\mathbf{U}_v \ T EN\ 22^{-1}(\mathbf{C})]^{-1} = TEN\ 22(\mathbf{C})\mathbf{U}_v$ establishing that Equation (3.21c) is equivalent to Equation (3.20c).

We now draw the immediate conclusion from (3.21b) and (3.21c) that if \mathbf{C} is totally symmetric

$$TEN\ 22(\mathbf{C}) = \mathbf{U}_v \ T EN\ 22(\mathbf{C})\mathbf{U}_v. \quad (3.23d)$$

We next prove the following.

Corollary 14: [12] For \mathbf{C} as defined above,

$$\mathbf{C}^{-1} \text{ is totally symmetric if and only if } \mathbf{C} \text{ is totally symmetric.} \quad (3.24)$$

Proof using KPA:

Note that $TEN22(\mathbf{C})\mathbf{U}_v = \mathbf{C}$ implies that $\mathbf{U}_v TEN22(\mathbf{C}^{-1}) = TEN22(\mathbf{C}^{-1})$ while $\mathbf{U}_v TEN22(\mathbf{C}) = TEN22(\mathbf{C})$ implies that $TEN22(\mathbf{C}^{-1})\mathbf{U}_v = TEN22(\mathbf{C}^{-1})$.

Corollary 15: [12] For \mathbf{C} as defined above, if \mathbf{G} is an $n \times n$ second order tensor, $\mathbf{G}\mathbf{C}\mathbf{G}^T$ is totally symmetric if \mathbf{C} is totally symmetric. (3.25)

Proof Using KPA:

First, Equation (3.5) implies that $TEN22(\mathbf{G}\mathbf{C}\mathbf{G}^T) = (\mathbf{I}_n \times_K \mathbf{G})TEN22(\mathbf{C})(\mathbf{I}_n \times_K \mathbf{G}^T)$, so that $TEN22(\mathbf{G}\mathbf{C}\mathbf{G}^T)$ is certainly symmetric if \mathbf{C} is totally symmetric. Next consider whether \mathbf{A}' given by

$$\mathbf{A}' = \mathbf{G}\mathbf{C}\mathbf{G}^T \mathbf{B}' \quad (3.26)$$

is symmetric in which \mathbf{B}' is a second order nonsingular $n \times n$ tensor. But we may write

$$\mathbf{G}^{-1}\mathbf{A}'\mathbf{G}^{-T} = \mathbf{C}\mathbf{G}^T \mathbf{B}'\mathbf{G}^{-T}. \quad (3.27)$$

We conclude that $\mathbf{G}^{-1}\mathbf{A}'\mathbf{G}^{-T}$ is symmetric since \mathbf{C} is totally symmetric, and therefore \mathbf{A}' is symmetric. Next consider whether \mathbf{B}' given by the following is symmetric

$$\mathbf{B}' = \mathbf{G}^{-T}\mathbf{C}^{-1}\mathbf{G}^{-1}\mathbf{A}'. \quad (3.28)$$

But we may write

$$\mathbf{G}^T \mathbf{B}' \mathbf{G} = \mathbf{C}^{-1}\mathbf{G}^{-1}\mathbf{A}'\mathbf{G}. \quad (3.29)$$

Since \mathbf{C}^{-1} is totally symmetric, it follows that $\mathbf{G}^T \mathbf{B}' \mathbf{G}$ is symmetric, and hence, \mathbf{B}' is symmetric. We conclude that $\mathbf{G}\mathbf{C}\mathbf{G}^T$ is totally symmetric.

4. CONCLUSION

Kronecker product algebra (KPA) is widely applied in control theory, signal processing, image processing and statistics. In broad terms the goal of the current investigation is to extend Kronecker product algebra to tensors, for application to continuum mechanics. The established properties of Kronecker product operators are summarized. Kronecker product operators on third and fourth order tensors are introduced, and several of their properties are established. Kronecker counterparts of quadratic products and of tensor outer products are presented. The tensor nature of Kronecker products of tensors and of the new operators is established. Finally, conditions for symmetry classes in fourth order tensors are stated in terms of Kronecker products. In several related studies on continuum mechanics, the KPA

extensions were shown to furnish compact expressions for elaborate quantities such as the tangent modulus tensor in thermohyperelasticity.

REFERENCES

- [1] Nicholson, D.W., Tangent stiffness matrix for finite element analysis of hyperelastic materials, *Acta Mech.*, 11, 1995, 187.
- [2] Nicholson, D.W. and Lin, B., Theory of Thermohyperelasticity for near-incompressible elastomers, *Acta Mech.*, 116, 1996, 15.
- [3] Nicholson, D.W. and Lin, B., Finite element method for thermomechanical response of near-incompressible elastomers, *Acta Mech.*, 124, 1997a, 181.
- [4] Nicholson, D.W. and Lin, B., Incremental finite element equations for thermomechanical contact of elastomers: Effect of boundary conditions including contact, *Acta Mech.*, 128, 1997b, 81.
- [5] Nicholson, D.W. and Lin, B., On the tangent modulus tensor in hyperelasticity, *Acta Mech.*, 131, 1997c, 121.
- [6] Lin, B., *Finite Element Methods for Thermohyperelastic Bodies Under Nonclassical Boundary Conditions*, Doctoral Dissertation, University of Central Florida, Orlando, Florida, 1996.
- [7] Nicholson, D.W. et. al, Finite element analysis of hyperelastic components, *Appl. Mech. Rev.*, 51, 1998, 303.
- [8] Nicholson, D.W. and Lin, B., Stable response of non-classically damped mechanical systems-II, *Appl. Mech. Rev.*, 49, 1996, S49.
- [9] Nicholson, D.W. and Lin, B., Tangent modulus tensor in plasticity under finite strain, *Acta Mech.*, 134, 1999, 199.
- [10] Nicholson, D.W. and Lin, B., Extensions of Kronecker product algebra with applications in continuum and computational mechanics, to appear in *Acta Mechanica*.
- [11] Graham, A., *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwood, Ltd., London, 1981.
- [12] Nicholson, D.W., On incremental equilibrium equations in solid mechanics, *Boundary Elements XX*, Computational Mechanics Press, A.J. Kassab and C.A. Brebbia, Eds., 1998.

19 GLOBAL BEHAVIOR OF SOLUTIONS OF A CERTAIN NTH ORDER DIFFERENTIAL EQUATION IN THE VICINITY OF AN IRREGULAR SINGULAR POINT

T.K. Puttaswamy
Department of Mathematical Sciences
Ball State University
Muncie, Indiana 47306

ABSTRACT

This paper is devoted to the global behavior of solutions of the n th order differential equation

$$z^n \left(a_n + b_n z^m \right) \frac{d^n y}{dz^n} + z^{n-1} \left(a_{n-1} + b_{n-1} z^m \right) \frac{d^{n-1} y}{dz^{n-1}} + \sum_{k=0}^{n-2} z_k \left(a_k + b_k z^m + c_k z^{2m} \right) \frac{d^k y}{dz^k} = 0.$$

Here, m is an arbitrary positive integer. The variable z and the constants $a_n, b_n, a_{n-1}, b_{n-1}$ and a_k, b_k, c_k ($k = 0, 1, 2, \dots, n-2$) are complex with $a_n \neq 0$, $b_n \neq 0$, $c_{n-2} \neq 0$. We shall also assume that the difference of no two roots of the indicial equation about the regular singular part $z = 0$ is congruent to zero modulo m .

1. INTRODUCTION

Analytic theory of linear differential equations in the extended complex plane has a local as well as a global aspect. The local part of the theory is satisfactorily developed, whereas global results which are exceedingly important, but extremely difficult are sparse and fragmentary. A typical global problem is the so called connection problem, the problem of decomposing a solution at a given point into a linear combination of fundamental solutions at some other point. To be specific, if

z_1 and z_2 are two distinct points and if $Q(z)$ is a solution at z_1 , how can we express $Q(z)$ as a linear combination of linearly independent solutions at z_2 ? If z_1 and z_2 are both ordinary point, the problem is simple. All that we have to do is to carry out analytic continuation. If z_1 or z_2 or both regular singular points, then the situation is totally different. Even for seemingly simple looking differential equations, the problem becomes very complicated. If in particular, z_1 or z_2 or both irregular singular points, then the degree of difficulty becomes multihold, for the neighborhood of an irregular singular point is not a homogeneous media with reference to the behavior of the solutions. In fact, the neighborhood breaks up into complimentary regions, usually sectorial in shape and in each of which the solutions will have a different asymptotic development. As a general rule, the degree of difficulty increases also with the number of singular points the differential equation has. Among the articles devoted to the general theory, we mention W.J. Tritzinsky [25] and [26] and H.L. Turrittin [27] and [28].

In 1936, Walter B. Ford [3] published several general theorems on asymptotic expansions, and then as an application of his theorems, tried to solve in the large the second order differential equation

$$\sum_{k=0}^2 z^k (a_k + b_k z + c_k z^2) \frac{d^k y}{dz^k} = 0$$

where the variable z is regarded as real or complex, as likewise the constants a_k, b_k, c_k ($k = 0, 1, 2$) with the restriction that the indicial exponents about the regular singular point $z = 0$ do not differ by an integer. A number of special cases came up and most of them Ford was able to handle by modification of his analysis. However, he was not entirely successful in all cases and was blocked completely in two cases:

Case I: The equation $a_2 + b_2 z + c_2 z^2 = 0$ has two equal roots $\mu \neq 0$ and $a_1 + b_1 \mu + c_1 \mu^2 \neq 0$.

Case II: $c_2 = b_2 = 0$, $a_2 \neq 0$ and $c_1 \neq 0$.

In Case I, Ford's differential equation will have two regular singular points at $z = 0$ and $z = \infty$ and an irregular singular point at a finite point $z = \mu$.

In Case II, Ford's differential equation will have a regular singular point at $z = 0$ and an irregular singular point of rank 2.

In order to succeed by Ford's method, one must be able to sum the asymptotic solutions of a certain second order different equation and express them as a convergent factorial series in a right half plane. T.K. Puttaswamy treated Ford's Case II. By using a paper by Harris and Culmer [2], in his master's thesis [8] written under the direction and guidance of late Professor H.L. Turrittin, he was able to do

that. He obtained the desired asymptotic expansions valid in the neighborhood of infinity using Ford's first general theorem [3, page 205]. But the sector did not cover completely the entire neighborhood of infinity. Four narrow sectors were left uncovered. But Later, he completed the problem using a paper of Wright [30] instead of Ford's first general theorem.

In Ford's Case I, despite efforts and advances in solving difference equations that have been made particularly by Culmer [1], Harris [4], [5], [6], and Sibuya [7], the asymptotic solutions of the second order difference equation cannot be summed by any known method.

Puttaswamy then considered the global behavior of the solutions of the n th order differential equation

$$\sum_{k=0}^n z^k (a_k + b_k z^m + c_k z^{2m}) \frac{d^k y}{dz^k} = 0,$$

where m is an arbitrary positive integer, the variable z and the constants a_k, b_k, c_k ($k = 0, 1, 2, \dots, n$) are complex. First he examined the case $n = 3$ and was able to solve the problem in all the cases except the case corresponding to Ford's Case I. Then, he treated the general case and has succeeded in completing the problem in most of the cases that arose. Some of these papers have been listed in the references at the end of this paper.

2. NATURE OF THE PRESENT PROBLEM

In order to introduce the investigation of this paper, let us take for consideration the linear homogeneous differential equation of n th order

$$\begin{aligned} z^n (a_n + b_n z^m) \frac{d^n y}{dz^n} + z^{n-1} (a_{n-1} + b_{n-1} z^m) \frac{d^{n-1} y}{dz^{n-1}} \\ + \sum_{k=0}^{n-2} z^k (a_k + b_k z^m + c_k z^{2m}) \frac{d^k y}{dz^k} = 0. \end{aligned} \quad (2.1)$$

Here, m is an arbitrary positive integer. The variable z and the constants $a_n, b_n, a_{n-1}, b_{n-1}$ and a_k, b_k, c_k ($k = 0, 1, 2, \dots, n-2$) are complex with $a_n \neq 0, b_n \neq 0$ and $c_{n-2} \neq 0$. If μ_k ($k = 1, 2, \dots, m$) are the roots of $a_n + b_n z^m = 0$, then in the language of Fuchs's theory (2.1) will have regular singular points at $z = 0, z = \mu_k$ ($k = 1, 2, \dots, m$) and an irregular singular point at $z = \infty$. In all, (2.1) will have $(m+2)$ singular points. Since m is an arbitrary positive integer, then (2.1) will have an arbitrary number of singular points.

If $\{h\}_k = h(h-1)(h-2)\dots(h-k+1) = \prod_{j=0}^{k-1} (h-j)$, $k \geq 1$ and $\{h\}_0 = 1$, then the

indicial equation about the regular singular point $z = 0$ is found to be

$$\sum_{k=0}^n a_k \{h\}_k = 0. \quad (2.2)$$

We shall also assume that the roots $h_i (i = 1, 2, \dots, n)$ of (2.2) are such that the difference of no two of them is congruent to zero modulom. Then the existing theory of linear differential equations assure us that (2.1) will have n linearly independent solutions $y_i(z), i = 1, 2, \dots, n$ about the regular singular point $z = 0$ of the form

$$y_i(z) = z^{h_i} \sum_{k=0}^{\infty} g_i(k) z^{mk} \quad (2.3)$$

with $g_i(0) = 1, i = 1, 2, \dots, n$.

Here, each $g_i(k), k = 1, 2, 3, \dots$ is a determinate function of k and each of these series solutions is convergent in a punctured circle drawn about $z = 0$ and extending up to the nearest of the singular points $z = \mu_k (k = 1, 2, \dots, m)$ and outside this circle, each becomes divergent. The behavior of these solutions $y_i(z), i = 1, 2, \dots, n$ in the neighborhood of $z = \infty$ is not available as far as Fuch's theory is concerned. It is to the study of the behavior of these solutions $y_i(z), i = 1, 2, \dots, n$ in the neighborhood of the irregular singular point $z = \infty$ that this paper has been addressed.

3. RECURRENCE RELATIONS SATISFIED BY EACH $g_i(k), i = 1, 2, \dots, n$

Substituting $y(z) = \sum_{k=0}^{\infty} g_i(k) z^{mk+h}$ into (2.1) and equating the coefficients of z^{mk+h+2} , we get

$$p_2(k) g(k+2) + p_1(k) g(k+1) + p_0(k) g(k) = 0 \quad (3.1)$$

where

$$p_2(k) = \sum_{i=0}^n a_i \{km + h + 2m\}_i \quad (3.2)$$

$$p_1(k) = \sum_{i=0}^n b_i \{km + h + m\}_i \quad (3.3)$$

and

$$p_0(k) = \sum_{i=0}^{n-2} c_i \{km + h\}_i. \quad (3.4)$$

We observe that $g(k), k = 1, 2, 3, \dots$ are completely and correctly determined by (3.1), if we arbitrarily take $g(-1) = 0$, as we do. Furthermore, we also make the assumption that when one of the $h_i (i = 1, 2, \dots, n)$ is used for h , no root of

$p_o(k-2) = 0$ is zero or a negative integer, in which case, we will have $g(-k) = 0$ for $k = 2, 3, \dots$. Since we wish to regard k as a complex variable, to emphasize this, we replace k by s .

Our first objective, then is to determine that analytic function $u(s)$ which is a solution of (3.1) and which takes on the initial values $u(0) = 1$ and $u(-1) = 0$.

4. FUNDAMENTAL SOLUTION OF (3.1)

Let

$$\begin{aligned} U_1(s) &= u(s) \\ U_2(s) &= u(s+1). \end{aligned} \quad (4.1)$$

Then (3.1) can be put in the matrix form

$$G(s+1) = A(s) G(s) \quad (4.2)$$

where, $G(s)$ is a 2×2 matrix which can be decomposed into two column vectors each with two components and $A(s)$ is a 2×2 matrix, representable as a power series in $\frac{1}{s}$ which is convergent in some neighborhood of $s = \infty$. Therefore, there is a finite s_0 such that the series

$$A(s) = A_0 + \frac{A_1}{s} + \frac{A_2}{s^2} + \dots \quad (4.3)$$

converges, if $|s| > s_0$, where

$$\begin{aligned} A_0 &= \begin{pmatrix} 0 & 1 \\ 0 & w_0 \end{pmatrix} \\ A_1 &= \begin{pmatrix} 0 & 0 \\ 0 & w_1 \end{pmatrix} \\ A_j &= \begin{pmatrix} 0 & 0 \\ \gamma_j & w_j \end{pmatrix}, \quad j = 2, 3, \dots \end{aligned} \quad (4.4)$$

We find that

$$\begin{aligned} w_0 &= -\frac{b_n}{a_n} \neq 0 \\ w_1 &= \frac{1}{m} \frac{b_n}{a_n} \left[mn - \frac{b_{n-1}}{b_n} - \frac{a_{n-1}}{a_n} \right] \\ \gamma_2 &= -\frac{c_{n-2}}{m^2 a_n} \neq 0 \\ \gamma_3 &= -\frac{c_{n-2}}{m^3 a_n} \left[\frac{c_{n-3}}{c_{n-2}} - \frac{a_{n-1}}{a_n} - (2d + 2mn) + (2n - 5) \right]. \end{aligned} \quad (4.5)$$

Using a paper by W.J.A. Culmer and W.A. Harris Jr., [2], we find that the two linearly independent solutions $u_1(s)$ and $u_2(s)$ of (3.1) are given in term of known factorial series as follows:

$$\begin{aligned} u_1(s) &= w_0^2 s^r \beta^s s^{\alpha-2s} [1 + \varepsilon_1(s)] \\ u_2(s) &= w_0^s s^r [1 + \varepsilon_2(s)] \end{aligned} \quad (4.6)$$

where

$$r = \frac{w_1}{w_0}$$

$$d = 1 + \frac{\gamma_3}{\gamma_2} - 2r$$

and

$$\beta = \frac{c^2 c_{n-2} a_n}{b_n^2}.$$

Here $u_1(s)$ and $u_2(s)$ are analytic at every point in some right half plane and as $|s| \rightarrow \infty$ in this right half plane $\varepsilon_i(s) \rightarrow 0$ ($i = 1, 2$).

5. MAIN RESULTS AND MAIN THEOREM

We define

$$f_0(k) = \sum_{i=0}^n a_i \{k\}_i \quad (5.1)$$

$$f_1(k) = \sum_{i=0}^n b_i \{k\}_i. \quad (5.2)$$

Then, we observe that in view of (2.2), $f_0(h) \equiv 0$.

Moreover from (3.2) and (3.3), we have

$$p_2(s) = f_0(m[s+2]+h) \quad (5.3)$$

$$p_1(s) = f_1(m[s+1]+h). \quad (5.4)$$

Again, in view of (2.2), $p_2(-2) = f_0(h) \equiv 0$ so 0 is a root of $p_2(s-2) = f_0(ms+h) = \sum_{i=0}^n a_i [ms+h]_i = 0$.

If β_i ($i = 1, 2, \dots, n-1$) are the other roots of $p_2(s-2) = 0$, we observe that none of the roots β_i ($i = 1, 2, \dots, n-1$) can be a positive integer. For, if $\beta_i = N$ ($N = 1, 2, \dots$), then $p_2(N-2) = f_0(mN+h) = 0$, which would imply that the difference of two roots of the indicial equation (2.2) about the regular singular point $z = 0$ is congruent to zero modulom, which contradicts hypothesis.

We further note that if none of β_i ($i = 1, 2, \dots, n-1$) is a positive integer, then none of the quantities $1 - \beta_i$ ($i = 1, 2, \dots, n-1$) can be zero or a negative integer.

The main theorem is the following:

If in the differential equation (2.1), $a_n \neq 0$, $b_n \neq 0$, $c_{n-2} \neq 0$ and if the indicial exponents h_i ($i = 1, 2, \dots, n$) about the regular singular point $z = 0$ and the roots k_i ($i = 1, 2, \dots, n-2$) of the equation $p_0\left(\frac{-s-h}{m}\right) = 0$ are such that the difference of no two of them is congruent to zero modulom, then the solutions $y_j(z)$ ($j = 1, 2, \dots, n$) about the regular singular point $z = 0$ as defined in (2.3), when $|z|$ is large and when z lies along any ray except for which $\arg z^m = \arg\left(\frac{1}{w_0}\right)$

may be expanded asymptotically in the form

$$y_j(z) \sim c_n(h, k_1, k_2, \dots, k_{n-2}) s_n(z) + \sum_{i=1}^{n-2} c_i(h, k_1, k_2, \dots, k_{n-2}) s_i(z) \quad (5.5)$$

in $\frac{\pi}{2} \leq \arg \sqrt{\mu z^m} < \frac{3\pi}{2}$

and

$$y_j(z) = \sum_{i=1}^{n-1} c_i(h, k_1, k_2, \dots, k_{n-2}) s_i(z) \text{ in } -\frac{\pi}{2} < \arg \sqrt{\mu z^m} < \frac{\pi}{2} \quad (5.6)$$

where $s_i(z)$, $i = 1, 2, \dots, n-2$ are given by

$$s_i(z) = \frac{1}{z^{k_i}} \left[1 + \frac{(\)}{z^m} + \frac{(\)}{z^{2m}} + \dots \right] \quad (5.7)$$

$$s_n(z) = z^{-\frac{m}{2}p} e^{-2\sqrt{\mu z^m}} \left[1 + \frac{(\)}{z^{\frac{m}{2}}} + \frac{(\)}{z^m} + \dots \right] \quad (5.8)$$

and

$$s_{n-1}(z) = z^{-\frac{m}{2}p} e^{2\sqrt{\mu z^m}} \left[1 + \frac{(\)}{z^{\frac{m}{2}}} + \frac{(\)}{z^m} + \dots \right] \quad (5.9)$$

with

$$p = \frac{w_1}{w_0} - \frac{\gamma_3}{\gamma_2} - \frac{3}{2} \quad (5.10)$$

$$\mu = w_0 \beta. \quad (5.11)$$

- (a) In case none of the quantities $\frac{h+k_i}{m}$, $i = 1, 2, \dots, n-2$ is an integer or if $\frac{h+k_i}{m}$ is an integer < 2 , the constants $c_n(h, k_1, k_2, \dots, k_{n-2})$ and $c_{n-1}(h, k_1, k_2, \dots, k_{n-2})$ are determined by

$$c_n(h, k_1, k_2, \dots, k_{n-2}) = (-1)^p \mu^{-\frac{p}{2}} u_2(-1) \frac{c_{n-1}}{m^2 a_n w_0} \frac{\prod_{k=1}^{n-2} \Gamma(1 - \alpha_k)}{\prod_{k=1}^{n-1} \Gamma(1 - \beta_k)} \quad (5.12)$$

$$c_{n-1}(h, k_1, k_2, \dots, k_{n-2}) = \mu^{-\frac{p}{2}} u_1(-1) \frac{c_{n-1}}{m^2 a_n w_0} \frac{\prod_{k=1}^{n-2} (1 - \alpha_k)}{\prod_{k=1}^{n-1} (1 - \beta_k)} \quad (5.13)$$

where as the constant $c_1(h, k_1, k_2, \dots, k_{n-2})$ is given by

$$c_1(h, k_1, k_2, \dots, k_{n-2}) = \frac{e^{\pi i \left(\frac{h+k_1}{m}\right)} \Gamma\left(2 - \frac{h+k_1}{m}\right) \prod_{k=1}^{n-1} \Gamma(1 - \beta_k) w\left(\frac{-h-k_1}{m}\right)}{\prod_{i=2}^{n-2} (k_i - k_1) \prod_{i=2}^{n-2} \Gamma\left(\frac{h+k_i}{m} - 1\right)} \quad (5.14)$$

with

$$w(s) = v(s+2) f_0(2m - k_1) + v(s+1) f_1(m - k_1) \quad (5.15)$$

and

$$v(s) = \begin{vmatrix} u_1(-1) & u_2(-1) \\ u_1(s) & u_2(s) \end{vmatrix}$$

and the constants $c_i(h, k_1, k_2, \dots, k_{n-2})$ for $i = 2, 3, \dots, n-2$ are obtained from (5.14) by interchanging k_1 and k_i .

(b) In case $\frac{h+k_1}{m} = N =$ an integer ≥ 2 , the constant $c_1(h, k_1, k_2, \dots, k_{n-2})$ is given

by

$$c_1(h, k_1, k_2, \dots, k_{n-2}) =$$

$$\frac{(-1)^{N-1} \prod_{k=1}^{n-1} \Gamma(1 - \beta_k) \left[f_0(2m - k_1) \frac{d}{ds} v(s+2) + f_1(m - k_1) \frac{d}{ds} v(s+1) \right]_{s=-N}}{\Gamma(N-1) \prod_{i=2}^{n-2} (k_i - k_1) \prod_{i=2}^{n-2} \Gamma\left(\frac{h+k_i}{m} - 1\right)}. \quad (5.16)$$

Proof of theorem will be given in Sections 6 and 7.

6. SOLUTIONS $y_i(z)$, $i = 1, 2, \dots, n$ ABOUT THE REGULAR SINGULAR POINT $z = 0$

By using well known asymptotic properties of gamma functions, the two fundamental solutions $u_1(s)$ and $u_2(s)$ of 3.1) as given by (4.6) can be written as

$$\begin{aligned} u_1(s) &= \mu^s G_1(s) \\ u_2(s) &= w_0^s G_2(s) \end{aligned} \quad (6.1)$$

where

$$\begin{aligned} G_1(s) &= \frac{2\pi}{\Gamma(s+k_1^*) \Gamma(s+k_2^*)} [1 + E_1(s)] \\ G_2(s) &= s^r [1 + E_2(s)] \end{aligned} \quad (6.2)$$

$$k_1^* + k_2^* = 1 - r - \alpha = r - \frac{\gamma_3}{\gamma_2} \quad (6.3)$$

$$\mu = w_0 \beta.$$

Here, the two solutions $u_1(s)$ and $u_2(s)$ are analytic in some right half plane and $E_i(s) \rightarrow 0$ for $i = 1, 2$ as $|s| \rightarrow \infty$ in this right half plane.

Let $\alpha_i (i = 1, 2, \dots, n-2)$ be the roots of $p_0(s) = 0$. By rendering precise meaning through certain convention and by analytic continuation by means of the different equation (3.1), when we write this equation in the form

$$u(s) = \frac{-[p_2(s) u(s+2) + p_1(s) u(s+1)]}{p_0(s)}. \quad (6.4)$$

We find that $u_1(s)$ and $u_2(s)$ as given by (6.1) are single valued and analytic throughout the finite s plane except for poles at the points

$$s = \alpha_i - j = \frac{-k_i - h}{m} - j, \quad i = 1, 2, \dots, n-2 \quad (6.5)$$

and $j = 0, 1, 2, \dots$, where $k_i = (i = 1, 2, \dots, n-2)$ are the roots of $p_0\left(\frac{s-h}{m}\right) = 0$.

Now that we have obtained the two linearly independent solutions $u_1(s)$ and $u_2(s)$ of (3.1), it must be possible to determine constants a and b such that

$$u(s) = a u_1(s) + b u_2(s) \quad (6.6)$$

will constitute the particular solution desired, that is one which will be single valued and analytic throughout the finite s -plane except for the poles (6.5) and which will be such that $u(0) = 1$ and $u(-1) = 0$, in which case if we assign the value $h_i (i = 1, 2, \dots, n)$ for h , we shall have $u(n) = g_i(n)$ for $n = 1, 2, \dots$ and $u(n) = 0$ for $n = -1, -2, \dots$

However, in order to obtain such a function $u(s)$, we shall assume that none of $\alpha_i (i = 1, 2, \dots, n-2)$ is an integer. The result of removing this restriction, we shall examine at a later point, but it must be observed that in part this restriction was made at the very outset, when we assumed that none of the roots of $p_0(s-2) = 0$ is zero or a negative integer.

The constants a and b are evidently to be determined from the equations

$$\begin{aligned} 0 &= a u_1(-1) + b u_2(-1) \\ 1 &= a u_1(0) + b u_2(0). \end{aligned} \quad (6.7)$$

If we let

$$D(s) = \begin{vmatrix} u_1(s) & u_2(s) \\ u_1(s+1) & u_2(s+1) \end{vmatrix}. \quad (6.8)$$

We find that

$$a = \frac{-u_2(-1)}{D(-1)} \text{ and } b = \frac{u_1(-1)}{D(-1)}. \quad (6.9)$$

Then, our next objective is to find the functional value of the determinant $D(s)$, known as the Casorati's determinant.

By Heyman's theorem, we know that $D(s)$ satisfies

$$\frac{D(s+1)}{D(s)} = \frac{p_0(s)}{p_2(s)} = \frac{c_{n-1} \prod_{k=1}^{n-2} (s - \alpha_k)}{m^2 a_n \prod_{k=1}^n (s - \beta_k + 2)}$$

where β_j , $j = 1, 2, \dots, n$ are the roots of $p_2(s-2) = 0$. Since 0 is a root of $p_2(s-2) = 0$, let β_n be this root, then,

$$D(s) = \frac{k^* \left(\frac{c_{n-1}}{m^2 a_n} \right)^s \Gamma(s - \alpha_k)}{\Gamma(s+2) \prod_{k=1}^{n-1} (s - \beta_k + 2)} \quad (6.10)$$

where $k^* = 2\pi w_0$.

Hence,

$$D(-1) = \frac{k^* m^2 a_n}{c_{n-1}} \frac{\prod_{k=1}^{n-2} \Gamma(-1 - \alpha_k)}{\prod_{k=1}^{n-1} \Gamma(1 - \beta_k)}. \quad (6.11)$$

Since none of the expressions $1 - \beta_k$, $k = 1, 2, \dots, n-1$ can be zero or a negative integer, $D(-1) \neq 0$. Again, the numerator in (6.11) is necessarily finite in as much as we have assumed that none of α_j , $j = 1, 2, \dots, n-2$ are integers.

Then, in view of (6.6), (6.9), and (6.11), we get

$$u(s) = \frac{v(s)}{D(-1)} \quad (6.12)$$

where

$$v(s) = \begin{vmatrix} u_1(-1) & u_2(-1) \\ u_1(s) & u_2(s) \end{vmatrix}. \quad (6.13)$$

It follows that each of the solutions $y_i(z)$, $i = 1, 2, \dots, n$ as given by (2.3) about the regular singular point $z = 0$ is found to be of the form

$$y(z) = z^h \left[\frac{-u_2(-1)}{D(-1)} \sum_{k=0}^{\infty} G_1(k) (\mu z^m)^k + \frac{u_1(-1)}{D(-1)} \sum_{k=0}^{\infty} G_2(k) (w_0 z^m)^k \right] \quad (6.14)$$

where $G_1(s)$ and $G_2(s)$ are given by (6.2) and (6.3) respectively.

7. ASYMPTOTIC BEHAVIOR OF SOLUTIONS $y_i(z)$, $i = 1, 2, \dots, n$ ABOUT THE REGULAR SINGULAR POINT $z = 0$

We apply W.B. Ford's theorem [3. Page 275] to the first series in (6.14) and the modified version of his first general theorem [11] to the second series in (6.14). Then, we find that the fundamental solution $y(z)$ about the regular singular point $z = 0$ as defined by (2.3), when $|z|$ is large and when z lies along any ray except for which $\arg z = \arg \left(\frac{1}{w_0} \right)$ may be developed asymptotically in the form

$$y(z) \sim c_n(h, k_1, k_2, \dots, k_{n-2}) s_n(z) - z^h R(z) \text{ in } \frac{\pi}{2} < \arg \sqrt{\mu z^m} < \frac{3\pi}{2} \quad (7.1)$$

and

$$y(z) \sim c_{n-1}(h, k_1, k_2, \dots, k_{n-2}) s_{n-1}(z) - z^h R(z) \text{ in } \frac{-\pi}{2} < \arg \sqrt{\mu z^m} < \frac{\pi}{2} \quad (7.2)$$

where

$$s_n(z) = z^{-\frac{mp}{2}} e^{-2\sqrt{\mu z^m}} \left[1 + \frac{\left(\frac{1}{z}\right)}{z^{\frac{m}{2}}} + \frac{\left(\frac{1}{z}\right)}{z^m} + \dots \right] \quad (7.3)$$

$$s_{n-1}(z) = z^{\frac{-mp}{2}} e^{2\sqrt{\mu z^m}} \left[1 + \frac{\left(\frac{1}{z}\right)}{z^{\frac{m}{2}}} + \frac{\left(\frac{1}{z}\right)}{z^m} + \dots \right] \quad (7.4)$$

$$\text{with } p = k_1 + k_2 - 3/2 = r - \frac{\gamma_3}{\gamma_2} - \frac{3}{2}$$

$$\begin{aligned} c_n(h, k_1, k_2, \dots, k_{n-2}) &= 2\pi(-1)^{1-p} \mu^{-\frac{p}{2}} \frac{u_2(-1)}{D(-1)} \\ &= (-1)^{1-p} \mu^{-\frac{p}{2}} u_2(-1) \frac{c_{n-1}}{m^2 a_n w_0} \frac{\prod_{k=1}^{n-2} (-1 - \alpha_k)}{\prod_{k=1}^{n-1} \Gamma(1 - \beta_k)} \end{aligned} \quad (7.5)$$

$$\begin{aligned}
c_{n-1}(h, k_1, k_2, \dots, k_{n-2}) &= 2\pi \mu^{-\frac{p}{2}} \frac{u_1(-1)}{D(-1)} \\
&= \mu^{-\frac{p}{2}} u_1(-1) \frac{c_{n-1}}{m^2 a_n w_0} \frac{\prod_{k=1}^{n-2} \Gamma(-1 - \alpha_k)}{\prod_{k=1}^{n-1} (1 - \beta_k)}
\end{aligned} \tag{7.6}$$

and $R(z)$ represents the sum of the residues of the function

$$\frac{\pi u(s) (-z^m)^s}{\sin \pi s} \tag{7.7}$$

at the poles (6.5).

Our next objective then is to obtain $R(z)$. If $R_{ij}(z)$ is the residue of (7.7) at the pole $s = \alpha_i - j = \frac{-h - k_1}{m} - j$, then

$$R(z) = \sum_{i=1}^{n-2} \sum_{j=0}^{\infty} R_{ij}(z). \tag{7.8}$$

At this stage, we shall assume that none of $\alpha_i - \alpha_j = \frac{k_j - k_i}{m}$ ($j \neq i, i, j = 1, 2, \dots, n-2$) is an integer, which means that none of $k_i = k_j$ ($i \neq j, i, j = 1, 2, \dots, n-2$) is congruent to zero modulom. Then the poles (6.5) will all be simple poles.

After computing $R_{ij}(z)$ and taking their sum in (7.8), we find

$$-R(z) = \sum_{i=1}^{n-2} c_i(h, k_1, k_2, \dots, k_{n-2}) \frac{1}{z^{\frac{h+k_1}{m}}} \left[1 + \frac{(\)}{z^m} + \frac{(\)}{z^{2m}} + \dots \right]$$

where,

$$c_1(h, k_1, k_2, \dots, k_{n-2}) = \frac{e^{\pi i \left(\frac{h+k_1}{m} \right)} \Gamma \left(2 - \frac{h+k_1}{m} \right) \prod_{k=1}^{n-1} \Gamma(1 - \beta_k) w \left(\frac{-h - k_1}{m} \right)}{\prod_{i=2}^{n-2} (k_i - k_1) \prod_{i=2}^{n-2} \Gamma \left(\frac{h+k_i}{m} - 1 \right)} \tag{7.9}$$

with

$$w(s) = v(s+2) f_0(2m - k_1) + v(s+1) f_1(m - k_2) \tag{7.10}$$

and $c_j(h, k_1, k_2, \dots, k_{n-2})$, $j = 2, 3, \dots, n-2$ are obtained from (7.9) by interchanging k_1 and k_j .

Thus, in view of (7.1) and (7.2), we observe that the fundamental solution $y(z)$ about the regular singular point $z=0$ as defined in (2.3), when $|z|$ is large and

when z lies along any ray except for which $\arg z^m = \arg \left(\frac{1}{w_0} \right)$ may be expanded asymptotically in the form

$$y(z) \sim c_n(h, k_1, k_2, \dots, k_{n-2}) s_n(z) + \sum_{i=1}^{n-2} c_i(h, k_1, k_2, \dots, k_{n-2}) s_i(z) \quad (7.11)$$

in $\frac{\pi}{2} < \arg \sqrt{\mu z^m} < \frac{3\pi}{2}$

and

$$y(z) \sim \sum_{i=1}^{n-1} c_i(h, k_1, k_2, \dots, k_{n-2}) s_i(z) \text{ in } \frac{-\pi}{2} < \arg \sqrt{\mu z^m} < \frac{\pi}{2} \quad (7.12)$$

where $s_i(z)$, $i = 1, 2, \dots, n-2$ are given by

$$s_i(z) = \frac{1}{z^{k_i}} \left[1 + \frac{(\)}{z^m} + \frac{(\)}{z^{2m}} + \dots \right] \quad (7.13)$$

and $s_n(z)$ and $s_{n-1}(z)$ are respectively given by (7.3) and (7.4). Here, $c_1(h, k_1, k_2, \dots, k_{n-2})$ is given by (7.9), $c_j(h, k_1, k_2, \dots, k_{n-2})$ for $j = 2, 3, \dots, n-2$ are given by interchanging k_1 and k_j in (7.9), while $c_n(h, k_1, k_2, \dots, k_{n-2})$ and $c_{n-1}(h, k_1, k_2, \dots, k_{n-2})$ are respectively given by (7.5) and (7.6).

In arriving at this result, we have placed several restrictions. In fact, we have assumed that none of the quantities $\alpha_j = \frac{-h - k_j}{m}$ and $\alpha_i - \alpha_j = \frac{k_j - k_i}{m}$ ($i \neq j$, $i, j = 1, 2, \dots, n-2$) is an integer.

Assuming that $\alpha_i - \alpha_j = \frac{k_j - k_i}{m}$ for $i \neq j$, $i, j = 1, 2, \dots, n-2$ is not an integer which means that none of $k_j - k_i$ ($i \neq j$, $i, j = 1, 2, \dots, n-2$) is congruent to zero modulo m , we shall now examine the effects upon (7.5), (7.6), and (7.9), when one of $\frac{h+k_i}{m}$, $i = 1, 2, \dots, n-2$ is an integer. In this case it is to be noted that in no case two of these quantities are integer, for this would imply that the difference $k_j - k_i$ ($i \neq j$) is congruent to zero modulo m , contradicting the hypothesis. Moreover, it will suffice to consider merely the case in which $\frac{h+k_1}{m}$ is an integer in as much as all possible cases may be covered by merely interchanging those of k_1 and k_j for $j = 2, 3, \dots, n-2$.

We shall suppose that $\frac{h+k_1}{m}$ is an integer. Let us divide the possible integral values of $\frac{h+k_1}{m}$ into two classes:

- (a) $\frac{h+k_1}{m}$ is an integer < 2
 (b) $\frac{h+k_1}{m}$ is an integer ≥ 2 .

Case (a) $\frac{h+k_1}{m}$ is an integer < 2 .

If we now form the differential equation in which a_n in (2.1) is changed to $a_n - E$, where E is arbitrarily small and positive, the resulting values of k_i ($i = 1, 2, \dots, n-2$) remains as before in as much as they depend only on c_j ($j = 0, 1, \dots, n-2$). The resulting value of h will be of the form $h - \eta$ where η becomes indefinitely small with E and at least after E has been taken sufficiently small, none of the quantities $\frac{h+k_i - \eta}{m}$ is an integer. The solution $y(z)$ corresponding to (2.3) of the new differential equation will therefore be developable asymptotically in the form obtained by (7.11) and (7.12) replacing $\frac{h}{m}$ by $\frac{h-\eta}{m}$ and the resulting $c_i(h - \eta, k_1, k_2, \dots, k_{n-2})$ for ($i = 1, 2, \dots, n$) determined from (7.5), (7.6), and (7.9).

After the expressions thus obtained have been examined, it is noted that when $\frac{h+k_1}{m}$ is an integer < 2 , each of the following limit exists:

$$\lim_{\eta \rightarrow 0} \frac{\Gamma\left(2 - \frac{h+k_j - \eta}{m}\right)}{\prod_{\substack{i=1 \\ i \neq j}}^{n-2} \Gamma\left(\frac{h+k_i - \eta}{m} - 1\right)} \quad \text{for } j = 1, 2, \dots, n-2. \quad (7.14)$$

Hence, we conclude that so long as $\frac{h+k_1}{m}$ is an integer belonging to class (a) mentioned above, the relations (7.11) and (7.12) continue as before.

Case (b) $\frac{h+k_1}{m}$ is an integer ≥ 2 .

First we consider the case $\frac{h+k_1}{m} = 2$. Proceeding as in Case (a), we obtain the new differential equation from (2.1) by replacing a_n by $a_n - E$ and hence $\frac{h}{m}$ by $\frac{h-\eta}{m}$ by (7.11) and (7.12). We endeavor to examine the limit as $\eta \rightarrow 0$ of $c_1(h, k_1, k_2, \dots, k_{n-2})$ as determined by (7.9).

In doing this, there is involved aside from the factor which clearly approaches a limit, the product

$$\Gamma\left(\frac{\eta}{m}\right) \left[f_0(h) v\left(\frac{\eta}{m}\right) + f_1(h-m) v\left(\frac{\eta}{m}-1\right) \right]. \quad (7.15)$$

Since $f_0(h) \equiv 0$ in view of (2.2), we have $f_0(h) v\left(\frac{\eta}{m}\right) = 0$.

Hence, the limit of (7.15) as $\eta \rightarrow 0$

$$= \lim_{\eta \rightarrow 0} \Gamma\left(\frac{\eta}{m}\right) \left[f_1(h-m) v\left(\frac{\eta}{m}-1\right) \right] \quad (7.16)$$

We observe that as $\eta \rightarrow 0$, $\Gamma\left(\frac{\eta}{m}\right)$ becomes infinite to the first order.

Also, $\lim_{\eta \rightarrow 0} f_1(h-m) v\left(\frac{\eta}{m}-1\right) = f_1(h-m) v(-1) = 0$. Thus, the limit of (7.16)

as $\eta \rightarrow 0$ is indeterminate. We shall however evaluate it by replacing

$$\Gamma\left(\frac{h}{m}\right) = \frac{\Gamma\left(1 + \frac{h}{m}\right)}{\frac{h}{m}}$$

the result being

$$m f_1(h-m) \left[\frac{d}{d\eta} v\left(\frac{\eta}{m}-1\right) \right]_{\eta=0} = f_1(h-m) \left[\frac{d}{ds} v(s-1) \right]_{s=0}$$

or as we shall prefer to write for future purposes

$$\left[f_0(2m-k_1) \frac{d}{ds} v(s+2) + f_1(m-k_1) \frac{d}{ds} v(s+1) \right]_{s=\frac{-h-k_1+2}{m}}. \quad (7.17)$$

Finally, let us suppose that $\frac{h+k_1}{m}$ is an integer $N > 2$, then, reasoning as in the case $\frac{h+k_1}{m} = 2$, we replace h in (7.9) by $h-\eta$ and attempt to obtain the limit of the resulting expressions as $\eta \rightarrow 0$. No difficulties are encountered except in connection with the product

$$\Gamma\left(2-N+\frac{\eta}{m}\right) \left[f_0(2m-k_1) v\left(2-N+\frac{\eta}{m}\right) + f_1(m-k_1) v\left(1-N+\frac{\eta}{m}\right) \right] \quad (7.18)$$

But

$$\Gamma\left(2-N+\frac{\eta}{m}\right) = \frac{(-1)^N \Gamma\left(1+\frac{\eta}{m}\right)}{\left(N-2-\frac{\eta}{m}\right)\left(N-3-\frac{\eta}{m}\right)\cdots\left(1-\frac{\eta}{m}\right)\left(\frac{\eta}{m}\right)}$$

since $N > 2$

$$\lim_{\eta \rightarrow 0} v\left(2-N+\frac{\eta}{m}\right) = v(2-N) = 0$$

and

$$\lim_{\eta \rightarrow 0} v \left(1 - N + \frac{\eta}{m} \right) = v(1 - N) = 0.$$

Hence, we see \lim of (7.18) as $\eta \rightarrow 0$

$$= (-1)^N \left[f_0(2m - k_1) \frac{d}{ds} v(s+2) + f_1(m - k_1) \frac{d}{ds} v(s+1) \right]_{s=-N} \quad (7.19)$$

Finally, we note that (7.17) may be obtained from (7.19) by putting $N = 2$ so that when $\frac{h+k_1}{m}$ is an integer ≥ 2 , we conclude that the relations (7.11) and (7.12) for which $c_1(h, k_1, k_2, \dots, k_{n-2})$ is no longer to be determined from (7.9) but by (7.19), while $c_j(h, k_1, k_2, \dots, k_{n-2})$ for $j = 2, 3, \dots, n-2$ are to be determined by interchanging k_1 and k_j in (7.9).

This completes the proof of the theorem.

Due to the limitation of space, the case when the roots of $p_0\left(\frac{-s-h}{m}\right) = 0$ are congruent to zero modulo m will not be discussed here and will be published elsewhere as a sequel to this paper. It must be noted that in this case, some of the poles (6.5) will be of higher order.

DIRECTION OF FURTHER RESEARCHES

Other directions of research are possible. For example, one could examine the global behavior of solutions of the differential equation

$$\sum_{k=0}^n z^\kappa p_k(z) \frac{d^k y}{dz^k} = 0$$

where $p_k(z)$ ($k = 0, 1, 2, \dots, n$) are polynomials given by $p_k(z) = \sum_0^r a_{k\ell} z^{\ell m}$, $a_{n_0} \neq 0$,

with m being an arbitrary positive integer and $r \geq 3$. Ford and Puttaswamy have carried throughout that the difference of no two roots of the indicial equation about the regular singular point is congruent to zero modulo 1 and m respectively. A likewise investigation is evidently possible, when this assumption is reversed. Finally, under either hypothesis, the limitation that has pervaded that the finite singular points of the differential equation shall be regular carries its own suggestions. These then are some of the problems for the future. By way of conclusion, one must emphasize that new ideas and new methods for solving differential equations in the large are much needed.

REFERENCES

- [1] Culmer, W.J.A., Convergent solutions of ordinary linear homogeneous difference equations in the neighborhood of an irregular singular point, Ph.D. thesis, University of Minnesota, August 1959.
- [2] Culmer, W.J.A. and Hanis, Jr., W.A., Convergent solutions of ordinary linear homogeneous difference equations, *Pac. Jour. of Math.*, 13 (1963), 1111-1138.
- [3] Ford, W.B., *Studies on Divergent Series and Summability and the Asymptotic Developments of Functions Defined by Maclaurion Series*, Chelsea Publishing Company, 1960.
- [4] Harris, Jr., W.A., Linear systems of difference equations solvable by factorial series, Tech. Summary Report No. 213, U.S. Army Math. Research Center, University of Wisconsin, Madison, December 1960.
- [5] Harris, Jr., W.A., Linear systems of difference equation, Contribution to differential equations, 1 (1963), 489-518.
- [6] Harris, Jr., W.A., Equivalent classes of difference equations, Contribution to differential equations, 1 (1964), 253-264.
- [7] Harris, Jr., W.A. and Sibuya, Y., Note on linear difference equations, *Bull. Amer. Math. Soc.*, 70 (1964), 123-127.
- [8] Puttaswamy, T.K., The solution of certain ordinary linear differential equation in the large, Master's thesis, University of Minnesota, May 1963.
- [9] Puttaswamy, T.K., Solution in the large of a certain n th order differential equation. Proceedings of Ramanujan Birth Centenary Year International Symposium on Analysis, Pune, India, 1989, 191-206, MacMillan (India) Ltd.
- [10] Puttaswamy, T.K., A connection problem for a certain n th order homogeneous differential equation, *Differential Equations and Applications*, Ohio University Press, (1989), 317-322.
- [11] Puttaswamy, T.K., A generalization of a theorem of W.B. Ford. Proceedings of the International Conference on New Trends in Geometric Function Theory and Applications, World Scientific Publishers, (1991), 90-95.
- [12] Puttaswamy, T.K., Stokes multipliers for a certain third order differential equaiton of W.B. Ford's type, *Journal of Ramanujan Math. Soc.*, 8, 1 and 2 (1993), 139-166.
- [13] Puttaswamy, T.K., Stokes multipliers for a certain n th order differential equation. Proceedings of the Third International Colloquium on Differential Equations, Plovdiv, Bulgaria, VSP, Netherlands (1993), 157-168.

- [14] Puttaswamy, T.K. and Sastry, T.V., Solution in the large of a certain third differential equation of Langer. Proceedings of the International Symposium on Trends and Developments in Ordinary Differential Equations, World Scientific Publishers, (1994), 271-282.
- [15] Puttaswamy, T.K., Connection coefficients for a certain n th order differential equation. Proceedings of the International Symposium on Trends and Developments in Ordinary Differential Equations, World Scientific Publishers, (1994), 243-254.
- [16] Puttaswamy, T.K., Asymptotic behavior of solutions of a certain n th order differential equation about an irregular singular point. Proceedings of the Fourth International Colloquium on Differential Equations, Impulse-7, (1994), 79-90.
- [17] Puttaswamy, T.K. and Sastry, T.V., A two point connection problem for a certain third order differential equation. Proceedings of the Fourth International Colloquium on Differential Equations, VSP, (1994), 235-244.
- [18] Puttaswamy, T.K. and Sastry, T.V., Asymptotic behavior of solutions of a certain third order differential equation near an irregular singular point, Revue Roumaine De Mathematiques Pures Et Appliquées, Vol. 39, Series 7, (1994), 739-748.
- [19] Puttaswamy, T.K., Connection coefficients for a certain third order differential equation containing an arbitrary number of regular singular points, Revue Roumaine De Mathematiques Pures Et Appliquées, Vol. 39, Series 7, (1994), 717-737.
- [20] Puttaswamy, T.K., Global behavior of solutions of a certain n th order differential equation in the neighborhood of an irregular singular point of arbitrary rank, Revue Roumaine De Mathematiques Pures Et Appliquées, Vol. 39, Series 7, (1994), 703-715.
- [21] Puttaswamy, T.K., Connection coefficients for a certain third order differential equation in the neighborhood of an irregular singular point, Indian Journal of Pure and Applied Mathematics, Vol. 27 (10), (1996), 945-957.
- [22] Puttaswamy, T.K., Solution in the large of a certain third order differential equation of W.B. Ford's type, To appear.
- [23] Puttaswamy, T.K., Stokes multipliers for a certain third order differential equation in the vicinity of an irregular singular point, To appear.
- [24] Puttaswamy, T.K., Asymptotic behavior of solutions of a certain n th order differential equation in the vicinity of an irregular singular point of arbitrary rank, To appear.
- [25] Tritzinsky, W.J., Analytic theory of linear differential equations, Acta Math, 62 (1934), 167-227.
- [26] Tritzinsky, W.J., Laplace integrals and factorial series in the theory of linear differential and linear difference equations, Trans. Amer. Math. Soc., 37 (1935), 80-146.
- [27] Turrrittin, H.L., Convergent solutions of linear homogeneous differential equations in the neighborhood of an irregular singular point, Acta Math., 93 (1955), 27-66.

- [28] Turrittin, H.L., Stokes multipliers for asymptotic solutions of a certain differential equation, Trans. Amer. Math. Soc., 68 (1950), 304-329.
- [29] Wright, E.M., The asymptotic expansion of generalized hypergeometric function, Jour. London Math. Soc., 10 (1935), 286-293.
- [30] Wright, E.M., The asymptotic expansion of the generalized hypergeometric function, Proc. London Math. Soc., (2) 46, (1940), 389-408.
- [31] Wright, E.M., The asymptotic expansion of integral functions defined by Taylor Series, Phil. Trans. Royal Soc. London, A 238 (1940), 423-451 and A 239 (1941), 217-232.
- [31] Wright, E.M., The asymptotic expansion of integral functions and of the coefficients in their Taylor Series, Trans. Amer. Math. Soc., 64 (1948), 409-438.

20 ON THE MODELLING OF DISSIPATIVE PROCESSES

K.R. Rajagopal

Department of Mechanical Engineering
and Department of Mathematics

Texas A&M University
College Station, TX 77845

1. INTRODUCTION

A basic premise in the classical theory of elasticity (see Truesdell and Noll [1]) is that the body has a unique “natural configuration” modulo rigid motion, the “natural configuration” being usually understood as the stress free state. Such a premise is not generally true for most if not all materials, there being numerous other “configurations” in which they can exist naturally. Not all such states may be accessed by a body in a specific process, but that is not to say that such natural configurations do not exist. For instance, a virgin specimen of metal (if such a specimen is possible for after all such a specimen is produced via a complicated manufacturing process), that is in a stress free state, could be subject to sufficiently small deformations by the application of sufficiently small loads, which on the removal of the loads could return to its original stress free state (or “natural configuration”). However, the same body when subjected to a sufficiently large homogenous deformation would not return to its original stress free configuration on the removal of the load, as it might have undergone “plastic” deformation. A piece of steel, for instance, can undergo classical slip, or twin, or undergo solid to solid phase change on the application of loads, or it could melt due to an increase in temperature. At the end of any one of these above mentioned processes, the piece of steel would be associated with a different natural configuration from its initial natural configuration. In order to model the response of materials undergoing such processes, it is imperative to explicitly take cognizance of the fact that different natural configurations are accessed during such processes.

Here, I shall restrict my discussion to isothermal processes, and discuss a framework that takes into account the fact that a variety of natural configurations are accessed by the body during the process. A common feature shared by all these bodies, subject to such processes, is that the rate of dissipation is non zero. Within a more general thermodynamic framework, we would have to worry about concepts such as entropy and its production. The theoretical framework that is presented has been used, with some amount of success, in describing a widely disparate class of problems: traditional plasticity (see Rajagopal and Srinivasa [2] Rajagopal and Srinivasa [3]), twinning (see Rajagopal and Srinivasa [4], [5], Srinivasa, Rajagopal and Armstrong [6], Lapczyk, Rajagopal and Srinivasa [7]), shape memory transformations (see Rajagopal and Srinivasa [8]), multi-network theory for polymers (Rajagopal and Wineman [9], Wineman and Rajagopal [10]), crystallization of polymers (Rao and Rajagopal [11], [12]), and Anisotropic liquids (Rajagopal and Srinivasa [13]). Here, I shall present the salient elements of the framework, a detailed account of the same can be found elsewhere (Rajagopal [14]).

2. KINEMATICS

Let us denote by B the set of particles belonging to a body and by $k(B)$ the configuration of the set of particles in a three dimensional Euclidean space. We shall refer to (B, k) as the body. Here we make a distinction from the usual definition of a body in that we refer to the order pair (B, k) as the body rather than just B , the latter being referred to as the abstract body. As usual, we can endow a measure and topology on B (see Truesdell [15]). Let $k_t(B)$ denote the configuration of the body at time t . Corresponding to every $k_t(B)$, there exists a natural configuration $k_{p(t)}(B)$ which we shall call the preferred natural configuration corresponding to that $k_t(B)$. It is important to recognize that the preferred natural configuration is not determined by just $k_t(B)$, but could depend on how the body got into that configuration, i.e., given two distinct processes by which means a body got to $k_t(B)$, the preferred natural configuration corresponding to $k_t(B)$ reached in the two distinct processes, could be different.

It is also important to recognize that the notion of a natural configuration is a local idea, that is, corresponding to each material point belonging to B , an infinitesimal neighborhood of B containing that point possesses a natural configuration, the reason for this being different points belonging to B , when B is subject to a deformation have associated with them different physical attributes. We could talk of a natural configuration corresponding to a homogeneous body that is subject to a homogeneous deformation, however, as a body can be inhomogeneous,

the natural configurations can at best be a local idea (see Rajagopal [14] for a more detailed discussion).

Given a reference configuration of a body $k(B)$, we can make measurements that are referred to this configuration. Let $X \in k(B)$. By a motion of a body, we mean an assignment $\mathbf{x} = \chi_t(X, t)$ belonging to the Euclidean space, for each $X \in k(B)$, as time progresses. Thus,

$$k_t(B) := \left\{ \mathbf{x} = \chi_k(\mathbf{x}, t) \mid X \in k(B), t \in \mathbb{R} \right\}.$$

For each t , we shall assume that $X(\mathbf{x}, t)$ is one to one.

The deformation gradient \mathbf{F}_k is defined through

$$\mathbf{F}_k(\mathbf{x}, t) := \frac{\partial \chi_k(\mathbf{x}, t)}{\partial X}. \quad (2.1)$$

We shall henceforth assume that the functions are sufficiently smooth to render all derivatives that are indicated to be meaningful. We shall also assume that our motion is such that

$$\det \mathbf{F}_k \neq 0. \quad (2.2)$$

The velocity, acceleration, etc., associated with the motion are defined as usual through

$$\mathbf{v} := \frac{\partial \chi_k}{\partial t}, \quad (2.3)$$

$$\mathbf{a} := \frac{\partial^2 \chi_k}{\partial t^2}, \quad (2.4)$$

and the velocity gradient \mathbf{L}_k through

$$\mathbf{L} := \frac{\partial \mathbf{v}}{\partial \mathbf{x}}. \quad (2.5)$$

We note that for the last definition to be meaningful \mathbf{v} should be expressed as a function of \mathbf{x} and t , i.e., in its Eulerian representation.

To illustrate the notion of natural configuration, let us consider the typical stress-strain curve for the one-dimensional response of a metal that is deformed so that plastic slip comes into play. On deforming the unstressed, unstrained specimen we follow the stress-strain curve along OA . If the specimen is unloaded at any state before the yield point A is reached, the specimen retraces its path, i.e., its response is elastic. However, deforming the body beyond A , say to B , and then unloading it causes the body to return to O_B rather than to O . The response of such a body can be thought of as a one-parameter family (in the case of the one-dimensional problem) of elastic responses but from a one-parameter family of natural configurations. The lines OA , OB need not be of the same shape, that is the response can be distinct. For any configuration corresponding to any point along the line OA ,

the configuration corresponding to O is the preferred natural configuration, while for a configuration corresponding to any point along the line $O_B B$, the configuration corresponding to O_B is the preferred natural configuration.

We notice that changing the natural configuration of the body from O to O_B is accompanied by a certain amount of dissipation or to be more precise, the rate of dissipation. The manner in which the natural configurations evolve during a process is determined by the rate of dissipation, for example it could be determined by maximizing the rate of dissipation.

A body is said to be a classical elastic body if the Cauchy stress \mathbf{T} in the body is determined completely by knowing the deformation gradient \mathbf{F}_k from a reference configuration k , i.e.,

$$\mathbf{T} = f_k(\mathbf{F}_k). \quad (2.6)$$

There is the underlying assumption that there is one and only one natural configuration (modulo rigid motions) for an elastic body, and thus the underlying natural configurations do not change during any process such a body is subject to.

Let ξ denote the position occupied by X at time τ . Then

$$\xi = \chi_k(x, \tau) = \chi_k(\chi_k^{-1}(x, t), \tau) := \chi_t^k(x, \tau). \quad (2.7)$$

The relative deformation gradient $\mathbf{F}_t^k(x, \tau)$ is defined through

$$\mathbf{F}_t^k(x, \tau) := \frac{\partial \chi_t^k(x, \tau)}{\partial x}. \quad (2.8)$$

$\chi_t^k(x, \tau)$ is referred to as the relative motion.

For the purpose of our illustration here, the above meager kinematical definitions suffice. We now turn to constitutive specifications.

3. CONSTITUTIVE RELATIONS

Most classical models that have been used to describe the response of materials like the Navier-Stokes model for fluids or various models such as the linearized elastic model, Neo-Hookean model and other models that describe the finite elastic response of solids fall under the category of simple materials.

A material is said to be a simple material (see Noll [16], Truesdell and Noll [1]) if the stress \mathbf{T} is determined by a functional of the history of $\mathbf{F}_t^k(x, \tau)$, i.e.,

$$\mathbf{T} = \sum_{s=0}^{\infty} \mathcal{F}_k \left(\mathbf{F}_t^k(x, t-s) \right). \quad (3.1)$$

We note the emphasis that the functional \mathcal{F} is indexed with a k to denote that the functional form while dependent on k is fixed as soon as k is fixed. This subtle point is pregnant with consequences, an important one being that classical plasticity theories are not theories of simple materials in the sense defined above. Plasticity

theories can be cast into a more general theory of simple materials that Noll introduced in 1972 (see Noll [17]), however, this definition essentially is a multiparameter family of simple materials in the earlier sense, the parameter being indexed by a state variable which is left undefined. This definition of Noll does little to shed any light on the underlying physics of the material (see Rajagopal [14] for a detailed discussion of these issues). We shall return to a discussion of the inadequacy of the model (3.1) to describe most dissipative phenomena. The elastic-plastic response depicted in Figure 1 allows the material to possess various elastic responses from different natural configurations. More importantly, it also allows the material to have different symmetries in these various natural configurations, and this possibility is of paramount importance in the modeling of many real materials.

For elastic materials of the class depicted by (2.6), we can define the symmetry group associated with k through (see Truesdell [15])

$$\mathcal{G}_k := \left\{ \mathbf{H} \in \mathcal{U} \mid f_k(\mathbf{F}_k) = f_k(\mathbf{F}_k \mathbf{H}) \right\}. \quad (3.2)$$

Then, if \hat{k} is any other reference configuration, $\mathcal{G}_{\hat{k}}$ the symmetry group with respect to \hat{k} is related to \mathcal{G}_k through

$$\mathcal{G}_{\hat{k}} = \mathbf{P} \mathcal{G}_k \mathbf{P}^{-1}, \quad (3.3)$$

where \mathbf{P} is the gradient of the mapping from $k(B)$ to $\hat{k}(B)$. The above relation (3.3) is called Noll's rule and it has to necessarily hold for materials belonging to the class (2.6). In fact, Noll's rule holds for simple materials that belong to the class (3.1).

Thus, if there are materials which when subject to a specific deformation change their underlying natural configuration and moreover, if the material symmetry in these natural configurations are not related by Noll's rule, then we can conclude that such materials are not simple materials in the sense of (3.1). This is indeed the case for a crystal undergoing slip, for a solid undergoing solid to solid phase transitions, for a material that is crystallizing, or for a crystalline solid that is melting. Unfortunately, each and every one of these above mentioned phenomena are repeatedly modeled using models that belong to the class of simple materials.

One way to model the various phenomena mentioned in the preceding paragraph is through a more general structure than that given by (3.1). For example, consider the class of materials in which

$$\mathbf{T} = f_{k_{p(t)}}(\mathbf{F}_{k_{p(t)}}). \quad (3.4)$$

Then, as t changes, $k_{p(t)}$ can change and the form of the function $f_{k_{p(t)}}$ can change as well as the symmetry group associated with $k_{p(t)}$. No longer are the symmetry groups as the process progresses related through Noll's rule, for now we have

$$\mathcal{G}_{k_{p(t)}} = \left\{ \mathbf{H} \in \mathcal{U} \mid \mathbf{f}_{k_p} \left(\mathbf{F}_{k_{p(t)}} \mathbf{H} \right) = \mathbf{f}_{k_{p(t)}} \left(\mathbf{F}_{k_{p(t)}} \right) \right\}. \quad (3.5)$$

In the above discussion $k_{p(t)}$ denotes the preferred configuration corresponding to that at time t . If $k_{p(t)}$ does not change during a part of the process, during that part of the process, Noll's rule would hold.

Models of the form (3.1) cannot be used to describe phenomena that involve phase change with an attendant change of symmetry that does not meet Noll's rule, and unfortunately this is often the case. Such phase transitions are dissipative and thus the model (3.1) cannot be used to describe materials that undergo such dissipative processes. Neither can the model be used in crystal plasticity, or for that matter rate type viscoelastic materials.

The material is characterized by the manner in which it stores energy, the manner in which such stored energy can be recovered (i.e., mechanical processes, or thermal processes such as annealing), the manner in which energy is dissipated (in general the manner in which entropy is produced), etc. The way in which the natural configuration evolves depends on the manner in which the material dissipates energy. For a certain class of processes, it seems that the assumption that the rate of dissipation be maximized leads to response characteristics that agree with experimentally observed results. Using such an assumption, Rajagopal and co-workers have studied a large class of material response that includes traditional plasticity, twinning, solid to solid phase transitions, response of multinetwork polymers, crystallization of polymers, viscoelasticity, anisotropic liquids and growth of biological materials.

It seems that the framework of materials with multiple natural configurations has the potential to model a diverse class of phenomena and a great deal can be done within such a framework.

REFERENCES

- [1] Truesdell, C. and Noll, W., The non-linear field theories of mechanics, 2nd edition, Springer-Verlag, Berlin 1992.
- [2] Rajagopal, K.R. and Srinivasa, A.R., Mechanics of the inelastic behavior of materials: Part I – Theoretical underpinnings, International Journal of Plasticity, 14, 1998, 945-967.
- [3] Rajagopal, K.R. and Srinivasa, A.R., Mechanics of the inelastic behavior of materials: Part II – Inelastic response, International Journal of Plasticity, 14, 1998, 969-995.
- [4] Rajagopal, K.R. and Srinivasa, A.R., On the inelastic behavior of solids: Part I – Twinning, International Journal of Plasticity, 11, 1995, 653-678.

- [5] Rajagopal, K.R. and Srinivasa, A.R., Inelastic behavior of materials: Part II – Energetic associated with discontinuous twinning, International Journal of Plasticity, 13, 1997, 1-35.
- [6] Srinivasa, A.R., Rajagopal, K.R., and Armstrong, R., A phenomenological model of twinning based on dual reference structures, Acta Materialia, 46, 1998, 1235-1248.
- [7] Lapczyk, E., Rajagopal, K.R., and Srinivasa, A.R., Deformation twinning during impact – Numerical calculations using a constitutive theory based on multiple natural configurations, Computational Mechanics, 21, 1998, 20-27.
- [8] Rajagopal, K.R. and Srinivasa, A.R., On the thermodynamics of shape memory wires, ZAMP, 50, 1999, 459-494.
- [9] Rajagopal, K.R. and Wineman, A.S., A constitutive equation for non-linear solids which undergo deformation induced microstructural changes, International Journal of Plasticity, 8, 1992, 385-395.
- [10] Wineman, A.S. and Rajagopal, K.R., On a constitutive theory for materials undergoing microstructural changes, Archives of Mechanics, 42, 1990, 53-75.
- [11] Rao, I.J. and Rajagopal, K.R., Phenomenological modelling of polymer crystallization using the notion of multiple natural configurations, Interfaces and Free Boundaries, 2, 2000, 73-94.
- [12] Rao, I.J. and Rajagopal, K.R., A study of strain induced crystallization in polymers, International Journal of Solids and Structures, in press.
- [13] Rajagopal, K.R. and Srinivasa, A.R., A thermodynamic framework for rate type fluid models, 88, 2000, 207-227.
- [14] Rajagopal, K.R., On the constitutive modeling of materials, to appear.
- [15] Truesdell, C., *A First Course in Rational Continuum Mechanics*, Academic Press, 19.
- [16] Noll, W., On the foundations of mechanics of continuous media, Carnegie Institute of Technology, Department of Mathematics, Report 7, 1957.
- [17] Noll, W., A new mathematical theory of simple materials, Archive for Rational Mechanics and Analysis, 48, 1972, 243-293.

21 PATHWISE AVERAGE COST PER UNIT TIME PROBLEM FOR STOCHASTIC DIFFERENTIAL GAMES WITH A SMALL PARAMETER

Kandethody M. Ramachandran

Department of Mathematics

University of South Florida

Tampa, FL 33620-5700

and

A.N.V. Rao

Department of Mathematics

University of South Florida

Tampa, FL 33620-5700

ABSTRACT

A two person zero-sum stochastic differential game problem for wideband noise driven system is considered. We will consider the payoff structure in the pathwise sense. Under sufficiently general conditions, it will be shown that the optimal equilibrium policies of the limit process when applied to the physical processes, will be δ -equilibrium as the parameters $\varepsilon \rightarrow 0$ and $T \rightarrow \infty$. Martingale convergence techniques will be used in the analysis. A Chattering result is also derived. The entire problem will be set in relaxed control framework.

1. INTRODUCTION

Average cost per unit time problem over an infinite time horizon for two person zero-sum stochastic differential games with diffusion model have been dealt with in the literature. For the diffusion models where payoff with expectations (not pathwise), existence of equilibrium has been proved in ([3]) and in the case of

discounted and average cost cases the existence of equilibria in Markov strategies was established in [1]. We treat such a problem for wideband noise driven systems, which are ‘close’ to diffusion. The average is in the pathwise but not necessarily in the expected value sense ([8]). The ‘pathwise’ convergence result is of particular importance in applications, since we often have a single realization, then expectation is not appropriate in the cost function. In a typical application, we have a particular process with a wideband noise driving forces. Our interest is in knowing how well good policies for the ‘limit’ problem do for the actual ‘physical’, problem as well as various qualitative properties of the ‘physical’ process. Physical problem is better modeled by a wideband width noise driven process than the white noise process. However, owing to the wideband noise and appearance of the two parameters ε and T , convergence results of the ‘almost sure’ type are often rather meaningless from a practical point of view as well as nearly impossible to obtain. It is important that the convergence result obtained should not depend on the way in which $\varepsilon \rightarrow 0$ and $T \rightarrow \infty$. Where this is not the case, it would be possible that as $\varepsilon \rightarrow 0$, a larger and larger T is needed to closely approximate the limit value. In that case, the white noise limit (1) would not be useful for predictive or control purposes when the true model is (7). It will be shown that the optimal equilibrium policies of the limit diffusion when applied to the wide bandwidth processes, will be δ -equilibrium as the parameters $\varepsilon \rightarrow 0$ and $T \rightarrow \infty$, irrespective of the order in which the limit takes place. It is also shown that the δ -optimal pathwise discounted payoffs converge to the δ -equilibrium as both the discounted factor $\lambda \rightarrow 0$ and bandwidth goes to ∞ . Apart from the fact that this gives a robustness statement for the diffusion model, one of the major advantages is by using the method of this work, it is enough to compute the optimal strategies for the limit diffusion and then use these strategies to the physical system to obtain near optimal strategies. The entire problem will be set in relaxed control framework. In the proofs, we will use the weak convergence theory.

In Section 2, we will describe the differential game problem for both diffusion model and for the wideband noise driven model. In Section 3, a chattering result will be derived. In Section 4, we will briefly mention some relevant results of weak convergence theory. The main weak convergence and δ -optimality result for (7) will be derived in Section 5. A stochastic game problem will be mentioned in Section 6. Some concluding remarks will be given in Section 7.

2. PROBLEM DESCRIPTION

Let the diffusion model be given in a non-anticipative relaxed control framework. Let $U_i, i = 1, 2$ be compact metric spaces (we can take U_i as compact subsets of

R^d), and $M_i = P(U_i)$, the space of probability measures on U_i with Prohorov topology.

For $m = (m_1, m_2) \in M = M_1 \times M_2$ and $U = U_1 \times U_2, x(\cdot) \in R^d$ by an R^d -valued process given by the following controlled stochastic differential equation

$$\begin{aligned} dx(t) &= \int_{U_1} a_1(x(t), \alpha_1) m_{1t}(d\alpha_1) + \int_{U_2} a_2((t), \alpha_2) m_{2t}(d\alpha_2) dt + \bar{g}((t)) dt + \sigma((t)) dw(t) \\ x(0) &= x_0 \end{aligned} \quad (2.1)$$

where x_0 is a prescribed random variable. The pathwise average payoff per unit time for player 1 is given by

$$J[m](x) = \liminf_{T \rightarrow \infty} \frac{1}{T} \int_0^T r(x(s), \alpha) m_x(d\alpha) ds \quad (2.2)$$

and for the initial law π in $P(\mathbb{R}^d)$, it is

$$J[m](\pi) = \int_{\mathbb{R}^d} J[m](x) \pi(dx). \quad (2.3)$$

Let $w(\cdot)$ in (2.1) be a Wiener process with respect to a filtration $\{\mathcal{T}_t\}$ and let $\Omega_i, i = 1, 2$ be a compact set in some Euclidean space. A measure valued random variable $m_i(\cdot)$ is an *admissible strategy* for the i^{th} player if $\int_0^t \int f_i(s, \alpha_i) m_i(ds d\alpha_i)$ is progressively measurable for each bounded continuous $f_i(\cdot)$ and $m_i([0, t] \times \Omega_i) = t$, for $t \geq 0$. If $m_i(\cdot)$ is admissible then there is a derivative $m_{it}(\cdot)$ (defined for almost all t) that is non-anticipative with respect to $w(\cdot)$ and

$$\int_0^t \int f_i(s, \alpha_i) m_i(ds d\alpha_i) = \int_0^t ds \int f_i(s, \alpha_i) m_{is}(d\alpha_i)$$

for all t with probability one (w.p.1). The results derived in this work are for so called *Markov strategies*, which is a measure on the Borel sets of Ω_i for each x , and $m_i(C)$ is Borel measurable for each Borel measurable set C . We will denote by A_i the set of admissible strategies and M_{ai} the set of Markov strategies for the player i . One can introduce appropriate metric topology under which M_{ai} is compact, [1].

In relaxed control settings, one chooses at time t a probability measure m_t on the control set M rather than an element $u(t)$ in U . We call the measure m_t the relaxed control at time t . Any ordinary control can be represented as a relaxed control via the definition of the derivative $m_t(d\alpha) = \delta_{u(t)}(\alpha) d\alpha$. Hence, if m_t is an atomic measure concentrated at a single point $m(t) \in M$ for each t , then the relaxed control will be called ordinary control. We will denote the ordinary control by $u_m(t) \in M$.

An admissible strategy $m_1^* \in A_1$ is said to be an *ergodic optimal* for initial law π if

$$\begin{aligned} J[m_1^*, \tilde{m}_2](\pi) &\geq \inf_{m_2 \in A_2} \sup_{m_1 \in A_1} J[m_1, m_2](\pi) \\ &= V^+(\pi) \end{aligned} \quad (2.4)$$

for any $\tilde{m}_2 \in A_2$. A strategy $m_1^* \in M_{a_1}$ is called discounted optimal for player I, if it is ergodic optimal for all initial laws. Similarly, $m_2^* \in A_2$ is discounted optimal for player II for an initial law π if

$$\begin{aligned} J(\tilde{m}_1, m_2^*)(\pi) &\leq \sup_{m_1 \in A_1} \inf_{m_2 \in A_2} J[m_1, m_2](\pi) \\ &= V^-(\pi) \end{aligned} \quad (2.5)$$

for any $\tilde{m}_1 \in A_1$, $m_2^* \in M_{a_2}$ is ergodic optimal for player II if (2.5) holds for all initial laws. If for any initial law π , $V^+(\pi) = V^-(\pi)$, then the game is said to have an ergodic equilibrium and we will denote it by $V(\pi)$. The policies $m_{1\delta}$ and $m_{2\delta}$ are said to be *δ -ergodic equilibrium* if

$$\sup_{m_1 \in A_1} J(m_1, m_{2\delta}) - \delta \leq V \leq \inf_{m_2 \in A_2} J(m_{1\delta}, m_2) + \delta. \quad (2.6)$$

The wide band noise system considered in this work is of the following type:

$$\begin{aligned} dx^\varepsilon &= \left[\int a_1(x^\varepsilon, \alpha_1) m_{1t}^\varepsilon(d\alpha_1) + \int a_2(x^\varepsilon, \alpha_2) m_{2t}^\varepsilon(d\alpha_2) dt + G(x^\varepsilon, \xi^\varepsilon(t)) \right. \\ &\quad \left. + \frac{1}{\varepsilon} g(x^\varepsilon, \xi^\varepsilon) dt \right] \end{aligned} \quad (2.7)$$

and pathwise average payoff per unit time for player k is given by

$$J^\varepsilon[m^\varepsilon] = \liminf_{T \rightarrow \infty} \frac{1}{T} \int_0^T \int r(x^\varepsilon(s), \alpha) m_s^\varepsilon(d\alpha) ds. \quad (2.8)$$

Player I aims to maximize his accumulated income, while player II will minimize the same. An *admissible relaxed strategy* $m_k^\varepsilon(\cdot)$ for the k^{th} player with

system (2.7) is a measure valued random variable satisfying $\int \int_0^t f(s, \alpha) m^\varepsilon(ds d\alpha)$

is progressively measurable with respect to $\{\mathcal{T}_t^\varepsilon\}$, where $\mathcal{T}_t^\varepsilon$ is the minimal σ -algebra generated by $\{\xi^\varepsilon(s), x^\varepsilon(s), s \leq t\}$. Also $m^\varepsilon([0, t] \times U) = t$ for all $t \geq 0$. Also, there is a derivative m_t^ε , where $m_t^\varepsilon(B)$ are $\mathcal{T}_t^\varepsilon$ measurable for Borel B . The concept of δ -ergodic equilibrium for $x^\varepsilon(\cdot)$ is similarly defined as in (2.6).

In [1], under the Lyapunov type stability condition (assumption A in [1]), following result is proved.

Theorem 2.1. For the stochastic differential game with ergodic payoff criterion has a value and both players have optimal strategies $m^* = (m_1^*, m_2^*) \in M_{a1} \times M_{a2}$.

3. CHATTERING LEMMA

In the relaxed control setting, each player chooses at time t a probability measure $m_i(t)$ on the control set M_i rather than an element $u_i(t) \in U_i, i = 1, 2$. Since relaxed controls are devices with primarily a mathematical use, it is desirable to have a chattering type result for the game problem. In order for the relaxed control problem to be true extension of the original problem, the equilibrium among the relaxed control strategies must be the same as the equilibrium taken among the ordinary strategies when it exists. For this purpose, we extend the chattering results obtained for control problems as in [4], to two person zero-sum stochastic differential games. We note that $U_i \subseteq M_i$, because, if $m_i(t)$ is an atomic measure, concentrated at a single point $u(t)$ for each t , then we get an ordinary control policy as a special case of a relaxed control policy. Let (m_1^*, m_2^*) be the equilibrium policy pair in the relaxed controls and (u_1^*, u_2^*) be the equilibrium policy pair (if exists) in the ordinary controls.

Theorem 3.1. Under the conditions of Theorem 2.1,

$$J(m_1^*, m_2^*) = J(u_1^*, u_2^*). \quad (3.1)$$

Proof: (a) Suppose $J(m_1^*, m_2^*) \geq J(u_1^*, u_2^*)$.

From ([4]), there exists a $u_{1\varepsilon} \in U_1$ such that

$$\left| J(m_1^*, u_2^*) - J(u_{1\varepsilon}, u_2^*) \right| < \varepsilon. \quad (3.2)$$

From the definition of $J(u_1^*, u_2^*)$ and $J(m_1^*, m_2^*)$, we have

$$J(u_1^*, u_2^*) \geq J(u_{1\varepsilon}, u_2^*) \quad (3.3)$$

$$J(m_1^*, u_2^*) \geq J(m_1^*, m_2^*). \quad (3.4)$$

Adding (3.3) and (3.4),

$$J(u_1^*, u_2^*) + J(m_1^*, u_2^*) \geq J(u_{1\varepsilon}, u_2^*) + J(m_1^*, m_2^*)$$

which implies

$$J(m_1^*, u_2^*) - J(u_{1\varepsilon}, u_2^*) \geq J(m_1^*, m_2^*) - J(u_1^*, u_2^*) \geq 0, \text{ by assumption}$$

implies

$$\varepsilon > \left| J(m_1^*, u_2^*) - J(u_{1\varepsilon}, u_2^*) \right| \geq \left| J(m_1^*, m_2^*) - J(u_1^*, u_2^*) \right|$$

which implies $J(m_1^*, m_2^*) = J(u_1^*, u_2^*)$, as ε is arbitrary.

(b) Suppose $J(m_1^*, m_2^*) \leq J(u_1^*, u_2^*)$.

Let $u_{2\varepsilon} \in U_2$, such that

$$\left| J(u_1^*, m_2^*) - J(u_1^*, u_{2\varepsilon}) \right| < \varepsilon$$

as before

$$J(u_1^*, u_2^*) \leq J(u_1^*, u_{2\varepsilon}) \quad (3.5)$$

$$J(u_1^*, m_2^*) \leq J(m_1^*, m_2^*) \quad (3.6)$$

implies

$$0 \leq J(u_1^*, u_2^*) - J(m_1^*, m_2^*) \leq J(u_1^*, u_{2\varepsilon}) - J(u_1^*, m_2^*) < \varepsilon$$

implies

$$J(m_1^*, m_2^*) = J(u_1^*, u_2^*).$$

Hence, the proof.

4. WEAK CONVERGENCE PRELIMINARIES

Let $D^d[0, \infty)$ denote the space of R^d valued functions which are right continuous and have left-hand limits endowed with the Skorohod topology. Let $\{\mathcal{F}_t^\varepsilon\}$ denote the minimal σ -algebra over which $\{x^\varepsilon(s), \xi^\varepsilon(s), s \leq t\}$ is measurable, and let E_t^ε denote the expectation conditioned on $\mathcal{F}_t^\varepsilon$. Let \tilde{M} denote the set of real valued functions of (ω, t) that are nonzero only on a bounded t -interval. Let

$$\bar{M}^\varepsilon = \left\{ f \in \tilde{M}; \sup_t E|f(t)| < \infty \text{ and } f(t) \text{ is } \mathcal{F}_t^\varepsilon \text{ measurable} \right\}.$$

Let $f(\cdot), f^\Delta(\cdot) \in \bar{M}^\varepsilon$, for each $\Delta > 0$. Then $f = p - \lim_\Delta f^\Delta$ if and only if

$$\sup_{t, \Delta} E|f^\Delta(t)| < \infty$$

and $\lim_{\Delta \rightarrow 0} E|f(t) - f^\Delta(t)| = 0$, for each t . $f(\cdot)$ is said to be in the domain of \hat{A}^ε ,

i.e., $f(\cdot) \in D(\hat{A}^\varepsilon)$, and $\hat{A}^\varepsilon f = g$ if

$$p - \lim_{\Delta \rightarrow 0} \left(\frac{E_t^\varepsilon f(t + \Delta) - f(t)}{\Delta} - g(t) \right) = 0.$$

If $f(\cdot) \in D(\hat{A}^\varepsilon)$, then

$$f(t) - \int_0^t \hat{A}^\varepsilon f(u) du \text{ is a martingale,}$$

and

$$E_t^\varepsilon f(t + s) - f(t) = \int_t^{t+s} E_u^\varepsilon \hat{A}^\varepsilon f(u) du, \text{ w.p.1.}$$

The following result from [6] will be used to conclude that various terms will go to zero in probability.

Lemma 4.1. Let $\xi(\cdot)$ be ϕ -mixing process with mixing rate $\phi(\cdot)$, and let $h(\cdot)$ be a function of ξ which is bounded and measurable on \mathcal{T}_t^∞ . Then, there exist $K_i, i = 1, 2, 3$ such that

$$\left| E\left(h(t+s)/\mathcal{T}_0^t\right) - Eh(t+s) \right| \leq K_1 \phi(s).$$

If $t < u < v$, and $Eh(s) = 0$ for all s , then,

$$\left| E\left(h(u)h(v)/\mathcal{T}_\tau^t\right) - Eh(u)h(v) \right| \leq \begin{cases} K_2 o(v-u), & u < \tau < v \\ K_3 o(u-t), & t < \tau < u \end{cases},$$

where $\mathcal{T}_\tau^t = \sigma\{\xi(s); \tau \leq s \leq t\}$.

The process $x^{\varepsilon,K}(t)$ is said to be the K -truncation of $x^\varepsilon(\cdot)$. Let

$$q^K(x) = \begin{cases} 1 & \text{for } x \in S_K \\ 0 & \text{for } x \in R^d - S_{K+1} \\ \text{smooth} & \text{otherwise.} \end{cases}$$

Define $a_K(x, \alpha) = a(x, \alpha)q^K(x)$ and $g_K(x, \xi) = g(x, \xi)q^K(x)$. Let $x^{\varepsilon,K}(\cdot)$ denote the solution of (2.7) corresponding to the use of truncated coefficients. Then $x^{\varepsilon,K}(\cdot)$ is bounded uniformly in t and $\varepsilon > 0$.

For proof of the main weak convergence result, Theorem 5.1, we will use following results from [6].

Lemma 4.2. Let $\{z^\varepsilon(\cdot)\}$ be tight on $D^d[0, \infty)$. Suppose that for each $f(\cdot) \in C_0^3$, and each $T < \infty$, there exist $f^\varepsilon(\cdot) \in D(\hat{A}^\varepsilon)$ such that

$$p - \lim\left(f^\varepsilon(\cdot) - f(z^\varepsilon(\cdot))\right) = 0 \quad (4.1)$$

and

$$p - \lim_\varepsilon \left(\hat{A}^\varepsilon f^\varepsilon(\cdot) - \hat{A} f(z^\varepsilon(\cdot))\right) = 0. \quad (4.2)$$

Then $z^\varepsilon(\cdot) \Rightarrow z(\cdot)$, the solution of the martingale problem for the operator A .

Lemma 4.3. Let the K -truncations $\{z^{\varepsilon,K}\}$ be tight for each K , and that the martingale problem for the diffusion operator A have a unique solution $z(\cdot)$ for each initial condition. Suppose that $z^K(\cdot)$ is a K -truncation of $z(\cdot)$ and it solves the martingale problem for operator A^K . For each K and $f(\cdot) \in D$, let there be $f^\varepsilon(\cdot) \in D(A^\varepsilon)$ such that (4.1) and (4.2) hold with $z^{\varepsilon,K}(\cdot)$ and A^K replacing z^ε and A , respectively. Then $z^\varepsilon(\cdot) \Rightarrow z(\cdot)$.

5. MAIN RESULT

In this section, we will prove the weak convergence of the wide-band system (2.7) to the diffusion system (2.1) and the δ -optimality of the equilibrium strategies of (2.1) applied to (2.7). We will use following assumptions, which are very general. For a detailed description on these types of assumptions, we refer to [6] and [7].

(A1) $a_i(\cdot, \cdot)$, $i = 1, 2$, $G(\cdot, \cdot)$, $g(\cdot, \cdot)$, $g_x(\cdot, \cdot)$ are continuous and are bounded by $O(1 + |x|)$. $G_x(\cdot, \xi)$ is continuous in x for each ξ and is bounded. $\xi(\cdot)$ is bounded, right continuous, and $EG(x, \xi(t)) \rightarrow 0$, $Eg(x, \xi(t)) \rightarrow 0$ as $t \rightarrow \infty$, for each x . Also, $r(\cdot, \cdot)$ is bounded and continuous.

(A2) $g_{xx}(\cdot, \xi)$ is continuous for each ξ , and is bounded.

(A3) Let $W(x, \xi)$ denote either $\varepsilon G(x, \xi)$, $G_x(x, \xi)$, $g(x, \xi)$ or $g_x(x, \xi)$. Then for compact Q ,

$$\varepsilon \sup_{x \in Q} \left| \int_t^\infty E_s^\varepsilon W(x, \xi(s)) ds \right| \xrightarrow{\varepsilon \rightarrow 0} 0$$

in the mean square sense, uniformly in t .

(A4) Let g_i denote the i th component of g . There are continuous $\bar{g}_i(\cdot)$, $b(\cdot) = \{b_{ij}(\cdot)\}$ such that

$$\int_t^\infty Eg_{i,x}(x, \xi(s)) g(x, \xi(t)) ds \rightarrow \bar{g}_i(x),$$

$$\int_t^\infty Eg_i(x, \xi(s)) g_j(x, \xi(t)) ds \rightarrow \frac{1}{2} b_{ij}(x),$$

as $t \rightarrow \infty$, and the convergence is uniform in any bounded x -set.

Note: Let $b(x) = \{b_{ij}(x)\}$. For $i \neq j$, it is not necessary that $b_{ij} = b_{ji}$. In that case define $\tilde{b}(x) = \frac{1}{2} [b(x) + b'(x)]$ as the symmetric covariance matrix, then use b for the new \tilde{b} . Hence, for notational simplicity, we will not distinguish between $b(x)$ and $\tilde{b}(x)$.

(A5) For each compact set Q and all i, j ,

$$(a) \sup_{x \in Q} \varepsilon^2 \left| \int_t^\infty d\tau \int_\tau^\infty ds \left[E_{t/\varepsilon^2} g'_{i,x}(x, \xi(s)) g(x, \xi(t)) - Eg'_{i,x}(x, \xi(s)) g(x, \xi(t)) \right] \right| \rightarrow 0;$$

$$(b) \sup_{x \in Q} \varepsilon^2 \left| \int_t^\infty d\tau \int_\tau^\infty ds \left[E_{t/\varepsilon^2} g_i(x, \xi(s)) g_j(x, \xi(t)) - Eg_i(x, \xi(s)) g_j(x, \xi(t)) \right] \right| \rightarrow 0;$$

in the mean square sense as $\varepsilon \rightarrow 0$, uniformly in t .

Define $\bar{a}(x, \alpha) = a_1(x, \alpha_1) + a_2(x, \alpha_2) + \bar{g}(x)$ and the operator A'' as

$$A^m f(x) = \int A^\alpha f(x) m_x(d\alpha),$$

where

$$A^\alpha f(x) = f'_x(x) \bar{a}(x, \alpha) + \frac{1}{2} \sum_{i,j} b_{ij}(x) f_{x_i x_j}(x).$$

For a fixed control α , A^α will be the operator of the process that is the weak limit of $\{x^\varepsilon(\cdot)\}$.

(A6) The martingale problem for operator A^m has a unique solution for each relaxed admissible Markov strategy $m_x(\cdot)$, and each initial condition. The process is a Feller process. The solution of (2.7) is unique in the weak sense for each $\varepsilon > 0$. Also $b(x) = \sigma(x) \sigma'(x)$ for some continuous finite dimensional matrix $\sigma(\cdot)$.

For an admissible relaxed policy for (2.7) and (2.1), respectively, define the occupation measure valued random variables $P_T^{m,\varepsilon}(\cdot)$ and $P_T^m(\cdot)$ by, respectively,

$$\begin{aligned} P_T^{m,\varepsilon}(B \times C) &= \frac{1}{T} \int_0^T I_{\{x^\varepsilon(t) \in B\}} m_t^\varepsilon(C) dt, \\ P_T^m(B \times C) &= \frac{1}{T} \int_0^T I_{\{x(t) \in B\}} m_t(C) dt \end{aligned}$$

where B and C are Borel subsets in \mathbb{R}^d and $[0, t] \times U$, respectively.

Let $\{m^\varepsilon(\cdot)\}$ be a given sequence of admissible relaxed controls.

(A7) For a fixed $\delta > 0$,

$$\{x^\varepsilon(t), \text{ small } \varepsilon > 0, t \in \text{dense set in } [0, \infty), m^\varepsilon \text{ used}\}$$

are tight.

Note: The assumption (A7) implies that the set of measure valued random variables

$$\{P_T^{m^\varepsilon, \varepsilon}(\cdot), \text{ small } \varepsilon > 0, T < \infty\}$$

are tight.

(A8) For the *ergodic equilibrium* pair of Markov strategies $m^* = (m_1^*, m_2^*)$ with initial law π for (2.1) and (2.2), the martingale problem has a unique solution. The solution is a Feller process and there is a unique invariant measure $\mu(m^*)$.

Note: Existence of such an invariant measure is assured if the process is positive recurrent. Also, under the conditions of Theorem 2.1, the assumption (A8) will follow.

The following result gives the main convergence and δ -optimality result for the ergodic payoff criterion.

Theorem 5.1. Assume (A1) to (A8). Let $(m_1^{*\varepsilon}, m_2^{*\varepsilon})$ be the policy pair (m_1^*, m_2^*) adaptively applied to (2.7) and (2.8). Then $\{x^\varepsilon(\cdot), m_1^{*\varepsilon}, m_2^{*\varepsilon}\} \Rightarrow (x(\cdot), m_1^*, m_2^*)$ (in the Skorohod topology) and there is a Wiener process $w(\cdot)$ such that $(x(\cdot), m_1^*, m_2^*)$ is nonanticipative with respect to $w(\cdot)$, and (2.1) holds. Also,

$$J^\varepsilon(m_1^{*\varepsilon}, m_2^{*\varepsilon}) \xrightarrow{P} J(m_1^*, m_2^*) = V(\pi). \quad (5.1)$$

In addition, let $(\hat{m}_1^\varepsilon(\cdot), m_2^\varepsilon(\cdot))$ be a δ -optimal strategy pair for player I and $(m_1^\varepsilon(\cdot), \hat{m}_2^\varepsilon(\cdot))$ be δ -optimal pair for player II for $x^\varepsilon(\cdot)$ of (2.7). Then

$$\lim_{\varepsilon, T} P \left\{ \left| J^\varepsilon(m_1^{*\varepsilon}, m_2^{*\varepsilon}) - J^\varepsilon(\hat{m}_1^\varepsilon(\cdot), m_2^\varepsilon(\cdot)) \right| < \delta \right\} = 1 \quad (5.2)$$

and

$$\lim_{\varepsilon, T} P \left\{ \left| J^\varepsilon(m_1^{*\varepsilon}, m_2^{*\varepsilon}) - J^\varepsilon(m_1^\varepsilon(\cdot), \hat{m}_2^\varepsilon(\cdot)) \right| < \delta \right\} = 1. \quad (5.3)$$

Proof. The correct procedure of proof is to work with the truncated processes $x^{\varepsilon, K}(\cdot)$ and to use the piecing together idea of Lemma 4.3 to get convergence of the original $x^\varepsilon(\cdot)$ sequence, unless $x^\varepsilon(\cdot)$ is bounded on each $[0, T]$, uniformly in ε . For notational simplicity, we ignore this technicality. Simply suppose that $x^\varepsilon(\cdot)$ is bounded in the following analysis. Otherwise, one can work with K -truncation. Let \hat{D} be a measure determining set of bounded real-valued continuous functions on R^d having continuous second partial derivatives and compact support. Let $m_i^\varepsilon(\cdot)$ be relaxed Markov policies of (A7). Whenever convenient, we write $x^\varepsilon(t) = x$. For the test function $f(\cdot) \in \hat{D}$, define the perturbed test functions (the change of variable $s/\varepsilon^2 \rightarrow s$ will be used through out the proofs)

$$\begin{aligned} f_0^\varepsilon(x, t) &= \int_t^\infty E_t^\varepsilon f'_x(x) G(x, \xi^\varepsilon(s)) ds \\ &= \varepsilon^2 \int_{t/\varepsilon^2}^\infty E_t^\varepsilon f'_x(x) G(x, \xi(s)) ds \\ f_1^\varepsilon(x, t) &= \frac{1}{\varepsilon} \int_t^\infty E_t^\varepsilon f'_x(x) g(x, \xi^\varepsilon(s)) ds \\ &= \varepsilon \int_{t/\varepsilon^2}^\infty E_t^\varepsilon f'_x(x) g(x, \xi(s)) ds \end{aligned}$$

$$\begin{aligned}
f_2^\varepsilon(x, t) &= \frac{1}{\varepsilon^2} \int_t^\infty ds \int_s^\infty d\tau \left\{ E_t^\varepsilon \left[f'_x(x) g(x, \xi^\varepsilon(\tau)) \right]_x' g(x, \xi^\varepsilon(s)) \right. \\
&\quad \left. - E \left[f'_x(x) g(x, \xi^\varepsilon(\tau)) \right]_x' g(x, \xi^\varepsilon(s)) \right\} \\
&= \varepsilon^2 \int_{t/\varepsilon^2}^\infty ds \int_s^\infty d\tau \left\{ E_t^\varepsilon \left[f'_x(x) g(x, \xi(\tau)) \right]_x' g(x, \xi(s)) \right. \\
&\quad \left. - E \left[f'_x(x) g(x, \xi(\tau)) \right]_x' g(x, \xi(s)) \right\}.
\end{aligned}$$

From (A1), (A2), (A3), and (A5), $f_i^\varepsilon(\cdot) \in D(A^\varepsilon)$ for $i = 0, 1, 2$. Define the perturbed test function

$$f^\varepsilon(t) = f(x^\varepsilon(t)) + \sum_{i=0}^2 f_i^\varepsilon(x^\varepsilon(t), t).$$

The reason for defining f_i^ε is to facilitate the averaging of the “noise” terms involving ξ^ε terms. By the definition of the operator A^ε and its domain $D(A^\varepsilon)$, we will obtain that $f(x^\varepsilon(\cdot))$ and the $f_i^\varepsilon(x^\varepsilon(\cdot), \cdot)$ are all in $D(A^\varepsilon)$, and

$$\begin{aligned}
A^{m^\varepsilon, \varepsilon} f(x^\varepsilon(t)) &= f'_x(x^\varepsilon(t)) \left[\sum_{i=1}^2 \int a_i(x^\varepsilon(t), \alpha) m_{ii}^\varepsilon(d\alpha) + G(x^\varepsilon(t), \xi^\varepsilon(t)) \right. \\
&\quad \left. + \frac{1}{\varepsilon} g(x^\varepsilon(t), \xi^\varepsilon(t)) \right]. \tag{5.4}
\end{aligned}$$

From this we can obtain,

$$\begin{aligned}
A^{m^\varepsilon, \varepsilon} f_0(x^\varepsilon(t)) &= -f'_x(x^\varepsilon(t)) G(x^\varepsilon(t), \xi^\varepsilon(t)) \\
&\quad + \int_t^\infty ds \left[E_t^\varepsilon f'_x(x^\varepsilon(t)) G(x^\varepsilon(t), \xi^\varepsilon(s)) \right]_x' x^\varepsilon(t) \\
&= -f'_x(x^\varepsilon(t)) G(x^\varepsilon(t), \xi^\varepsilon(t)) \\
&\quad + \varepsilon^2 \int_{t/\varepsilon^2}^\infty ds \left[E_t^\varepsilon f'_x(x^\varepsilon(t)) G(x^\varepsilon(t), \xi(s)) \right]_x' x^\varepsilon(t). \tag{5.5}
\end{aligned}$$

Note that the first term in (5.5) will cancel with $f'_x G$ term of (5.4). The $p - \varepsilon$ limit of the last term in (5.5) is zero.

$$\begin{aligned}
A^{m^\varepsilon, \varepsilon} f_1(x^\varepsilon(t)) &= -\frac{1}{\varepsilon} f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(t)) \\
&\quad + \frac{1}{\varepsilon} \int_t^\infty ds \left[E_t^\varepsilon f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(s)) \right]_x' x^\varepsilon(t) \\
&= -\frac{1}{\varepsilon} f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(t)) \\
&\quad + \varepsilon \int_{t/\varepsilon^2}^\infty ds \left[E_t^\varepsilon f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(s)) \right]_x' x^\varepsilon(t).
\end{aligned} \tag{5.6}$$

The first term on the right of (5.6) will cancel with the $\frac{f'_x g}{\varepsilon}$ term in (5.4). The only component of the second term on the right of (5.6) whose $p - \varepsilon$ lim is not zero is

$$\frac{1}{\varepsilon^2} \int_t^\infty ds \left\{ E_t^\varepsilon \left[f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(s)) \right]_x' g(x^\varepsilon(t), \xi^\varepsilon(t)) \right\}$$

This term will cancel with the first term of (5.7).

$$\begin{aligned}
A^{m^\varepsilon, \varepsilon} f_2(x^\varepsilon(t)) &= -\frac{1}{\varepsilon^2} \int_t^\infty ds \left\{ E_t^\varepsilon \left[f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(s)) \right]_x' g(x^\varepsilon(t), \xi^\varepsilon(t)) \right. \\
&\quad \left. - E \left[f'_x(x^\varepsilon(t)) g(x, \xi^\varepsilon(s)) \right]_x' g(x, \xi^\varepsilon(t)) \Big|_{x=x^\varepsilon(t)} \right\} \\
&\quad + \left[f_2^\varepsilon(x, t) \right]_x' x^\varepsilon \Big|_{x=x^\varepsilon(t)} \\
&= - \int_{t/\varepsilon^2}^\infty ds \left\{ E_t^\varepsilon \left[f'_x(x^\varepsilon(t)) g(x^\varepsilon(t), \xi^\varepsilon(x)) \right]_x' g(x^\varepsilon(t), \xi^\varepsilon(t)) \right. \\
&\quad \left. - E \left[f'_x(x^\varepsilon(t)) g(x, \xi^\varepsilon(s)) \right]_x' g(x, \xi^\varepsilon(t)) \Big|_{x=x^\varepsilon(t)} \right\} \\
&\quad + \left[f_2^\varepsilon(x, t) \right]_x' x^\varepsilon \Big|_{x=x^\varepsilon(t)}.
\end{aligned} \tag{5.7}$$

The $p - \varepsilon$ lim of the last term of the right side of (5.7) is zero.

Evaluating $A^{m^\varepsilon, \varepsilon} f^\varepsilon(t) = A^{m^\varepsilon, \varepsilon} \left[f(x^\varepsilon(t)) + \sum_{i=0}^2 f_i^\varepsilon(x^\varepsilon(t), t) \right]$ and by deleting

terms that cancel yields

$$\begin{aligned}
A^{m^\varepsilon, \varepsilon} f^\varepsilon(t) &= f'_x(x^\varepsilon(t)) \sum_{i=1}^2 \int a_i(x^\varepsilon(t), \alpha) m_{ii}^\varepsilon(d\alpha) \\
&\quad + \int_{t/\varepsilon^2}^\infty E \left[f'_x(x^\varepsilon(t)) g(x, \xi^\varepsilon(s)) \right]_x' g(x, \xi^\varepsilon(t/\varepsilon^2)) ds.
\end{aligned} \tag{5.8}$$

As a result, we get

$$p - \lim \left(f^\varepsilon(t) - f(x^\varepsilon(\cdot)) \right) = 0 \quad (5.9)$$

$$p - \lim_{\varepsilon} \left| A^{m^\varepsilon, \varepsilon} f(x^\varepsilon(t)) - A^{m^\varepsilon, \varepsilon} f^\varepsilon(t) \right| = 0. \quad (5.10)$$

Hence, by Lemma 4.2,

$$M_f^\varepsilon(t) = f^\varepsilon(t) - f^\varepsilon(0) - \int_0^t A^{m^\varepsilon} f^\varepsilon(s) ds$$

is a zero mean martingale.

Let $[t]$ denote the greatest integer part of t . Write

$$\frac{M_f^\varepsilon(t)}{t} = \frac{1}{t} \left[(M_f^\varepsilon(t) - M_f^\varepsilon([t])) + M_f^\varepsilon(0) \right] + \frac{1}{t} \sum_{k=0}^{[t]-1} [M_f^\varepsilon(k+1) - M_f^\varepsilon(k)].$$

Using the fact that $f(\cdot)$ is bounded and (5.10), and martingale property of $M_f^\varepsilon(\cdot)$,

we get $E \left[\frac{M_f^\varepsilon(t)}{t} \right]^2 \rightarrow 0$ as $t \rightarrow \infty$ and $\varepsilon \rightarrow 0$, which in turn implies that

$\frac{M_f^\varepsilon(t)}{t} \xrightarrow{P} 0$ as $t \rightarrow \infty$ and $\varepsilon \rightarrow 0$ in any way at all. From (5.10), and the fact that $\frac{M_f^\varepsilon(t)}{t}$, $\frac{f^\varepsilon(t)}{t}$, and $\frac{f^\varepsilon(0)}{t}$ all go to zero in probability implies that as $t \rightarrow \infty$ and $\varepsilon \rightarrow 0$,

$$\frac{1}{t} \int_0^t A^{m^\varepsilon} f(x^\varepsilon(s)) ds \xrightarrow{P} 0. \quad (5.11)$$

By the definition of $P_T^{m^\varepsilon, \varepsilon}(\cdot)$, (5.11) can be written as

$$\int A^\alpha f(x) P_T^{m^\varepsilon, \varepsilon}(dx d\alpha) \xrightarrow{P} 0 \text{ as } T \rightarrow \infty \text{ and } \varepsilon \rightarrow 0. \quad (5.12)$$

For the policy $m^*(\cdot)$, choose a weakly convergent subsequence of set of random variables $\{P_T^{m^*, \varepsilon}(\cdot, \varepsilon, T)\}$, indexed by ε_n, T_n , with limit $\hat{\mu}(\cdot)$. Let this limit $\hat{P}(\cdot)$ be defined on some probability space $(\tilde{\Omega}, \tilde{P}, \tilde{\mathcal{F}})$ with generic variable $\tilde{\omega}$. Factor $\hat{P}(\cdot)$ as $\hat{P}(dx d\alpha) = m_x^*(d\alpha) \mu(dx)$. We can suppose that $m_x(C)$ are x -measurable for each Borel C and $\tilde{\omega}$. Now (5.12) implies that for all $f(\cdot) \in \hat{D}$,

$$\iint A^\alpha f(x) m_x^*(d\alpha) \hat{\mu}(dx) = 0 \text{ for } \tilde{P} \text{-almost all } \tilde{\omega}. \quad (5.13)$$

Since $f(\cdot)$ is measure determining, (5.13) implies that almost all realizations of $\hat{\mu}$ are invariant measures for (2.1) under the relaxed policies m^* . By uniqueness of the invariant measure, we can take $\mu(m^*, \cdot) = \hat{\mu}(\cdot)$ does not depend on the chosen subsequence ε_n, T_n . By the definition of $P_T^{m^*, \varepsilon}(\cdot)$,

$$\begin{aligned} \frac{1}{t} \int_0^t \int r(x^\varepsilon(s), \alpha) m^{*\varepsilon}(d\alpha) ds &= \int_0^t \int r_k(x^\varepsilon(s), \alpha) P_T^{m^{*\varepsilon}}(d\alpha dx) \\ &\xrightarrow{P} \int_0^t \int r(x, \alpha) m_x^*(d\alpha) \hat{\mu}(dx) = J(m^*). \end{aligned}$$

Hence, we have (5.1). Let $\tilde{m}^{\delta_i, \varepsilon} = (\hat{m}_1^\varepsilon(\cdot), m_2^\varepsilon(\cdot))$ and $\tilde{m}^{\delta_i, \varepsilon} = (m_1^\varepsilon(\cdot), \hat{m}_2^\varepsilon(\cdot))$ are the δ -optimal strategies for players I and II respectively. Now (5.2) and (5.3) follows using the fact that (5.1) holds for all the limits of the tight sets $\{P_T^{m^{\delta_i, \varepsilon}}(\cdot; \varepsilon, T)\}, i = 1, 2$, the assumed uniqueness in (A8), and the definition of δ -optimality.

It is important to note that, as a result of Theorem 5.1, if one needs a δ -optimal policy for the physical system, it is enough to compute for the diffusion model and use it to the physical system. There is no need to compute optimal policies for each ε .

6. STOCHASTIC GAMES

For the stochastic or the discrete parameter games, the system is given by

$$X_{n+1}^\varepsilon = X_n^\varepsilon + \varepsilon G(X_n^\varepsilon) + \varepsilon \sum_{i=1}^N \int a_i(X_n^\varepsilon, \alpha_i) m_{in}(d\alpha_i) + \sqrt{\varepsilon g}(X_n^\varepsilon, \xi_n^\varepsilon) \quad (6.1)$$

where $\{\xi_n^\varepsilon\}$ satisfies the discrete parameter version of (A2) and $m_{in}(\cdot), i = 1, \dots, N$ be the relaxed control strategies depending only on $\{X_i, \xi_{i-1}, i \leq n\}$. It should be noted that, in the discrete case, strategies would not be relaxed, one need to interpret this in asymptotic sense, i.e., the limiting strategies will be relaxed. Let E_n^ε denote the conditional expectation with respect to $\{X_i, \xi_{i-1}, i \leq n\}$. Define $x^\varepsilon(\cdot)$ by

$$\begin{aligned} x^\varepsilon(t) &= X_n^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon) \text{ and } m_i(\cdot) \text{ by} \\ m_i(B_i \times [0, t]) &= \varepsilon \sum_{n=0}^{\lfloor t/\varepsilon \rfloor - 1} m_{in}(B_i) + \varepsilon(t - \varepsilon[\lfloor t/\varepsilon \rfloor]) m_{\lfloor t/\varepsilon \rfloor}(B_i), i = 1, \dots, N. \end{aligned}$$

(B1)

- (i) For V equal either $a(\cdot, \cdot)$, g or g_x , and for Q compact,

$$E \sup_x \left| \sum_{n=L_1}^L E_n^\varepsilon V(x, \xi_n^\varepsilon) \right| \rightarrow 0, \text{ as } L, n \text{ and } L_1 \rightarrow \infty, \text{ with } L > n + L_1$$

and $L - (n + L_1) \rightarrow \infty$.
- (ii) There are continuous functions $c(i, x)$ and $c_0(i, x)$ such that for each x

$$\frac{1}{L} \sum_{n=\ell}^{\ell+L} E_\varepsilon^\varepsilon g\left(x, \xi_{n+i}^\varepsilon\right) g'\left(x, \xi_n^\varepsilon\right) \xrightarrow{P} c(i, x)$$

$$\frac{1}{L} \sum_{n=\ell}^{\ell+L} E_\varepsilon^\varepsilon g'_x\left(x, \xi_{n+i}^\varepsilon\right) g\left(x, \xi_n^\varepsilon\right) \xrightarrow{P} c_0(i, x)$$

as ℓ and $L \rightarrow \infty$.

(iii) For each $T < \infty$ and compact Q ,

$$\varepsilon \sup_{x \in Q} \left| \sum_{j=n}^{T/\varepsilon} \sum_{k=j+1}^{T/\varepsilon} \left[E_n^\varepsilon g'_{i,x}\left(x, \xi_k\right) g\left(x, \xi_j\right) - E g'_{i,x}\left(x, \xi_k\right) g\left(x, \xi_j\right) \right] \right| \rightarrow 0 \quad i \leq n,$$

$$\varepsilon \sup_{x \in Q} \left| \sum_{j=n}^{T/\varepsilon} \sum_{k=j+1}^{T/\varepsilon} \left[E_n^\varepsilon g'\left(x, \xi_k\right) g\left(x, \xi_j\right) - E g'\left(x, \xi_k\right) g\left(x, \xi_j\right) \right] \right| \rightarrow 0,$$

in the mean as $\varepsilon \rightarrow 0$ uniformly in $n \leq T/\varepsilon$. Also, the limits hold when the bracketed terms are replaced by their x -gradient/ $\sqrt{\varepsilon}$.

Define

$$\tilde{a}(x) = \sum_1^\infty c_0(i, x)$$

and

$$\bar{c}(x) = c(0, x) + 2 \sum_1^\infty c(i, x) = \sum_{-\infty}^\infty c(i, x).$$

With some minor modifications in the proof of Theorem (5.1), we can obtain the following result (refer to [6] and [9], for convergence proofs in similar situation).

Theorem 6.1. Assume (A1) to (A3), (A6) to (A8) and (B1). Then the conclusions of Theorem (5.1) hold for model (6.1).

7. OTHER EXTENSIONS

The results of this work can be directly applied to two person zero-sum differential games with pathwise discounted payoff structure, analogous to the results in [8]. Also, other payoff structures, such as finite horizon payoff, and payoff up to exit time can be handled by some minor modifications. If the coefficients in (6.1) are state dependent or even discontinuous, still we can obtain the results of this paper by adapting the methods of [9].

REFERENCES

- [1] Borkar, V.S. and Ghosh, M.K., Stochastic differential games: An occupation measure based approach, Journal of Optimization Theory and Applications, 73, 1992, 359-385.

- [2] Chitashvili, R.J. and Elbakidze, N.V., Optimal stopping by two players, Statistics and Control of Stochastic Processes, Steklov Seminar, Ed. N.V. Krylov, R. Sh. Lipster, and A.A. Novikov, 1984, 10-53.
- [3] Elliott, R.J. and Davis, M.H.A., Optimal play in a stochastic differential game, SIAM J. Control Opt., 19, 1981, 543-554.
- [4] Fleming, W.H., Generalized solutions in optimal stochastic control, Proc. URI Conf. On Control, 1982, 147-165.
- [5] Krylov, N.V., *Controlled Diffusion Processes*, Springer-Verlag, 1980.
- [6] Kushner, H.J., *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*, MIT Press, 1984.
- [7] Kushner, H.J. and Dupuis, P.G., *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, 1992.
- [8] Ramachandran, K.M., Stochastic differential games with a small parameter, Stochastics and Stochastics Reports, 43, 1993, 73-91.
- [9] Ramachandran, K.M., Discrete parameter singular control problem with state dependent noise and non-smooth dynamics, Stochastic Analysis and Applications, 12, 1994, 261-276.
- [10] Krasovskii, N.N. and Subbotin, A.I., *Game Theoretical Control Problems*, Springer-Verlag, 1988.
- [11] Kushner, H.J., *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhauser, 1990.
- [12] Leitmann, G., *Multicriteria Decision Making and Differential Games*, Plenum Press, 1976.

22 INTERACTION OF SURFACE RADIATION WITH NATURAL CONVECTION

N. Ramesh, C. Balaji

and

S.P. Venkateshan

Heat Transfer and Thermal Power Laboratory
Department of Mechanical Engineering
Indian Institute of Technology Madras
Chennai 600 036 INDIA

INTRODUCTION

Natural convection heat transfer occupies a pride of place in contemporary convective heat transfer research. The applications of natural convection are seemingly endless. Electronic cooling, nuclear reactor cooling and insulation design are a few of them. The importance of natural convection stems from the essential fact that natural convection systems are noiseless, reliable and are capable of performing as stand-alone systems, although the heat fluxes associated with them are low, typically a few hundred Watts per square meter.

Natural convection flows arise because of self-induced forces, due either to temperature or concentration gradients. However, attention will be restricted on the former in the present paper. Buoyancy induced laminar flows over vertical surfaces have been considered previously by a number of researchers. Such flows inside enclosures have also been widely studied. However, in most of the studies, boundary conditions of the first and second kind have been assumed at the walls. In many practical applications, like the double pane window for example, the walls are under thermal balance between more than one mode of heat transfer. Particularly, when the wall under consideration is made of a metal with a non-zero surface emissivity, there is an interaction between all the three modes of heat transfer – conduction inside the wall, convection from the surface and radiation from the surface. In view of this, it is imperative that the interaction of the various modes of heat transfer be accounted for. Correlations to determine the overall heat transfer must, in general, account for the interaction so that the analysis will be realistic and meaningful. Attention will be focused on this aspect, in the following geometries: (a) closed cavities, (b) open cavities and (c) L shaped corners. Comparison with experimental results will be

provided wherever possible. A new method called the semi-experimental method to solve this class of problems is also discussed in detail.

REVIEW OF LITERATURE

Research focusing on the interaction of all the modes of heat transfer in closed cavities and L corners are scarce. Scarcer still are studies, that provide, physical insights into the mechanisms of such interactions and those which give useful correlations for determining the overall heat transfer. Schematic of the geometries considered here along with the flow patterns are indicated in Figures 1-3.

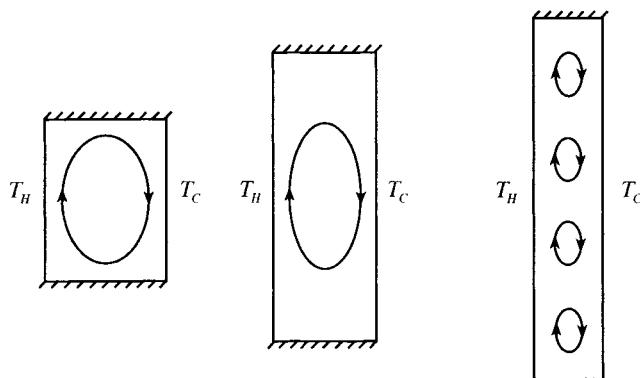


Figure 1. Schematic of flow patterns encountered in closed cavity flows.

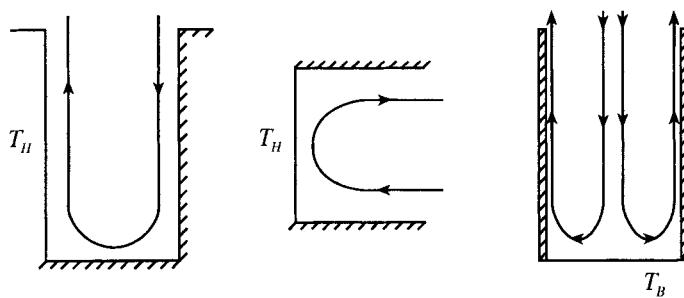


Figure 2. Schematic of flow patterns encountered in open cavity flows.

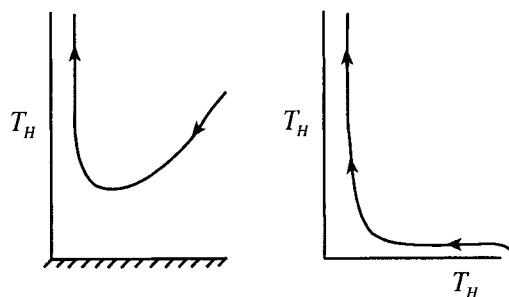


Figure 3. Schematic of flow patterns encountered in L shaped corners.

The literature on free convection in closed cavities is very vast. Therefore, the literature review is restricted to those that deal with studies focusing on interaction of free convection with radiation and/or conduction. Asako and Nakamura (1982) considered the effect of surface radiation in parallelogram shaped enclosures. Various emissivities were considered and the qualitative effect of radiation was brought out by a parametric study. However, a relatively coarse uniform grid (10x10) was employed in the computations. Also, correlations for determining the overall Nusselt number were not presented. It was observed that radiation tends to suppress natural convection, while compensating to some extent, by direct radiant heat transfer across the cavity. This feature was noted also by Yang and Lloyd (1985) and Lauriat (1991). Kim and Viskanta (1984) studied the effect of wall conduction and surface radiation in a square annulus with very thick walls on all sides. However, comprehensive correlations to predict overall heat transfer were not presented. The review of the literature on closed cavities suggests that a comprehensive study of interaction of radiation and free convection, with a view to develop correlations, could be a potentially useful addition to the literature.

The term open cavity can refer to a cavity open at the top or the side. Among the various studies on this geometry, only a few have considered the interaction of convection with radiation and/or conduction. Behnia and de Vahl Davis (1990), for instance, considered the cooling of a micro-component placed in a cavity with open top end. The qualitative effect of the thermal conductivity of the wall material on heat transfer was brought out from a parametric study. Lage et al. (1992) investigated the effect of surface radiation on free convection from a cavity with open top end, simulating an ash hopper commonly used in power plants. A simple model was employed for radiation which was assumed to be decoupled from convection. They concluded that radiation contribution was insignificant. However, Balaji and Venkateshan (1993) in their discussion on Lage et al. have concluded that radiation can contribute up to about 50% of the total heat transfer. Recently Lin et al. (1994) investigated the effect of surface radiation on turbulent free convection in a cavity open at the side. They concluded that surface radiation cannot be neglected, since it contributed substantially to the overall heat transfer.

Under the influence of both conduction and radiation, a cavity with open top end may simulate a typical section in an extruded heat sink or an intermediate fin in a fin array consisting of vertical fins on a horizontal base. Among the various studies, those of Starner and McManus (1963) and Harahap and McManus (1967) are of relevance and merit attention. These studies were similar in conception and they arrived at a broad conclusion that the height of the fin array and the ratio of height to depth were the important parameters that decided the nature of flow. For very low values of height to depth ratio, an up and down flow pattern was observed. Sobhan et al. (1990) investigated the effect of fin material for a single fin and an array of four vertical fins standing on a horizontal base using differential interferometry. The height to depth ratio was 1.4 and, as expected, the flow is three dimensional. Rammohan Rao and Venkateshan (1996) extended the study of Sobhan et al. to investigate the interaction of radiation by coating the fins to obtain four different emissivities ranging from 0.05 to 0.85. The study demonstrated that radiation can contribute up to 45% of the total heat transfer in the temperature range from room temperature to 100°C.

Smith et al. (1991) considered combined mode heat transfer in an electronic chassis and concluded that radiation was the dominant mode of heat transfer.

However, it can be seen that a comprehensive treatment of the interaction of free convection and/or conduction remains largely unexplored. Also, comprehensive correlations for overall heat transfer in such situations are conspicuous by their absence in the literature. With regard to the literature on L corners, studies reporting the influence of radiation and conduction are not available. However, the work of Rodigheiro and de Socio (1983) is worth mentioning as it provides a correlation for calculating free convection heat transfer in L corners through their experimental investigation. Angirasa and Mahajan (1993) too provided a correlation based on a numerical study of free convection in L corner with adiabatic and cold isothermal horizontal walls, using the method of finite differences.

In conclusion, it is clear that a considerable amount of work needs to be done in the case of open cavities and L corners. Even the classical closed cavity problem offers new vistas for research, in that, the combined mode problem has hitherto been ignored. Hence, an earnest attempt to explore the interaction between the various modes of heat transfer in the above mentioned geometries in order to gain insight into the nature of interaction, and, develop useful correlations for computing the overall heat transfer are the main thrusts of the present paper, which is based on the research carried out by our group in the Heat Transfer and Thermal Power Laboratory, Indian Institute of Technology Madras, India.

NUMERICAL PROCEDURE

Formulation for Convection

The governing equations for two-dimensional, steady, incompressible, laminar, constant property flow under the Boussinesq approximation are the well known Navier – Stokes equations. In the present study, the vorticity stream function (ω, Ψ) approach has been used in view of recirculating nature of the flow in the problems considered. The geometry and the co-ordinate axes for a typical problem, a closed cavity, are given in Figure 4.

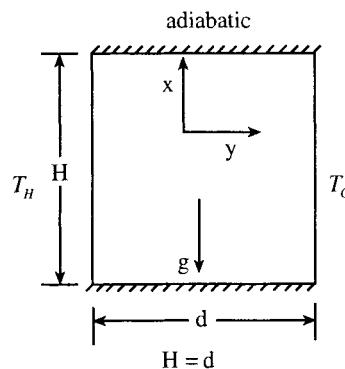


Figure 4. Geometry for the sidewall heated square cavity problem.

The governing equations in dimensionless form are:

$$U \frac{\partial \omega}{\partial X} + V \frac{\nabla \omega}{\partial Y} = \text{Pr} \left[\frac{\partial^2 \omega}{\partial X^2} + \frac{\partial^2 \omega}{\partial Y^2} \right] - Ra_d \frac{\partial \phi}{\partial Y} \quad (1)$$

$$\frac{\partial^2 \psi}{\partial X^2} + \frac{\partial^2 \psi}{\partial Y^2} = -\Pr \omega \quad (2)$$

$$U \frac{\partial \phi}{\partial X} + V \frac{\partial \phi}{\partial Y} = \frac{\partial^2 \phi}{\partial X^2} + \frac{\partial^2 \phi}{\partial Y^2}. \quad (3)$$

Formulation for Radiation

Since the medium inside the enclosure is air, which is radiatively non-participating, the present study considers only surface or wall radiation. The radiosity-irradiation technique is used for evaluating radiative heat fluxes. In the radiosity – irradiation method, each wall is divided into a finite number of regions or zones. Each such zone is treated to be isothermal and is considered as a separate entity, involved in exchange of radiation with other such zones. The leaving radiant heat flux of a wall element or radiosity as it is rightly called, is given by

$$J'_i = \varepsilon_i \sigma T_i^4 + (1 - \varepsilon_i) \sum_{j=1}^N F_{ij} J'_j \quad (4)$$

where i represents the i^{th} wall element and j represents any other wall element. The first term on the right hand side represents the emission component and the second represents the reflected component. ε_i refers to the total hemispherical emissivity. Radiation is assumed to be diffuse and independent of direction. Also, the surfaces are assumed to be gray. The summation term on the right hand side is commonly referred to as irradiation or incident radiant flux. F_{ij} represents the view factor from the i^{th} element to the j^{th} element. The view factors are evaluated using the well known Hottel's crossed string method (Hottel and Sarofim, 1967). The grids used for convection are retained for radiation to assure grid compatibility. This helps in obtaining stable and convergent solutions.

Boundary Conditions

There are two types of boundaries which are commonly encountered in heat and fluid flow problems. One is a solid wall and the other is a free or open boundary. The open boundary can be further classified as inflow or outflow boundary. For each of these, both momentum and temperature conditions will have to be specified to complete the problem formulation.

SOLID WALL

Momentum Conditions

The boundary conditions for vorticity and stream function have to be specified. Stream function is a constant on solid walls and is taken to be zero in the present case. Therefore $\psi = 0$ (on all solid walls).

As regards vorticity, the first assumption made is that the stream function equation is satisfied on the walls. The second assumption that is made is that gradients parallel to the wall are negligible in comparison with gradients perpendicular to the wall. By invoking such an assumption, the stream function equation reduces to

$$\frac{\partial^2 \psi}{\partial Y^2} = -\omega \text{Pr}. \quad (5)$$

Temperature Conditions

A very generalized form of the thermal boundary condition on a wall is derived in this section. The adiabatic wall, isothermal wall and all other thermal conditions actually turn out to be special cases of this. Consider a typical slice of length Δx on a wall of finite thickness $2t$ as shown in Figure 5.

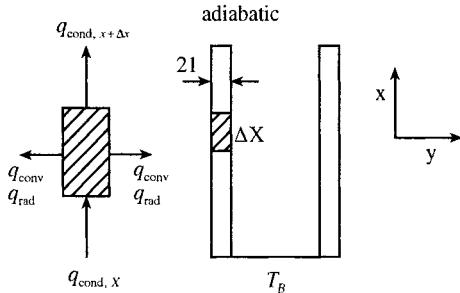


Figure 5. Schematic showing energy balance for a nodal element in the conjugate problem.

The wall is assumed very deep in the third direction. An energy balance on the element yields the following.

$$q_{cond,x} = q_{cond,x+\Delta x} + q_{conv} + q_{rad}. \quad (6)$$

Applying Fourier's law of heat conduction and noting that $q_{rad} = \varepsilon(\sigma T^4 - I')$, where I' refers to the elemental irradiation, the above equation in non-dimensional form becomes

$$\frac{\partial^2 \phi_s}{\partial X^2} = -\gamma \frac{\partial \phi}{\partial Y} + \varepsilon \gamma N_{RC} \left[\left(\frac{T}{T_{ref}} \right)^4 - I \right] \quad (7)$$

where T_{ref} is an appropriately chosen reference temperature that is specific to the problem. For example, in the case indicated in Figure 5, T_{ref} would be T_B and a reference temperature difference would be $T_B - T_\infty$ so that $\gamma = \frac{k_f d}{k_s t}$

$N_{RC} = \frac{\sigma T_B^4 d}{k_f (T_B - T_\infty)}$. γ represents the ratio of fluid conductance to solid conductance

and can be termed as the thermal conductivity parameter. N_{RC} , the radiation-conduction interaction parameter is the ratio of a representative black body emissive power to a representative conductive heat flux across the fluid layer. Several interesting, limiting cases of Equation (7) are possible and are given below.

1. In the absence of radiation ($\varepsilon = 0$) and when γ tends to zero (which implies that the solid is of very high thermal conductivity), the right hand side of Equation (7) vanishes and so the wall will be isothermal.

2. In the absence of radiation ($\varepsilon = 0$) and when γ tends to infinity, Equation (7) reduces to $\frac{\partial \phi}{\partial Y} = 0$. This represents an adiabatic wall.
3. In the absence of radiation ($\varepsilon = 0$) for finite values of γ , Equation (7) reduces to the classical conjugate convection problem.
4. In the presence of radiation ($\varepsilon \neq 0$) for γ tending to infinity, the wall conduction coupling becomes insignificant. The problem then becomes a convection radiation interaction problem. Such interactions are common for adiabatic walls and can be found in the analysis of such geometries as cavities, solar collectors, etc.
5. In the presence of radiation ($\varepsilon \neq 0$) for $\gamma = 0$, when there is no fluid, such as in outer space, the problem reduces to one of radiation conduction interaction. Such interactions occur in space radiators.
6. When $\varepsilon \neq 0$ and γ is finite, the interaction between all the three modes of heat transfer becomes important. It is apparent that in view of this, it represents the most general case in the class of interaction problems.

In the present study, various cases of this generalized interaction phenomenon are considered. The details of the thermal condition on the walls employed for the problems taken up for investigation will be clearly elucidated at appropriate places.

METHOD OF SOLUTION

Convection

The governing equations of the problem are non-linear partial differential equations. In the present study, a finite volume method based on Gosman et. al (1969) is employed. A detailed procedure highlighting the salient features of this method is given in Balaji (1994). Upwinding was used for the convection terms for obtaining stable solutions. The upwinding used here is analogous to a first order scheme of upwinding (Roache, 1982). Under-relaxation was used for the equations with a relaxation parameter of 0.5. This represents a balanced explicit – implicit scheme.

Radiation

The equation for i^{th} element is given in Equation (4). For a grid size of $N \times N$, there will be N^2 algebraic equations, and hence, N^2 unknown radiosities. The above equations were solved by the Gauss elimination technique, or Gauss Seidel method of iteration, depending upon the grid size employed in the problem. Generally, if the problem necessitates the use of more grids, the latter was employed, as the time consumed was found to be less and, round off errors were smaller, when compared with the elimination technique. The radiosity equation requires temperatures along various walls and these are taken as the previous iterate values. To this extent, the procedure can be deemed to be explicit. With the new

values of radiosities, the new values of temperatures are obtained. The procedure is repeated till convergence. Equation representing the energy balance in the wall (Equation (7)) shows that the wall temperature is dependent upon the convective and radiative heat transfer rates, and, in view of this, conduction, convection, and radiation are coupled.

RESULTS AND DISCUSSION

In this section, a few important results in each of the geometries considered will be presented. The first geometry taken up will be the square cavity.

Square Cavity

The problem geometry can be seen in Figure 4. The left wall is isothermal at a temperature T_H , the right wall at T_C , the top and bottom "float" at a temperature governed by a balance between convection and radiation. The appropriate boundary condition at the top and bottom are

$$\frac{\partial \phi}{\partial X} = \varepsilon N_{RC} \left[\left(\frac{T}{T_H} \right)^4 - I \right]. \quad (8)$$

This condition may be realized when the thermal conductivity of the wall material is very low. For the case of pure natural convection, however, $\varepsilon = 0$, and hence, Equation (8) reduces to $\frac{\partial \phi}{\partial X} = 0$ which represents an adiabatic wall. It is clear that in view of the above mentioned statements, steep gradients will be encountered near all the walls. Hence, non-uniform grids were employed for computation, as shown in Figure 6.

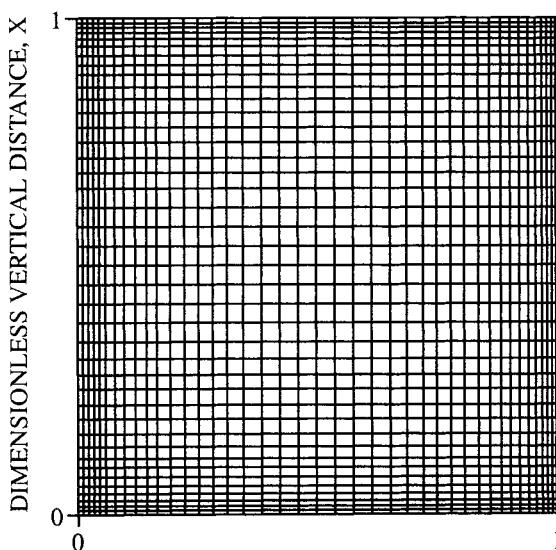


Figure 6. Grid pattern used in the square cavity problem.

A convergence criterion of 0.01% for all variables was used for $Ra_d = 1000$. The effect of surface radiation on convective heat transfer and the overall heat transfer is clear from Figures 7 and 8 respectively.

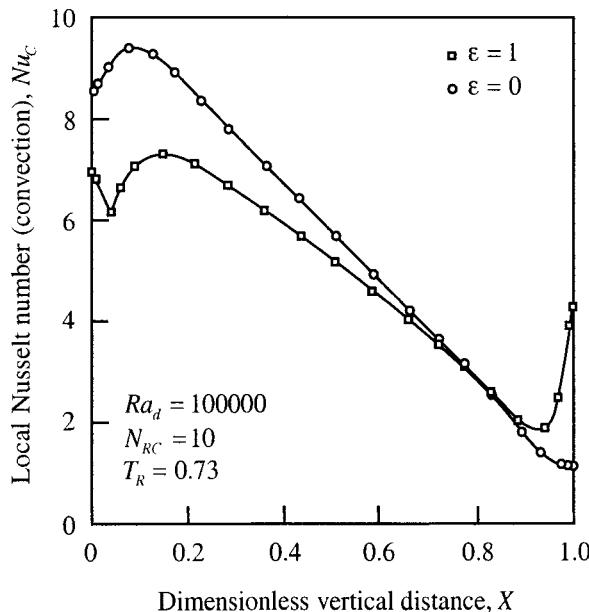


Figure 7. Effect of surface radiation on convection Nusselt number for a square cavity.

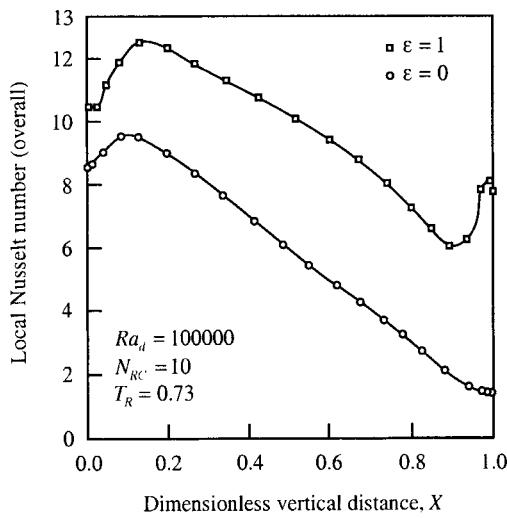


Figure 8. Effect of surface radiation on overall Nusselt number for a square cavity.

Radiation is found to reduce convective heat transfer but at the same time, the overall heat transfer is substantially increased because of surface radiation. This is the so called duality or non-linear effect of radiation. The variation of radiative Nusselt number along the hot wall for different values of emissivity is shown in Figure 9.

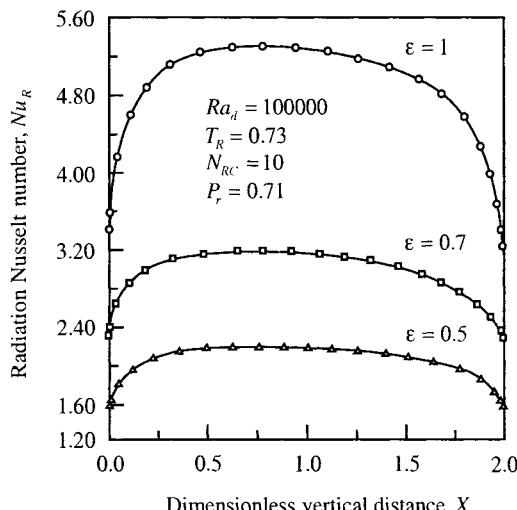


Figure 9. Variation of radiation Nusselt number with emissivity for a square cavity.

From the figure, it is clear that radiation Nusselt number varies strongly with position, contrary to the usual assumption of a uniform radiation Nusselt number for an isothermal surface.

SEMI - EXPERIMENTAL METHOD

In this method, the square cavity problem was again numerically solved using the finite volume method. However, the boundary temperatures along the top and bottom walls which are in a balance between convection, radiation and conduction are determined through experiments. Stated more explicitly, the temperature coupling along the top and bottom walls are actually taken care of by the experimentally determined temperatures. The temperatures are measured at several locations along the top and bottom walls. Curve fitting was used to convert these temperatures into polynomial functions. The order of the polynomial was varied from one to three, and in all cases the correlation coefficient of the fit was 98% and above. These measured temperatures were used as input data for the numerical solution, in order to obtain estimates of Nusselt numbers.

Two specific cases were considered for the present analysis : (a) the hot and cold vertical walls were highly polished ($\epsilon = 0.05$) and (b) the hot and cold vertical walls were coated with blackboard paint ($\epsilon = 0.85$). In both the cases, the top and bottom walls were coated with blackboard paint. Figures 10 and 11 show the variation of

dimensionless temperature along the top and bottom walls of the enclosure, for cases (a) and (b) respectively.

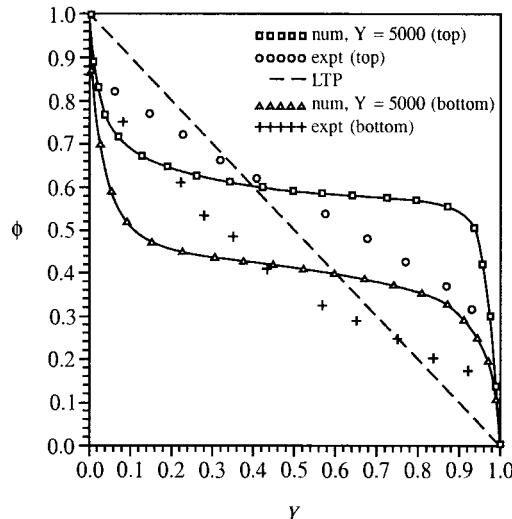


Figure 10. Variation of dimensionless temperature along width of the square enclosure for an enclosure with highly reflecting walls.

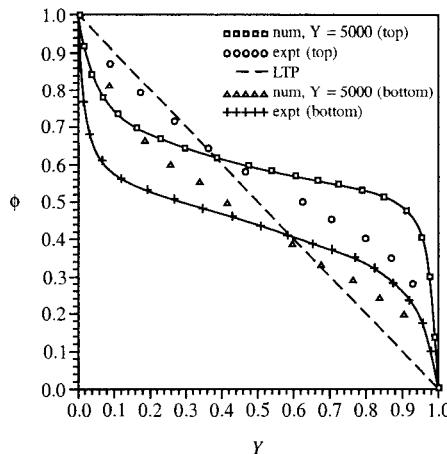


Figure 11. Variation of dimensionless temperature along width of the square enclosure for an enclosure with highly emissive walls.

Experimentally obtained values of temperature of the top and bottom walls are also shown in these graphs along with temperature profiles obtained from the numerical solution. The temperature variation for a linearly varying temperature (LTP) along the top and bottom horizontal walls is also shown. It is to be borne in mind that the first thermocouple in the top or bottom wall is located at a distance of 5 mm from the hot wall. Similarly, the last thermocouple in these walls is located at a distance of 5 mm from the cold wall. This means that the temperature variation

from 0 to 5 mm as well as 55 to 60 mm along the top and bottom walls is not obtained in the experiments. From the point of view of calculation of convective heat transfer coefficient through the experimental method, this poses no difficulties as the convection heat transfer coefficient is estimated directly from interferograms. Radiation heat transfer calculations are found to be insensitive to variation in temperature within the 5 mm zone where the temperatures are not measured. In both the cases shown in Figures 10 and 11, it can be seen that the temperatures obtained through experiments tend more towards adiabatic conditions rather than LTP.

In order to quantify the departure from adiabaticity, plots that show the variation of average convective Nusselt number with Grashof number for various boundary conditions is made use of. Figures 12 and 13 represent the variation of mean convective Nusselt number with Grashof number for the two cases considered in the present study.

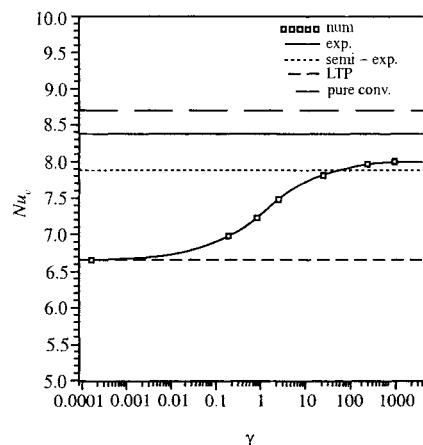


Figure 12. Variation of average convective Nusselt number with thermal conductivity parameter for an enclosure with highly polished walls.

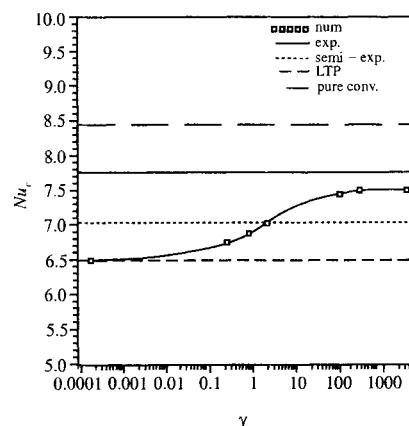


Figure 13. Variation of average convective Nusselt number with thermal conductivity parameter for an enclosure with highly emissive walls.

For the low emissivity case (case (a)), Nu_c for $Gr = 9.73 \times 10^5$ was determined to be 8.38 from experiments. Using the semi-experimental method, Nu_c was found to be 7.86, indicating a deviation of 6% from the experiments. Similarly for the high emissivity case (case (b)), for $Gr = 8.95 \times 10^5$, Nu_c was found to be 7.78 from experiments, and the semi-experimental method gave a value of 7.05, indicating a deviation of 9%. Though the agreement in both the cases is reasonable, the limitation in using this technique lies in the fact that temperatures are measured at a finite number of locations, and measurement of temperatures very close to the interfaces, for example, regions near the four corners of the enclosure becomes difficult. However, the temperature gradients may be steep in these regions, and therefore there does occur some error in reproducing the temperature profile using measured temperatures. Nevertheless, this combined technique reduces the complexity of both numerical and experimental procedures for investigating this type of heat transfer problem. For example, for the square cavity problem, this approach obviates the need for direct heat transfer measurement (calorimetry or optical methods) and at the same time, the modeller is relieved of the question of reasonableness and realizability of boundary conditions along the top and bottom walls, which are experimentally determined.

The nonlinear variation of Nu_c with γ can be seen in Figures 12 and 13 and the increase peters out at higher values of γ ($\gamma > 100$). It can be seen that the numerical results for a high γ ($\gamma > 100$) agree very well with the experimental results. It needs to be mentioned that larger the γ the closer are the top and bottom walls to adiabaticity. Conditions of very low γ correspond to thin metallic walls and the wall temperature actually approaches a linear temperature profile (LTP). Numerical calculations have also been done for (a) LTP, where the profile has been imposed on the top and bottom walls and (b) the pure convection case ($\varepsilon = 0; \gamma = \infty$) as well. Nu_c for LTP boundary conditions was determined to be 6.65. Nu_c (pure convection) in this case turned out to be 8.68. All the values of Nu_c , obtained whether numerical, experimental or semi-experimental will have to fall between these two limits. Similarly for the high emissivity case (case (b)), for $Gr = 8.95 \times 10^5$, Nu_c (LTP) was 6.48 and Nu_c (pure convection) was 8.46. The variation of Nu_c , when computed using a numerical scheme with conjugate wall boundaries with γ was similar to that of the low emissivity case. Also clear is the fact that the experimental conditions are closer to adiabaticity rather than LTP. However, it is of interest to note that reduction of Nu_c by radiation becomes more pronounced for the higher emissivity case, a trend noted in an earlier study by Balaji and Venkateshan (1993).

Another interesting feature of the square cavity problem is the relative insensitivity of radiation to the coupling on the top and bottom wall. For example, with LTP along the top and bottom walls, Nu_r for the high emissivity case turned out to be 10.35, while Nu_r determined by using the experimentally measured temperatures on these walls gave Nu_r of 10.39. With the fully coupled solution, Nu_r varied from 10.39 to 10.67 for γ varying from 0.0002 to 8400. All these variations are within the $\pm 5\%$ error band. This establishes the fact that Nu_r is not as sensitive to temperature profiles along the top and bottom walls as is Nu_c .

Recently, the problem of combined free convection and radiation in a square cavity partitioned in the middle was, also, taken up for a numerical investigations (Figure 14).

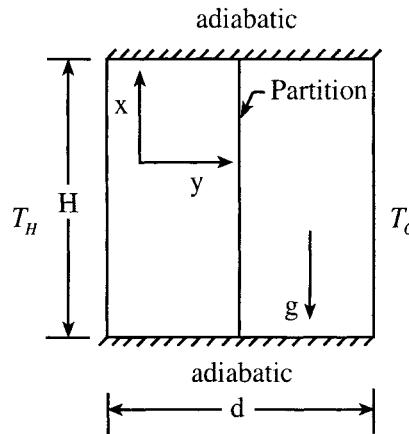


Figure 14. Schematic of a partitioned enclosure.

For the case of $\varepsilon_H = \varepsilon_C = 0.95$; $\varepsilon_T = \varepsilon_B = 0.85$ and $Ra_H = 1.16 \times 10^4$, it was found that the emissivity of the partition plays a crucial role in determining the overall Nusselt number. The overall Nusselt number with partition, can be as low as 32% of that without partition for $\varepsilon_p = 0.05$, whereas it is around 56% of the overall Nusselt number without partition for $\varepsilon_p = 0.95$. For a typical case, the variation of Nu_c with ε_p (with one partition) is as shown in Figure 15.

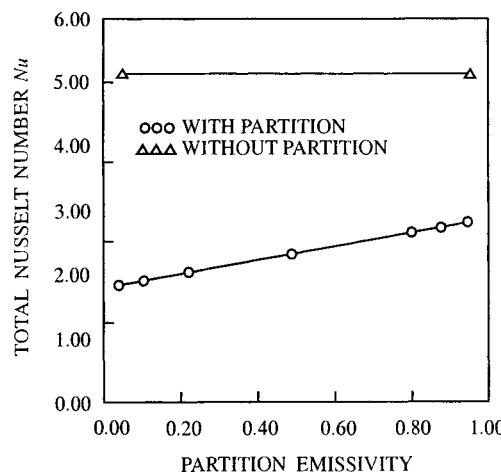


Figure 15. Variation of total Nusselt number with partition emissivity.

It can be seen that Nusselt number varies more or less linearly with ε_p .

Tall Cavity

The next problem taken up for investigation is the problem of combined free convection and radiation in tall cavities ($A \geq 2$). Figure 16 shows the details of the geometry for the tall cavity problem with side wall heating.

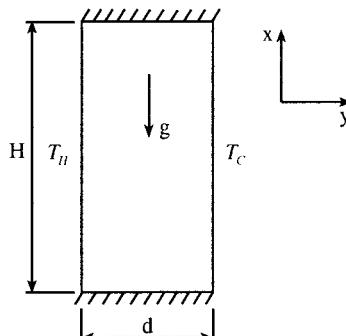


Figure 16. Details of the geometry of the tall cavity problem with side wall heating.

Two variations are possible in this problem. a) $\epsilon_h = \epsilon_c$ b) $\epsilon_h \neq \epsilon_c$. Interestingly, for the case a), it was observed that the coupling between free convection and radiation is weak, as might be expected. As the aspect ratio increases, the effect of the top and bottom walls diminish, as more of the radiation leaving the left wall will be received by the right wall directly. Table 1 shows the percentage error on the convection and radiation Nusselt numbers which arise because of decoupling. For this purpose, the most severe case of $\epsilon = 1$ for all the walls was taken. For $A > 2$, the error in decoupling is very well within the limits of numerical accuracy and decoupling is justified. However, it is seen that for $A \leq 1$, the error is not only significant, but also increases with decreasing A and it is around 20% for a shallow cavity with $A = 0.25$. The correlations and the range of parameters are presented towards the end (Table 3).

Table 1
Error in convection and radiation Nusselt numbers on account of decoupling for the case with $\epsilon_h = \epsilon_c = \epsilon_b = \epsilon_t = 1$, $Gr_d = 0.8 \times 10^5$, $Pr = 0.71$, $T_R = 0.73$

A	Error in Nu_c %	Error in Nu_R %
0.25	20.2	12.8
0.5	18.5	3.8
1.00	13.3	1.4
1.50	9.2	0.6
2.00	7.5	0.3
2.50	6.4	0.1
3.00	5.6	0.07

Figure 17 shows the top temperature profile of the tall cavity with and without radiation.

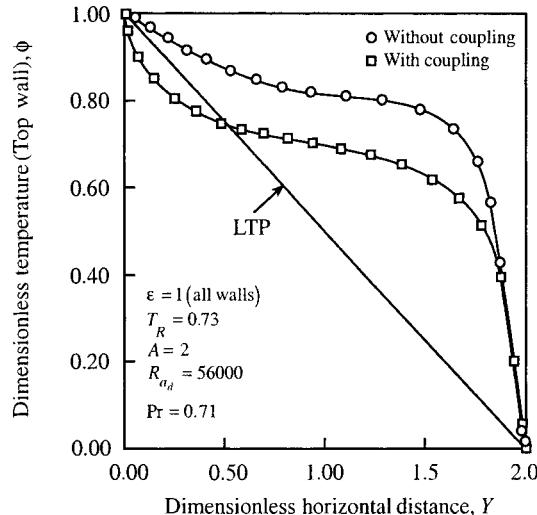


Figure 17. Temperature distribution in the top wall of the tall cavity highlighting the influence of surface radiation ($\varepsilon_H = \varepsilon_C, \varepsilon_T = \varepsilon_B$).

The basic point which emerges is that, the nature of the top temperature profile is relatively unaffected by radiation. In other words, the top temperature profile is essentially a “convection dominant” profile. The same is true for the bottom wall also. Hence, use of such a profile in the radiation calculations is superior to the use of a linear temperature profile (LTP), which is obtained when the wall is made up of a perfectly conducting material. Figure 18 shows the radiation Nusselt number as a function of aspect ratio with the other parameters fixed at the values shown in the figure.

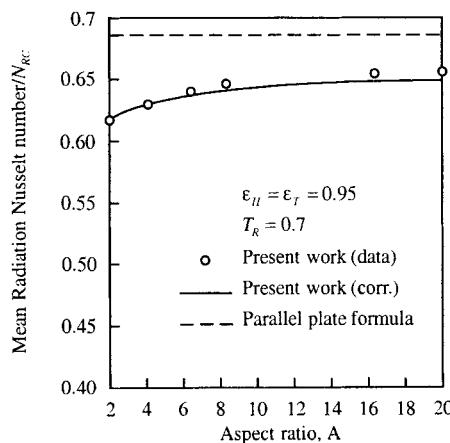
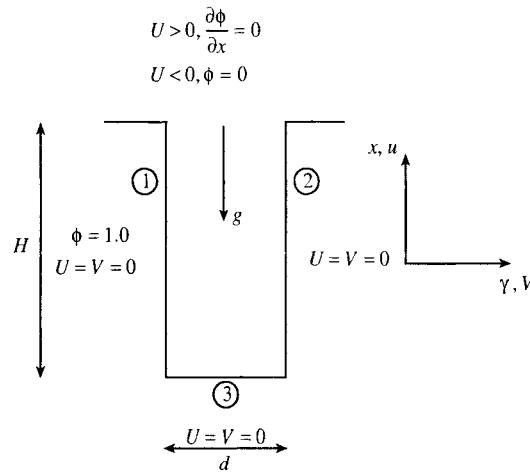


Figure 18. Effect of aspect ratio on radiation Nusselt number for the tall cavity ($\varepsilon_H = \varepsilon_C, \varepsilon_T = \varepsilon_B$).

It is seen that, the radiation Nusselt number approaches the value given by the parallel plate formula with increasing aspect ratio, and therefore, at high aspect ratio the parallel plate formula may be a reasonable approximation. The parallel plate formula is available in any text book on heat transfer (see, for example, Hottel and Sarofim 1967).

Open Cavity: Combined Convection and Radiation

The problems of free convection with radiation in an open cavity and, the problem of combined conduction, convection and radiation in an open cavity are considered next. The problem geometry for the former problem is given in Figure 19.



$$\textcircled{2} \text{ & } \textcircled{3} \quad \begin{cases} q_{\text{cond.}} + q_{\text{rad.}} = 0 \\ (\text{i.e.}) \frac{\partial \phi}{\partial X} \text{ or } \frac{\partial \phi}{\partial Y} = q_r \end{cases}$$

Figure 19. Details of the geometry for the combined convection and radiation case of the open cavity problem.

The thermal condition on the bottom and right walls given in the figure can be realised if they are made of a material of low thermal conductivity and if they are sufficiently thick. This, specifically, represents the case of $\gamma = 0$, in Equation (7). The influence of radiation on the vertical velocity, U , at the top can be seen in Figures 20 and 21.

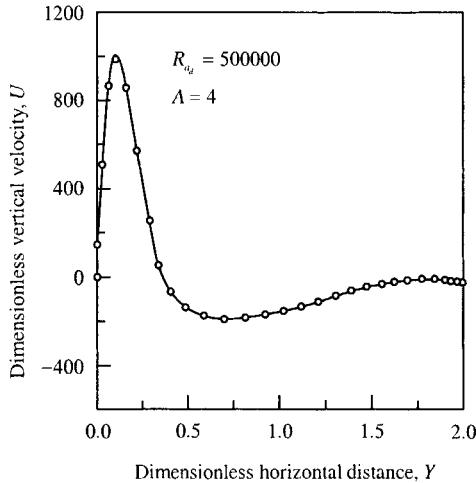


Figure 20. Variation of vertical velocity at the top of the open cavity for $\varepsilon = 0$.

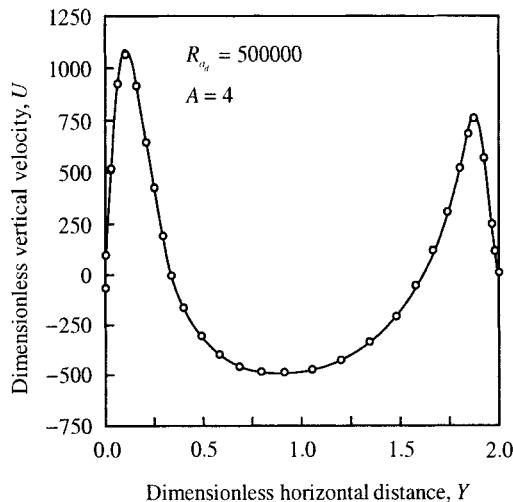


Figure 21. Variation of vertical velocity at the top of the open cavity for $\varepsilon = 1$.

It can be seen that instead of one flow loop occurring in the pure convection case ($\varepsilon = 0$), there are two such loops when radiation is present. The radiation heat transfer from the left wall has heated both the right and bottom walls, and, under equilibrium, these two walls lose heat convectively to the air. Thus, radiation changes the basic flow physics associated with the problem completely. It must be borne in mind that the fluid under consideration (air) is non-participating and all the above mentioned effects are purely due to wall radiation. The effect of radiation is explicitly brought out by the streamline pattern for the cases of $\varepsilon = 0$ and $\varepsilon = 1$, shown in Figures 22 and 23.

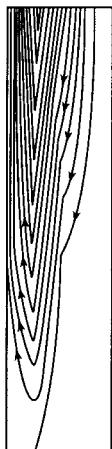


Figure 22

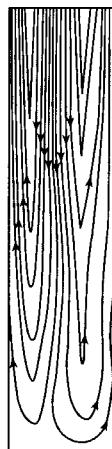


Figure 23

Figure 22. Streamline pattern for $Ra_d = 500,000$, $A = 4$, $T_R = 0.8$, $\varepsilon = 0$, for the open cavity.

Figure 23. Streamline pattern for $Ra_d = 500,000$, $A = 4$, $T_R = 0.8$, $\varepsilon = 1$, for the open cavity.

For this problem of combined radiation and free convection, it was found that radiation could contribute as much as 50% to the total heat transfer.

Based on a large number of numerical data, correlations were developed to predict Nu_C and Nu_R . The correlation coefficients for the two correlations were over 99% and the correlations, with the range of parameters are summarised towards the end (Table 3).

Open Cavity: Combined Conduction, Convection and Radiation

The logical extension of this problem would be the case in which the vertical walls are considered to be fins. This, then, will simulate the classical fin array problem of vertical fins standing on a horizontal, isothermal base. The problem geometry with the boundary conditions is given in Figure 24.

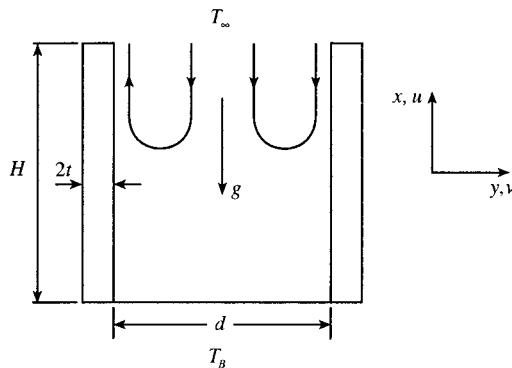


Figure 24. Details of the geometry for the combined conduction, convection and radiation problem of the open cavity.

Figure 25 shows the variation of vertical velocity at the top of the open cavity, for the parameters indicated in the figure.

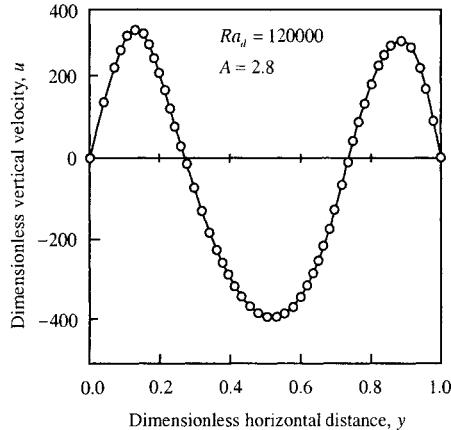


Figure 25. Variation of vertical velocity at the top of the open cavity for $\gamma = 0.005$ and $\varepsilon = 0$.

It can be seen that the flow enters from the top in the middle of the open cavity (also referred to as a slot) and rises symmetrically near the vertical walls. The velocity profile satisfies the mass balance across the opening i.e., fluid flow into the slot is equal to fluid flow out of the slot. The effect of radiation on the heat transfer at the vertical wall can be seen in Figure 26.

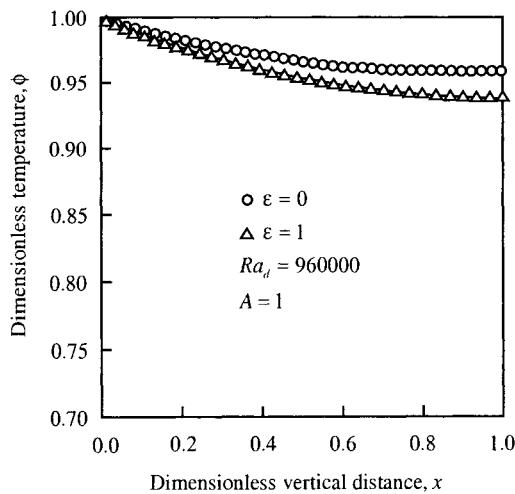


Figure 26. Temperature distribution in the vertical wall of the open cavity for $\varepsilon = 0$ and $\varepsilon = 1$ ($\gamma = 0.005$).

It is clear that radiation has a relatively weak influence on temperature distribution and, hence, also on the convective heat transfer. From a physical view point, this is justifiable. The thermal conductivity of the vertical wall being high, the problem

becomes more a conduction – convection interaction problem. In view of this, radiation can be decoupled, however, the conduction convection coupling cannot be ignored. For example, Figure 27 indicates that the tip temperature is 18–20°C less than the base temperature.

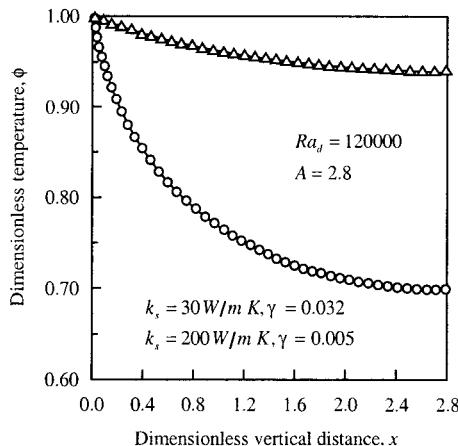


Figure 27. Temperature distribution in the vertical wall of the open cavity for two wall materials ($2t = 0.0015\text{ m}$, $\varepsilon = 0$).

Because of this Nu_C is about 30% smaller than that for an isothermal wall. The radiation heat transfer in this case can contribute as much as 30% to the total heat transfer.

For this problem, two correlations each for Nu_C and Nu_r are developed. That is because, convection takes place both from the vertical walls and the base and same is the case with radiation. The correlations are summarised towards the end of this section. The correlations were chosen in the present form to bring to focus, the physics of the interaction phenomenon occurring in the system.

Solution Using Semi-Experimental Method

In the semi-experimental method, the open cavity problem was solved numerically using the experimentally determined values of boundary temperatures for all walls of the cavity. It should be noted that the boundary temperatures along the bottom and the right walls are obtained as a consequence of the balance between conduction, convection and radiation heat transfer in the cavity. In other words, the numerical solver utilises the actual temperatures of those walls, which are subject to more than one mode of heat transfer. Thus the coupling along the right and bottom walls are actually taken care of by the experimentally determined temperatures.

Figure 28 shows the variation of dimensionless vertical velocity with dimensionless horizontal distance, or width of the cavity, from which, the existence of a right wall boundary layer is evident.

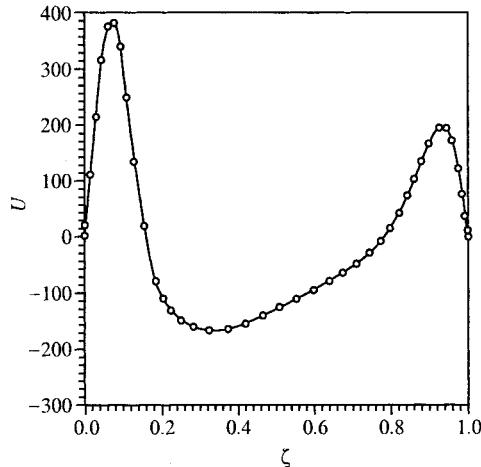


Figure 28. Variation of vertical velocity along the width of the open cavity, obtained from the semi-experimental method.

The corresponding stream line plot is shown in Figure 29.

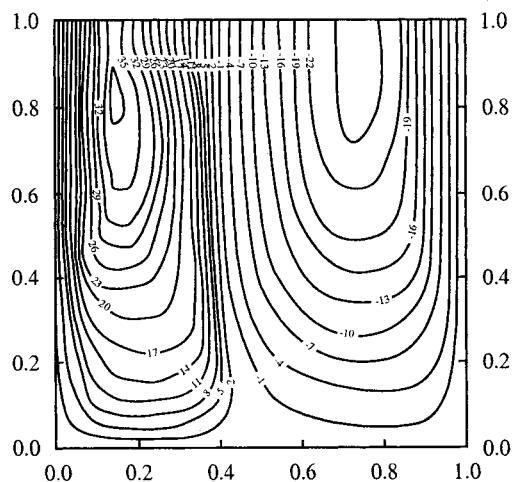


Figure 29. Isotherm plot of the case of the open cavity obtained by semi-experimental method.

For a few cases, the results of average convective Nusselt number (Nu_c) obtained from the semi-experimental method were compared with experimental values. Table 2 shows the results obtained from both the methods. It can be seen that the results from both the methods are in excellent agreement.

Table 2
Comparison of results obtained from semi-experimental method and experiments.

Sl. No.	Ra	Nu_c (semi-expt.)	Nu_c (expt.)
1	5.44×10^5	14.51	13.82
2	6.04×10^5	14.63	14.08
3	6.79×10^5	14.43	15.11
4	7.11×10^5	14.60	14.85
5	7.86×10^5	14.91	14.96

L Corner

The last problem taken up for investigation is the *L*-corner, i.e. a cavity open at the top and the side. Two *L*-corners can make a single fin and so in the discussion that follows, the two terms will be used interchangeably. This geometry has extensive applications in cooling of electronic packages, heat sinks etc. In such problems, there is, invariably, an interaction between all the three modes of heat transfer. Such an approach, strangely enough, is conspicuous by its absence in the literature. The details of the boundary conditions and the computational domain are given in Figure 30.

$$\omega = 0, \frac{\partial \phi}{\partial x} = 0, \frac{\partial \psi}{\partial x} = 0$$

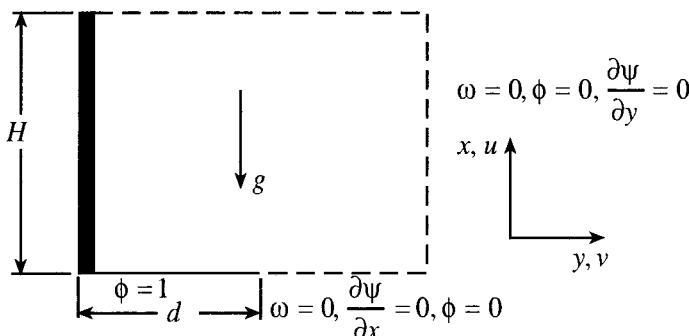


Figure 30. Details of the boundary conditions and computational domain for the *L* corner problem.

The range of parameters for this problem were chosen so that in the laminar regime, the γ values considered will cover commonly encountered materials like aluminum,

brass, steel, when the thicknesses associated with them is, typically, between 1 and 3 mm. Figure 31 shows the local Nusselt number distribution along the vertical wall for two values of γ .

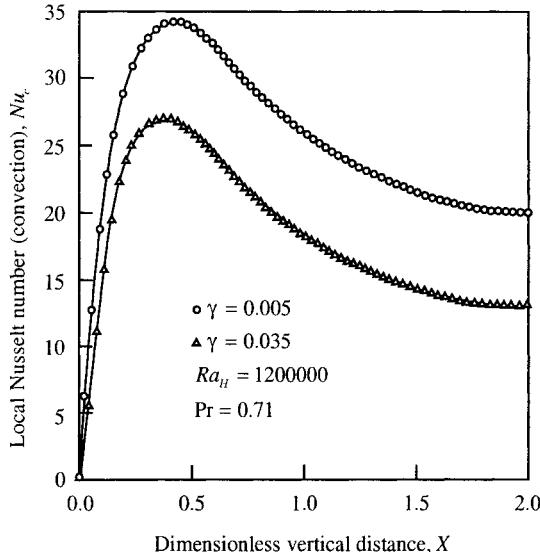


Figure 31. Convection Nusselt number distribution in the vertical wall of the L corner for two wall materials ($2t = 0.0015\text{m}$, $\varepsilon = 0$).

A higher γ represents a material of lower thermal conductivity, when the other geometric parameters are held the same. The local Nusselt number shows a peak around 27% from the base. If one considers the physics of the flow, this is actually because of the interaction of two boundary layers, one originating from the leading edge of the base and, the other originating at the intersection of the base and the vertical wall. This result is in qualitative agreement with the experimental observations of Sobhan et al. (1990) and Rammohan Rao and Venkateshan (1996).

Similar to the case of the open cavity with conduction, convection, and radiation, it was observed that for an L -corner too, the radiation coupling is negligibly small vis-à-vis the conduction-convection coupling. However, the radiation heat transfer can contribute as much as 49% to the total heat transfer when $\varepsilon = 1$, $Ra_d = 4 \times 10^6$, $T_R = 0.8$ and $\gamma = 0.05$. The numerical results obtained in the present study were also found to be in very good agreement, quantitatively with the results of Rammohan Rao and Venkateshan (1996). The latter, however, considered the interaction in a three dimensional fin and so for purposes of comparison we have chosen the case with $\varepsilon = 0.05$. The reason for this is, radiation will be three dimensional, but convection will, in general, not be, in view of the "chimney" type of up and down flow pattern. Specifically for $Ra_H = 10^6$, $T_H = 380\text{ K}$, $T_\infty = 303\text{ K}$, Nu_C for Rammohan Rao will be 12.68 whereas the same from the correlation developed in the present study is found to be 11.66. So, a two dimensional model seems reasonably accurate and is also able to reproduce the finer aspects of the flow, like the occurrence of the peak which was referred to earlier, experimentally.

The correlations and the range of parameters are now presented for all the geometries. These will be of use to the designer, to get an estimate of the overall heat transfer from the various systems.

Table 3
Summary of Correlations

SQUARE CAVITY

COMBINED CONVECTION AND SURFACE RADIATION $\text{Pr} = 0.71$

$$Nu_C = 0.149 Gr_d^{0.294} [1 + \varepsilon_H]^{-0.279} [1 + \varepsilon_C]^{0.182} [1 + \varepsilon_B]^{-0.135} [1 + \varepsilon_T]^{0.115} \left[N_{RC} / (N_{RC} + 1) \right]^{0.272}$$

$$Nu_R = 0.657 Gr_d^{-0.0093} \varepsilon_H^{0.808} \varepsilon_C^{0.342} [1 + \varepsilon_B]^{0.199} [1 + \varepsilon_T]^{-0.039} [1 - T_R^4]^{1.149} N_{RC}^{1.051}$$

Parameter Range

$$10^3 \leq Gr_d \leq 10^6; 0 \leq \varepsilon_H, \varepsilon_C, \varepsilon_T, \varepsilon_B \leq 1$$

$$0.73 \leq T_R \leq 0.95; 4 \leq N_{RC} \leq 22$$

TALL CAVITY

COMBINED CONVECTION AND SURFACE RADIATION

$$\varepsilon_H \neq \varepsilon_C, \text{Pr} = 0.71$$

$$Nu_C = 0.247 Ra_d^{0.264} A^{-0.21} [1 + \varepsilon_H]^{-0.06} [1 + \varepsilon_C]^{0.12} [1 + \varepsilon_T]^{-0.04} [1 + \varepsilon_B]^{-0.04} \left\{ N_{RC} / (N_{RC} + 1) \right\}^{-0.08}$$

$$Nu_R = 0.948 \left\{ A / (1 + A) \right\}^{0.04} [1 - T_R^4]^{0.97} [1 + \varepsilon_T]^{-0.01} [1 + \varepsilon_B]^{-0.01} N_{RC}^{1.01} \left\{ 1 / \varepsilon_H + 1 / \varepsilon_C - 1 \right\}^{-0.087}$$

Parameter Range

$$2 \leq A \leq 40, 10^2 \leq Ra_d \leq 10^5$$

$$0.5 \leq \varepsilon_H, \varepsilon_C, \varepsilon_T, \varepsilon_B \leq 0.95; 0.70 \leq T_R \leq 0.90$$

$$5 \leq N_{RC} \leq 30; \text{Pr} = 0.71$$

OPEN CAVITY

COMBINED CONVECTION AND SURFACE RADIATION, ε OF ALL WALLS EQUAL, $\text{Pr} = 0.71$

$$Nu_C = 0.426 Gr_H^{0.254} \left\{ N_{RC} / (N_{RC} + 1) \right\}^{1.181} [1 + \varepsilon]^{-0.039}$$

$$Nu_R = 1.306 Gr_H^{-0.079} [1 - T_R^4]^{1.223} N_{RC}^{1.21} \varepsilon^{0.875} A$$

Parameter Range

$$10^4 \leq Ra_H \leq 10^8; 0 \leq \varepsilon \leq 1$$

$$0.75 \leq T_R \leq 0.85; 9 \leq N_{RC} \leq 30$$

$$1 \leq A \leq 5$$

OPEN CAVITY**COMBINED CONDUCTION, CONVECTION AND SURFACE RADIATION, ε OF ALL WALLS THE SAME, $Pr = 0.71$** **Side Wall Heat Transfer**

$$Nu_{CW} = 0.36 Ra_d^{0.273} [1 + \gamma]^{-13.02} A^{-0.215}$$

$$Nu_{RW} = 1.66 [1 + \gamma]^{-4.25} [1 - T_R^4]^{0.33} \varepsilon^{0.36} N_{RC}^{0.36} A^{-1.46}$$

Base Heat Transfer

$$NU_{CB} = 0.626 Ra_d^{0.215} [1 + \gamma]^{-0.019} A^{0.168}$$

$$Nu_{RB} = 0.188 [1 + \gamma]^{-6.15} [1 - T_R^4]^{1.15} \varepsilon^{0.66} N_{RC}^{1.26} A^{-0.286}$$

Parameter Range

$$0.5 \leq \gamma \leq 0.07; 10^5 \leq Ra_d \leq 5 \times 10^6$$

$$1 \leq A \leq 3; 0 \leq \varepsilon \leq 1; 0.75 \leq T_R \leq 0.90$$

L SHAPED CORNER**COMBINED CONDUCTION, CONVECTION AND SURFACE RADIATION, WALL AND BASE HAVE SAME ε , $Pr = 0.71$** **Wall Heat Transfer**

$$Nu_{CW} = 0.226 Ra_H^{0.291} [1 - 8.03\gamma] [1 + \varepsilon]^{-0.202} \left\{ N_{RC} / (N_{RC} + 1) \right\}^{1.06}$$

$$Nu_{RW} = 0.516 [1 + \gamma]^{-2.56} [1 - T_R^4]^{0.672} \varepsilon^{0.841} N_{RC}^{0.872} A^{0.889}$$

Base Heat Transfer

$$Nu_{CB} = 0.586 Ra_H^{0.204} \gamma^{-0.001}$$

$$Nu_{RB} = 0.336 [1 + \gamma]^{-0.528} [1 - T_R^4]^{0.966} \varepsilon^{0.813} N_{RC}^{0.998} A^{0.954}$$

Parameter Range

$$0.005 \leq \gamma \leq 0.035; 5 \times 10^5 \leq Ra_H \leq 5 \times 10^7$$

$$1 \leq A \leq 2; 0 \leq \varepsilon \leq 1; 0.75 \leq T_R \leq 0.85$$

CONCLUSIONS

From the investigation on the interaction of convection with radiation and/or conduction, the following broad conclusions can be arrived at:

1. At low temperature levels (room temperature to 100°C) radiation heat transfer in the problems taken up for investigation was found to be significant and, hence, cannot be neglected.
2. For a square cavity, the convection-radiation coupling reduces Nu_C by 10-15%, and, radiation itself can contribute more than 50% to the total heat transfer.
3. In the case of a tall cavity ($A > 2$) the convection radiation coupling was negligible for the case of $\varepsilon_H = \varepsilon_C$. However, the coupling was severe for $\varepsilon_H \neq \varepsilon_C$ and, hence, for this case, decoupling of convection from radiation cannot be done.
4. In the case of an open cavity, it was found that radiation can contribute as much as 50% to the overall heat transfer. Radiation changes the basic flow pattern completely. In the case of the open cavity with finite thermal conductivity walls, the radiation coupling was found to be less severe, as opposed to the conduction-convection coupling. However, radiation heat transfer was found to contribute as much as 30% of the total heat transfer.
5. Finally, the problem of combined free convection, conduction and radiation in an L corner was considered. Here again, the convective heat transfer was found to be relatively insensitive to radiation. As in the case of an open cavity, it was found that the conductive coupling cannot be ignored. Radiation can contribute up to 50% to the overall heat transfer.
6. A possible extension of this work could be the study of turbulent three dimensional flows in such geometries. The problem of combined conduction, convection and radiation in closed cavities holds promise for the future, in view of its application in design.
7. The semi-experimental approach appears to be a viable option in modelling transport of heat and fluid flow in order to obtain estimates of heat transfer rates. This approach finds application in solving problems where the boundary conditions become complex because of interactions, and also because of the uncertainty in thermophysical properties of wall materials and fluid.

NOMENCLATURE

A	Aspect ratio, H/d
d	Cavity width or spacing in the case of open cavity or width of base of L corner, m
g	Acceleration due to gravity, m/s^2
Gr	Grashof number
H	Cavity height or wall height or height of the vertical leg of L corner, m
I'	Elemental irradiation, W/m^2
I	Elemental dimensionless irradiation, $I'/\sigma T_1^4$
J'	Elemental radiosity, W/m^2
J	Elemental dimensionless radiosity, $J'/\sigma T_1^4$
k	Thermal conductivity, $W/m K$
N_{RC}	Radiation conduction parameter $\sigma T_1^4 d/k_1 (T_1 - T_2)$
Nu_C	Average convection Nusselt number
Nu_R	Average radiation Nusselt number
Pr	Prandtl number of the fluid, ν/α
Ra	Rayleigh number
t	Semi-thickness of the vertical leg of L corner, m
T	Temperature, K
u	vertical velocity, m/s
U	Dimensionless vertical velocity, ud/α
v	Horizontal or cross velocity, m/s
V	Non-dimensional horizontal velocity, vd/α
x	Vertical co-ordinate, m
X	Non-dimensional vertical co-ordinate, x/d
y	Horizontal co-ordinate, m
Y	Non-dimensional horizontal co-ordinate, y/H

Greek Symbols

α	Thermal diffusivity of the fluid, m^2/s
γ	Non-dimensional thermal conductivity parameter, $k_f d/k_s t$
ε	Total hemispherical emissivity
ν	Kinematic viscosity of fluid, m^2/s
ϕ	Dimensionless temperature, $(T - T_2)/(T_1 - T_2)$
ψ	Dimensionless stream function ψ'/α
σ	Stefan Boltzmann constant, $5.67 \times 10^{-8} W/m^2 K^4$
ω	Dimensionless vorticity, $\omega' d^2/v$

Subscripts

B	Bottom wall in the closed cavity, base in the open cavity or the L -corner
C	Cold wall in the closed cavity problem
H	Hot wall in the closed or open cavity problem

- P* Partition in the case of the partitioned enclosure
T Top wall in the case of the closed cavity or open cavity
 1 Reference condition: *H* in the case of closed cavity or open cavity, *B* in the case of *L* corner
 2 Reference condition: *C* in the case of closed cavity, ambient in the case of open cavity or *L* corner
d Based on *d* as the characteristic dimension
H Based on *H* as the characteristic dimension
W Pertaining to the vertical wall in the open cavity or the vertical leg in the *L*-corner

REFERENCES

- [1] Angirasa, D. and Mahajan, R.L., Natural convection from *L* shaped corners with adiabatic and cold isothermal walls, ASME Journal of Heat Transfer, 116, 1993, 149-157.
- [2] Asako, Y. and Nakamura, H., Heat transfer across a parallelogram shaped enclosure, Bulletin JSME, 25, 1982, 1412-1424.
- [3] Balaji, C., Laminar free convection with conduction and surface radiation in open and closed cavities, Ph.D Thesis, Department of Mechanical Engineering, IIT Madras, 1994.
- [4] Balaji, C. and Venkateshan, S.P., Discussion on natural convection with radiation in a cavity with open top end, ASME Journal of Heat Transfer, 115, 1993, 1085-1086.
- [5] Behnia, M. and de Vahl Davis, G., Natural convection cooling of an electronic component in a slot, Proceedings of the Ninth International Heat Transfer Conference, Jerusalem, 2, 1990, 209-233.
- [6] Eckert, E.R.G. and Carlson, W.O., Natural convection in an air layer, International Journal of Heat and Mass Transfer, 2, 1961, 106-120.
- [7] Gosman, A.D., Pun, W.M., Runchal, A.K., Spalding D.B. and Wolfshtein, M., *Heat and Mass Transfer in Recirculating Flows*, Academic Press, London, 1969.
- [8] Harahap, F and McManus, H, Natural convection heat transfer from horizontal rectangular fin arrays, ASME Journal of Heat Transfer, 89, 1967, 32-38.
- [9] Hoogendoorn, C.J., *Experimental Methods in Natural Convection*, in Natural Convection Fundamentals and Applications, Kakac, S., Aung, W., and Viskanta, R. Eds., Hemisphere, Washington, 381-400, 1985.
- [10] Hottel, H.C. and Sarofim, A.F., *Radiative Heat Transfer*, McGraw Hill, New York, 1967.
- [11] Kim, D.M. and Viskanta, R., Effect of wall conduction and radiation on natural convection in a rectangular cavity, Numerical Heat Transfer, 7, 1984, 449-470.
- [12] Lage, J.L., Lim, J.S., and Bejan, A., Natural convection with radiation in a cavity with open top end, ASME Journal of Heat Transfer, 114, 1992, 479-486.
- [13] Lauriat, G, The effects of radiation on natural convection, International Journal of Chemical Engineering, 31, 1991, 693-700.
- [14] Lin, C.X., Ko, S.U., and Xin, M.D., Effects of surface radiation on turbulent free convection in an open ended cavity, International Communications in Heat and Mass Transfer, 21, 1994, 117-130.
- [15] Rammohan Rao, V. and Venkateshan, S.P., Experimental study of free convection and radiation in horizontal fin arrays, International Journal of Heat and Mass Transfer, 39, 1996, 779-789.
- [16] Roache, P.J., *Computational Fluid Dynamics*, Hermosa, Albuquerque, New Mexico, 1982.

- [17] Smith, T.F., Beckermann, C., and Weber, S.M., Combined conduction, natural convection and radiation in an electronic chassis., *Journal of Electronic Packaging*, 113, 1991, 382-391.
- [18] Sobhan, C.B., Venkateshan, S.P., and Seetharamu, K.N., Experimental studies on steady free convection heat transfer from fins and fin arrays, *Wärme*, 25, 1990, 345-352.
- [19] Starner, K.E. and McManus, H.N., An experimental investigation of free convection from rectangular fin arrays, *ASME Journal of Heat Transfer*, 85, 1963, 273-278.
- [20] Yang, K.T. and Lloyd, J.R., *Natural Convection Radiation Interaction in Enclosures*, in *Natural Convection—Fundamentals and Applications*, Kakac, S., Aung, W., and Viskanta, R., Eds., Hemisphere, Washington D.C., 381-400, 1985.
- [21] Zhong, Z.Y., Yang, K.T., and Lloyd, J.R., Variable property effects in laminar natural convection in a square enclosure, *ASME Journal of Heat Transfer*, 107, 1985, 133-138.

23 MATHEMATICAL RESULTS AND NUMERICAL METHODS FOR STEADY INCOMPRESSIBLE VISCOELASTIC FLUID FLOWS

Adélia Sequeira

and

Juha H. Videman

Centro de Matemática Aplicada
And Departamento de Matemática
Instituto Superior Técnico
Av. Rovisco Pais
1049-001 Lisbon, Portugal

ABSTRACT

The aim of this paper is to present a short survey of recent mathematical results for some models of incompressible, homogeneous, viscoelastic non-Newtonian fluids, namely for the grade type Rivlin-Ericksen and Oldroyd type models. We address the questions of existence, uniqueness and asymptotic behavior of steady solutions, in several physically relevant flow geometries, by suitably decoupling the elliptic and hyperbolic parts in the systems of equations. Finally, we discuss a numerical scheme using a mixed finite element method for the simulation of the models in two-dimensional bounded domains.

1. INTRODUCTION

The equations governing the motion of viscoelastic non-Newtonian fluids of differential and rate type have been extensively studied over the past decades, see, e.g., [7, 13, 41, 3, 20, 16, 4, 49] and all the references cited therein. Concerning the mechanical properties of these fluids, we refer to [48, 21, 44, 40]. As is well known,

for viscoelastic fluid the constitutive equations relating the stress tensor to the rate of deformation lead to highly nonlinear systems of PDE's of mixed type, being elliptic-hyperbolic in the steady state and parabolic-hyperbolic in the unsteady state. The complexity of these systems requires the use of special techniques of nonlinear analysis to investigate questions of existence, uniqueness, and stability of solutions. On the other hand, efficient and accurate numerical schemes for solving the equations must be based on their mixed mathematical structure in order to prevent numerical instabilities in problems which are mathematically well-posed, see, e.g., [6, 25, 10, 11].

In this survey article, we collect some recent results of our investigation on the mathematical theory and numerical methods for the equations governing the steady flow of incompressible, homogeneous, viscoelastic fluids of differential type. As the mathematical theory is concerned, we study second- and third-grade Rivlin-Ericksen fluids and fluids of the Oldroyd type in two- or three-dimensional unbounded domains. The numerical approximation is focused on the second-grade fluid model in a two-dimensional bounded domain.

The outline of the paper is as follows. In Section 2, we recall the equations describing the motion of second- and third-grade Rivlin-Ericksen fluids and of fluids of the Oldroyd type. We also present decomposed forms of these equations which reduce the original nonlinear problems to coupled equations, composed of a Stokes system and of a scalar transport equation, that at the first stage are studied as two separate linear systems. The solvability for the nonlinear coupled equations, and consequently for the original problems, is obtained using a suitable fixed point argument. Sections 3 and 4 are devoted to the mathematical study of these equations in certain physically interesting unbounded domains. In Section 3, we consider flows in an exterior domain and collect results on existence, uniqueness, and asymptotic behavior of solutions, presented in [17, 31, 32, 30]. Section 4 is concerned with flows in domains with cylindrical outlets to infinity. Besides proving the well-posedness of the equations, we study the existence of rectilinear flow fields of Poiseuille type and of secondary flows in straight channels and pipes. These results appeared in [36, 37]. In [37] more general outlets of paraboloidal type were also considered. Finally, in Section 5, we discuss a numerical scheme for the simulation of the flow of a second-grade fluid. The method is based on the ideas of [26] (see also the references cited therein) and has been presented in [45].

Most of our results are obtained in usual Sobolev, Hilbert, or Hölder spaces. For these spaces we use the standard notations $W^{m,p}(\Omega)$, $H^m(\Omega)$ and $C^{m,\delta}(\Omega)$, with the respective norms $\|\cdot\|_{m,p}$, $\|\cdot\|_m$ and $\|\cdot; C^{m,\delta}(\Omega)\|$. Whenever other functional settings are needed, the corresponding spaces and norms are introduced

with all details. The function spaces for vector- and tensor-valued functions are denoted by a boldface letter.

2. THE EQUATIONS OF MOTION AND THEIR DECOMPOSED FORMS

The equations governing the steady motion of an incompressible fluid are

$$\rho \mathbf{v} \cdot \nabla \mathbf{v} + \nabla p = \nabla \cdot \mathbf{T}_E + \rho \mathbf{f}, \quad \nabla \cdot \mathbf{v} = 0, \quad \text{in } \Omega \quad (2.1)$$

with \mathbf{v} denoting the velocity field, p the hydrostatic pressure, \mathbf{T}_E the extra-stress tensor ($\mathbf{T} = -p\mathbf{I} + \mathbf{T}_E$ is the usual Cauchy stress), \mathbf{f} the exterior body force, and ρ the constant density of the fluid. Here, $\Omega \subset \mathbb{R}^n$, $n = 2, 3$ denotes an open and connected set with a sufficiently smooth boundary $\partial\Omega$. The material properties of the fluid are defined by a constitutive law relating the extra-stress tensor to the kinematic variables.

2.1 Rivlin-Ericksen Fluids of Second- and Third-Grade.

In an incompressible Rivlin-Ericksen fluid of grade 3, the extra-stress tensor is given by, see [43]

$$\mathbf{T}_E = \eta \mathbf{A}_1(\mathbf{v}) + \alpha_1 \mathbf{A}_2(\mathbf{v}) + \alpha_2 \mathbf{A}_1^2(\mathbf{v}) + \beta (\operatorname{tr} \mathbf{A}_1^2(\mathbf{v})) \mathbf{A}_1(\mathbf{v}), \quad (2.2)$$

where $\mathbf{A}_1(\mathbf{v})$ and $\mathbf{A}_2(\mathbf{v})$ denote the first two Rivlin-Ericksen tensors

$$\begin{aligned} \mathbf{A}_1(\mathbf{v}) &= \nabla \mathbf{v} + (\nabla \mathbf{v})^T, \\ \mathbf{A}_2(\mathbf{v}) &= \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{A}_1(\mathbf{v}) + \mathbf{A}_1(\mathbf{v}) \nabla \mathbf{v} + (\nabla \mathbf{v})^T \mathbf{A}_1(\mathbf{v}), \end{aligned} \quad (2.3)$$

and η , α_1 , α_2 , and β stand for material constants. In fact, the constitutive relation (2.2) is a degenerate form of a more general Rivlin-Ericksen fluid of grade 3 defined by $\mathbf{T}_E = \eta \mathbf{A}_1 + \alpha_1 \mathbf{A}_2 + \alpha_2 \mathbf{A}_1^2 + \beta_1 \mathbf{A}_3 + \beta_2 (\mathbf{A}_1 \mathbf{A}_2 + \mathbf{A}_2 \mathbf{A}_1) + \beta_3 (\operatorname{tr} \mathbf{A}_1^2) \mathbf{A}_1$ obtained by assuming, in view of thermodynamics, that $\beta_1 = \beta_2 = 0$, see [13].

A third-grade fluid is compatible with thermodynamics if the material constants in (2.2) satisfy the conditions, cf. [13]

$$\begin{aligned} \eta &\geq 0, \quad \alpha_1 \geq 0, \quad \beta \geq 0 \\ -\sqrt{24\eta\beta} &\leq \alpha_1 + \alpha_2 \leq \sqrt{24\eta\beta}. \end{aligned} \quad (2.4)$$

The constitutive law (2.2), includes as special cases the fluids of second-grade ($\beta = 0$), and the Newtonian fluids ($\beta = \alpha_1 = \alpha_2 = 0$). In particular, the second-grade fluids are consistent with thermodynamics, if the material constants satisfy, cf. [7]

$$\eta \geq 0, \quad \alpha_1 \geq 0, \quad \alpha_1 + \alpha_2 = 0. \quad (2.5)$$

The constitutive relation (2.2), together with equations (2.1), yield the following system

$$\begin{cases} -v\Delta v - \alpha_1 v \cdot \nabla \Delta v + \nabla p = \nabla \cdot N(v) + f & \text{in } \Omega \\ \nabla \cdot v = 0 \end{cases} \quad (2.6)$$

where all the material constants are divided by the constant density ρ ($v = \eta/\rho$ denotes the kinematical viscosity coefficient) and

$$N(v) = \alpha_1 (\nabla v)^T A_1(v) + (\alpha_1 + \alpha_2) A_1^2(v) + \beta (tr A_1^2(v)) A_1(v) - v \otimes v, \quad (2.7)$$

with \otimes denoting the dyadic product.

Equations (2.6) become well set while complemented with the adherence boundary condition

$$v = v_*, \quad v_* \cdot n = 0 \quad \text{on } \partial\Omega, \quad (2.8)$$

where n stands for the outward unit normal vector. In unbounded domains, conditions on the asymptotic behavior of the solution at large distances must also be imposed.

Problems (2.6), (2.8) can be studied by considering the following coupled set of equations for (v, π, w)

$$\begin{cases} -\Delta v + \nabla \pi = w, \quad \nabla \cdot v = 0 & \text{in } \Omega \\ v = v_*, \quad v_* \cdot n = 0 & \text{on } \partial\Omega \\ vw + \alpha_1 v \cdot \nabla w = \nabla \cdot N(v) - \alpha_1 (\nabla v)^T \nabla \pi + f & \text{in } \Omega. \end{cases} \quad (2.9)$$

Sometimes it is more convenient to consider, instead of (2.9), the following coupled system for (v, π, z)

$$\begin{cases} -\Delta v + \nabla \pi = \nabla \cdot z, \quad \nabla \cdot v = 0 & \text{in } \Omega \\ v = v_*, \quad v_* \cdot n = 0 & \text{on } \partial\Omega \\ vz + \alpha_1 v \cdot \nabla z = \alpha_1 z (\nabla v)^T + N(v) - \alpha_1 \pi (\nabla v)^T + F & \text{in } \Omega. \end{cases} \quad (2.10)$$

where F is a tensor function such that $f = \nabla \cdot F$.

One readily shows that the solvability either of system (2.9) or (2.10) implies the existence of a solution (v, p) to problem (2.6) with

$$p = v \pi + \alpha_1 v \cdot \nabla \pi. \quad (2.11)$$

2.2 Fluids of the Oldroyd type.

In an *Oldroyd-type fluid* the extra-stress tensor T_E obeys the constitutive law, see [33]

$$T_E + \lambda_1 \frac{\mathcal{D}_a T_E}{\mathcal{D}t} = \mu_0 \left(A_1(v) + \lambda_2 \frac{\mathcal{D}_a A_1(v)}{\mathcal{D}t} \right), \quad (2.12)$$

where $\mu_0 = 0$ denotes the viscosity coefficient, $\lambda_1 > 0$ the relaxation time and λ_2 the retardation time ($0 \leq \lambda_2 < \lambda_1$). The symbol $\mathcal{D}_a/\mathcal{D}t$ stands for an objective derivative of Oldroyd type, see [33], defined by

$$\frac{\mathcal{D}_a \mathbf{T}_E}{\mathcal{D}t} = \frac{d}{dt} \mathbf{T}_E + \frac{1-a}{2} (\mathbf{T}_E (\nabla \mathbf{v}) + (\nabla \mathbf{v})^T \mathbf{T}_E) - \frac{1+a}{2} (\mathbf{T}_E (\nabla \mathbf{v})^T + (\nabla \mathbf{v}) \mathbf{T}_E), \quad (2.13)$$

where d/dt denotes the material time derivative and $a \in [-1,1]$. With $a=1$, the Oldroyd derivative (2.13) reduces to the upper convected derivative and the corresponding constitutive law to the constitutive law of the Oldroyd-B fluid. The special case $\lambda_2 = 0$ in (2.12) corresponds to the Maxwell fluid.

Now, defining the *Newtonian* and the *elastic viscosity coefficients*, v_n and v_e , by

$$v_n = \frac{\mu_0 \lambda_2}{\rho \lambda_1}, \quad v_e = \mu_0 \left(1 - \frac{\lambda_2}{\rho \lambda_1} \right), \quad (2.14)$$

and setting

$$\mathbf{T}_E = v_n \mathbf{A}_1(\mathbf{v}) + \boldsymbol{\tau}, \quad (2.15)$$

where \mathbf{T}_E and $\boldsymbol{\tau}$ ($\boldsymbol{\tau}$ denotes the “elastic part” of the extra stress tensor) have been redefined after division by the constant density ρ , one obtains from (2.1), (2.8), and (2.12) the following system of equations

$$\begin{cases} -v_0(1-\omega) \Delta \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} + \nabla p = \nabla \cdot \boldsymbol{\tau} + \mathbf{f} & \text{in } \Omega \\ \nabla \cdot \mathbf{v} = 0 \\ \mathbf{v} = \mathbf{v}_*, \quad \mathbf{v}_* \cdot \mathbf{n} = 0 & \text{on } \partial \Omega \\ \boldsymbol{\tau} + \lambda_1 \mathbf{v} \cdot \nabla \boldsymbol{\tau} + \lambda_1 \mathbf{g}(\boldsymbol{\tau}, \nabla \mathbf{v}) = v_0 \omega \mathbf{A}_1(\mathbf{v}) & \text{in } \Omega \end{cases} \quad (2.16)$$

where $\omega = (1 - \lambda_2 / \lambda_1)$ and

$$\mathbf{g}(\boldsymbol{\tau}, \nabla \mathbf{v}) = \frac{1-a}{2} (\boldsymbol{\tau} (\nabla \mathbf{v}) + (\nabla \mathbf{v})^T \boldsymbol{\tau}) - \frac{1+a}{2} (\boldsymbol{\tau} (\nabla \mathbf{v})^T + (\nabla \mathbf{v}) \boldsymbol{\tau}).$$

3. RESULTS IN AN EXTERIOR DOMAIN

Flows in an exterior domain correspond to the physical situation of flows around or past an obstacle. In all that follows, we assume that the obstacle is solid and rigid. In mathematical terms, the flow domain Ω is the complement of a compact set (the obstacle) $\mathcal{B} \subset \mathbb{R}^n$, $n = 2, 3$. Since the fluid adheres to the impermeable wall of the body \mathcal{B} , the velocity field satisfies

$$\mathbf{v} = \mathbf{v}_*, \quad \mathbf{v}_* \cdot \mathbf{n} = 0 \quad \text{on } \partial \Omega.$$

It is also natural to require that the velocity field tends to some constant vector field \mathbf{v}_∞ at infinity. In case $\mathbf{v}_\infty = 0$, this means that the motion of the fluid is driven by

the notation of the body and the fluid is at rest at large distances. If v_∞ does not vanish, the situation corresponds, after a suitable change of coordinates, to the study of the fluid motion caused by a body moving in the fluid with the constant velocity v_∞ .

3.1 The case $v_\infty = 0$.

Consider the equations governing the steady motion of an incompressible second-grade fluid around a rigid and smooth three-dimensional body \mathcal{B} . Attaching the system of coordinates to \mathcal{B} and assuming that the fluid is at rest at infinity, the problem is to determine the velocity field $v = (v_1, v_2, v_3)$ and the associated pressure field p from

$$\begin{cases} -v\Delta v - \alpha_1 v \cdot \nabla \Delta v + \nabla p = f + \nabla \cdot N(v) \\ \nabla \cdot v = 0 \\ v = v_*, \quad v_* \cdot n = 0 \\ \lim_{|x| \rightarrow \infty} v(x) = 0 \end{cases} \quad \begin{matrix} & \text{in } \Omega \\ & \text{on } \partial\Omega, \end{matrix} \quad (3.1)$$

where $\Omega = \mathbb{R}^3 \setminus \overline{\mathcal{B}}$, $x = (x_1, x_2, x_3) \in \Omega$ and

$$N(v) = \alpha_1 (\nabla v)^T A_1(v) - v \otimes v.$$

The first existence result is achieved by approximating problem (3.1) by a sequence of suitable problems in bounded domains, contained in, and eventually invading Ω . The solutions obtained should be considered as the analogue of the so called D -solutions for the Navier-Stokes equations having a finite Dirichlet integral, see, e.g., [22, 15]. The proof of the following theorem can be found in [17].

Theorem 3.1 Let $\partial\Omega \in C^3$ and assume that $f \in H^2(\Omega) \cap L^{6/5}(\Omega)$ and $v_* \in H^{5/2}(\partial\Omega)$ are given, with

$$\|f\|_2 + \|f\|_{0,\frac{6}{5}} + \|v_*\|_{\frac{5}{2},\partial\Omega} \leq \varepsilon.$$

For $\varepsilon > 0$ sufficiently small, problem (3.1) admits a solution (v, p) satisfying

$$v \in L^6(\Omega), \quad \nabla v \in H^2(\Omega), \quad \nabla p \in L^2(\Omega), \quad p - p_0 \in L^6(\Omega),$$

with some $p_0 \in \mathbb{R}$. Moreover, it holds

$$\|\nabla v\|_2 + \|\nabla p\|_0 \leq c \left(\|f\|_2 + \|f\|_{0,\frac{6}{5}} + \|v_*\|_{\frac{5}{2},\partial\Omega} \right).$$

As far as the uniqueness and asymptotic behavior are concerned, problem (3.1) has to be investigated in a more restricted class of solutions. Following the ideas and terminology introduced by Finn in his study of the Navier-Stokes equations [8, 9], we look for *physically reasonable* solutions, i.e., solutions which are sufficiently regular and have the decay

$$\mathbf{v}(x) = O(|x|^{-1}), \quad \nabla \mathbf{v}(x) = O(|x|^{-2}), \quad p(x) = O(|x|^{-2}). \quad (3.2)$$

Hence, for $k \geq 1$, let us define the following Banach spaces

$$V_k = \left\{ \mathbf{v} \in \mathbf{L}^6(\Omega), \nabla \mathbf{v} \in \mathbf{H}^{k+1}(\Omega) : |x|\mathbf{v}, |x|^2\nabla \mathbf{v}, |x|\nabla^2 \mathbf{v} \in \mathbf{L}^\infty(\Omega) \right\}$$

$$P_k = \left\{ \pi \in H^k(\Omega) : |x|^2\pi \in L^\infty(\Omega), |x|\nabla \pi \in L^\infty(\Omega) \right\},$$

and the corresponding norms $\|\cdot\|_{V_k}$ and $\|\cdot\|_{P_k}$. To simplify the presentation, we assume here that the only data of the problem is the boundary velocity \mathbf{v}_* . It is clear that the results presented in the following theorem remain the same also in the presence of a body force term f , provided it belongs to a suitable function space, see [32] for more details.

Theorem 3.2 Let $\partial\Omega \in C^{k+2}$, with $k \geq 5$, be given and let $\mathbf{v}_* \in \mathbf{H}^{\frac{k+3}{2}}(\partial\Omega)$. There exists a constant $\gamma > 0$ such that for $\|\mathbf{v}_*\|_{k+\frac{3}{2}, \partial\Omega} \leq \gamma$, problem (3.1) admits a unique solution $(\mathbf{v}, p) \in V_k \times P_k$ satisfying

$$\|\mathbf{v}\|_{V_k} + \|p\|_{P_k} \leq c \|\mathbf{v}_*\|_{k+\frac{3}{2}, \partial\Omega}.$$

The proof of Theorem 3.2 is based on the decomposition (2.10) and it uses the integral representation of the solution of the Stokes problem (2.10)_{1,2} in terms of the fundamental matrix of the Stokes operator and the problem data. The decay estimates for the transport equation (2.10)₃ can be obtained directly from the equation. The additional decay assumptions $D^2 \mathbf{v}(x) = O(|x|^{-1})$ and $\nabla p(x) = O(|x|^{-1})$, see spaces V_k and P_k above, although not being optimal, are needed to close the contraction scheme. Observe that by estimating the (weakly) singular integrals of the integral representation formulas of the Stokes problem one does not obtain the decay rates, $D^{k+1} \mathbf{v}(x) = O(|x|^{-(k+2)})$ and $D^k p(x) = O(|x|^{-(k+2)})$ predicted by the fundamental solution, for $k \geq 1$.

As noted above, the decay properties of the solution obtained in Theorem 3.2 are not optimal. This problem can be overcome by considering the equations in weighted Sobolev spaces with detached asymptotics. We refer to [27], see also [28, 29], for an overview of results on related problems in fluid mechanics, e.g., Stokes system and Navier-Stokes equations, in weighted function spaces with detached asymptotics. The idea of detaching the (main) asymptotics is to look for a solution having the asymptotic form

$$\mathbf{v}(x) = r^{-1} \mathbf{V}(\theta) + \tilde{\mathbf{v}}(x), \quad p(x) = r^{-2} P(\theta) + \tilde{p}(x), \quad (3.3)$$

where \mathbf{V} and P are functions on the sphere \mathbb{S}^2 , $r = |x|$, $\theta = |x|^{-1}x \in \mathbb{S}^2$ and the remainder parts $\tilde{\mathbf{v}}(x)$ and $\tilde{p}(x)$ have better decay properties than the main

asymptotic terms $r^{-1} V(\theta)$ and $r^{-2} P(\theta)$, the behavior of which is defined by the fundamental matrix of the Stokes operator in \mathbb{R}^3 .

The weighted Sobolev spaces $V_\gamma^l(\Omega)$ are defined as the closure of $C_0^\infty(\overline{\Omega})$ -functions with respect to the weighted norm

$$\|u; V_\gamma^l(\Omega)\| = \sum_{k=0}^l \|r^{\gamma-l+k} D^k u; L^2(\Omega)\|,$$

where $l \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\gamma \in \mathbb{R}$, $q \in (1, \infty)$, and $D^k u$ stands for the system of all k -th order derivatives of the function u . In order to abbreviate the notation, let us define the product space $\mathcal{R}_\gamma^l V(\Omega)$ as a space of functions (f, v_*) such that f has the asymptotic form

$$f(x) = r^{-3} F(\theta) + \tilde{f}(x), \quad (3.4)$$

with $F \in H^l(\mathbb{S}^2)$, $\tilde{f} \in V_\gamma^{l-1}(\Omega)$, and $v_* \in H^{\frac{l+1}{2}}(\partial\Omega)$. The space $\mathcal{R}_\gamma^l V(\Omega)$ is equipped with the norm

$$\|(f, v_*); \mathcal{R}_\gamma^l V(\Omega)\| = \|F\|_{l, \mathbb{S}^2} + \|\tilde{f}; V_\gamma^{l-1}(\Omega)\| + \|v_*\|_{l+\frac{1}{2}, \partial\Omega}.$$

Moreover, let $\mathcal{R}_\gamma^l V(\Omega)_\perp$ be the space of functions in $\mathcal{R}_\gamma^l V(\Omega)$ such that F satisfies the compatibility condition

$$\int_{\mathbb{S}^2} F(\theta) ds_\theta = 0.$$

Now, we can recall the following theorem, cf. [30].

Theorem 3.3 Assume that $(f, v_*) \in \mathcal{R}_\gamma^l V(\Omega)_\perp$, with $l \geq 2$ and $\gamma - l \in \left(\frac{1}{2}, \frac{3}{2}\right)$.

There exists a positive constant $\varepsilon > 0$ such that if

$$\|(f, v_*); \mathcal{R}_\gamma^l V(\Omega)\| \leq \varepsilon,$$

then problem (3.1) has a unique solution (v, p) admitting the asymptotic representation (3.3), with $V \in H^{l+2}(\mathbb{S}^2)$, $\tilde{v} \in V_\gamma^{l+1}(\Omega)$, $P \in H^{l+1}(\mathbb{S}^2)$ and $\tilde{p} \in V_{\gamma-1}^{l-1}(\Omega)$. Moreover, the following estimate holds

$$\|V\|_{l+2, \mathbb{S}^2} + \|\tilde{v}; V_\gamma^{l+1}(\Omega)\| + \|P\|_{l+1, \mathbb{S}^2} + \|\tilde{p}; V_{\gamma-1}^{l-1}(\Omega)\| \leq c \|(f, v_*); \mathcal{R}_\gamma^l V(\Omega)\|.$$

Remark 3.1 In all the previous theorems, p is defined by (2.11) in terms of the solution of the decomposed problem (2.10). Hence, the “loss” of regularity for p in comparison, e.g., with the solution (v, p) of the Navier-Stokes equations. However, notice that in the solution obtained in Theorem 3.3, one loses regularity only in the remainder part \tilde{p} and not in the main asymptotic term.

The assumption that the functions defined on the sphere \mathbb{S}^2 are more regular than the functions defining the remainder terms is needed for the solvability of the transport equation, see [30].

Next, consider the system of equations (2.16) governing the motion of a fluid of the Oldroyd type in an three-dimensional exterior domain, together with the condition

$$\lim_{|x| \rightarrow \infty} v(x) = 0. \quad (3.5)$$

If v and p have the form (3.3), it seems reasonable to look for the elastic part of the stress tensor τ in the asymptotic form

$$\tau(x) = r^{-2} T(\theta) + \tilde{\tau}(x). \quad (3.6)$$

Let us recall the following result, see [30] for the proof.

Theorem 3.4 Let $(f, v_*) \in \mathcal{R}_\gamma^l V(\Omega)_\perp$, with $l \geq 2$, $\gamma - l \in \left(\frac{1}{2}, \frac{3}{3}\right)$ and

$$\|(f, v_*); \mathcal{R}_\gamma^l V(\Omega)\| \leq \varepsilon.$$

There exists $\omega_0 \in (0, 1)$ such that for all $\omega \in (0, \omega_0]$ and for sufficiently small $\varepsilon > 0$, problem (2.16), (3.5) admits a unique solution (v, p, τ) of the form (3.3), (3.6) such that

$$\begin{aligned} V &\in H^{l+2}(\mathbb{S}^2), \quad \tilde{v} \in V_\gamma^{l+1}(\Omega), \quad P \in H^{l+1}(\mathbb{S}^2), \quad \tilde{p} \in V_\gamma^l(\Omega), \\ T &\in H^{l+1}(\mathbb{S}^2), \quad \tilde{\tau} \in V_\gamma^l(\Omega). \end{aligned}$$

Moreover, this solution satisfies the estimate

$$\begin{aligned} \|V\|_{l+2, \mathbb{S}^2} + \|\tilde{v}; V_\gamma^{l+1}(\Omega)\| + \|P\|_{l+1, \mathbb{S}^2} + \|\tilde{p}; V_\gamma^l(\Omega)\| + \\ + \|T\|_{l+1, \mathbb{S}^2} + \|\tilde{\tau}; V_\gamma^l(\Omega)\| \leq c \|(f, v_*); \mathcal{R}_\gamma^l V(\Omega)\|. \end{aligned}$$

3.2 The case $v_\infty \neq 0$.

Next, we shall investigate the motion of a compact body $\mathcal{B} \subset \mathbb{R}^n$, $n = 2, 3$, moving steadily in an incompressible viscoelastic fluid. Attaching again the system of coordinates to \mathcal{B} , one may reformulate the problem and study the motion of the liquid moving past the fixed obstacle \mathcal{B} and having a constant, non-zero velocity v_∞ at infinity.

First, let us consider the three-dimensional case and give existence results both for the second-grade and for the Oldroyd-B fluid. Although the corresponding decomposed systems for both problems are uniquely solvable in a function space where, in particular, $(v - v_\infty) \in L^4(\Omega)$, this is not enough to show uniqueness for the original system of equations of a second-grade fluid. Therefore, in this case one searches for a solution satisfying $(v - v_\infty) \in L^3(\Omega)$. Finally, we shall consider the

problem of the plane flow past an obstacle. For simplicity, we assume throughout this section that $v \equiv 0$ on $\partial\Omega$.

We shall present the equations of motion in a nondimensional form so that it becomes evident that the smallness assumption is needed only on the data, and not on the material coefficients. Rotating the coordinate axes in such a way that $v_\infty = v_\infty(1, 0, 0)$, $v_\infty > 0$, we obtain the following nondimensional system for $v = (v_1, v_2, v_3)$ and p , see e.g., [12]

$$\begin{cases} -\Delta v - \mathcal{W} v \cdot \nabla \Delta v + \nabla p = f + \nabla \cdot N(v) & \text{in } \Omega \\ \nabla \cdot v = 0 \\ v = 0 \\ \lim_{|x| \rightarrow \infty} v(x) = e_1 \end{cases} \quad (3.7)$$

Here, $\Omega = \mathbb{R}^3 / \overline{\mathcal{B}}$ and the nonlinear term $N(v)$ is given by

$$N(v) = \mathcal{W} (\nabla v)^T A_1(v) - \mathcal{R} v \otimes v.$$

Moreover, the nondimensional numbers \mathcal{R} , the *Reynolds number*, and \mathcal{W} , the *Weissenberg number*, are defined by

$$\mathcal{R} = \frac{\rho U d}{\eta}, \quad \mathcal{W} = \frac{\alpha_1 U}{\eta d},$$

where U is the characteristic velocity and d is the characteristic length. We may take $U = v_\infty$ and d to be the diameter of \mathcal{B} . Further, p and f are relabeled appropriately.

Setting $u = v - e_1$, we consider the following decoupled problem for (u, π, w)

$$\begin{cases} -\Delta u + \mathcal{R} \frac{\partial u}{\partial x_1} + \nabla \pi = w \\ \nabla \cdot u = 0 \\ w + \mathcal{W} (u + e_1) \cdot \nabla w = f + \nabla \cdot N(u, \pi) \\ u = -e_1 \\ \lim_{|x| \rightarrow \infty} u(x) = 0 \end{cases} \quad (3.8)$$

where $N(u, \pi)$ denotes a nonlinear term given by

$$N(u, \pi) = \mathcal{W} ((\nabla u)^T A_1(u) - \pi (\nabla u)^T) - \mathcal{R} u \otimes u + \mathcal{W} \mathcal{R} \frac{\partial u}{\partial x_1} \otimes u + \mathcal{W} \mathcal{R} \frac{\partial u}{\partial x_1} \otimes e_1.$$

Proving that (3.8) admits a solution (u, π, w) in some suitable function space, implies that problem (3.7) is solvable with $v = u + e_1$ and the pressure p given by

$$p = \pi + \mathcal{W} v \cdot \nabla \pi.$$

Note that the linearization of (3.8) corresponds to a coupled problem composed of the Oseen system and the transport equation. The following result is proven in [31], see also [49].

Theorem 3.5 Let $\partial\Omega \in C^{k+2}$, with $k \geq 1$ an integer, and let $f \in H^k(\Omega) \cap L^{\frac{6}{5}}(\Omega)$, with

$$\|f\|_k + \|f\|_{0,\frac{6}{5}} \leq \varepsilon$$

be given. Then, there exist constants $\mathcal{R}_0, \mathcal{W}_0, \varepsilon_0 \geq 0$ such that, for all $\mathcal{R} \in (0, \mathcal{R}_0)$, $\mathcal{W} \in (0, \mathcal{W}_0)$ and $\varepsilon \in (0, \varepsilon_0)$, problem (3.7) admits a unique solution (v, p) such that

$$\begin{aligned} (v - e_1) &\in L^3(\Omega), \quad \nabla v \in L^{\frac{12}{7}}(\Omega), \quad D^2 v \in L^{\frac{6}{5}}(\Omega) \cap H^k(\Omega) \\ p &\in H^k(\Omega). \end{aligned}$$

Moreover, it holds

$$\begin{aligned} \mathcal{R}^{\frac{1}{2}} \|v - v_\infty\|_{0,3} + \mathcal{R}^{\frac{1}{4}} \|\nabla v\|_{0,\frac{12}{7}} + \|D^2 v\|_{0,\frac{6}{5}} + \|D^2 v\|_k + \|p\|_k &\leq \\ \leq c \left(1 + \|f\|_k + \|f\|_{0,\frac{6}{5}} \right), \end{aligned} \tag{3.9}$$

for some constant $c = c(\Omega, \mathcal{R}_0, \mathcal{W}_0, \varepsilon_0) > 0$.

For the Oldroyd-B fluid, assuming again that $v_\infty = v_\infty e_1$, $v_\infty > 0$, the equations of motion take the following nondimensional form.

$$\begin{cases} -(1-\omega)\Delta v + \mathcal{R} v \cdot \nabla v + \nabla p = f + \nabla \cdot \tau \\ \tau + \mathcal{W} \left((v \cdot \nabla) \tau - \nabla v \tau - \tau (\nabla v)^T \right) = \omega A_1(v) \quad \text{in } \Omega \\ \nabla \cdot v = 0 \\ v = 0 \quad \text{on } \partial\Omega \\ \lim_{|x| \rightarrow \infty} v(x) = e_1, \end{cases} \tag{3.10}$$

where \mathcal{R} and \mathcal{W} are the corresponding Reynolds and Weissenberg numbers. See [49] for the proof of the following result.

Theorem 3.6 Let $\partial\Omega \in C^{k+2}$, with $k \geq 1$, and let $f \in \mathcal{H}^{-1}(\Omega) \cap H^k(\Omega)$, with

$$\|f; \mathcal{H}^{-1}(\Omega)\| + \|f\|_k \leq \varepsilon$$

be given. Then, there exist $\mathcal{R}_0 > 0$, $\mathcal{W}_0 > 0$ and $\varepsilon_0 > 0$ such that for all $\mathcal{R} \in (0, \mathcal{R}_0)$, $\mathcal{W} \in (0, \mathcal{W}_0)$ and $\varepsilon \in (0, \varepsilon_0)$, one can find $\omega_0 \in (0, 1)$, with the property that for all $0 < \omega < \omega_0$, problem (3.10) admits a unique solution (v, p, τ) such that

$$(v - e_1) \in L^4(\Omega), \quad \nabla v \in H^{k+1}(\Omega), \quad p \in H^{k+1}(\Omega), \quad \tau \in H^{k+1}(\Omega).$$

Moreover, this solution satisfies the estimate

$$\begin{aligned} \mathcal{R}^{\frac{1}{4}} \|\boldsymbol{v} - \boldsymbol{e}_1\|_{0,4} + \|\nabla \boldsymbol{v}\|_{k+1} + \|p\|_{k+1} + \|\boldsymbol{\tau}\|_{k+1} &\leq \\ &\leq c \left(1 + \|f; \mathcal{H}^{-1}(\Omega)\| + \|f\|_k \right) \end{aligned} \quad (3.11)$$

with the constant $c = c(\Omega, \mathcal{R}_0, \mathcal{W}_0, \varepsilon_0, w_0) > 0$.

Above $\mathcal{H}^{-1}(\Omega)$ stands for the dual space of the homogeneous Sobolev space $\mathcal{H}_0^1(\Omega)$ obtained as the closure of $C_0^\infty(\overline{\Omega})$ -functions in the norm $\|\cdot; \mathcal{H}_0^1(\Omega)\| = \|\nabla \cdot\|_0$.

We shall end our discussion on the flows past an obstacle by considering problem (3.7) in a two-dimensional exterior domain. The functional setting is somewhat more technical, due to the solvability of the Oseen system, cf. [15].

The main theorem reads as follows, cf. [49].

Theorem 3.7 Let $\Omega \in C^3$ be an exterior domain in \mathbb{R}^2 . For $1 < q < \frac{6}{5}$ and $2 < r < \infty$, let $f \in W^{1,q}(\Omega) \cap W^{1,r}(\Omega)$, with

$$\|f\|_{1,q} + \|f\|_{1,r} \leq \varepsilon$$

be given. Then, there exist constants $\mathcal{R}_0, \varepsilon_0 > 0$ such that, for all $\mathcal{R} \in (0, R_0)$ and $\varepsilon \in (0, \varepsilon_0)$, problem (3.7) admits a solution (\boldsymbol{v}, p) such that

$$\begin{aligned} (\boldsymbol{v} - \boldsymbol{e}_1) &\in L^{\frac{3q}{3-2q}}(\Omega), \quad \nabla \boldsymbol{v} \in L^{\frac{3q}{3-q}}(\Omega) \cap W^{2,r}(\Omega), \quad \nabla v_2, \frac{\partial \boldsymbol{v}}{\partial x_1} \in L^q(\Omega) \\ D^2 \boldsymbol{v} &\in W^{1,q}(\Omega), \quad p, v_2 \in L^{\frac{2q}{2-q}}(\Omega), \quad \nabla p \in L^q(\Omega) \cap L'(\Omega). \end{aligned}$$

Moreover, it holds

$$\begin{aligned} \mathcal{R}^{\frac{2}{3}} \|\boldsymbol{v} - \boldsymbol{e}_1\|_{0,\frac{3q}{3-2q}} + \mathcal{R} \left(\left\| \frac{\partial \boldsymbol{v}}{\partial x_1} \right\|_{0,q} + \|v_2\|_{0,\frac{2q}{2-q}} + \|\nabla v_2\|_{0,q} \right) + \\ + \mathcal{R}^{\frac{1}{3}} \left(\|\boldsymbol{v}\|_{0,\frac{6q}{6-5q}} + \|\nabla \boldsymbol{v}\|_{0,\frac{3q}{3-q}} + \|\nabla \boldsymbol{v}\|_{2,r} + \|D^2 \boldsymbol{v}\|_{1,q} + \right. \\ \left. + \|p\|_{0,\frac{2q}{2-q}} + \|\nabla p\|_{0,q} + \|\nabla p\|_{0,r} \right) &\leq c \left(1 + \|f\|_{1,q} + \|f\|_{1,r} \right), \end{aligned} \quad (3.12)$$

for some constant $c = c(\Omega, \mathcal{R}_0, q, r, \mathcal{W}, \varepsilon_0) > 0$.

Remark 3.2. The uniqueness remains an open question.

Remark 3.3 The asymptotic behavior at large distances of the solutions established in Theorems 3.5, 3.6, and 3.7, as well as the asymptotic structure of the solution to

the equations governing the motion of Maxwell fluid has been recently studied in [38].

4. RESULTS IN DOMAINS WITH OUTLETS TO INFINITY

In this section, we study second- and third-grade Rivlin-Ericksen fluids and fluids of the Oldroyd type in a domain $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, having m ($m > 1$) cylindrical outlets to infinity. We assume that outside a certain ball $|x| = R_0$, $R_0 > 0$, the domain Ω splits into $m > 1$ connected components Ω_j , $j = 1, \dots, m$ (outlets to infinity), i.e.,

$$\Omega = \Omega_0 \cup \left(\bigcup_{j=1}^m \Omega_j \right), \quad \Omega_0 = \Omega \cap \{x : |x| < R_0\}.$$

By $\sum_j \subset \mathbb{R}^{n-1}$ we denote an arbitrary, constant cross-section of Ω_j with smooth boundary $\partial \Sigma_j$. Concerning the Rivlin-Ericksen fluids, the problem can be formulated as follows.

Given the fluxes $F_j \in \mathbb{R}$, $j = 1, \dots, m$, with $\sum_{j=1}^m F_j = 0$, find the velocity field v

and the pressure function p satisfying the equations

$$\begin{cases} -v\Delta v - \alpha_1 v \cdot \nabla \Delta v + \nabla p = \nabla \cdot N(v) \\ \nabla \cdot v = 0 & \text{in } \Omega \\ v = 0 & \text{on } \partial \Omega \\ \int_{\Sigma_j} v \cdot n = F_j, \quad j = 1, \dots, m, \end{cases} \quad (4.1)$$

with $N(v)$ given by (2.7).

It is a standard procedure to construct a flux carrier v_0 , i.e., a solenoidal function v_0 satisfying the boundary condition $v_0 = 0$ on $\partial \Omega$ and the flux conditions (4.1)₄ so that the problem can be studied with zero fluxes. The existence of such a flux carrier is known for very general domains with outlets to infinity [46, 47], see also [35] (and all the references quoted therein) where the Navier-Stokes equations in similar domains have been considered. However, here we wish to choose the flux carrier v_0 and the corresponding “artificial pressure function” p_0 in such a way that (v_0, p_0) satisfies the original nonlinear equations (4.1) for large $|x|$, i.e., (v_0, p_0) solves (4.1) in each cylindrical outlet Ω_j . Consequently, we look for the solution (v, p) in the form

$$v = u + v_0, \quad p = q + p_0.$$

This yields the following system for (u, q)

$$\begin{cases} -v \Delta \mathbf{u} - \alpha_1 (\mathbf{u} + \mathbf{v}_0) \cdot \nabla \Delta \mathbf{u} + \nabla q = \mathbf{M}(\mathbf{v}_0, p_0) + \\ \quad + \nabla \cdot \mathbf{N}(\mathbf{u}) + \nabla \cdot \mathbf{N}(\mathbf{u}, \mathbf{v}_0) + \alpha_1 \mathbf{u} \cdot \nabla \Delta \mathbf{v}_0 & \text{in } \Omega \\ \nabla \cdot \mathbf{u} = 0 & \\ \mathbf{u} = 0 & \text{on } \partial\Omega \\ \int_{\Sigma_j} \mathbf{u} \cdot \mathbf{n} = 0, \quad j = 1, \dots, m, & \end{cases} \quad (4.2)$$

where

$$\mathbf{M}(\mathbf{v}_0, p_0) = v \Delta \mathbf{v}_0 + \alpha_1 \mathbf{v}_0 \cdot \nabla \Delta \mathbf{v}_0 + \nabla \cdot \mathbf{N}(\mathbf{v}_0) - \nabla p_0 \quad (4.3)$$

and $\mathbf{N}(\mathbf{u}, \mathbf{v}_0)$ denotes the part of $\mathbf{N}(\mathbf{u} + \mathbf{v}_0)$ containing only mixed terms of \mathbf{u} and \mathbf{v}_0 . Note that $\mathbf{M}(\mathbf{v}_0, p_0)$ has compact support.

As it comes to the fluids of the Oldroyd type, we look for the solution $(\mathbf{v}, p, \boldsymbol{\tau})$ in the form

$$\mathbf{v} = \mathbf{u} + \mathbf{v}_0, \quad p = q + p_0, \quad \boldsymbol{\tau} = \boldsymbol{\xi} + \boldsymbol{\tau}_0,$$

where $\boldsymbol{\tau}_0 = v_0 \omega A_1(\mathbf{v}_0)$ and (\mathbf{v}_0, p_0) is the flux carrier. Substituting these sums into (2.16), we obtain for $(\mathbf{u}, q, \boldsymbol{\xi})$ the following set of equations

$$\begin{cases} -v_0(1-\omega) \Delta \mathbf{u} + \nabla q = \mathbf{M}(\mathbf{v}_0, p_0) + \mathbf{F}_1(\mathbf{u}, \mathbf{v}_0) + \nabla \cdot \boldsymbol{\xi} \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega \\ \mathbf{u} = 0 & \text{on } \partial\Omega \\ \int_{\Sigma_j} \mathbf{u} \cdot \mathbf{n} = 0, \quad j = 1, \dots, m \\ \boldsymbol{\xi} + \lambda_1 (\mathbf{u} + \mathbf{v}_0) \cdot \nabla \boldsymbol{\xi} = \mathbf{F}_2(\boldsymbol{\tau}_0, \mathbf{v}_0) + \mathbf{F}_3(\boldsymbol{\tau}_0, \mathbf{v}_0, \boldsymbol{\xi} \mathbf{u}) & \text{in } \Omega, \end{cases} \quad (4.4)$$

where

$$\mathbf{M}(\mathbf{v}_0, p_0) = v_0 \Delta \mathbf{v}_0 - \mathbf{v}_0 \cdot \nabla \mathbf{v}_0 - \nabla p_0,$$

$$\mathbf{F}_1(\mathbf{u}, \mathbf{v}_0) = -\mathbf{u} \cdot \nabla \mathbf{u} - \mathbf{u} \cdot \nabla \mathbf{v}_0 - \mathbf{v}_0 \cdot \nabla \mathbf{u},$$

$$\mathbf{F}_2(\boldsymbol{\tau}_0, \mathbf{v}_0) = -\lambda_1 \mathbf{v}_0 \cdot \nabla \boldsymbol{\tau}_0 - \lambda_1 \mathbf{g}(\boldsymbol{\tau}_0, \nabla \mathbf{v}_0),$$

$$\begin{aligned} \mathbf{F}_3(\boldsymbol{\tau}_0, \mathbf{v}_0, \boldsymbol{\xi} \mathbf{u}) = & -\lambda_1 (\mathbf{u} \cdot \nabla \boldsymbol{\tau}_0 - \mathbf{g}(\boldsymbol{\xi} \nabla \mathbf{u}) - \mathbf{g}(\boldsymbol{\tau}_0, \nabla \mathbf{u}) - \mathbf{g}(\boldsymbol{\xi} \nabla \mathbf{v}_0) + \\ & + v_0 \omega A_1(\mathbf{u})). \end{aligned}$$

Note that if (\mathbf{v}_0, p_0) solves the original problem for large $|x|$, then the terms $\mathbf{M}(\mathbf{v}_0, p_0)$ and $\mathbf{F}_2(\boldsymbol{\tau}_0, \mathbf{v}_0)$ have compact supports.

Now, whether studying the Rivlin-Ericksen fluids via the decomposition scheme (2.9) or considering the coupled system (4.4) for the Oldroyd type fluids, one ends up with a coupled problem composed of a Stokes system with zero flux conditions and of a transport equation. Since it is well known that the solution of the Stokes problem in a domain with cylindrical outlets vanishes exponentially, provided the data have either a compact support or vanish exponentially, cf. [23],

29], it is reasonable to study the equations in spaces of functions having exponential decay at infinity. Before introducing the spaces, let us construct the flux carriers (v_0, p_0) .

4.1 Construction of the Flux Carriers

First, we consider the problems (4.1) and (2.16) in an infinite straight cylinder

$$\Pi = \{x \in \mathbb{R}^n : (x', x_n) \in \Sigma \times \mathbb{R}\},$$

where $\Sigma \subset \mathbb{R}^{n-1}$ is an arbitrary bounded cross-section.

4.1.1 Rivlin-Ericksen Fluids

The equations governing the motion of a second-grade fluid admit an exact solution (v, p) of Poiseuille type in Π , see e.g., [42, 5], i.e., a solution of the form

$$\begin{aligned} v &= (0', v_n(x')), \quad v_n(x') = \frac{F z_0(x')}{\kappa_0}, \\ p(x', x_n) &= -\frac{v F}{\kappa_0} x_n - \frac{\alpha_1 F}{\kappa_0} v_n + \frac{\alpha_1}{2} |\nabla' v_n|^2, \end{aligned} \tag{4.5}$$

where F is the given flux and $\kappa_0 = \int_{\Sigma} z_0(x') dx' \neq 0$, with z_0 satisfying

$$\begin{cases} -\Delta' z_0 = 1 & \text{in } \Sigma \\ z_0 = 0 & \text{on } \partial \Sigma. \end{cases} \tag{4.6}$$

Here, Δ' and ∇' denote the Laplace operator and the gradient with respect to the first $n-1$ variables. As is well known, the smoothness of the unique solution z_0 of (4.6) depends on the smoothness of the boundary $\partial \Sigma$. In a two-dimensional channel of width d , one has the classical Poiseuille flow with the corresponding modified pressure field

$$v_2(x_1) = \frac{3F}{2d} \left(1 - 4 \frac{x_1^2}{d^2} \right), \quad p(x_1, x_2) = -\frac{12vF}{d^3} x_2 + \frac{18\alpha_1 F^2}{d^4} \left(8 \frac{x_1^2}{d^2} - 1 \right)$$

and in a circular pipe of radius $R > 0$, the solution is given by

$$v_3(|x'|) = \frac{2F}{\pi R^2} \left(1 - \frac{|x'|^2}{R^2} \right), \quad p(|x'|, x_3) = -\frac{8vF}{\pi R^4} x_3 + \frac{8\alpha_1 F^2}{\pi^2 R^6} \left(3 \frac{|x'|^2}{R^2} - 2 \right).$$

In a third-grade fluid a steady rectilinear motion cannot take place in general, apart from the exceptional cases of the channel flow and of the flow in a straight pipe where the cross-section is either circular or the annulus between two concentric circles, cf. [14]. In \mathbb{R}^2 , with

$$v = (0, v_2(x_1)), \quad p = -G x_2 + q(x_1), \tag{4.7}$$

where G is a given pressure drop, one obtains from (4.1)₂

$$-\nu v_2'' - 2\beta(v_2')' = G, \quad (4.8)$$

where $v_2' = dv_2/dx_1$. It is easy to verify that (4.8), supplemented with the boundary conditions $v_2(-d) = v_2(d) = 0$, admits a unique solution and that the associated pressure field has the form

$$p(x) = -G x_2 + (2\alpha_1 + \alpha_2) |v_2'|^2.$$

Moreover, the flux F can be shown to be related to the pressure drop G in such a way that for each prescribed flux $F \in \mathbb{R}$, there exists a solution of the form (4.7) satisfying the estimate, cf. [37]

$$\left| \frac{d^k v_2(x_1)}{dx_1^k} \right| \leq c |F|, \quad k = 0, 1, \dots$$

Next, consider the equations for the third-grade fluid and let Π be a straight tube in \mathbb{R}^3 with a constant non-circular cross-section Σ . Denoting by $v_3(x')$, $x' = (x_1, x_2) = (x_1, x_2)$ the axial velocity, it seems reasonable to look for a velocity field $v(x')$ composed of $v_3(x')$ and of a secondary velocity component, say $v'(x') = (v_1(x'), v_2(x'))$, orthogonal to the axis of the pipe. Hence, one assumes that the solution has the form

$$v(x') = (v'(x'), v_3(x')), \quad p(x) = -G x_3 + q(x'), \quad (4.9)$$

where G denotes the constant pressure gradient.

Substituting (4.9) into (4.1), written in a single outlet Π , yields the following two problems

$$\begin{cases} -\nu \Delta' v' - \alpha_1 v' \cdot \nabla' \Delta' v' + \nabla' q = L'(v', v_3) & \text{in } \Sigma \\ \nabla' \cdot v' = 0 & \text{on } \partial \Sigma, \end{cases} \quad (4.10)$$

$$\begin{cases} -\nu \Delta' v_3 - \alpha_1 v' \cdot \nabla' \Delta' v_3 = G(F) + L_3(v', v_3) & \text{in } \Sigma \\ v_3 = 0 \quad \text{on } \partial \Sigma, \quad \int_{\Sigma} v_3 dx' = F, \end{cases} \quad (4.11)$$

where we have set $\nabla \cdot N(v) = (L'(v), L_3(v))$.

Studying the coupled set of equations (4.10) and (4.11), one arrives at the following conclusion, cf. [37, 49].

Theorem 4.1 Let $\Pi = \Sigma \times \mathbb{R}$ be an infinite cylinder with a bounded, connected cross section Σ with $\partial \Sigma \in C^{l+4,\delta}$, $l \geq 0, \delta \in (0, 1)$, and let $F \in \mathbb{R}$, with $|F| < \gamma_0$, be given. For $\gamma_0 > 0$ sufficiently small, problem (4.1) admits a solution of the form $v = (v(x'), v_3(x'))$, $p = -G(F)x_3 + q(x')$ such that $v \in C^{l+4,\delta}(\Sigma)$,

$\nabla' q \in C^{l+1,\delta}(\Sigma)$. Moreover, the solution is unique within this class of solutions and satisfies

$$\begin{aligned}\|v_3; C^{l+4,\delta}(\Sigma)\| &\leq c|F|, \\ \|v'; C^{l+4,\delta}(\Sigma)\| + \|\nabla' q; C^{l+1,\delta}(\Sigma)\| &\leq c|F|^2.\end{aligned}\quad (4.12)$$

Remark 4.1 Estimates (4.12) show, in particular, that the velocity components $v' = (v_1, v_2)$ are secondary, in comparison with the axial velocity v_3 .

Now, it is a standard matter to define a global flux carrier v_0 in the domain Ω with $m > 1$ cylindrical outlets to infinity, cf. [37]. In fact, if $\partial\Omega \in C^{l+5,\delta}$, then there exists a solenoidal velocity field $v_0 \in C^{l+4,\delta}(\Omega)$ vanishing on $\partial\Omega$ and coinciding with the previously constructed exact solutions in each outlet Ω_j , $j = 1, \dots, m$. Moreover, it holds

$$\|v_0; C^{l+4,\delta}(\Omega)\| \leq c|\vec{F}|,$$

where $|\vec{F}| = (F_1^2 + \dots + F_m^2)^{1/2}$.

4.1.2 Fluids of the Oldroyd Type

Let us first consider Oldroyd-type fluids with $a \neq 1$ in the expression of the objective derivative $\mathcal{D}_a/\mathcal{D}_t$, see (2.13). In this case, according to the Fosdick-Serrin conditions, cf. [14], a steady rectilinear flow cannot happen in an infinite pipe with a non-circular cross-section. To prove existence of secondary flows, we search for a solution (v, p, τ) to equations (2.16) in the form

$$\tau(x') = \begin{bmatrix} S(x') & \sigma(x') \\ \sigma^T(x') & t_{33}(x') \end{bmatrix}, \quad v(x') = \begin{bmatrix} u(x') \\ v_3(x') \end{bmatrix}, \quad p(x) = -Gx_3 + q(x'),$$

where

$$S(x') = \begin{bmatrix} \tau_{11}(x') & \tau_{12}(x') \\ \tau_{21}(x') & \tau_{22}(x') \end{bmatrix}, \quad \sigma(x') = \begin{bmatrix} \tau_{31}(x') \\ \tau_{32}(x') \end{bmatrix}.$$

Substituting (4.13) into (2.16), one manages to split the components of the stress tensor and velocity into a set of five coupled equations. Analysis of this system leads to the following result, cf. [37].

Theorem 4.2 Let $\Pi = \Sigma \times \mathbb{R}$ be an infinite cylinder with a bounded, connected cross section Σ with $\partial\Sigma \in C^{l+4,\delta}$, $l \geq -1$, $\delta \in (0, 1)$, and let $F \in \mathbb{R}$, with $|F| < \gamma_0$, for some constant $\gamma_0 > 0$.

There exists $\omega_0 \in (0, 1)$ such that if $\gamma_0 > 0$ is sufficiently small, then problem (2.16) admits for all $\omega \in (0, \omega_0]$ a solution of the form (4.13) such that

$v \in C^{l+4,\delta}(\Sigma)$, $\nabla' q \in C^{l+2,\delta}(\Sigma)$ and $\tau \in C^{l+3,\delta}(\Sigma)$. Moreover, the solution is unique within this class of solutions and satisfies

$$\begin{aligned} \|v_3; C^{l+4,\delta}(\Sigma)\| + \|\sigma; C^{l+3,\delta}(\Sigma)\| &\leq c |F|, \\ \|u; C^{l+4,\delta}(\Sigma)\| + \|\nabla' q; C^{l+2,\delta}(\Sigma)\| + \|S; C^{l+3,\delta}(\Sigma)\| + \\ &+ \|C_{33}; C^{l+3,\delta}(\Sigma)\| \leq c |F|^2. \end{aligned} \quad (4.14)$$

Remark 4.2 If $a = 1$ or $a = -1$ in the expression of the objective derivative $\mathcal{D}_a/\mathcal{D}_t$ in the constitutive equation (2.13), hence, in particular if one considers the Oldroyd-B fluid, a rectilinear flow of Poiseuille type takes place in a straight cylinder with an arbitrary, bounded cross-section, cf. [34]. If $a \neq \pm 1$, a steady rectilinear flow is only possible in a channel or in a pipe with a constant circular cross-section. In fact, in this case, assuming that the velocity field has the form $v = (0, 0, v_3(r))$ and that the pressure is given by $p = -Gx_3 + q(r)$, one easily shows that $u_r = \partial u / \partial r$ satisfies the following cubic equation

$$v_0 \lambda_1 \lambda_2 (1 - a^2) u_r^3 + \frac{Gr}{2} \lambda_1^2 (1 - a^2) u_r^2 + v_0 u_r + \frac{Gr}{2} = 0. \quad (4.15)$$

Now, if $9\lambda_2 > \lambda_1$, i.e., for $\omega \in \left(0, \frac{8}{9}\right)$, the cubic equation (4.15) has one and only one real root for any G . For G sufficiently small, one can solve (4.15) for all values of G , cf. [19]. Furthermore, one obtains estimates for the solution u in terms of the flux $|F|$. Analogous results are valid for the two-dimensional channel flow.

The existence of a global flux carrier defined in all of Ω follows from standard reasoning as explained above for the Rivlin-Ericksen fluids.

4.2 Existence in Weighted Hölder Spaces

Since the terms $M(v_0, p_0)$ and $F_2(\tau_0, v_0)$ have compact support, it is convenient to study the nonlinear problems (4.1) and (2.16) in weighted Hölder spaces with exponential weights. These spaces, denoted by $\Lambda^{l,\delta}(\Omega; \beta)$, $l \geq 0$ an integer, $\delta \in (0, 1)$, $\beta \in \mathbb{R}$, are obtained as a closure of $C_0^\infty(\overline{\Omega})$ in the norm

$$\begin{aligned} \|u; \Lambda^{l,\delta}(\Omega; \beta)\| = \|u; C^{l,\delta}(\Omega_0)\| + \sum_{j=1}^m \left(\sum_{|\alpha| \leq l} \sup \exp \left(\beta \left(1 + |x_n^{(j)}|^2 \right)^{1/2} |D^\alpha u(x)| + \right. \right. \\ \left. \left. + \sum_{|\alpha|=l} \sup_{x \in \Omega_j} \left\{ \exp \left(\beta \left(1 + |x_n^{(j)}|^2 \right)^{1/2} \langle D^\alpha u \rangle_{\delta, \Omega_j}(x) \right) \right\} \right) \right)^{1/2}. \end{aligned}$$

with $\langle D^\alpha u \rangle_{\delta, \Omega_j}(x)$ denoting the usual Hölder quotient. Note that for $\beta = 0$ the space $\Lambda^{l,\delta}(\Omega; \beta)$ coincides with $C_0^{l,\delta}(\Omega)$ and that for $\beta > 0$ the elements of $\Lambda^{l,\delta}(\Omega; \beta)$, together with their derivatives up to the order l , vanish exponentially.

Let us recall the following main results, cf. [37].

Theorem 4.3 Let $l \geq 0$, $\delta \in (0, 1)$, $|\beta| \leq \beta_0$, $\partial\Omega \in C^{l+5,\delta}$ and $|\vec{F}| \leq \gamma_0$, with $\gamma_0 > 0$ sufficiently small. Problem (4.2) admits a unique solution

$$(\mathbf{u}, \nabla q) \in \Lambda^{l+3,\delta}(\Omega; \beta) \times \Lambda^{l,\delta}(\Omega; \beta)$$

and the following estimate holds

$$\|\mathbf{u}; \Lambda^{l+3,\delta}(\Omega; \beta)\| + \|\nabla q; \Lambda^{l,\delta}(\Omega; \beta)\| \leq c |\vec{F}|.$$

Theorem 4.4 Let $l \geq -1$, $\delta \in (0, 1)$, $|\beta| \leq \beta_0$, $\partial\Omega \in C^{l+5,\delta}$ and $|\vec{F}| \leq \gamma_0$, with $\gamma_0 > 0$ sufficiently small. There exists $\omega_0 \in (0, 1)$ such that for any $\omega \in (0, \omega_0]$ problem (4.4) admits a unique solution

$$(\mathbf{u}, \nabla q, \xi) \in \Lambda^{l+3,\delta}(\Omega; \beta) \times \Lambda^{l+1,\delta}(\Omega; \beta) \times \Lambda^{l+2,\delta}(\Omega; \beta)$$

and the following estimate holds

$$\|\mathbf{u}; \Lambda^{l+3,\delta}(\Omega; \beta)\| + \|\nabla q; \Lambda^{l+1,\delta}(\Omega; \beta)\| + \|\xi; \Lambda^{l+2,\delta}(\Omega; \beta)\| \leq c |\vec{F}|.$$

Remark 4.3 It is evident that all the results of this section remain valid in the case where the equations are nonhomogeneous. In other words, one can add to the right-hand side of the equations a given function f , expressing an external body force, belonging to an appropriate function space and having a sufficiently small norm.

It is also clear that all the results could be obtained similarly in weighted $L^q(\Omega)$ spaces, see also [36, 49].

Remark 4.4 Notice that the perturbation (\mathbf{u}, q) tends exponentially to the flux carrier (v_0, p_0) for all considered fluid models. Therefore, the leading asymptotic term of the solution is (v_0, p_0) . However, it should be mentioned that, when $|\vec{F}| \rightarrow 0$, the asymptotic behavior of (\mathbf{u}, q) is the same, $O(|\vec{F}|)$, as that of the flux carrier (v_0, p_0) .

5. FINITE ELEMENT APPROXIMATION

In this section, we present a mixed finite element approximation scheme for the second-grade fluid equations based on the decomposition (2.10) of the problem into

a Stokes system and a transport equation. The discussion is focused on the existence and uniqueness of solutions for the decoupled discretized problem and on the corresponding error estimates. The results are obtained by a fixed-point iteration which requires, as other classical iterative numerical methods, either an initial guess or the simulation of a transient problem. For the complete proofs of the results, we refer to [45]. Similar studies for the Oldroyd-type fluids have been recently carried out by other authors, see e.g., [26] and the references quoted therein.

When the discretization is based on the Galerkin variational form, the finite element spaces for the velocity and pressure in the Stokes problem cannot be chosen independently. An abstract theory for this type of saddle-point problems has been developed by Babuska [1] and Brezzi [2] (see also, e.g., [39], [18]). On the other hand, it is well-known that hyperbolic problems are difficult to solve by means of finite elements. Usually, highly refined meshes and specific upwinding techniques or artificial diffusivity must be used.

5.1 Formulation of the Problem

Let $\Omega \subset \mathbb{R}^2$ denote a convex polygonal domain and consider the following system of equations

$$\begin{cases} -\nu \Delta \mathbf{v} - \alpha_1 \mathbf{v} \cdot \nabla \Delta \mathbf{v} + \nabla p = \nabla \cdot N(\mathbf{v}) + \mathbf{f} \\ \nabla \cdot \mathbf{v} = 0 & \text{in } \Omega \\ \mathbf{v} = 0 & \text{on } \partial\Omega, \end{cases} \quad (5.1)$$

with $N(\mathbf{v}) = \alpha_1 (\nabla \mathbf{v})^T A_1(\mathbf{v}) - \mathbf{v} \otimes \mathbf{v}$. First, let us recall the following existence and uniqueness result, see e.g., [49].

Theorem 5.1 Let $\mathbf{f} \in W^{k,q}(\Omega)$, $k \geq 1$, $1 < q < \infty$ with $\|\mathbf{f}\|_{k,q} \leq \gamma$ and let $\partial\Omega \in C^{k+2}$. For sufficiently small $\gamma > 0$, problem (5.1) admits a unique solution $(\mathbf{v}, p) \in W^{k+2,q}(\Omega) \times [W^{k,q}(\Omega) \cap L_0^2(\Omega)]$ such that

$$\|\mathbf{v}\|_{k+2,q} + \|p\|_{k,q} \leq c \|\mathbf{f}\|_{k,q}.$$

Next, let us define the spaces $X = H_0^1(\Omega)$, $Q = L_0^2(\Omega)$ and $T = L^2(\Omega)$, with the norms $\|\mathbf{v}\|_x = \|\mathbf{D}(\mathbf{v})\|_0$, $\|q\|_Q = \|q\|_0$ and $\|\boldsymbol{\tau}\|_T = \|\boldsymbol{\tau}\|_0$, and consider the following variational formulation of the decomposed form of problem (5.1) (see (2.10) with $\mathbf{v} = 0$ on $\partial\Omega$).

Find $(\mathbf{v}, \boldsymbol{\pi}, z) \in X \times Q \times T$ solution of

$$\begin{cases} 2(\mathbf{D}(\mathbf{v}), \mathbf{D}(\mathbf{u})) - (\boldsymbol{\pi}, \nabla \cdot \mathbf{u}) = (z, \mathbf{D}(\mathbf{u})), \quad \forall \mathbf{u} \in X \\ (\nabla \cdot \mathbf{v}, q) = 0, \quad \forall q \in Q \\ \mathbf{v}(z, \boldsymbol{\tau}) + \alpha_1 (\mathbf{v} \cdot \nabla z, \boldsymbol{\tau}) - \alpha_1 (z \cdot \nabla \mathbf{v}^T, \boldsymbol{\tau}) = (N(\mathbf{v}, \boldsymbol{\pi}) + \mathbf{F}, \boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in T, \end{cases} \quad (5.2)$$

where $N(\mathbf{v}, \boldsymbol{\pi}) = N(\mathbf{v}) - \alpha_1 \boldsymbol{\pi} (\nabla \mathbf{v})^T$ and $\mathbf{D}(\mathbf{v}) = \frac{1}{2} \mathbf{A}_1(\mathbf{v})$.

Let \mathcal{T}_h be a uniformly regular family of triangulations made of triangles K with diameter h_K and $h = \max_{K \in \mathcal{T}_h} h_K$. Finally, let $\mathbb{P}^l(K)$ denote the space of all polynomials of degree less than or equal to l on $K \in \mathcal{T}_h$.

For fixed z , equations (5.2)_{1,2} form a Stokes system in the variables \mathbf{v} and $\boldsymbol{\pi}$. In order to use the well-known Hood-Taylor FEM, based on an adequate partition of $\overline{\Omega}$ into macro-elements, for the approximation of $(\mathbf{v}, \boldsymbol{\pi})$, we introduce the finite element spaces

$$X_h = \left\{ \mathbf{u}_h \in \mathbf{C}(\overline{\Omega}) : \mathbf{u}_h|_K \in \mathbb{P}^2(K), \forall K \in \mathcal{T}_h, \mathbf{u}_h|_{\partial\Omega} = 0 \right\} \subset X$$

$$\mathcal{Q}_h = \left\{ q_h \in L_0^2(\Omega) \cap C(\overline{\Omega}) : q_h|_K \in \mathbb{P}^1(K), \forall K \in \mathcal{T}_h \right\} \subset Q.$$

It can be shown that the spaces (X_h, \mathcal{Q}_h) satisfy the discrete inf-sup condition: $\exists \beta^* > 0$ (independent of h) such that

$$\inf_{q_h \in \mathcal{Q}_h - \{0\}} \sup_{\mathbf{u}_h \in X_h - \{0\}} \frac{|b(\mathbf{u}_h, q_h)|}{\|\mathbf{u}_h\|_X \|q_h\|_Q} \geq \beta^*,$$

where

$$b(\mathbf{u}, q) = - \int_{\Omega} q \nabla \cdot \mathbf{u} \, dx.$$

For fixed \mathbf{v} and $\boldsymbol{\pi}$, equation (5.2)₃ is a transport equation in the auxiliary variable z . Since no continuity requirement is needed on z , an upwinding technique based on a discontinuous Galerkin FEM introduced by Lesaint and Raviart [24] for the neutron transport equation, will be used. This allows the computation of z on an element by element basis (see also [26]). It is natural to define

$$\mathbf{T}_h = \left\{ \boldsymbol{\tau}_h \in \mathbf{T} : \boldsymbol{\tau}_h|_k \in \mathbb{P}^1(K), \forall K \in \mathcal{T}_h \right\}.$$

In order to describe this approximation, we introduce the notations

$$\partial K^-(\mathbf{v}) = \left\{ x \in \partial K : \mathbf{n}(x) \cdot \mathbf{v} < 0 \right\} \quad (\text{inflow boundary of } K)$$

$$\partial K^+(\mathbf{v}) = \left\{ x \in \partial K : \mathbf{x}(x) \cdot \mathbf{v} > 0 \right\} \quad (\text{outflow boundary of } K)$$

$$\boldsymbol{\tau}^-(\mathbf{v})(x) = \lim_{\varepsilon \rightarrow 0^-} \boldsymbol{\tau}(x + \varepsilon \mathbf{v}), \quad \boldsymbol{\tau}^+(\mathbf{v})(x) = \lim_{\varepsilon \rightarrow 0^+} \boldsymbol{\tau}(x + \varepsilon \mathbf{v}),$$

where ∂K is the boundary of $K \in \mathcal{T}_h$ and \mathbf{n} is the outward unit normal to K . We often drop the index h to simplify the notation. Further, let us introduce the trilinear form G_h on $X_h \times \mathbf{T}_h \times \mathbf{T}_h$ defined by

$$G_h(\mathbf{v}, \boldsymbol{\sigma}, \boldsymbol{\tau}) = \left((\mathbf{v} \cdot \nabla) \boldsymbol{\sigma}, \boldsymbol{\tau} \right)_h + \frac{1}{2} (\nabla \cdot \mathbf{v} \boldsymbol{\sigma}, \boldsymbol{\tau}) + \langle\langle \boldsymbol{\sigma}^+ - \boldsymbol{\sigma}^-, \boldsymbol{\tau}^+ \rangle\rangle_{h,v},$$

where

$$(\boldsymbol{\sigma}, \boldsymbol{\tau})_h = \sum_{K \in \mathcal{T}_h} (\boldsymbol{\sigma}, \boldsymbol{\tau})_K,$$

$$\ll \boldsymbol{\sigma}^\pm, \boldsymbol{\tau}^\pm \gg_{h,v} = \sum_{K \in \mathcal{T}_h} \int_{\partial K-(v)} (\boldsymbol{\sigma}^\pm(v) : \boldsymbol{\tau}^\pm(v)) |\mathbf{n} \cdot \mathbf{v}| ds.$$

Now, one can formulate the following discrete problem.

Find $(\mathbf{v}_h, \boldsymbol{\pi}_h, \mathbf{z}_h) \in X_h \times Q_h \times T_h$ such that

$$\begin{cases} 2(\mathbf{D}(\mathbf{v}_h), \mathbf{D}(\mathbf{u})) - (\boldsymbol{\pi}_h, \nabla \cdot \mathbf{u}) = (\mathbf{z}_h, \mathbf{D}(\mathbf{u})), & \forall \mathbf{u} \in X_h \\ (\nabla \cdot \mathbf{v}_h, q) = 0, & \forall q \in Q_h \\ v(\mathbf{z}_h, \boldsymbol{\tau}) + \alpha_1 G_h(\mathbf{v}_h, \mathbf{z}_h, \boldsymbol{\tau}) - \alpha_1 (\mathbf{z}_h \cdot \nabla \mathbf{v}_h^T, \boldsymbol{\tau}) = \\ = (N(\mathbf{v}_h, \boldsymbol{\pi}_h), \boldsymbol{\tau}) + (\mathbf{F}, \boldsymbol{\tau}), & \forall \boldsymbol{\tau} \in T_h. \end{cases} \quad (5.3)$$

5.2 Main Results

Since the continuous problem (5.2) is well-posed, having a sufficiently regular solution, the approximate problem (5.3) becomes also well-posed and can be solved using a fixed point iteration. Moreover, the discrete solution is close to the continuous one and error estimates are available. These results can be formulated as follows, cf. [45] for the proof.

Theorem 5.2 There exist two constants $\gamma_0 > 0$ and $h_0 > 0$ (independent of h , α and v) such that if problem (5.2) admits a solution

$$(\mathbf{v}, \boldsymbol{\pi}, \mathbf{z}) \in [\mathbf{H}^3(\Omega) \cap \mathbf{H}_0^1(\Omega)] \times [H^2(\Omega) \cap L_0^2(\Omega)] \times \mathbf{H}^2(\Omega)$$

satisfying

$$\|\mathbf{v}\|_3 + \|\boldsymbol{\pi}\|_2 + \|\mathbf{z}\|_2 \leq \gamma_0,$$

then for all $h \leq h_0$, problem (5.3) admits a unique solution $(\mathbf{v}_h, \boldsymbol{\pi}_h, \mathbf{z}_h) \in X_h \times Q_h \times T_h$ in a certain neighborhood of $(\mathbf{v}, \boldsymbol{\pi}, \mathbf{z})$ and the following error bound holds

$$\|\mathbf{v} - \mathbf{v}_h\|_1 + \|\boldsymbol{\pi} - \boldsymbol{\pi}_h\|_0 + \|\mathbf{z} - \mathbf{z}_h\|_0 \leq c h^{3/2}. \quad (5.4)$$

Moreover, for an initial approximation $(\mathbf{v}_h^0, \mathbf{z}_h^0)$ close to (\mathbf{v}, \mathbf{z}) , the discrete solution $(\mathbf{v}_h, \boldsymbol{\pi}_h, \mathbf{z}_h)$ can be obtained as the limit of the decoupled fixed point iteration scheme.

Given $(\mathbf{v}_h^n, \boldsymbol{\pi}_h^n, \mathbf{z}_h^n)$ find $(\mathbf{v}_h^{n+1}, \boldsymbol{\pi}_h^{n+1}, \mathbf{z}_h^{n+1}) \in X_h \times Q_h \times T_h$ such that

$$\begin{cases} 2\left(\mathbf{D}(\mathbf{v}_h^{m+1}), \mathbf{D}(\mathbf{u})\right) - \left(\boldsymbol{\pi}_h^{n+1}, \nabla \cdot \mathbf{u}\right) = (\mathbf{z}_h^n, \mathbf{D}(\mathbf{u})), \quad \forall \mathbf{u} \in X_h \\ (\nabla \cdot \mathbf{v}_h^{n+1}, q) = 0, \quad \forall q \in Q_h \\ v(z_h^{n+1}, \boldsymbol{\tau}) + \alpha_1 G_h(\mathbf{v}_h^n, z_h^{n+1}, \boldsymbol{\tau}) - \alpha_1 \left(z_h^n \cdot (\nabla \mathbf{v}_h^n)^T, \boldsymbol{\tau}\right) = \\ = (N(\mathbf{v}_h^n, \boldsymbol{\pi}_h^{n+1}), \boldsymbol{\tau}) + (\mathbf{F}, \boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in T_h \end{cases} \quad (5.5)$$

Corollary 5.1 Under the conditions of the preceding theorem and since

$$p_h = v \boldsymbol{\pi}_h - \alpha_1 \mathbf{v}_h \cdot \nabla \boldsymbol{\pi}_h, \quad (5.6)$$

we get the error estimate

$$|\mathbf{v} - \mathbf{v}_h|_1 + \|p - p_h\|_0 \leq ch, \quad (5.7)$$

where $(\mathbf{v}, p) \in [H^3(\Omega) \cap H_0^1(\Omega)] \times [H^1(\Omega) \cap L_0^2(\Omega)]$ is the solution to problem (5.1).

The above decoupled iterative method consists of solving alternatively a Stokes system and a transport equation. For the Stokes problem in $(\mathbf{v}, \boldsymbol{\pi})$ the Hood-Taylor FEM gives the error estimate

$$|\mathbf{v} - \mathbf{v}_h|_1 + \|\boldsymbol{\pi} - \boldsymbol{\pi}_h\|_0 \leq ch^2$$

and for a pure convection equation the discontinuous Galerkin FEM gives

$$\|z - z_h\|_0 \leq ch^{3/2}.$$

Collecting these error estimates we derive the expected error bound (5.4) given in the main Theorem 5.2. Moreover, taking into account relation (5.6) for the pressure $\boldsymbol{\pi}$ of the original problem (5.1), we get the error estimate (5.7).

Remark 5.1 Here, we have been interested in the development of numerical methods for the approximation of a steady solution of problem (5.2). However, in many flow situations time dependent problems must be solved, even in the simulation of steady flows, when transient algorithms are used to obtain steady solutions. In fact, this is one way to initialize the iterative method (5.5) with (\mathbf{v}_h^0, z_h^0) sufficiently close to the exact solution (\mathbf{v}, z) .

Remark 5.2 Similar results can be obtained using other common stabilization methods: \mathbb{P}_2 plus bubble for the velocity \mathbf{v} , discontinuous \mathbb{P}_1 for the pressure $\boldsymbol{\pi}$ and discontinuous \mathbb{P}_2 for z ; $\mathbb{P}_2 - \mathbb{P}_1$ Hood-Taylor for $(\mathbf{v}, \boldsymbol{\pi})$ and continuous approximation of z (streamline upwinding Petrov-Galerkin method – SUPG). Moreover it is possible to extend the previous results to finite elements of higher order and to use special quadrangular meshes (see [10, 11]).

The results of the numerical implementation of the method described here will be presented in a forthcoming work.

Acknowledgement: The authors are greatful to the financial support from European Union FEDER/PRAXIS (Project Nr. 2 (2.1/MAT/380/94).

REFERENCES

- [1] Babuska, I., The finite element method with Lagrangian multipliers, *Numer. Math.*, 20, 1973, 179-192.
- [2] Brezzi, F., On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers, *RAIRP, Anal. Numer.*, R2, 1974, 129-151.
- [3] Cioranescu, D. and Ouazar, E.H., Existence and uniqueness for fluids of second grade, *Coll. France Sem.*, Pitman Res. Notes Math., 109, 1984, 178-197.
- [4] Cioranescu, D., Girault, V., Glowinski, R., and Scott, L.R., Some theoretical and numerical aspects of grade-two fluid models, *Partial Differential Equations*, CRC Research Notes Math., 406, 1999, Chapman & Hall, 99-110.
- [5] Criminale, W.O., Erickson, J.L., and Filbey, G.I., Steady shear flow of non-Newtonian fluids, *Arch. Rational Mech. Anal.*, 1, 1957, 410-417.
- [6] Crochet, M.J., Davies, A.R., and Walters, K., *Numerical Simulation of Non-Newtonian Flow*, Elsevier, New York, 1984.
- [7] Dunn, J.E. and Fosdick, R.L., Thermodynamics, stability and boundedness of fluids of complexity 2 and fluids of second grade, *Arch. Rational Mech. Anal.*, 56, 1974, 191-252.
- [8] Finn, R., Estimates at infinity for stationary solutions of the Navier-Stokes equations, *Bull. Math. Soc. Sci. Math. Phys. R.P. Roumaine (N.S.)*, 3, 1959, 387-418.
- [9] Finn, R., On the exterior stationary problem for the Navier-Stokes equations, and associated perturbation problems, *Arch. Rational Mech. Anal.*, 19, 1965, 363-406.
- [10] Fortin, M. and Essellaoui, D., A finite element procedure for viscoelastic flows, *Int. J. Numer. Meth. Fluid*, 7, 1987, 1035-1052.
- [11] Fortin, M. and Fortin, A., A new approach for the FEM simulation of viscoelastic flows, *J. Non-Newtonian Fluid Mech.*, 32, 1989, 295-310.
- [12] Fosdick, R.L. and Rajagopal, K.R., Uniqueness and drag for fluids of second-grade in steady motion, *Int. J. Non-Linear Mechanics*, 13, 1978, 131-137.
- [13] Fosdick, R.L. and Rajagopal, K.R., Thermodynamics and stability of fluids of third grade, *Proc. Roy. Soc. London*, A339, 1980, 351-377.

- [14] Fosdick, R.L. and Serrin, J., Rectilinear steady flow of simple fluids, Proc. Roy. Soc. London, A332, 1973, 311-333.
- [15] Galdi, G.P., An introduction to the mathematical theory of the Navier-Stokes equations, *Springer Tracts in Natural Philosophy*, 38, 39, Springer, 1994.
- [16] Galdi, G.P., Grobelaar-Van Dalsen, M., and Sauer, N., Existence and uniqueness of classical solutions of the equations of motion for second-grade fluids, Arch. Rational Mech. Anal., 124, 1993, 221-237.
- [17] Galdi, G.P., Sequeira, A., and Videman, J.H., Steady motions of a second-grade fluid in an exterior domain, Adv. Math. Sci. Appl., 7, 1997, 977-995.
- [18] Girault, V. and Raviart, P.A., *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, Berlin, 1986.
- [19] Guillopé, C. and Saut, J.-C., Global existence and one-dimensional nonlinear stability of shearing motions of viscoelastic fluids of Oldroyd type, M²AN, 24, 1990, 369-401.
- [20] Guillopé, C. and Saut, J.-C., Existence results for the flow of viscoelastic fluids with a differential constitutive law, Nonlinear Analysis, Theory, Methods & Applications, 15, 1990, 849-869.
- [21] Huigol, R.R., *Continuum Mechanics of Viscoelastic Liquids*, Hindustan Publishing Co., Delhi, 1975.
- [22] Ladyzhenskaya, O.A., *The Mathematical Theory of Viscous Incompressible Flow*, Gordon and Breach, New York, 1969.
- [23] Ladyzhenskaya, O.A. and Solonnikov, V.A., Determination of the solutions of boundary value problems for stationary Stokes and Navier-Stokes equaitons having an unbounded Dirichlet integral, Zapiski Nauchn. Sem. LOMI, 96, 1980, 117-160. English Transl.: J. Sov. Math., 21, 1983, 728-761.
- [24] Lesaint, P. and Raviart, P.A., On a finite element method for solving the neutron transport equation, *Mathematical Aspects of Finite Elements in Partial Differential Equations*, C. Boor (ed.), Academic Press, 1974, 89-122.
- [25] Marchal, J.M. and Crochet, M.J., A new finite element for calculating viscoelastic flow, J. Non-Newtonian Fluid Mech., 26, 1987, 77-114.
- [26] Najib, K. and Sandri, D., On a decoupled algorithm for solving a finite element problem for the approximation of viscoelastic fluid flow, Numer. Math., 72, 1995, 223-238.
- [27] Nazarov, S.A., Weighted spaces with detached asymptotics in applications to the Navier-Stokes equations, Proceedings of the Sixth Winter School, Paseky, Czech Republic, 1999, Pitman Res. Notes Math., to appear.

- [28] Nazarov, S.A. and Pileckas, K., On steady Stokes and Navier-Stokes problems with zero velocity at infinity in a three-dimensional exterior domain, *J. Math. Kyoto Univ.*, to appear.
- [29] Nazarov, S.A. and Plamenevskii, B.A., *Elliptic Boundary Value Problems in Domains with Piecewise Smooth Boundaries*, Walter de Gruyter, 1994.
- [30] Nazarov, S.A., Sequeira, A., and Videman, J.H., Asymptotic behavior at infinity of three-dimensional steady viscoelastic flows, submitted.
- [31] Novotny, A., Sequeira, A., and Videman, J.H., Existence of three-dimensional flows of second-grade fluids past an obstacle, *Nonlinear Analysis, Theory, Methods & Applications*, 30, 1997, 3051-3058.
- [32] Novotny, A., Sequeira, A., and Videman, J.H., Steady motions of viscoelastic fluids in 3-D exterior domains – Existence, uniqueness and asymptotic behavior, *Arch. Rational Mech. Analysis*, 149, 1999, 49-67.
- [33] Oldroyd, J.G., On the formulation of rheological equations of state, *Proc. Roy. Soc. London*, A200, 1950, 523-541.
- [34] Oldroyd, J.G., Non-Newtonian effects in steady motion of some idealized elasto-viscous liquids, *Proc. Roy. Soc. London*, A245, 1958, 278-297.
- [35] Pileckas, K., Recent advances in the theory of Stokes and Navier-Stokes equations in domains with non-compact boundaries, *Mathematical Theory in Fluid Mechanics*, Pitman Res. Notes Math., 354, 1996, 30-85.
- [36] Pileckas, K., Sequeira, A., and Videman, J.H., A note on steady flows of non-Newtonian fluids in channels and pipes, Magalhaes, L., Sanchez, L., and Rocha, C. (eds.), *EQUADIFF-95* World Scientific, 1998, 458-467.
- [37] Pileckas, K., Sequeira, A., and Videman, J.H., Steady flows of viscoelastic fluids in domains with outlets to infinity, submitted.
- [38] Pokorny, M., *Asymptotic Behavior of Solutions to Certain Partial Differential Equations Describing the Flow of Fluids in Unbounded Domains*, Ph.D. Thesis, University of Toulon and Charles University of Prague, 1999.
- [39] Quarteroni, A. and Valli, A., *Numerical Approximation of Partial Differential Equations*, Springer-Verlag, Heidelberg, 1994.
- [40] Rajagopal, K.R., Mechanics of non-Newtonian fluids, Galdi, G.P. and Nečas, J. (eds.), *Recent Developments in Theoretical Fluid Mechanics*, Pitman Res. Notes Math., 291, 1993, 129-162.
- [41] Renardy, M., Recent advances in the mathematical theory of steady flow of viscoelastic fluids, *J. Non-Newtonian Fluid Mech.*, 29, 1988, 11-24.
- [42] Rivlin, R.S., Solution of some problems in the exact theory of visco-elasticity, *J. Rational Mech. Anal.*, 5, 1956, 179-188.

- [43] Rivlin, R.S. and Ericksen, J.L., Stress-deformation relations for isotropic materials, *J. Rational Mech. Anal.*, 4, 1955, 323-425.
- [44] Schowalter, W.R., *Mechanics of Non-Newtonian Fluids*, Pergamon Press, 1978.
- [45] Sequeira, A. and Baía, M., A finite element approximation for the steady solution of a second-grade fluid model, *J. Comp. Appl. Math.*, 111, 1999, 281-295.
- [46] Solonnikov, V.A. and Pileckas, K., Certain spaces of solenoidal vectors and the boundary value problem for Navier-Stokes system of equations in domains with non-compact boundaries, *Zapiski Nauchn. Sem. LOMI*, 73, 1977, 136-151. English Transl.: *J. Sov. Math.*, 34, No. 5, 1986, 2101-2111.
- [47] Solonnikov, V.A., Stokes and Navier-Stokes equations in domains with non-compact boundaries, *College de France Seminars*, 4, 1983, 240-349.
- [48] Truesdell, C. and Noll, W., *The Nonlinear Field Theories of Mechanics*, 2nd edition, Springer, Berlin, 1992.
- [49] Videman, J.H., *Mathematical Analysis of Viscoelastic Non-Newtonian Fluids*, Ph.D. Thesis, Instituto Superior Técnico, Lisbon, 1997.

24 FULL CONVERSION IN GAS-SOLID REACTIONS

Ivar Stakgold

Department of Mathematical Sciences
University of Delaware
Newark, Delaware 19716

ABSTRACT

As a gas diffuses through a porous solid, a reaction takes place between the gas and a species of the solid. The mathematical formulation consists of a coupled PDE and ODE. The reaction rate is taken as proportional $C^p S^m$, where C and S are the respective concentrations of the gas and reacting solid species. The case of constant porosity has been studied by Diaz and Stakgold [2], but here some of the results are extended to the case where the porosity increases as the solid is consumed. When $m < 1$ the solid is fully consumed in a finite time T . The principal goal of this paper is to obtain estimates for T .

1. INTRODUCTION

A gas diffusing through a porous solid reacts with some species in the solid matrix, the porosity increasing as the solid is being consumed. The reaction proceeds isothermally and irreversibly with a rate proportional to $C^p S^m$, where C is a nondimensional gas concentration and S is a nondimensional concentration of the reacting solid species. The somewhat simpler case $p = 1$ was treated in [6]. Porosity changes have also been taken into account in the study of traveling waves for geophysical problems (see [1], for instance). The constant porosity case for gas-solid reactions has been investigated in the comprehensive paper [2].

Existence and uniqueness for the boundary value problem with porosity change can be handled by combining quasimonotone methods (see [3], [4]) with a variant of the iteration scheme in [2]. As was shown in [2] and [6], an interesting feature of

the nonlipschitz case $m < 1$ (which occurs in many practical applications) is that the solid is fully consumed in a finite time T , known as the *time to full conversion*. Our principal goal in this article is to provide estimates for T .

2. FORMULATION AND PRELIMINARY RESULTS

We consider a reaction-diffusion problem in an initially homogeneous porous solid occupying a bounded domain Ω in R^n , when the pores are initially gas-free and the boundary gas concentration is maintained at a constant positive value for $t > 0$. These rather simple side conditions retain the essential character of the problem while simplifying the analysis.

Mass balances for the immobile solid and the diffusing gas yield the system

$$S_t = -S^m C^p, \quad x \in \Omega, \quad t > 0, \quad (2.1)$$

$$(\varepsilon C)_t - \Delta C = -\lambda S^m C^p (= \lambda S_t), \quad x \in \Omega, \quad t > 0, \quad (2.2)$$

where λ is a positive constant and the porosity ε is related to S through

$$\varepsilon = \varepsilon_0 + \varepsilon_1(1 - S) \quad (2.3)$$

with $\varepsilon_0 > \varepsilon_1 \geq 0$ being given constants. The nondimensional initial and boundary conditions are

$$S(x, 0) = 1, \quad C(x, 0) = 0, \quad x \in \Omega \quad (2.4)$$

$$C(x, t) = 1, \quad x \in \partial\Omega, \quad t > 0. \quad (2.5)$$

We seek solutions of (2.1)-(2.5) with $C \geq 0$, $S \geq 0$. It is clear from (2.1) that S is monotonically decreasing in time. By noting that

$$(\varepsilon C)_t = \varepsilon C_t - \varepsilon_1 S_t C = \varepsilon C_t + \varepsilon_1 S^m C^{p+1},$$

we can rewrite (2.2) as

$$\varepsilon C_t - \Delta C = -\lambda S^m C^p - \varepsilon_1 S^m C^{p+1} \quad (2.6)$$

which leads to the inequality

$$C_t - \frac{1}{\varepsilon} \Delta C \leq 0,$$

where $0 < \varepsilon_0 \leq \varepsilon \leq \varepsilon_0 + \varepsilon_1$. Thus the maximum principle for parabolic problems applies (see [5]) so that $C(x, t) \leq 1$ for $x \in \overline{\Omega}$ and $t \geq 0$.

Next we would like to show that, as $t \rightarrow \infty$, the solution of (2.1)-(2.5) tends to the steady state $S = 0$, $C = 1$. In anticipation of this result we find it convenient to introduce (as in [7]) the time integral of the difference between the gas concentration and its projected steady state:

$$\eta(x, t) = \int_0^t [1 - C(x, \tau)] d\tau, \quad \eta_t = 1 - C. \quad (2.7)$$

We note that η and η_t are both nonnegative.

Integration of (2.2) from time 0 to time t yields

$$[\varepsilon_0 + \varepsilon_1(1 - S)] \eta_t - \Delta \eta = \varepsilon_0 + (\varepsilon_1 + \lambda)(1 - S), \quad \eta(x, 0) = \eta(\partial\Omega, t) = 0. \quad (2.8)$$

Hence η satisfies the inequality

$$L\eta = \varepsilon_0 \eta_t - \Delta \eta - \varepsilon_0 - \varepsilon_1 - \lambda \leq 0,$$

so that η is a subsolution to the problem $Lu = 0$ with vanishing initial and boundary conditions. From the classical theory of linear parabolic equations with constant coefficients we know that u increases monotonically in time to the steady state $(\varepsilon_0 + \varepsilon_1 + \lambda) w(x)$ where $w(x)$ is the solution of the Poisson problem

$$-\Delta w = 1, \quad x \in \Omega, \quad w(\partial\Omega) = 0. \quad (2.9)$$

It follows that

$$\eta(x, t) \leq u(x, t) \leq (\varepsilon_0 + \varepsilon_1 + \lambda) w(x)$$

so that η is uniformly bounded on $\overline{\Omega} \times (0, \infty)$. The definition (2.7) then shows that

$$\lim_{t \rightarrow \infty} C(x, t) = 1, \quad \text{uniformly on } \overline{\Omega}. \quad (2.10)$$

We conclude from (2.1) that

$$S(x, t) = B_m \left(\int_0^t C^p(x, \tau) d\tau \right), \quad (2.11)$$

where $B_m(t)$ is the solution of

$$S_t = -S^m, \quad S(0) = 1, \quad (2.12)$$

which describes the evolution of the solid concentration on $\partial\Omega$. The function $B_m(t)$ is given explicitly by

$$B_m(t) = \begin{cases} [1 - (1-m)t]_+^{1/(1-m)} & m \neq 1 \\ e^{-t} & m = 1 \end{cases} \quad (2.13)$$

where z_+ is defined as the greater of 0 and z . For $m \geq 1$, $B_m(t)$ is positive for all t and decreases to 0 as $t \rightarrow \infty$; it follows from (2.11) that $S(x, t) \rightarrow 0$ as $t \rightarrow \infty$. Thus, the solution of (2.1)-(2.5) approaches the steady state $S = 0$, $C = 1$ as $t \rightarrow \infty$.

Returning to (2.8), we now see that

$$\lim_{t \rightarrow \infty} \eta(x, t) = \eta_\infty(x)$$

where

$$\eta_\infty(x) = (\varepsilon_0 + \varepsilon_1 + \lambda) w(x). \quad (2.14)$$

3. BOUNDS FOR THE TIME T TO FULL CONVERSION

Since $C \rightarrow 1$ as $t \rightarrow \infty$, (2.11) and (2.13) show that the solid is fully consumed in finite time if and only if $m < 1$. This time T is known as the *time to full conversion* and is characterized by

$$\frac{1}{1-m} = \min_{x \in \Omega} \int_0^T C^p(x, \tau) d\tau. \quad (3.1)$$

In the present section we confine ourselves to the case $m < 1$ and obtain various estimates for T . We shall use the following elementary inequalities which hold in the interval $0 \leq C \leq 1$:

$$\text{For } p \leq 1, \quad C \leq C^p \leq 1 - p(1-C) \quad (3.2a)$$

$$\text{For } p \geq 1, \quad 1 - p(1-C) \leq C^p \leq C. \quad (3.2b)$$

- a) Let us first consider the case $p \leq 1$ when

$$\int_0^T C^p(x, \tau) d\tau \geq \int_0^T C(x, \tau) d\tau = T - \eta(x, T) \geq T - (\varepsilon_0 + \varepsilon_1 + \lambda) \|w\|$$

where $\| \cdot \|$ stands for the sup norm. In view of (3.1), we have

$$T \leq \frac{1}{1-m} + (\varepsilon_0 + \varepsilon_1 + \lambda) \|w\|. \quad (3.3)$$

Good estimates for $\|w\|$ are available (see [1]). For simple domains like balls w is even known explicitly. To obtain a lower bound we use the other inequality in (3.2a):

$$\int_0^T C^p(x, \tau) d\tau \leq \int_0^T [1 - p(1-C)] d\tau = T - p\eta(x, T).$$

It follows that

$$T \geq p\eta(x, T) + \int_0^T C^p d\tau \geq p\eta(x, T) + \frac{1}{1-m},$$

which, being valid for every x , yields

$$T \geq p \|\eta(x, T)\| + \frac{1}{1-m}. \quad (3.4)$$

Another lower bound stems from the observation that $t - \eta(x, t)$ is nonnegative and monotone increasing in time:

$$T \geq \|\eta(x, T)\|. \quad (3.5)$$

We can use (2.8) to estimate $\eta(x, T)$. At time T and beyond, $S \equiv 0$; since $0 \leq \eta_t \leq 1$, (2.8) yields $-\Delta \eta(x, T) \geq \lambda$ and hence

$$\|\eta(x, T)\| \geq \lambda \|w\|. \quad (3.6)$$

- b) For $p \geq 1$, we use (3.2b) to obtain, after calculations similar to those of (a),

$$\frac{1}{1-m} + \|\eta(x, T)\| \leq T \leq \frac{1}{1-m} + p(\varepsilon_0 + \varepsilon_1 + \lambda) \|w\|. \quad (3.7)$$

Using (3.6) in (3.4) and recalling (3.3) we have the estimates

$$\frac{1}{1-m} + p\lambda \|w\| \leq T \leq \frac{1}{1-m} + (\varepsilon_0 + \varepsilon_1 + \lambda) \|w\| \quad (p \leq 1)$$

whereas (3.7) yields

$$\frac{1}{1-m} + \lambda \|w\| \leq T \leq \frac{1}{1-m} + p(\varepsilon_0 + \varepsilon_1 + \lambda) \|w\| \quad (p \geq 1)$$

Improved estimates can be found using Jensen's inequality as in [2].

REFERENCES

- [1] Chadam, J., Chen, X., Comparini, E. and Ricci, R., Travelling wave solutions of a reaction-infiltration problem and a related free boundary problem, *Euro. J. Appl. Math.*, 5 (1994), 255-265.
- [2] Diaz, J. I. and Stakgold, I., Mathematical aspects of the combustion of a solid by a distributed isothermal gas reaction, *SIAM J. Math. Anal.*, 26 (1995), 305-328.
- [3] Ladde, G. S., Lakshmikantham, V. and Vatsala, A. S., *Monotone Iterative Techniques for Nonlinear Differential Equations*, Pitman, London, 1985.
- [4] Pao, C. V., *Nonlinear Parabolic and Elliptic Equations*, Plenum, 1992.
- [5] Protter, M. H. and Weinberger, H., *Maximum Principles in Differential Equations*, Prentice-Hall, 1967.
- [6] Stakgold, I., Gas-solid reaction with porosity change, *J. Diff. Eq.*, to appear.
- [7] Stakgold, I. and McNabb, A., Conversion estimates for gas-solid reactions, *Math. Modelling*, 5 (1984), 325-330.

25 NEW ANALYSIS PROCEDURE IN PREDICTING ROTOR VIBRATION

Rajagopal Subbiah
Siemens Westinghouse Power Corporation
A Siemens Company
Orlando, FL 32826-2399

ABSTRACT

Vibration is one of the primary concerns in rotating machinery. Vibration reduction will help extend turbo-machinery life and is viewed as one of the major winning points for turbine-generator (T-G) manufacturers. To better predict the level of machine vibration requires a great deal of analysis. The success of different analyses mainly rely on reliable models that are accurate and consistently verifiable by testing. In this paper, another new modeling approach that uses strain energy is reported (Subbiah, 1999). This modeling technique was successfully applied for rotor torsional and lateral analyses with greater accuracy. Further, a faster and more stable solution method which uses Riccati algorithm is discussed.

1. INTRODUCTION

Rotor vibration is a measure of responses due to external disturbances, primarily due to mass imbalances and unbalance torques. In general, rotor vibration can be lowered by balancing rotor shafts in the factory and/or on site. Prior knowledge of the balancing behaviors of rotors will be helpful when similar or identical rotor configurations are balanced at site or at the factory. However, when shaft configuration changes due to the evolution of a new product design, only a reliable model will be able to provide some idea of the degree to which the rotor can be balanced. Besides, an accurate model will help predicting the dynamics very early in the design cycle and will buy more time to fine tune or modify the design, if required. Further, accurate rotor models set the stage for successful performance of different important analyses such as bearing and coupling alignments, rotor rubs, oil whirl instability, seal dynamics, and rotor stresses etc.

Numerous papers were published, over the years, to report the step improvements made in rotor modeling. It is voluminous to report all of them here. Therefore, the author refers to Nelson's (1994) comprehensive historical survey on this subject. In this paper, a methodology for accurate modeling of rotor discs using strain energy is presented. This approach has been extensively verified by testing.

Also reported in this paper is another analysis approach that combines the finite element (FE) beam models (derived by strain energy) with the "Riccati" recursive solution algorithm by Horner(1978). The combined analysis method is called, "Ric-FEM" was developed in the frequency domain using coupled lateral and axial degrees of freedom. Since this approach utilizes small matrix arrays compared to the conventional FE and transfer matrix (TM) models, it works faster. Since the algorithm is well damped, the solution converges faster than any conventional method (even for long and flexible rotors). Ratan and Rodriguez (1992) developed similar procedure in time-domain using four lateral degrees of freedom. In this paper, modeling and analysis of the rotor are discussed.

2. MODELING OF ROTOR-DISCS

Strain energy modeling approach for rotor-discs is presented. Usually, the rotor disc sections are modeled for stiffness using empirical angle rule (α) as shown in Figure 1. This rule can be applied only to verified rotor geometry and it can not be used as an universal rule to model any rotor configuration that was not previously verified. Therefore, when the rotor configuration changes, it may be necessary to update "the angle" to accurately determine the rotor stiffness models.

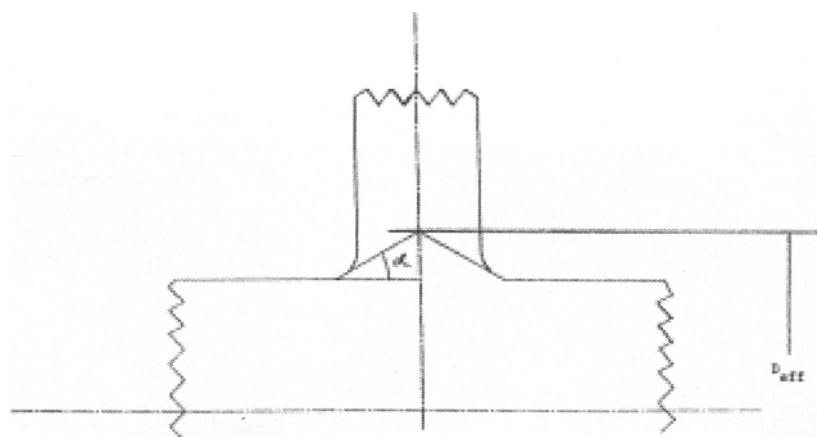


Figure 1. Angle Rule Applied for the Determination of Rotor Effective Stiffness.

Since the strain energy distribution automatically assumes the intricate geometrical shapes of the rotor, the effective rotor stiffnesses calculated using this approach is reliable, accurate and consistent regardless of rotor configuration. This has been verified by many tests. Therefore, strain energy approach in considered to provide a powerful scientific tool for modeling irregular rotor sections.

3. EVALUATING STIFFNESS DIAMETERS

Consider a circular shaft of length L with an applied torque T at one end with the other end fixed as shown in Figure 2. The assumption here is the torque (T) - twist angle (ϕ) for the shaft under the applied load is linear, indicating that the rotation and the angular twist behaves linear with the applied torque. This linearity assumption is valid for the material that follows Hooke's law and if the strain in the structure is small. The corresponding elastic strain energy of the shaft is

$$U = T\phi/2 \quad (3.1)$$

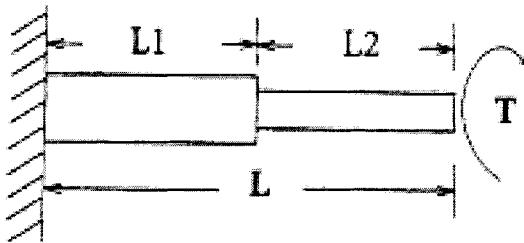


Figure 2. Circular Beam With Boundary Conditions.

Combining equation (1) with the equation $\phi = TL/GJ$ gives expressions for strain energy as shown:

$$U = T^2 L / 2GJ \quad (3.2)$$

where G is the shear modulus and J is polar moment of inertia of the cross-section of the shaft. J for circular solid section is $\pi D_{et}^4 / 32$, where D_{et} is the effective stiffness diameter of an equivalent beam section in torsion. Equation (3.2) provides the relationship between the strain energy and the effective stiffness diameters for a uniform shaft section.

If the shaft has axi-symmetric circular cross sections with non-uniform planar shapes, or if the torque changes along the axis of the shaft, then we need more general relationship for strain energy in torsion. To accomplish this result, consider an elemental disk of length dx at a distance x from one end of the shaft. Assuming that the torque acting on this element is $T(x)$ and the polar moment of inertia of its cross-section is $J(x)$, the strain energy of the element according to equation (3.2) is

$$dU = [T(x)]^2 dx / 2GJ(x) \quad (3.3)$$

When $T(x)$ is constant and does not vary along the length, equation (3.3) can be written as:

$$U = \int_0^{L_1} [T^2 dx / 2GJ(x)] + \int_{L_1}^{L_2} [T^2 dx / 2GJ(x)] + \dots \quad (3.4)$$

where $L = L_1 + L_2 + \dots$.

In equation (3.4), the strain energy for each section can be calculated and the appropriate section effective stiffness diameters can be obtained as shown below:

$$D_{et} = \sqrt[4]{\frac{16T^2 L}{\pi G U}} \quad (3.5)$$

where D_{et} is the effective diameter for the stiffness of a rotor section in torsion.

Another approach using Boundary Element Method complements this procedure (Chen, 1998).

Similar technique can be applied to obtain the rotor stiffness diameters in bending for lateral analysis (Gieck, 1986). For bending motion, equation (3.4) becomes:

$$U = \int_0^{L_1} [M^2 dx / 2EI(x)] + \int_{L_1}^{L_2} [M^2 dx / 2EI(x)] + \dots \quad (3.6)$$

where $L = L_1 + L_2 + \dots$ and the bending moment M is constant.

The final D_{eb} for bending stiffness is

$$D_{eb} = \sqrt[4]{\frac{32M^2 L}{\pi E U}} \quad (3.7)$$

where M is the bending moment applied at the ends and I is the transverse moment of inertia of the rotor disc. Equations (3.5) and (3.7) can be used to calculate effective diameters for the stiffness of the shaft section respectively for the torsional and the lateral models of a rotor.

The author used the strain energy approach to calculate stiffness diameters for a variety of rotors that were used in steam and gas turbine applications. The calculation process begins by developing a finite element model of a rotor configuration as shown in Figure 3a. This figure shows a portion of a symmetric double-flow low pressure rotor as applied to a steam turbine. Torque was applied at one end of the rotor while the other end was fixed. The resulting strain energy was evaluated at all sections using ANSYS general purpose finite element code and the

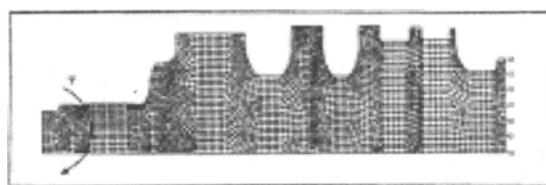


Figure 3a. Finite Element Mesh for a Portion of an LP Rotor.

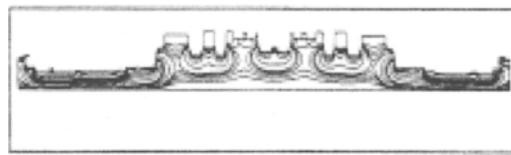


Figure 3b. Strain Energy Plots for the Complete LP Rotor.

effective stiffness diameters were computed at the respective sections. The strain energy distribution for a typical double flow LP rotor is shown in Figure 3b. It was observed from these plots that for the uniform diameter sections of the rotor, the strain energy reached its maximum value at the outer diameter of the section. Whereas, for the varying diameter sections of the rotor, (such as the blade disc geometry) the maximum strain energy occurred somewhere in between the adjacent shaft diameter and the outer diameter of the disc.

The rotor frequencies, obtained by strain energy modeling approach, were compared with those measured by testing. The difference between calculations and tests was found to stay within 2%.

Similar technique can be used to obtain stiffness diameters for bending using equation (3.7).

Analysis

Analysis is another important element of design which could reduce calculation effort and help bring down the total design cycle cost and time. Response analysis by the conventional Transfer Matrix Method (TMM) requires a dynamic matrix size as large as 17×17 (Lund, 1965). The improved transfer matrix method (Subbiah and Rieger, 1988) reduces the matrix size to 9×9 . Whereas, FEM requires variable matrix size which increases proportionately with the size of the problem. Therefore, another approach, which combined the strengths of FEM and TMM, was developed and reported (Subbiah et. al, 1988). These algorithms are very popular and are extensively used in developing machinery monitoring systems (Song et. al, 1998). Further improvements in computational effort can be achieved by combining the consistent shaft properties of the FEM with the recursive solution using Riccati method. The mathematical formulation of this method is referred to as Ric-FEM in this paper. The solution is obtained in the frequency domain. The details are given below:

Ric-FEM

Riccati-finite element approach has been utilized in the modeling and analysis of rotor systems. The rotor continuum is modeled using beam finite elements (FEM) which use the consistent properties of the rotor to formulate the mass and the stiffness matrices (Nelson, 1976). Riccati method is then adopted by solving the initial value problem. Figure 5 shows a typical rotor shaft supported at two

bearings. This shaft has been discretized into finite number of elements. For the present work, each element is connected by two nodes with each node representing four dynamic degrees of freedom (DOF) such as two translations along x and y and the two rotations θ_x and θ_y about x and y axes as shown in the Figure 4. Therefore, each element has eight DOF in all. Concentrated forces such as disk inertia, unbalance forces and bearing dynamic forces are lumped at the appropriate node locations. The very first element of the rotor continuum model is a fictitious zero parameter element. This element 0 has two nodes 0 and 1.

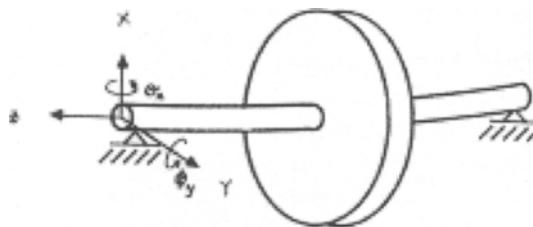


Figure 4. Rotor Modeling Convention.

The dynamical equations of motion for an i -th element can be written as follows:

$$\begin{bmatrix} m_{11}^i & m_{12}^i \\ m_{21}^i & m_{22}^i \end{bmatrix} \begin{Bmatrix} \ddot{x}_i \\ \ddot{x}_{i+1} \end{Bmatrix} + \begin{bmatrix} c_{11}^i & c_{12}^i \\ c_{21}^i & c_{22}^i \end{bmatrix} \begin{Bmatrix} \dot{x}_i \\ \dot{x}_{i+1} \end{Bmatrix} + \begin{bmatrix} k_{11}^i & k_{12}^i \\ k_{21}^i & k_{22}^i \end{bmatrix} \begin{Bmatrix} x_i \\ x_{i+1} \end{Bmatrix} = \begin{Bmatrix} f^{R_i} \\ f^{L_{i+1}} \end{Bmatrix} \quad (3.8)$$

where $m_{11}^i, m_{12}^i, \dots$ etc., represent elements of the i -th mass matrix of size (4×4) ; $c_{11}^i, c_{12}^i, \dots$ etc., represent elements of the i -th damping matrix of size (4×4) ; $k_{11}^i, k_{12}^i, \dots$ etc., represent elements of the i -th stiffness matrix of size (4×4) ; $x_i = \{X_i, Y_i, \theta_x, \theta_y\}$ represents vector of amplitudes (4×1) ; \dot{x}, \ddot{x} are respectively velocity and acceleration components; $\{f_i\}$ represents force vector (4×1) . Superscripts R and L represent the right and left side of the node point.

Assume a solution of

$$\begin{aligned} Q_i &= \{x_i\} e^{j\omega t} \text{ and} \\ F_i &= \{f_i\} e^{j\omega t} \end{aligned} \quad (3.9)$$

where $j = \sqrt{-1}$, ω = angular velocity of the rotor, t = time instant.

Now, let us assume the following relationship at the I -th node

$$[S^i] Q_i = \{F^{L_i} - R_i\} \quad (3.10)$$

where $[S^i]$ is the Riccati matrix at the i -th node, $\{R_i\}$ is the Riccati vector at the i -th node, and S and R have known initial conditions.

Substituting equation (3.9) in equation (3.8) and adding to equation (3.10) yields,

$$[a_{11}^i + S^i] \cdot Q_i + a_{12}^i \cdot Q_{i+1} = F^{L_i} + F^{R_i} - R_i \quad (3.11a)$$

$$a_{21}^i \cdot Q_i + a_{22}^i \cdot Q_{i+1} = F^{L_{i+1}} \quad (3.11b)$$

where

$$a_{11}^i = -\omega^2 m_{11} + j\omega c_{11} + k_{11}.$$

Other elements can be obtained in a similar fashion.

Now, equation (11a) can be written as,

$$A_i = [a_{11}^i + S^i]^{-1} \{ (F_i - R_i) \} - a_{12}^i \cdot Q_{i+1} \quad (3.12a)$$

where $F_i = F^{L_i} + F^{R_i}$.

Use of equation (3.12a) in equation (3.11b) results,

$$a_{21}^i [[a_{11}^i + S^i]^{-1} \{ F_i - R_i \} - a_{12}^i \cdot Q_{i+1}] + a_{22}^i Q_{i+1} = F^{L_{i+1}}. \quad (3.12b)$$

Equation (3.12b) can be rewritten after reordering as,

$$[S^{i+1}] Q_{i+1} = F^{L_{i+1}} - R_{i+1} \quad (3.13)$$

where

$$[S^{i+1}] = [a_{21}^i [a_{11}^i + S^i]^{-1} \cdot (-a_{12}^i) + a_{22}^i] \quad (3.13a)$$

and

$$\{R_{i+1}\} = -a_{21}^i [a_{11}^i + S^i]^{-1} \{F_i - R_i\}. \quad (3.13b)$$

It can be noticed that equation (3.13) is of the same form as equation (3.10).

At the fictitious element 0, the nodal S^0 and R_0 are zero per initial condition. Using this in equations (3.13a) and (3.13b), the value of S^1 and R_1 can be calculated and thereupon equations (3.13a) and (3.13b) are used recursively to calculate $S^2, S^3, S^4, \dots, S^N$. When a concentrated disk, bearing or an unbalance location is encountered along the rotor model, it is added at the appropriate nodal points.

In general, the Riccati algorithm is found well damped. Consequently, numerical instability did not occur when it was used to analyze flexible rotor systems (Lund and Wang, 1985). For small to medium size rotor systems (such as 50 to 100 stations), the computational efforts for TMM and Ric-FEM were almost identical. Rotor models that consist of 300 stations and above, Ric-FEM was found about 10 times faster than TMM when calculations were performed in a Sun SPARC 20 workstation. The Ric-FEM was tested for up to 1000 stations, surprisingly, no numerical divergence was noticed. This shows that the Ric-FEM method is well damped and stays numerically stable for longer and flexible rotor trains.

Several critical speed calculations were carried out using Ric-FEM on a variety of rotors. In all cases, the predicted results were found very close to those measured (within 2-3% difference). Two examples are discussed to show how powerful the Ric-FEM is over the other methods. The first example is a Low Pressure rotor as shown in Figure 5. The numerical data for one half of the rotor is provided in Table 1. This was modeled using both TMM and Ric-FEM. The first bending critical speed was calculated at 2475 RPM and 2450 RPM respectively by Ric-FEM and TMM. The measured critical speed for the machine was observed at 2467 RPM which was

very close to that calculated using Ric-FEM. This trend has been observed for many rotor critical speed calculations. Moreover, higher modes predicted using Ric-FEM were better correlated with measurements than those by TMM.

Table 1 Rotor Dimensions (given for one half of the rotor).

<u>Transverse Moment of Inertia, in⁴</u>	<u>Weight, Lbs.</u>	<u>Length, Inch</u>	<u>Concentrated Mass, Lbs.</u>
13772.53	1411.62	11.99	
5715.12	342.80	4.52	
5715.12	521.79	6.88	
5715.12	521.79	6.88	
5715.12	101.63	1.34	
7468.58	162.13	1.87	
7853.98	88.91	1.00	
7468.58	184.67	2.13	
8669.33	1236.07	13.23	
12362.81	2061.09	8.38	
128681.30	5117.60	6.67	2753.0
128681.30	5424.45	7.07	
136022.42	2012.20	5.44	
231658.97	4897.40	6.44	1126.6
231658.97	4124.60	5.42	
244668.00	1330.37	2.68	
251575.14	5647.00	6.73	804.2
251575.14	2859.24	3.64	
251575.14	2404.50	2.40	237.6
251575.14	2418.86	2.97	
251575.14	2308.00	2.40	179.0
251575.14	2898.02	3.57	
251575.14	1675.20	1.81	137.4
251575.14	2391.40	2.94	
251575.14	1191.20	1.31	139.2
251575.14	2287.40	2.81	
251575.14	1966.09	3.06	77.62
181025.03	916.87	2.04	
181025.03	1892.92	3.75	

Bearing: Sleeve Type L/D = 0.62 Oil: Standard Turbine Oil

Support Stiffnesses: Bearing Location #3

Vertical: = 10×10^6 lbf/in

Horizontal: = 5×10^6 lbf/in.

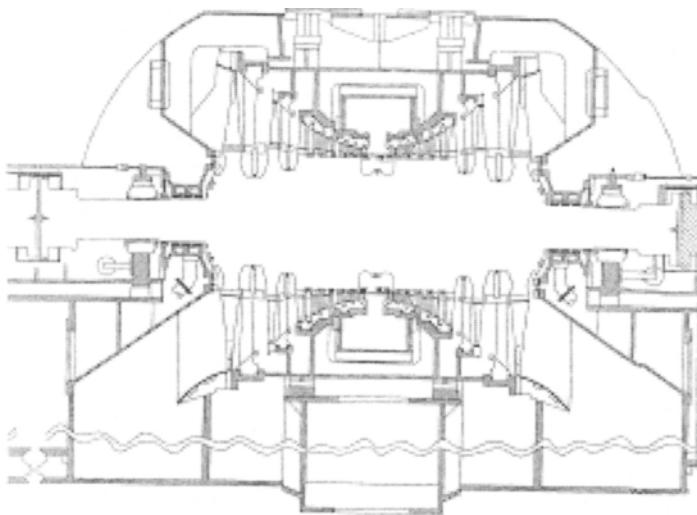


Figure 5. Low Pressure Rotor.

A second example is a machine tool spindle as shown in Figure 6 and was modeled using FEM including one additional degree of freedom in the axial direction. This formulation had, therefore, a total of five degrees of freedom per node as described by Anwar (1987). Transient dynamic analysis was performed on the machine tool rotor using the data provided by Anwar. Figure 7a shows the dynamic responses obtained using Ric-FEM and Figure 7b shows the response plots obtained by Anwar. One can observe the qualitative agreement between the two. However, the amplitudes obtained using Ric-FEM are slightly different because the unbalance magnitude and phase information used in the present work are not exactly same as in Anwar's work. In the present work, Houbolt time-marching algorithm referenced in (Subbiah, 1988) was utilized in the time-domain analysis. The problem size has been dramatically reduced in the present work than those (originally used 4225 size matrix) reported in (Anwar, 1987). The main emphasis here is the speed and stability of the Ric-FEM compared to traditional methods. Comprehensive support displacement models (Subbiah, 1985) that were used in earth-quake and ship motion studies can be efficiently built using Ric-FEM.

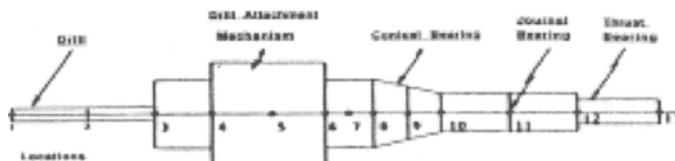


Figure 6. Machine Tool Spindle (5 DOF).

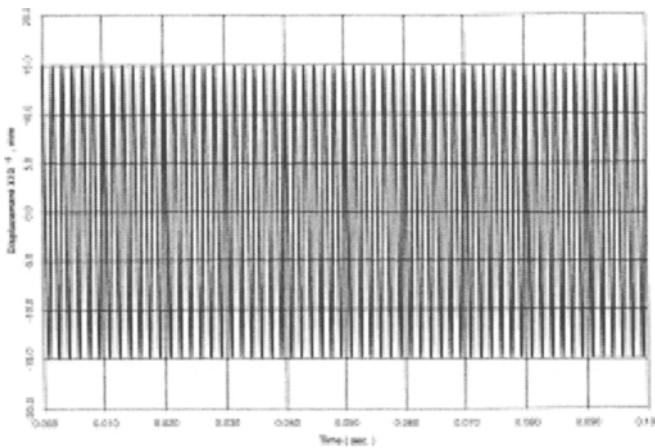


Figure 7a. Dynamic Response of Machine Tool Spindle Using Ric-FEM.

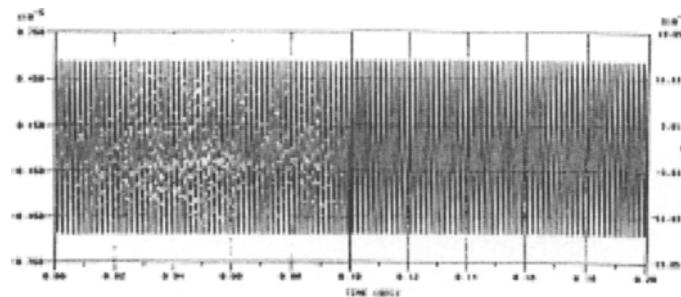


Figure 7b. Dynamic Response of Machine Tool Spindle Obtained in Ref. [18].

4. CONCLUSIONS

Reliable rotor shaft, bearing support and seal models are required for better prediction of turbo-machinery dynamics. Strain energy modeling approach is used to model irregular axisymmetric rotor discs and has been verified by factory tests. Further, a Riccati-FEM solution method, which is faster and more stable, has been used to analyze rotor-bearing systems. The following conclusions are drawn:

1. Strain energy approach provides reliable tool for modeling rotor-discs.
2. Ric-FEM provides faster and more stable solutions for rotor systems regardless of the number of stations and the degree of shaft flexibility. This is found very useful to model and analyze the dynamics and vibration of flexible systems such as machine tool spindles.

3. Comprehensive rotor system models that include support systems and seals are required for accurate dynamic predictions.

REFERENCES

- [1] ANWAR, I. and COLSHER, R., Rotor dynamic analysis of a axially coupled system" Proceedings of ASME Rotating Machinery Dynamics, DE-Vol. 2, 1987, 327-336.
- [2] CHEN, L., KASSAB, A.J., CHOPRA, M.B., and SUBBIAH R., Strain energy density based BEM for rotor dynamic analysis, Proceedings of the 12th Engineering Mechanics Conference, ASCE, La Jolla, 1998, 142-145.
- [3] CHILDS, D.W. and SCHARRER, J.K., Theory versus experiment for the rotor dynamic coefficients of labyrinth gas seals: Part II A comparison to experimental results, Journal of Vibration, Acoustics, Stress, and Reliability in Design, Transactions of ASME, 110, No. 3, 1998, 281-287.
- [4] CHILDS, D., *Turbomachinery Rotordynamics*, Wiley Interscience, (1993).
- [5] GIECK, K., *Engineering Formulas*, McGraw Hill Book Company, (1986).
- [6] GLIENICKE, J., Experimental investigation of the stiffness and damping coefficients of turbine bearings and their application to instability prediction, Proceedings of I. Mech. E., 181, Part 3B, Paper 13, 1966-67, 116-129
- [7] HORNER, G.C. and PILKEY, W.D., The Riccati transfer matrix method, Journal of Mechanical Design, Transactions of ASME, 100, 1978, 297-302.
- [8] LUND, J.W., Rotor bearing dynamics design technology, Part V: Computer program manual for rotor response and stability, Mechanical Technology Inc., Latham, N.Y., 1965, AFAPL-Tr-65-45.
- [9] LUND, J.W. and WANG, Z., Application of the Riccati method to rotor dynamic analysis of long shafts in a flexible foundation, Journal of Vibration, Acoustics, Stress and Reliability in Design, Transactions of ASME, 108, No. 2, 1985, 177-181.
- [10] MUSZYNSKA, A., Whirl and whip-rotor/bearing stability problems, Journal of Sound and Vibrations, 110, 1986, 443-462.
- [11] NELSON, H.D. and McVAUGH, J.M., The dynamics of rotor-bearing systems using finite elements, Journal of Engineering for Industry, Transactions of ASME, 98, No. 2, 1976, 593-600.
- [12] NELSON, H.D., Modeling, analysis and computation in rotor dynamics: A historical perspective, Proceedings of IFTOMM Fourth International Conference on Rotor Dynamics, 1994, 171-178.
- [13] RATAN, S. and RODRIGUEZ, J., Transient dynamic analysis of rotors using SMAC techniques: Parts 1 and 2, Journal of Vibration and Acoustics, Transactions of ASME, 114, 1992, 477-488.
- [14] SONG, C.C., WEI, Z., YU, Y., and XIANG LEI, CF., An introduction to the condition monitoring, analysis and diagnosis system of a 300 MW pumped storage power unit, Proceedings of ISROMAC-7, 22-26 A, 1998, 363-372.
- [15] SUBBIAH, R., BHAT, R.B., and SANKAR, T.S., Rotational stiffness and damping coefficients of fluid film in a finite cylindrical bearing, Transactions of ASME, # 85-Am-2E-2, 1985, 1-9.
- [16] SUBBIAH, R. and RIEGER, N.F., On the transient analysis of rRotor-bearing systems, Journal of Vibration, Acoustics, Stress and Reliability in Design, Transactions of ASME, 110, No. 4, 1988, 515-520.

- [17] SUBBIAH, R., KUMAR, A.S., and SANKAR, T.S., Transient dynamic analysis of rotors using the combined methodologies of finite elements and transfer matrix. *Journal of Applied Mechanics, Transactions of ASME*, 55, 1988, 448-452.
- [18] SUBBIAH, R., BHAT, R.B., and SANKAR, T.S., Response of rotors subjected to random support excitations, *Journal of Vibration, Acoustics, Stress and Reliability in Design, Transactions of ASME*, 108, No. 2, 1985, 177-181.
- [19] SUBBIAH, R., On further improvements in predicting rotor vibration, *ASME Vibration Conference in Las Vegas*, Paper No.DETC99/VIB-8265, 1999.

26 EQUIVALENT CONDITIONS FOR DISCONJUGACY IN SELF-ADJOINT SYSTEMS

Betty Travis

Division of Mathematics and Statistics
University of Texas at San Antonio
San Antonio, Texas 78249

and

Ramón Navarro
UPEL – Universidad Simón Bolívar
Dpto. De Matemáticas Puras y Aplicadas
Caracas – Venezuela

ABSTRACT

In this paper we consider second order self-adjoint systems of differential equations. A fairly straight-forward and self-contained proof of the equivalence of three known conditions for disconjugacy is established.

Consider the second order self-adjoint system

$$x''(t) + A(t)x(t) = 0 \quad (1)$$

where $A(t)$ is an $n \times n$ continuous matrix such that $A^T = A$, i.e. $A(t)$ is symmetric on the interval $[c, d]$.

Definition 1. *b is a conjugate point of c, $c > b$, if there exists a nontrivial solution $y(t)$ of (1) such that $y(b) = y(c) = 0$.*

Definition 2. *The system (1) is disconjugate on $[c, d]$ if no nontrivial solution of (1) vanishes more than once in $[c, d]$.*

Properties of conjugate points for self-adjoint systems have been studied extensively, see [3] and [4]. Such properties dealing with comparison and separation theorems have been studied by many mathematicians over the years, most notably by M. Morse [2] and W. Reid [3]. However, the proofs are often quite complicated and not easily accessible. In this paper, we consider three equivalent conditions for

disconjugacy, which are known, and give an elementary and self-contained proof of their equivalences. The proof is based on lecture notes by Ahmad and Lazer [1].

Definition 3. *The functional $J[y]$ is defined as*

$$J[y] = \int_c^d (\langle y', y' \rangle - \langle y, Ay \rangle) dt$$

where $y \in H = \{x : [c, d] \rightarrow R^n \mid x \text{ is continuous, } x' \text{ is piecewise continuous, and } x(c) = x(d) = 0\}$. The set H is referred to as the set of admissible functions.

Theorem 1. *The following three conditions are equivalent:*

- (A) *The Riccati equation $W' = W^2 + A(t)$ has a symmetric matrix solution.*
- (B) *$J[y] > 0$ for all $y \in H$, $y \neq 0$.*
- (C) *$x'' + A(t)x = 0$ is disconjugate on $[c, d]$.*

Proof: Assume (A), and let $v \in H$. It follows that

$$\int_c^d (2\langle v', Wv \rangle + \langle v, W'v \rangle) dt = 0$$

since

$$\langle v, Wv \rangle' = \langle v', Wv \rangle + \langle v, Wv' \rangle + \langle v, W'v \rangle = 2\langle v', Wv \rangle + \langle v, W'v \rangle$$

by symmetry of W , and

$$\langle v, Wv \rangle \Big|_c^d = 0$$

by the fundamental theorem of calculus. Therefore,

$$\begin{aligned}
J[v] &= \int_c^d (\langle v', v' \rangle - \langle v, Av \rangle + 2\langle v', Wv \rangle + \langle v, W'v \rangle) dt \\
&= \int_c^d (\langle v', v' \rangle + 2\langle v', Wv \rangle + \langle v, (W' - A)v \rangle) dt \\
&= \int_c^d (\langle v', v' \rangle + 2\langle v', Wv \rangle + \langle v, W^2v \rangle) dt \\
&= \int_c^d (\langle v', v' \rangle + 2\langle v', Wv \rangle + \langle Wv, Wv \rangle) dt \\
&= \int_c^d (\langle v' + Wv, v' + Wv \rangle) dt = \int_c^d \|v' + Wv\|^2 dt \geq 0.
\end{aligned}$$

We have shown that $J[v] \geq 0$. We note that $J[v] = 0$ if $v' = -Wv$ a.e. It is easy to see that v' is continuous, and $v(c) = 0$ since $v \in H$. Hence $v \equiv 0$. This proves that (A) implies (B).

Assume that (B) holds and the system $x'' + A(t)x = 0$ is not disconjugate on $[c, d]$. Then there exist numbers $t_0, t_1, c \leq t_0 \leq t_1 \leq d$, and a nontrivial solution $\mu(t)$ of (1) such that $\mu(t_0) = \mu(t_1) = 0$, and $\mu(t) \neq 0$ on (t_0, t_1) . Define $v(t)$ as

$$v(t) = \begin{cases} 0 & \text{if } c \leq t \leq t_0 \\ \mu(t) & \text{if } t_0 \leq t \leq t_1 \\ 0 & \text{if } t_1 \leq t \leq d. \end{cases}$$

Then $v \in H$,

$$\begin{aligned}
J[v] &= \int_c^d (\langle v', v' \rangle - \langle v, Av \rangle) dt \\
&= \int_{t_0}^{t_1} (\langle \mu', \mu' \rangle - \langle \mu, A\mu \rangle) dt
\end{aligned}$$

and $v \neq 0$.

From $\mu'' + A(t)\mu = 0$ we obtain $-\langle \mu, \mu'' \rangle - \langle \mu, A\mu \rangle = 0$. Integrating by parts, we obtain

$$\begin{aligned}
\int_{t_0}^{t_1} -\langle \mu, \mu'' \rangle dt &= -\langle \mu, \mu' \rangle \Big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \langle \mu', \mu' \rangle dt \\
&= \int_{t_0}^{t_1} \langle \mu', \mu' \rangle dt
\end{aligned}$$

since $\mu(t_1) = \mu(t_0) = 0$. Therefore, we have

$$\begin{aligned}
0 &= \int_{t_0}^{t_1} (-\langle \mu, \mu'' \rangle - \langle \mu, A\mu \rangle) dt \\
&= \int_{t_0}^{t_1} (\langle \mu', \mu' \rangle - \langle \mu, A\mu \rangle) dt \\
&= \int_c^d (\langle v', v' \rangle - \langle -v, A v \rangle) dt = J[v]
\end{aligned}$$

contracting (B). This shows that (B) implies (C).

Now, we assume that (C) holds. Let $X(t)$ be a solution of the matrix differential equation

$$X'' + A(t)X = 0$$

satisfying the initial conditions $X(c) = 0$ and $X'(c) = I$, where I is the $n \times n$ identity matrix. It follows that $X(t)$ is nonsingular on $[c, d]$. For, if $X(s)$ is singular for some $s, c < s \leq d$, then there exists a vector \bar{a} in R^n such that $X(s)\bar{a} = 0$, $\bar{a} \neq 0$. But this implies that $x(t) = X(t)\bar{a}$ is a solution of (1) satisfying $x(c) = x(s) = 0$. Since $x'(c) = X'(c)\bar{a} \neq 0$, $x(t)$ is nontrivial. This contradicts the assumption of disconjugacy.

This shows that $X^{-1}(t)$ exists on $(c, d]$. Now let $Y(t, \varepsilon)$ be the solution of (1) satisfying the initial conditions

$$Y(c, \varepsilon) = \varepsilon I, Y'(c, \varepsilon) = I.$$

We wish to show that for $\varepsilon > 0$ and sufficiently small, $Y(c, \varepsilon)$ is nonsingular on $[c, d]$. For $\bar{a} \in R^n$, $\|\bar{a}\| = 1$, define

$$f(t, \bar{a}, \varepsilon) = \langle \bar{a}, Y'(t, \varepsilon)\bar{a} \rangle.$$

Then f is uniformly continuous on $[c, d] \times S_n \times [0, 1]$ where $S_n = \{\bar{a} \in R^n \mid \|\bar{a}\| = 1\}$. Since $f(c, \bar{a}, \varepsilon) = \langle \bar{a}, \bar{a} \rangle = 1$ for all $\bar{a} \in S_n$, there exists $\delta > 0$ such that $f(t, \bar{a}, \varepsilon) \geq \frac{1}{2}$ for $c \leq t \leq c + \delta$, $0 \leq \varepsilon \leq \delta$, and $\bar{a} \in S_n$. Note that $Y(t, 0) = X(t)$, and we have shown that $X(t)$ is nonsingular on $(c, d]$. Thus, $|X(t)| \neq 0$ on $[c + \delta, d]$; and for $\varepsilon > 0$ sufficiently small, there exists $\delta' > 0$ such that if $0 \leq \varepsilon \leq \delta'$, then $|Y(t, \varepsilon)| \neq 0$ on $[c + \delta, d]$. We now show that for $0 < \varepsilon < \min\{\delta, \delta'\}$, $Y(t, \varepsilon)$ is nonsingular on $[c, d]$. We have that $Y(t, \varepsilon)$ is nonsingular on $[c + \delta, d]$. Consider the interval $(c, c + \delta]$. If $s \in (c, c + \delta)$ and $Y(s, \varepsilon)$ is singular, then $Y(s, \varepsilon)\bar{a} = 0$ for some $\bar{a} \in R^n$, $\|\bar{a}\| = 1$. Let $g(t) = \langle \bar{a}, Y(t, \varepsilon)\bar{a} \rangle$. We have $g(c) = \varepsilon \|\bar{a}\|^2 = \varepsilon > 0$, and

$$g'(t) = \langle \bar{a}, Y'(t, \varepsilon) \bar{a} \rangle = f(t, \bar{a}, \varepsilon) \geq \frac{1}{2}$$

for $c \leq t \leq c + \delta$. Therefore, $g(t) > 0$ on $c \leq t \leq c + \delta$. But $g(s) = \langle \bar{a}, Y(s, \varepsilon) \bar{a} \rangle = 0$, a contradiction. Hence $Y(t, \varepsilon)$ is nonsingular on $(c, c + \delta]$. But, $Y(c, \varepsilon) = \varepsilon I$ is also nonsingular. We have shown that for $\varepsilon > 0$ and sufficiently small, if $Y(t, \varepsilon)$ satisfies $Y'' + A(t)Y = 0$, $Y(c, \varepsilon) = \varepsilon I$ and $Y'(c, \varepsilon) = I$, then Y^{-1} exists on $[c, d]$, assuming disconjugacy on $[c, d]$. Now, if we let $W = -Y' Y^{-1}$, then $W' = W^2 + A$ on $[c, d]$. Furthermore, $W(c) = -\varepsilon^{-1} I$ is symmetric implies (by uniqueness) that $W(t)$ is symmetric on the whole interval $[c, d]$. This completes the proof of the theorem.

REFERENCES

- [1] Ahmad, S. and Lazer, A.C., *Sturmian Theory*, Lecture Notes, Oklahoma State University, 1972.
- [2] Morse, M., A generalization of the Sturm separation and comparison theorems in n-space, math. Ann., 103 (1930).
- [3] Reid, W.T., *Ordinary Differential Equations*, John Wiley, N.Y., 1973.

27 SOME RESULTS ON REACTION DIFFUSION EQUATIONS WITH INITIAL TIME DIFFERENCE

Aghalaya S. Vatsala

Department of Mathematics

University of Southwestern Louisiana

Lafayette, LA 70504-1010

1. INTRODUCTION

In the qualitative study of initial value problems as well as initial boundary value problems, we have been partial to the independent variable in the sense that we only perturb the dependent variable or the space variable and keep the initial time unchanged [1, 2, 3]. However, it is important to vary the initial time as well because it is impossible not to make errors in the starting time. For example, the solutions of the perturbed and unperturbed system may start at different initial time. Also, there are several ways of comparing any two solutions which differ in time. To each choice of measuring the difference, we will end up with different results. Recently, some results are developed for first order ordinary differential equations with initial time difference [4, 5].

In this paper we develop a comparison theorem, existence results by the method of upper and lower solutions and the monotone iterative technique respectively for reaction diffusion equations with the initial time difference and Dirichlet boundary condition. Here, we use the type of measure that is used in [4, 5] for initial value problems.

2. PRELIMINARIES

Consider the following reaction diffusion equation with initial and boundary conditions:

$$\begin{aligned} \frac{\partial u}{\partial t} - \mathcal{L}u &= f(t, x, u) \text{ on } Q_{T, t_0} \\ u(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ u(t_0, x) &= u_0(x) \text{ on } \overline{\Omega}, \end{aligned} \tag{2.1}$$

where Ω is a bounded domain in R^N , $Q_{T, t_0} = (t_0, t_0 + T] \times \Omega$, for $T > 0$,

$\Gamma_{T, t_0} = (t_0, t_0 + T) \times \partial\Omega$ for any $T < \infty$. Here $\mathcal{L}u$ is given by

$$\mathcal{L}u = \sum_{i,j=1}^N a_{ij}(t, x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{j=1}^N b_j(t, x) \frac{\partial u}{\partial x_j}.$$

We list the following assumptions for convenience.

(A0)

- (i) \mathcal{L} is a uniformly elliptic operator in $R_+ \times \overline{\Omega}$ and the coefficients in \mathcal{L} belongs to $C^{\alpha/2, \alpha}[R_+ \times \overline{\Omega}, R]$;
- (ii) $\partial\Omega$ belongs to $C^{2+\alpha}$;
- (iii) $f \in C^{\alpha/2, \alpha}[R_+ \times \overline{\Omega} \times R, r]$;
- (iv) $\varphi(t, x) \in C^{1+\alpha/2, 1+\alpha}[R_+ \times \partial\Omega, R]$ and $u_0(x) \in C^{2+\alpha}[\overline{\Omega}, R]$;
- (v) the initial boundary value problem (2.1) satisfies the compatibility condition of order $[(1+\alpha)/2]$. See [1] for definition.

We recall a comparison result, and known existence results relative to (2.1). Throughout this paper Γ_{T, t_0} is regular for every $t_0 \geq 0$. The first result we recall is a comparison result in terms of upper and lower solution of (2.1).

Theorem 2.1 Assume that

- (i) $v, w \in C^{1,2}[R_+ \times \Omega, R]$, $f \in [R_+ \times \Omega \times R, R]$, where v satisfies

$$\begin{aligned} \frac{\partial v}{\partial t} - \mathcal{L}v &\leq f(t, x, v) \text{ in } Q_{T, t_0} \\ v(t, x) &\leq \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ v(t_0, x) &\leq \varphi_0(x) \text{ on } \overline{\Omega}, \end{aligned} \tag{2.2}$$

and w satisfies

$$\begin{aligned} \frac{\partial w}{\partial t} - \mathcal{L}w &\geq f(t, x, w) \text{ in } Q_{T, t_0} \\ w(t, x) &\geq \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ w(t_0, x) &\geq u_0(x) \text{ on } \overline{\Omega}, \end{aligned} \tag{2.3}$$

- (ii) $f(t, x, u_1) - f(t, x, u_2) \leq \mathcal{L}(u_1 - u_2)$, $u_1 \geq u_2$, $M > 0$ for (t, x) in Q_{T, t_0} . Then $v(t, x) \leq w(t, x)$ on $[t_0, t_0 + T] \times \overline{\Omega}$.

Proof. See [1] for the proof. This is a special case of Theorem 4.1.1 of [1].

The next result is an existence result by the method of upper and lower solutions of (2.1). Here replace $\mathcal{L}u$ by $\tilde{\mathcal{L}}u = \mathcal{L}u + c(t, x)u$, where $c(t, x) \in C^{\alpha/2, \alpha} [R_+ \times \bar{\Omega}, R]$ is a bounded function. Further we assume t belongs to a bounded interval in R_+ .

Theorem 2.2 Assume that all of (A_0) are satisfied. Assume further that $v, w \in C^{1,2} [R \times \Omega, R]$ are upper and lower solutions of (2.1), such that $v(t, x) \leq w(t, x)$ on \bar{Q}_{T, t_0} . $\mathcal{L}u$ is replaced by $\mathcal{L}u = \mathcal{L}u + cu$, where $c(t, x) \in C^{\alpha/2, \alpha} [\bar{Q}_{T, t_0}, R]$. Then the initial boundary value problem (2.1) has a solution $u(t, x)$ belonging to $C^{1+\alpha/2, 2+\alpha} [\bar{Q}_{T, t_0}, R]$, such that $v(t, x) \leq u(t, x) \leq w(t, x)$ on \bar{Q}_{T, t_0} .

This is a special case of Theorem 4.2.2 of [1]. The next known existence result is by monotone method combined with the method of upper and lower solutions. This is a constructive method of showing the existence of solution of (2.1). For this purpose, we assume the following relative to $f(t, x, u)$ in (2.1). That is,

$$f(t, x, u_1) - f(t, x, u_2) \geq -M(u_1 - u_2) \quad (2.4)$$

for $v(t, x) \leq u_2 \leq u_1 \leq w(t, x)$, $t, x \in \bar{Q}_{T, t_0}$.

Theorem 2.3 Let all assumptions of Theorem 2.1 hold. Assume further that f satisfies (2.4) on the sector defined by v and w , the lower and upper solutions of (2.1), with \mathcal{L} replaced by $\tilde{\mathcal{L}}$. Then there exist monotone sequences $\{v_n\}, \{w_n\}$ which converge in $C^{1,2} [\bar{Q}_{T, t_0}, R]$ to ρ and r respectively on \bar{Q}_{T, t_0} . Moreover ρ and r are minimal and maximal solutions of the initial boundary value problem (2.1).

Proof. See [1] for details of the proof. This is a special case of Theorem 4.3.1 of [1].

3. MAIN RESULTS

In this section we develop results analogous to Theorem 2.1, 2.2, and 2.3 relative to equation (2.1) with initial time difference and Dirichlet boundary conditions. We introduce the following notations to develop our main results. Let $t_0 \geq 0$ and $\tau_0 \geq t_0$ be such that $\tau_0 - t_0 = \eta > 0$. Further, we denote $\bar{Q}_{T, \tau_0} = [\tau_0, \tau_0 + T] \times \Omega$, $\Gamma_{T, \tau_0} = (\tau_0, \tau_0 + T) \times \partial\Omega$, and $\bar{Q}_{t_0, T} = [t_0, t_0 + T] \times \Omega$,

$\Gamma_{T,t_0} = (t_0, t_0 + T) \times \partial\Omega$. Our first result is a comparison result which is analogous to Theorem 2.1 for (2.1) with initial time difference.

Theorem 3.1 Assume that

(A1) $v, w \in C^{1,2}[R_+ \times \Omega, R]$, $f \in C[R_+ \times \Omega \times R, R]$, where v satisfies

$$\frac{\partial v}{\partial t} - \mathcal{L}v \leq f(t, x, v) \text{ in } Q_{T,t_0}$$

$$v(t, x) \leq \varphi(t, x) \text{ on } \Gamma_{T,t_0}$$

$$v(t_0, x) \leq u_0(x) \text{ on } \bar{\Omega},$$

and w satisfies

$$\frac{\partial w}{\partial t} - \mathcal{L}w \geq f(t, x, w) \text{ in } Q_{T,t_0}$$

$$w(t, x) \geq \varphi(t, x) \text{ on } \Gamma_{T,t_0}$$

$$w(\tau_0, x) \geq u_0(x) \text{ on } \bar{\Omega};$$

(A2) $f(t, x, u_1) - f(t, x, u_2) \leq \mathcal{L}(u_1 - u_2)$, $u_1 \geq u_2$, $\mathcal{L} > 0$;

(A3) $\tau_0 > t_0$ and $f(t, x, u)$ is nondecreasing in t for each $x, u \in \Omega \times R$, $\varphi(t, x)$ is nondecreasing in t for each $x \in \Omega$. Then (a) $v(t, x) \leq w(t + \eta, x)$ on \bar{Q}_{T,t_0} and (b) $v(t - \eta, x) \leq w(t, x)$ on \bar{Q}_{T,t_0} where $\eta = \tau_0 - t_0 > 0$.

Proof. Define $w_0(t, x) = w(t + \eta, x)$ on \bar{Q}_{T,t_0} , then $w_0(t_0, x) = w(t_0 + \eta, x) = w(\tau_0, x) \geq \varphi_0 \geq v(t_0, x)$. Also $w_0(t, x) = w(t + \eta, x) \geq \varphi(t + \eta, x)$ on Γ_{T,t_0} . Since $\varphi(t, x)$ is nondecreasing in t for each $x \in \partial\Omega$, we get $w_0(t, x) \geq \varphi(t, x) \geq v(t, x)$ on Γ_{T,t_0} . Let $\tilde{w}_0(t, x) = w_0(t, x) + \varepsilon e^{2L(t-t_0)}$ for some $\varepsilon > 0$ small. Certainly $\tilde{w}_0(t, x) > w_0(t, x) \geq \varphi(t, x) \geq v(t, x)$ on Γ_{T,t_0} . Also $\tilde{w}_0(t_0, x) = w_0(t_0, x) + \varepsilon > w_0(t_0, x) \geq u_0(x) \geq v(t_0, x)$ on $\bar{\Omega}$. Further using (A2) and (A3) we get

$$\begin{aligned}
\frac{\partial \tilde{w}_0}{\partial t} - L \tilde{w}_0 &= \frac{\partial w_0}{\partial t} - L w_0 + 2L\epsilon e^{2L(t-t_0)} \\
&= \frac{\partial w(t+\eta, x)}{\partial t} - L w(t+\eta, x) + 2L\epsilon e^{2L(t-t_0)} \\
&\geq f(t+\eta, x, w(t+\eta, x)) + 2L\epsilon e^{2L(t-t_0)} \\
&\geq f(t, x, w_0(t, x)) - f(t, x, \tilde{w}_0(t, x)) + f(t, x, \tilde{w}_0(t, x)) + 2L\epsilon e^{2L(t-t_0)} \\
&\geq -L\epsilon e^{2L(t-t_0)} + f(t, x, \tilde{w}_0(t, x)) + 2L\epsilon e^{2L(t-t_0)} \\
&> L\epsilon e^{2L(t-t_0)} + f(t, x, \tilde{w}_0(t, x)) \\
&> f(t, x, \tilde{w}_0(t, x)) \text{ on } Q_{T,t_0}.
\end{aligned}$$

This proves that v and \tilde{w}_0 are lower and upper solutions of (2.1) on \overline{Q}_{T,t_0} with \tilde{w}_0 satisfying strict inequality. Applying Theorem 2.1, we get

$$v(t, x) < \tilde{w}_0(t, x) \text{ on } \overline{Q}_{T,t_0}.$$

Taking the limit as $\epsilon \rightarrow \infty$, we get

$$v(t, x) \leq w_0(t, x) \leq w(t+\eta, x) \text{ for } t \geq t_0.$$

In order to prove (b) set $v_0(t, x) = v(t-\eta, x)$ and $\tilde{v}_0(t, x) = v(t, x) - \epsilon e^{2L(t-\tau_0)}$

and carrying out a similar argument as in (a) and using (A2) and (A3), we get

$$v_0(t, \eta) \leq w(t, x) \text{ on } Q_{T,\tau_0}.$$

Hence, the conclusion (b) follows.

The next result is analogous to Theorem 2.2

Theorem 3.2 Let all of (A0) and (A3) of Theorem 3.1 hold. Further, assume that

(A4)

$$v, w \in C^{1,2}[R_+ \times \Omega, R], f \in C[R_+ \times \Omega \times R, R]$$

where v satisfies

$$\frac{\partial v}{\partial t} - \tilde{\mathcal{L}}v \leq f(t, x, v) \text{ in } Q_{T,t_0}$$

$$v(t, x) \leq \varphi(t, x) \text{ on } \Gamma_{T,t_0}$$

$$v(t_0, x) \leq u_0(x) \text{ on } \overline{\Omega},$$

and w satisfies

$$\frac{\partial w}{\partial t} - \tilde{\mathcal{L}}w \geq f(t, x, w) \text{ in } Q_{T,\tau_0}$$

$$w(t, x) \geq \varphi(t, x) \text{ on } \Gamma_{T,t_0}$$

$$w(\tau_0, x) \geq \varphi_0(x) \text{ on } \overline{\Omega},$$

such that (i) $v(t, x) \leq w(t + \eta, x)$ on Q_{T, t_0} and (ii) $v(t - \eta, x) \leq w(t, x)$ on Q_{T, τ_0} holds. Then the initial boundary value problem (2.1), where $\mathcal{L}u$ replaced by $\tilde{\mathcal{L}}u$ has a solution $u(t, x)$ such that

$$(a) \quad v(t, x) \leq u(t, x) \leq w(t + \eta, x) \text{ on } [t_0, t_0 + \tau] \times \bar{\Omega},$$

and

$$(b) \quad v(t - \eta, x) \leq u(t, x) \leq w(t, x) \text{ on } [\tau_0, \tau_0 + T] \times \bar{\Omega}.$$

Proof. We prove (a) here and the proof of (b) follows on the same lines. Certainly $v(t, x)$ is a lower solution of

$$\begin{aligned} \frac{\partial u}{\partial t} - \tilde{\mathcal{L}}u &= f(t, x, u) \text{ on } Q_{T, t_0} \\ u(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ u(t_0, x) &= u_0(x) \text{ on } \bar{\Omega}. \end{aligned} \tag{3.1}$$

We prove that $w(t + \eta, x)$ is an upper solution of (3.1) on Q_{T, t_0} . For that purpose set $w_0(t, x) = w(t + \eta, x)$. Then

$$\begin{aligned} \frac{\partial w}{\partial t} - \tilde{\mathcal{L}}w_0 &= \frac{\partial w(t + \eta, x)}{\partial t} - \tilde{\mathcal{L}}w(t + \eta, x) \\ &\geq f(t + \eta, x, w(t + \eta, x)) \\ &= f(t + \eta, x, w_0(t, x)) \\ &\geq f(t, x, w_0(t, x)) \text{ on } Q_{T, t_0}. \end{aligned}$$

Also $w_0(t, x) = w(t + \eta, x) = \varphi(t + \eta, x) \geq \varphi(t, x)$ on Γ_{T, t_0} , since $\varphi(t, x)$ is nondecreasing in t . Further, $w_0(t, x) = w(t_0 + \eta, x) = w(\tau_0, x) \geq u_0(x)$ on $\bar{\Omega}$. From (A4), we have $v(t, x) \leq w(t + \eta, x) = w_0(t, x)$ on Q_{T, t_0} . Applying Theorem 2.2 on v, w_0 , it easily follows (3.1) has a solution $u(t, x)$ belonging to $C^{1+\alpha/2, 2+\alpha}[\bar{Q}_{T, t_0}, R]$ such that $v(t, x) \leq u(t, x) \leq w_0(t, x)$ on \bar{Q}_{T, t_0} . Thus (a) follows. One can prove (b) on the same lines by proving $v(t - \eta, x) = v_0(t, x)$ is a lower solution of (3.1) on Q_{T, τ_0} .

Theorem 3.3 Let all the hypothesis of Theorem 2.2 hold. Further let

$$f(t, x, u_1) - f(t, x, u_2) \geq -M(u_1 - u_2)$$

whenever

$$v(t, x) \leq u_2 \leq u_1 \leq w(t + \eta, x).$$

Then there exists monotone sequences $\{v_n(t, x)\}, \{\tilde{w}_n(t, x)\}$ such that $v_n(t, x) \rightarrow \rho(t, x)$ and $\tilde{w}_n(t, x) \rightarrow \tilde{r}(t, x)$ as $n \rightarrow \infty$, uniformly and monotonically on $[t_0, t_0 + T] \times \bar{\Omega}$, where $\tilde{w}_0(t, x) = w(t + \eta, x)$, $v_0(t, x) = v(t, x)$. Moreover, $\rho(t, x), \tilde{r}(t, x)$ are minimal and maximal solutions of

$$\begin{aligned} \frac{\partial u}{\partial t} - \mathcal{L}u &= f(t, x, u) \text{ on } Q_{T, t_0} \\ u(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ u(t_0, x) &= u_0(x) \text{ on } \bar{\Omega}, \end{aligned} \quad (3.2)$$

and

$$\begin{aligned} \frac{\partial u}{\partial t} - \mathcal{L}u &= f(t, x, u) \text{ on } Q_{T, \tau_0} \\ u(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, \tau_0} \\ u(\tau_0, x) &= u_0(x) \end{aligned} \quad (3.3)$$

respectively in the sector $[v, \tilde{w}]$.

Proof. We recall that $\tilde{w}_0(t, x) = w(t + n, x)$ implies $\tilde{w}_0(t, x) = w(t + \eta, x) = \varphi(t + \eta, x) \geq \varphi(t, x)$ on Γ_{T, t_0} and $\tilde{w}_0(t_0, x) = w(t_0 + \eta, x) = w(\tau_0, x) \geq \varphi_0(x)$ on $\bar{\Omega}$. We define $\{v_n(t, x)\}$ and $\{w_n(t, x)\}$ as solutions of the following initial value problems respectively,

$$\begin{aligned} \frac{\partial v_{n+1}}{\partial t} - \mathcal{L}v_{n+1} &= f(t, x, v_n(t, x)) - M(v_{n+1} - v_n), \text{ on } Q_{T, t_0} \\ v_{n+1}(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ v_{n+1}(t_0, x) &= \varphi_0(x) \text{ on } \bar{\Omega}, \end{aligned} \quad (3.4)$$

$$\begin{aligned} \frac{\partial \tilde{w}_{n+1}}{\partial t} - \mathcal{L}\tilde{w}_{n+1} &= f(t + \eta, x, \tilde{w}_n) - M(\tilde{w}_{n+1} - \tilde{w}_n) \\ \tilde{w}_{n+1}(t, x) &= \varphi(t + \eta, x) \\ \tilde{w}_{n+1}(t_0, x) &= \varphi_0(x). \end{aligned} \quad (3.5)$$

Let

$$\begin{aligned} p(t, x) &= w_1(t, x) - w_0(t, x), \text{ so that} \\ p(t, x) &= \varphi(t + \eta, x) - \varphi(t + \eta, x) = 0 \text{ on } \Gamma_{T, t_0} \\ p(t_0, x) &= \varphi_0(x) - \varphi_0(x) = 0 \text{ on } \bar{\Omega} \\ \frac{\partial p}{\partial t} - \mathcal{L}p &\geq f(t + \eta, x, \varphi_0(x)) - M(w_1 - w_0) - f(t + \eta, x, w_0). \end{aligned}$$

Using the linear comparison theorem, it easily follows $w_1(t, x) \geq w_0(t, x)$ on Q_{T, t_0} .

In order to prove $v_1(t, x) \leq \tilde{w}_1(t, x)$, set $p(t, x) = v_1 - \tilde{w}_1$. Then

$$\begin{aligned} \frac{\partial p}{\partial t} - \mathcal{L}p &= f(t, x, v_0) - M(v_1 - v_0) - [f(t + \eta, x, \tilde{w}_0) - M(\tilde{w}_1 - \tilde{w}_0)] \\ &\leq f(t, x, v_0) - f(t, x, \tilde{w}_0) - M p + M(v_0 - \tilde{w}_0) \\ &\leq M(\tilde{w}_0 - v_0) - M p + M(v_0 - \tilde{w}_0) = -M p. \end{aligned}$$

Also, $p(t, x) = v_1(t, x) - \tilde{w}_1(t, x) = \varphi(t, x) - \varphi(t + \eta, x) \leq 0$ on Γ_{T, t_0} , using the nondecreasing nature of $\varphi(t, x)$ in t , $p(t_0, x) = v_1(t_0, x) - \tilde{w}_1(t_0, x) = \varphi_0(x) - \varphi_0(x) = 0$. Again, using the linear parabolic comparison theorem it follows $v_1(t, x) \leq \tilde{w}_1(t, x)$ on $[t_0, t_0 + T] \times \bar{\Omega}$. Thus, we have $v_0 \leq v_1 \leq \tilde{w}_1 \leq \tilde{w}_0$ on $[t_0, t_0 + T] \times \bar{\Omega}$. Following similar arguments and using the method of mathematical induction, one can easily prove that $v_0 \leq v_1 \leq \dots \leq v_n \leq \tilde{w}_n \leq \dots \leq \tilde{w}_1 \leq \tilde{w}_0$ on $[t_0, t_0 + T] \times \bar{\Omega}$. Using the standard arguments as in [1], we can prove that $v_n(t, x) \rightarrow \rho(t, x)$ and $\tilde{w}_n(t, x) \rightarrow \tilde{r}(t, x)$ as $n \rightarrow \infty$, uniformly and monotonically on $[t_0, t_0 + T] \times \bar{\Omega}$. Further, ρ and \tilde{r} satisfies the following reaction diffusion equations respectively. That is, $\rho(t, x)$ satisfies

$$\begin{aligned} \frac{\partial \rho}{\partial t} - \rho &= f(t, x, \rho) \text{ on } Q_{T, t_0} \\ \rho(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, t_0} \\ \rho(t_0, x) &= \varphi_0(x) \text{ on } \bar{\Omega}, \end{aligned}$$

and $\tilde{r}(t, x)$ satisfies

$$\begin{aligned} \frac{\partial \tilde{r}}{\partial t} - \mathcal{L}\tilde{r} &= f(t + \eta, x, r) \text{ on } Q_{T, t_0} \\ \tilde{r}(t, x) &= \varphi(t + \eta, x) \text{ on } \Gamma_{T, t_0} \\ \tilde{r}(t_0, x) &= \varphi_0(x) \text{ on } \bar{\Omega}. \end{aligned}$$

Let $u(t, x)$ be any solution of (3.2) such that $v_0(t, x) \leq u(t, x) \leq \tilde{w}_0(t, x)$. Then by the standard proof of [1], it follows $\rho(t, x) \leq u(t, x)$ on $[t_0, t_0 + T] \times \bar{\Omega}$. Since $u(t, x)$ is any solution of (3.2), we also get

$$\begin{aligned} \frac{\partial u}{\partial t} - \mathcal{L}u &= f(t, x, u) \leq f(t + \eta, x, u) \text{ on } Q_{T, t_0} \\ u(t, x) &= \varphi(t, x) \leq \varphi(t + \eta, x) \text{ on } \Gamma_{T, t_0} \\ u(t_0, x) &= \varphi_0(x) \text{ on } \bar{\Omega}. \end{aligned}$$

Since $\tilde{w}_n(t, x)$ is a decreasing sequence, one can easily prove $u(t, x) \leq \tilde{w}_n(t, x)$ for all n . Taking the limit as $n \rightarrow \infty$, it follows $u(t, x) \leq \tilde{r}(t, x)$ on $[t_0, t_0 + T] \times \bar{\Omega}$. Let $u(t, x)$ be any solution of

$$\begin{aligned}\frac{\partial u}{\partial t} - \mathcal{L}u &= f(t, x, u) \text{ on } Q_{T, \tau_0} \\ u(t, x) &= \varphi(t, x) \text{ on } \Gamma_{T, \tau_0} \\ u(\tau_0, x) &= \varphi_0(x) \text{ on } \bar{\Omega}.\end{aligned}$$

Then setting $\tilde{u}(t, x) = u(t + \eta, x)$, we see that

$$\begin{aligned}\frac{\partial \tilde{u}}{\partial t} - \mathcal{L}\tilde{u} &= f(t + \eta, x, u) \\ \tilde{u}(t, x) &= \varphi(t + \eta, x) \geq \varphi(t, x) \\ \tilde{u}(\tau_0, x) &= u(\tau_0, x) = \varphi_0(x) \text{ on } \bar{\Omega}.\end{aligned}$$

Suppose that $v_0(t, x) \leq \tilde{u}(t, x) \leq \tilde{w}_0(t, x)$ on $[t_0, t_0 + T] \times \Omega$. Then as before, we can prove by induction $v_n(t, x) \leq \tilde{u}(t, x) \leq \tilde{w}_n(t, x)$ on $[t_0, t_0 + T] \times \bar{\Omega}$. This proves $\rho(t, x) \leq \tilde{u}(t, x) \leq \tilde{r}(t, x)$ on $[t_0, t_0 + T] \times \bar{\Omega}$. This proves ρ and \tilde{r} are extremal solutions of the IBVP's (3.2) and (3.3). This completes the proof.

REFERENCES

- [1] Ladde, G.S., Lakshmikantham, V., and Vatsala, A.S., *Monotone Iterative Technique for Nonlinear Differential Equations*, Pitman, Boston, 1985.
- [2] Lakshmikantham, V. and Leela, S., *Differential and Integral Inequalities*, Vol. I, II, Academic Press, New York, 1969.
- [3] Lakshmikantham, V., Leela, S., and Martynyuk, *Stability Analysis of Nonlinear Systems*, Marcel Dekker, Inc., New York, 1989.
- [4] Lakshmikantham, V., and Vatsala, A.S., Differential inequalities with initial time difference and applications, *Journal of Inequalities and Applications*, Vol. 3, 1999, 233-244.
- [5] Lakshmikantham, V. and Vatsala, A.S., Theory of differential and integral inequalities with initial time difference and applications, *Analytic and Geometric Inequalities and Their Applications*, Kluwer Academic Press, Rassias and Srivastav, Eds., 191-203, 1999.

28 DYNAMICS OF NEURAL NETWORKS WITH DELAY: ATTRACTORS AND CONTENT- ADDRESSABLE MEMORY

Jianhong Wu

Department of Mathematics and Statistics
York University
Toronto, Canada M3J 1P3

In the design of a neural network, either for biological modeling, cognitive simulation, numerical computation or engineering applications, it is important to describe the dynamics of the network. The success in this area in the early 1980's was one of the main sources for the resurgence of interest in neural networks, and the current progress towards understanding neural dynamics has been part of exhaustive efforts to lay down a solid theoretical foundation for this fast growing theory and for the applications of neural networks.

The purpose of this note is to give a short overview about the impact of the signal delay on the dynamics of the network, and in particular, on the applications to content-addressable memory. We will start with a brief introduction to the basic knowledge in neurobiology and physiology which is necessary in order to understand the mathematical models to be discussed. In particular, we will recall some of the basic structures of a single neuron, the connection topology of networks of neurons and the main mechanisms of the neural signal transmission. We will then present the derivation of additive equations and a brief description of the popular signal functions. A central subject of this note is the long-term behaviors of the network, and in particular, we will address the connection of the convergence and the global attractor to the important property of content-addressable memory of networks, in conjunction with the convergence theorem due to Cohen, Grossberg, and Hopfield based on LaSalle's invariance principle. We will then emphasize on the effect of signal delays on the long-term performance of the networks under

consideration. As will be illustrated, such time lags exist due to the finite propagation speed of neural signals along axons and the finite speed of the neurotransmitters across synaptic gaps in a biological neural network and due to the finite switching speed of amplifiers (neurons) in artificial neural networks. We will report various phenomena associated with the signal delays: delay-induced instability, nonlinear periodic solutions, transient oscillations, phase-locked oscillations, and changes of the basin of attractions.

1. NEURONS, SIGNAL TRANSMISSIONS AND ADDITIVE MODELS

As we know, the human central nervous system is composed of a vast number of simple but interconnected cellular units called neurons. While neurons have a wide variety of shapes, sizes, and location, most neurons are of rather uniform structures and neural signals are transmitted on the same basic electrical and chemical principles.

The central part of a neuron is called the cell body which contains nucleus and other organelles that are essential to the function of all cells. From the cell body project, many root-like extensions (the dendrites) as well as a singular tubular fiber (the axon). The axon ramifies at its end into a number of small branches.

Dendrites are branch-like protrusions from the neuron cell body. A typical cell has many dendrites that are highly branched. The receiving zones of signals, called synapses, are on the cell body and dendrites. Some neurons have spines on the dendrites, thus creating more receiving sites.

The axon is a long fiber-like extension of the cell body. The axon's purpose is the signal conduction, i.e., transmission of the impulses to other neurons. The axonal mechanism that carries signals over the axon is called the action potential, a self-regenerating electrical wave that propagates from its point of initiation at the cell body to the terminals of the axon. Axon terminals are highly specialized to convey the signal to target neurons. These terminal specializations are called synaptic endings, and the contacts they make with the target neurons are called chemical synapses. Each synaptic ending contains secretory organelles called synaptic vesicles. The same signal encoded by action potentials propagates along each branch with varying time delays. As mentioned above, each branch end has a synapse which is typically of a bulb-like structure. The synapses are on the cell bodies, dendrites and spines of target neurons. Between the synapses and the target neurons is a narrow gap (called synaptic gap), typically 20 nanometers wide. Special molecules called neurotransmitters are released from synaptic vesicles and cross the synaptic gap to receptor sites on the target neuron. The transmission of

signals between synapses and target neurons is the flow of neurotransmitter molecules.

Neuron signals are transmitted either electrically or chemically. Electrical transmission prevails in the interior of a neuron, whereas chemical mechanisms operate at the synapses. These two types of signaling mechanisms are the basis for all the information-processing capabilities of the brain.

An electrochemical mechanism produces and propagates signals along the axon. The axon signal pulses are the fundamental electrical signals of neurons. They are called action potential and are described electrically by current-voltage characteristics. Injecting a current pulse into the axon causes the potential across the membrane to vary. When injected an inhibitory or excitatory signal pulse, the response is a RC-type (exponential) followed by relaxation. When the injected current causes the voltage to exceed the threshold, a special electrochemical process generates a rapid increase in the potential and this injected current then produces a single pulse with a peak-to-peak amplitude of about 120mV and a duration of 1ms.

The pulse signal traveling along the axon comes to a halt at the synapse due to the synaptic gap. The signal is transferred to the target neuron across the synaptic gap mainly via a special chemical mechanism, called synaptic transmission. In synaptic transmission, when a pulse train signal arrives at the presynaptic site, special substances called neurotransmitters are released. The neurotransmitter molecules travel across the synaptic gap, reaching the postsynaptic neuron (or muscle fiber) within about 0.5ms. Upon their arrival at special receptors, these substances modify the permeability of the postsynaptic membrane for certain ions. These ions then flow in or out of the neurons, causing a polarization or depolarization of the local postsynaptic potential. If the induced polarization potential is positive (resp. negative), then the synapse is termed excitatory (resp. inhibitory).

The polarization potential caused by a single synapse might or might not be large enough to depolarize the postsynaptic neuron to its firing threshold. In fact, the postsynaptic neuron typically has thousands of dendrites receiving synapses from different presynaptic cells (neurons). Hence, its firing depends on the sum of depolarizing effects from these different dendrites. These effects decay with a characteristic time of 5-10ms, but if signals arrive at the same synapse over such a period, then excitatory effects accumulate. When the total magnitude of the depolarization potential in the cell body exceeds the critical threshold, the neuron fires.

The complexity of the human nervous system rests on not only the complicated structures of single nerve cells and the complicated mechanisms of nerve signal

transmission, but also the vast number of neurons and their mutual connections. Connectivity of neurons is essential for the brain to perform complex tasks.

Plasticity is another essential feature of (developing) neural networks. There is a great deal of evidence showing that the strength of a synaptic coupling between two given neurons is not fixed once and for all. As originally postulated by Hebb [1949], the strength of a synaptic connection can be adjusted if its level of activity changes. An active synapse, which repeatedly triggers the activation of its postsynaptic neuron, will grow in strength, while others will gradually weaken. This plasticity permits the modification of synaptic coupling and connection topology, which is important in the network's ability of learning from and adaptive to its environments.

We now formulate the so-called additive STM equation for the time evolution of a network of neurons. Assume that the network consists of n neurons, denoted by v_1, \dots, v_n . We will introduce a variable x_i to describe the neuron's state and a variable Z_{ij} to describe the coupling between two neurons v_i and v_j .

Let

$x_i(t)$ = deviation of the i th neuron's potential from its equilibrium.

This variable describes the activation level of the i th neuron. It is called the short-term memory (STM) trace. Let

Z_{ij} = neurotransmitter average release rate per unit axon signal frequency.

This is called the synaptic coupling coefficient or the long-term memory (LTM) trace.

Assume a change in neurons potential from equilibrium occurs. Then we have the following form for the STM trace:

$$\frac{dx_i}{dt} = -A_i(x_i)x_i + \sum_{k \neq i} S_{ki}Z_{ki} - \sum_{k \neq i} C_{ki} - I_i(t), \quad 1 \leq i \leq n,$$

where

- $A_i(x_i) > 0$ and the terms $-A_i(x_i)x_i$ describes the stability of internal neuron processes (that is, the neuron's potential decays exponentially to its equilibrium without external processes);
- $\sum_{k \neq i} S_{ki}Z_{ki}$ describes the additive synaptic excitation which is assumed to be proportional to the pulse train frequency, where S_{ki} is the average frequency of signal evaluated at v_i in the axon from v_k to v_i . This is called the signal function. In general, S_{ki} depends on the propagation time delay τ_{ki} from v_k to v_i and the threshold Γ_k for firing of v_k such that $S_{ki}(t) = f_k(x_k(t - \tau_{ki}) - \Gamma_k)$ for a given nonnegative function $f_k : \mathbb{R} \rightarrow [0, \infty)$.

Commonly used forms of the signal function will be described later;

- the term $-\sum_{k \neq i}^n C_{ki}$ describes the hardwiring of the inhibitory inputs from other neurons, with $C_{ki}(t) = c_{ki} f_k(x_k(t - \tau_{ki}) - \Gamma_k)$, where $c_{ki} \geq 0$ are constants;
- the term $I_i(t)$ is the external input.

It should be mentioned that we should have another set of equations for LTM trace if the excitatory coupling strength varies with time. A common model based on the Hebb's law adds an additional $n \times n$ equations for the evolution of Z_{ij} , but we are not going to discuss this in this note.

Typical functions of f_k include step function, piecewise linear function, and sigmoid function. Models involving a step signal function: $f_k(x) = 1$ for $x \geq 0$ and $f_k(x) = 0$ for $x < 0$, are referred as the McCullon-Pitts models, in recognition of the pioneering work of McCullon and Pitts [1943] (the function describes the all-or-none property of a neuron in McCullon-Pitts model). A piecewise linear function:

$f_k(x) = 0$ for $x < 0$, $f_k(x) = \beta x$ for $x \in \left[0, \frac{1}{\beta}\right]$ and $f_k(x) = 1$ for $x \geq 1$, describes

the nonlinear off-on characteristic of neurons. β is called the neuron gain and a piecewise linear function reduces to the step function if β is infinity. Such a function has been widely used in cellular neural network models (see Chua and Yang [1988]). The sigmoid function is by far the most common form of a signal function. It is defined as a strictly increasing smooth bounded function satisfying certain concavity and asymptotic properties. An example of a sigmoid function is the logistic function given by

$$f_k(v) = \frac{1}{1 + e^{-4\beta v}}, \quad v \in \mathbb{R},$$

where $\beta = f'_k(0) > 0$ is the neuron gain. Other examples include inverse tangent function and hyperbolic tangent function. As $\beta \rightarrow \infty$, the sigmoid function becomes the step function. Whereas a step function assumes the values of 0 or 1, a sigmoid function takes continuous values in the interval $[0, 1]$. This is important. It allows analog signal processing and it makes many mathematical theories applicable. Another important feature of the bounded sigmoid function is to limit the magnitude of a nerval signal's impact to its receiving neurons. It was noted, see Levine [1991], that if the firing threshold of an all-or-none neuron is described by a random variable with a Gaussian (normal) distribution, then the expected value of its output signal is a sigmoid function of activity. For this and other reasons, sigmoids have become increasingly popular in recent neural network models. Also, there has been some physiological verification of sigmoid signal functions at the neuron level, see Wu [1999] for related references.

Other signal functions are also possible. Many global neural network properties are not sensitive to the choice of particular signal functions, but some are. Choice of a signal function also depends on the applications considered.

In this note, we will always assume the signal function is a sigmoid function with necessary smoothness and boundedness. Readers can always assume the signal function as the logistic function given above.

Assuming the LTM traces stabilize at constant values and lumping together the excitatory and inhibitory feedback terms in the additive STM, we get

$$\frac{d}{dt} x_i = -A_i x_i + \sum_{j=1}^n w_{ij} f_j(x_j) + I_i, \quad 1 \leq i \leq n.$$

Such a model is also called Hopfield's net, as it also describes the evolution of an electronic net invented by Hopfield. In fact, it is interesting to see how easy it is to implement the above model in an artificial neural network consisting of electronic neurons (amplifiers) interconnected through a matrix of resistors (see Müller and Reinhardt [1991] and Hopfield [1984] for more detailed discussions). Here an electronic neuron, the building block of the network, consists of a nonlinear amplifier which transforms an input signal u_i into the output signal v_i , and the input impedance of the amplifier unit is described by the combination of a resistor ρ_i and a capacitor C_i . We assume the input-output relation is completely characterized by a voltage amplification function $v_i = f_i(u_i)$. The synaptic connections of the network are represented by resistors R_{ij} which connect the output terminal of the amplifier j with the input part of the neuron i . In order that the network can function properly the resistances R_{ij} must be able to take on negative values. This can be realized through a slight modification of the connection matrix, namely, the amplifiers are supplied with an inverting output line which produces the signal $-v_j$. The number of rows in the resistor matrix is doubled, and whenever a negative value of R_{ij} is needed this is realized by using an ordinary resistor which is connected to the inverting output line.

The time evolution of the signals of the network is described by the Kirchhoff's law. Namely, the strength of the incoming and outgoing current at the amplifier input port must balance. Consequently, we arrive at

$$C_i \frac{du_i}{dt} + \frac{u_i}{\rho_i} = \sum_{j=1}^n \frac{1}{R_{ij}} (v_j - u_i),$$

and thus,

$$T_i \frac{du_i}{dt} + u_i = \sum_{j=1}^n w_{ij} f_j(u_j),$$

with $\frac{1}{R_i} = \frac{1}{\rho_i} + \sum_{j=1}^n \frac{1}{R_{ij}}$, $T_i = C_i R_i$ and $w_{ij} = \frac{R_i}{R_{ij}}$.

2. CONVERGENCE AND CAM

To illustrate the essential idea behind the connection between CAM and convergence of networks, let us first recall that in a dynamical system generated by a differential equation, an equilibrium represents a steady state of the system which does not change in time. A (asymptotically) stable steady state (equilibrium) is one which has a neighborhood such that the trajectory converges to the equilibrium if it initially starts from a point in such a neighborhood. The set of all initial values, where the orbits converge to this equilibrium, is called the basin of attraction.

Content-addressable memory (CAM) is one of the simplest properties of a wide variety of neural network models. This property involves the change of the network with time where the motion of the states of neurons describes the computation that the network is performing. In mathematical terms, the evolution of a network defines a dynamical system in a phase space. The network is developing stable steady states which are related to stored information. States near to the particular stable equilibrium point contain partial information about the memory, from an initial state of partial information about the memory, a final stable steady state with all the information of the memory can be found due to the aforementioned convergence. The memory is reached not by knowing an address, but rather by supplying in the initial state some subpart of the memory. Any subpart of adequate size will do as long as it is in the basin of attraction of the memory – the memory is truly addressable by constant rather than location. Of course, a network may contain many memories simultaneously, and this is reflected by the fact that the associated dynamical system exhibits coexistence of multiple stable equilibria.

From a mathematical perspective, the question of CAM in a neural network can be formulated as follows:

- (a) Under what conditions does a network approach an equilibrium point in response to the arbitrary initial data?
- (b) How many equilibria exist and which are stable?
- (c) How does the system in response to a given initial data approach an equilibrium?
- (d) What is the basin of attraction of a stable equilibrium and how does this depend on the parameters?
- (e) Can a network store a memory as a stable periodic orbit or other spatio-temporal patterns?
- (f) What is the structure of the global attractor?

This is clearly related to the global analysis of the dynamics of the network and, in most cases, is extremely difficult. Also, answers to the above questions depend heavily on the connection topology, the synaptic coefficients, and the signal delay of the network.

Some progress has been achieved over the past decade towards the above questions in the case where signal delay is ignored. In this case, the additive STM trace model equation becomes

$$\frac{d}{dt} x_i = -A_i x_i + \sum_{j=1}^n w_{ij} f_j(x_j) + I_i, \quad 1 \leq i \leq n, \quad (2.1)$$

with $A_i > 0$, $f_i : R \rightarrow R$ being C^1 -smooth and $f'_i(x_i) > 0$.

In Cohen and Grossberg [1983] and in Hopfield [1984], the following convergence theorem, based on the construction of a Liapunov (energy) function and the application of the LaSalle's invariance principle is established.

Theorem 2.1. If the connection matrix $[w_{ij}]$ is symmetric and if each signal function is bounded, then every solution of (2.1) is convergent to the set of equilibria.

While Hopfield's convergence theorem was stated specifically for the model equation (2.1), the paper of Cohen and Grossberg deals with a much more general system which includes some Volterra-Lotka equations, the Gilpin and Ayala system from population biology, the dynamical model of the Hartline-Ratliff-Miller equation which describes the output of the Limulus retina, the Eigen and Schuster equation from the theory of macro-molecular evolution, the transformed dynamical analogue of the Brain-State-in-a Box (BSB) model, as well as the dynamic McCulloch-Pitts model as special cases.

In some applications of the general Cohen-Grossberg model, the coefficient matrix $[w_{ij}]$ may be asymmetric, thereby rendering the general convergence theorem inapplicable. Asymmetric coefficients typically occur in problems relating to the learning and recognition of temporal order in behaviors. On the other hand, certain network models may have asymmetric interaction coefficients, yet be reducible to the Cohen-Grossberg system with symmetric interaction coefficients through a suitable change of variables. A typical example is the masking field model introduced by Grossberg to explain data about speech learning, word recognition, and the learning of adaptive sensory-motor plans.

Let us emphasize that the convergence theorem, however, requires the interconnection matrix $W = [w_{ij}]$ be symmetric. This requirement is justified when the synaptic strengths w_{ij} between cell v_i and cell v_j depend only on the intercellular distance. However, examples exist wherein the matrix W may be

chosen as close to a symmetric matrix as one pleases, yet almost all trajectories persistently oscillate. A famous example is the May and Leonard's model of the voting paradox (see May and Leonard [1975]) described by a three-dimensional system of ordinary differential equations.

Considerable effort has been spent to remove or to weaken the symmetry condition in the Cohen-Grossberg-Hopfield convergence theorem. In particular, Gedeon [1999] and Fiedler and Gedeon [1998] have been able to drop the symmetry assumption on the structure matrix $W = [w_{ij}]$, by using results from combinatorial matrix theory.

Using the theory of monotone dynamical systems (see, Smith [1987] and Hirsch [1988]), we can obtain some generic convergence theorem, in particular, if we know (possibly after a change of variables) W is cooperative (that is, $w_{ij} \geq 0$ if $i \neq j$) and irreducible, and if the set of equilibria of (2.1) is discrete, and that all trajectories are bounded (this is the case if $A_i > 0$ and if each signal function f_i is bounded for $1 \leq i \leq n$), then the set Y of points $x \in \mathbb{R}^n$ for which the corresponding orbit does not converge to an equilibrium of (2.1) has Lebesgue measure 0, and $\bigcup_{e \in E} \text{int}(B(e))$ is open and dense in \mathbb{R}^n , where E is the set of equilibria and $B(e)$ is the basin of attraction of e . See the monograph of Wu [1999] for details.

3. IMPACT OF THE SIGNAL DELAYS

Time delays always occur in the signal transmission between neurons due to the finite propagation velocity of the electrical signals along axons, due to the finite speed of neurotransmitters across the synaptic gap and due to the finite switching speed of amplifiers (neurons) in the electronic hardware implementation (see Herz et al. [1989]).

For the sake of simplicity, we are going to consider a network of two identical neurons. Two-neuron networks have played an important role in the analysis of delay-induced changes in the dynamics of neural networks. Despite the small number of neurons in the system, in many instances, two-neuron networks with delay display the same behavior as larger networks and many techniques developed to deal with two-neuron networks can carry over to large size networks such as those used in practical applications.

The dynamics of the network of two identical neurons is describe by the following system of delay differential equations

$$\begin{cases} \dot{u}(t) = -\mu u(t) + f(v(t-\tau)), \\ \dot{v}(t) = -\mu v(t) + f(u(t-\tau)), \end{cases}$$

or equivalently (after rescaling the time variable)

$$\begin{cases} \dot{u}(t) = -\mu \tau u(t) + \tau f(v(t-1)), \\ \dot{v}(t) = -\mu \tau v(t) + \tau f(u(t-1)), \end{cases} \quad (3.1)$$

where τ and μ are given positive constants and $f : \mathbb{R} \rightarrow \mathbb{R}$ is a C^1 -smooth sigmoid signal function.

Note that the above model describes the case where the interaction of two neurons are excitatory. There are two other connection topologies. The mutually inhibitory interaction, where increasing the activity of one neuron tends to inhibits the activity of another, can be reduced to the excitatory interaction case after a simple change of variables. Another connection is significantly different. In this case, the feedback from one neuron (A) to another (B) is inhibitory, while the feedback from neuron B back to A is excitatory. If one looks this network as a loop, then the network's feedback from one neuron back to itself after the full loop, is negative. Such a network performs the computational task very much like a single neuron with negative self-feedback. In fact, Braptistini and Taboas [1996] studied the existence of periodic solutions due to the existence of delay, Chen and Wu [1999c] studied the stability and domain of attraction of this periodic orbit and Ruan and Wei [1999] studied the case where two different delays are involved (one for the self-feedback and another for the feedback). This is a special case of the so-called frustrated network introduced by Bélair, Campbell, and van den Driessche [1996], where they showed that the existence of a negative feedback loop is essential for a network to have a stable periodic solution, confirming the numerical observation in Marcus and Westervelt [1988, 1989].

We first consider the case where $f'(0) < \mu$. In this case, $(0,0)$ is the only equilibrium and every solution of the network is convergent to this equilibrium, no matter whether delay is considered or not and no matter how large this signal delay is. This is the special case for the so-called contractive network where the delay does not alter the asymptotic behavior of the system. See Gopalsamy and He [1999], Cao and Zhou [1998], and van den Driessche and Zou [1998] for some global asymptotic stability results for general additive models with delay. We should also mention that the global attractivity of a unique equilibrium is quite important in applications of neural networks to optimization problems.

We now consider the case where $f'(0) > \mu$. For the sake of comparison, let us first consider the case when the signal delay τ is ignored ($\tau = 0$), the dynamics of system (3.1) is very simple. It has three equilibria $E_0 = (0,0)$ and $E_{\pm} = (\xi^{\pm}, \xi^{\pm})$,

where ξ^+ and ξ^- are the nontrivial zeros of $\mu\xi = f(\xi)$, and every trajectory is convergent to $E_0 \cup E_- \cup E_+$. The unstable manifold $W^u(E_0)$ is important since the union $W^u(E_0) \cup E_- \cup E_+$ gives the global attractor. The stable manifold $W^s(E_0)$ is exactly the boundary of the basins of attraction of E_- and E_+ and thus, plays an important role in applications to content-addressable memory.

Introducing the delay τ does not change the equilibrium structure, the stability of the equilibria and the generic convergence. System (3.1) with positive τ generates a monotone semiflow on the phase space $C([-1, 0]; R^2)$ and thus, almost every trajectory is convergent to either E_+ or E_- (Smith [1987] and Hirsch [1988]). This represents the special case of almost (generic) convergence in cooperative irreducible networks, for which the delay has limited effect on the long-term behavior of the system. Recall that an irreducible network is one in which there is a directed path connecting any pair of neurons. It is referred to as cooperative when, possibly after a change of variables, all interactions are excitatory. Such systems, with or without delay, generate an eventually strongly monotone semiflow, so that the asymptotic behavior of the network with delay is essentially the same as that of the corresponding network without delay. See also Olien and Bélair [1997] and Wu [1999].

For system (3.1), the set of initial points whose trajectories do not converge to E_\pm is of codimension 1 (Pakdaman et al. [1998]) and in fact, is shown in Chen and Wu [1998] to be the graph of a Lipschitz map over a closed subspace of codimension 1. This is a separating surface in the sense that it is exactly the boundary of the basins of the attraction of E_- and E_+ . Interesting dynamics occurs only on this separating surface and thus, it is important to describe the dynamics of the semiflow restricted to this separating surface and to describe the global attractor.

Due to the symmetry of system (3.1), we can easily see that if the two components u and v of the initial value are identical, then these two components of the solution remain identical for all $t \geq 0$. Such a solution is said to be synchronous. The above convergence result for system (3.1) in case $\tau = 0$ shows that in the absence of delay, every trajectory is eventually synchronized since all equilibria are synchronous. Note also that the common component $u = v = x$ is completely described by the following scalar delay differential equation

$$\tau^{-1} \dot{x}(t) = -\mu x(t) + f(x(t-1)). \quad (3.2)$$

Describing the global attractor of equation (3.2) is the main subject of the recent monograph by Krisztin, Walther, and Wu [1999]. Namely, in the case where delay is ignored, we know that the global attractor of the system is nothing but the three zeros of $\mu x = f(x)$ and the connecting orbits. This is still the case when the delay

is small. A special case of the so-called harmless delay, and the problem of small harmless delay in neural networks was studied in the work of van den Driessche and Zou [1998]. However, when we increase the delay τ to pass a critical value τ_s , the dynamics of the network is changed significantly. In particular, it was shown that there exists a 3-dimensional manifold at the zero solution whose global forward extension under the semiflow of (3.2) has a nonempty intersection with the separating surface. The closure of such an intersection is a smooth disk borded by a periodic orbit and such that each orbit inside the disk is a heteroclinic orbit from the zero solution to the periodic orbit. The closure of the global forward extension of the above 3-dimensional manifold at the origin is geometrically a 3-dimensional smooth manifold with boundary (except the singularity at the two tips: the two nontrivial equilibria), and topologically a 3-dimensional solid ball, the upper half of the spindle consists of connecting orbits from either the periodic orbit or the zero equilibrium to the positive equilibrium, and the lower half of the spindle consists of connecting orbits from either the periodic orbit or the zero equilibrium to the negative equilibrium. The singularity of the spindle at the two tips was studied in Walther [1998] and it was shown in Krisztin and Walther [1999] that this spindle coincides with the global attractor if the delay is moderate.

The calculation of the critical value τ_s is related to the stability analysis of the zero solution of the scalar delay differential equation which is determined by the zero of the so-called characteristic equation

$$\lambda + \mu\tau - \tau f'(0) e^{-\lambda} = 0.$$

It is exactly at

$$\tau_s = \frac{2\pi - \arccos \frac{\mu}{f'(0)}}{\sqrt{[f'(0)]^2 - \mu^2}},$$

the characteristic equation has a pair of purely imaginary zeros and thus, a local Hopf bifurcation of periodic solutions occur. It should be mentioned that a detailed local Hopf bifurcation analysis for (3.2) is given in Giannakopoulos and Zapp [1998].

It is interesting to note that the characteristic equation for the full system (3.1) at zero solution can be decoupled as

$$[\lambda + \mu\tau - \tau f'(0) e^{-\lambda}] [\lambda + \mu\tau + \tau f'(0) e^{-\lambda}] = 0,$$

where the first factor corresponds to the characteristic equation for the scalar equation. It is surprising to note that at

$$\tau_d = \frac{\pi - \arccos \frac{\mu}{f'(0)}}{\sqrt{[f'(0)]^2 - \mu^2}},$$

the equation $\lambda + \mu\tau + \tau f'(0) e^{-\lambda} = 0$ has a pair of purely imaginary zeros and a Hopf bifurcation of periodic solutions occur near $\tau = \tau_d$. Obviously, these periodic solutions are not synchronized. In Chen and Wu [1998], we proved that this periodic orbit must be a phase-locked oscillation. Namely, each neuron oscillates like the other, but in different phase. The phase-difference is half of the period.

In Chen and Wu [1998], it was proved that when $\tau > \tau_d$, there exists a complete analogue of the aforementioned spindle for the full system, except that now the boarding periodic orbit of the smooth disk is phase-locked, not synchronized.

In general, it is known that the presence of long delays may render unstable an otherwise stable equilibrium point through a Hopf bifurcation, thus leading to stable periodic oscillations. The existence of phase-locked oscillation for (3.1) indicates the possibility of delay-induced desynchronization. In particular, in a general network of identical neurons and in the absence of synaptic delays, the asymptotic behaviors of the network can be characterized by the convergence to the set of synchronized equilibria. Delay, however, can cause stable phase-locked oscillations and other asynchronous behaviors. See also Wu [1998, 1999].

The existence of the two spindles for system (3.1) is then guaranteed for all $\tau > \tau_s$ and $\tau > \tau_d$. Surprisingly, we have $\tau_s > \tau_d$. So when $\tau > \tau_s$, the unstable space of the generator of the C_0 -semigroup generated by the linearization of (3.1) at the zero solution is at least 5-dimensional and there exists a 5-dimensional local unstable manifold $W_{5,loc}$. The global forward extension W_5 of $W_{5,loc}$ contains the aforementioned two spindles, and in particular, $\overline{W_5}$ contains a phase-locked periodic orbit and a synchronous periodic orbit. Chen, Krisztin, and Wu [1999] described completely the global dynamics of the semiflow of system (3.1) restricted to $\overline{W_5}$ and established the existence of heteroclinic orbits from the synchronous periodic orbit to the phase-locked periodic orbit, which suggests a mechanism for desynchronization of the network.

It should be emphasized that the above phase-locked oscillation and synchronized periodic orbit are all unstable. While unstable periodic orbits can be hardly observed in numerical simulations, the existence of these orbits may have significant impact on the transit behaviors of the network, and transit oscillations may last sufficiently long that it is hard to distinguish them from sustained stable oscillations. See Pakdaman et al. [1997] and Huang and Wu [1999a,b]. Examples of delay-induced transient oscillations were first reported in two mutually exciting neurons and in rings by Babcock and Westervelt [1987] and Baldi and Attyia [1994]. In Pakdaman, Grotta-Ragazzo, Malta, and Vibert [1997], it was observed that when delayed-induced transit oscillations appear, the duration of the transients,

that is, the time required for the system to stabilize at an equilibrium, increases as the characteristic charge-discharge time of the neurons tends to zero, or equivalently, as the delay is increased.

Limiting profiles of the synchronous periodic orbit and the phase-locked periodic solution have been described in Chen and Wu [1999b], pulses and square waves are both observed. This is related to the discussion of delay-induced change of the domain of attraction, and in general, Pakadaman et al. [1998] has indicated that even for a cooperative irreducible system whose quasiconvergence is guaranteed, delay may affect the shape and structure of the basins of attraction of equilibria.

As shown above, signal delays can change the stability of equilibria, causing nonlinear oscillations and inducing periodic solutions. Increasing the delay is among many mechanisms to create a network that exhibits periodic oscillations. While the existence of periodic solutions creates some problems in neural network applications such as CAM, periodic sequences of neural impulses are of fundamental importance for the control of motor body functions (see Müller and Reinhart [1991] and Smith [1991]) such as the heartbeat, which occurs with great regularity almost three billion times during an average person's life. It should be emphasized that delay-induced oscillation need not be periodic, and even a two-neuron network can provide a tractable prototype for the study of delay-induced chaos in neural networks (Gilli [1993]).

REFERENCES

- [1] Babcock, K.L. and Westervelt, R.M., Dynamics of simple electronic neural networks, *Physica D*, 28, 1987, 305-316.
- [2] Baldi, P. and Attyia, A.F., How delays effect neural dynamics and learning, *IEEE Trans. Neural Networks*, 5, 1994, 612-621.
- [3] Baptistini, M.Z. and Táboas, P.Z., On the existence and global bifurcation of periodic solutions to planar differential delay equations, *J. Differential Equations*, 127, 1996, 391-425.
- [4] Bélair, J., Campbell, S.A., and van den Driessche, P., Frustration, stability and delay-induced oscillations in a neural network model, *SIAM J. Appl. Math.*, 46, 1996, 245-255.
- [5] Cao, J. and Zhou, D., Stability analysis of delayed cellular neural networks, *Neural Networks*, 11, 1998, 1601-1605.
- [6] Chen, Y., Krisztin, T., and Wu, J., Connecting orbits from synchronous periodic solutions to phase-locked periodic solutions in a delay differential system, *J. Differential Equations*, 1999, to appear.

- [7] Chen, Y. and Wu, J., Existence and attraction of a phase-locked oscillation in a delayed network of two neurons, *Advances in Differential Equations*, 1998, to appear.
- [8] Chen, Y. and Wu, J., Minimal instability and unstable set of a phase-locked periodic orbit in a delayed neural network, *Physics D*, 134, 1999a, 185-199.
- [9] Chen, Y. and Wu, J., The asymptotic shapes of periodic solutions of a singular delay differential system, *J. Differential Equations*, 1999b, to appear.
- [10] Chen, Y. and Wu, J., Slowly oscillating periodic solutions in a delayed frustrated network of two neurons, 1999c, preprint.
- [11] Chua, L.O. and Yang, L., Cellular neural networks: Theory, *IEEE Trans. Circuits Syst. I*, 35, 1988, 1257-1272.
- [12] Cohen, M.A. and Grossberg, S., Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 1983, 815-826.
- [13] Fiedler, B. and Gedeon, T., A class of convergence neural network dynamics, *Physica D*, 111, 1998, 288-294.
- [14] Gedeon, T., Structure and dynamics of artificial neural networks, in *Fields Institute Communications: Differential Equations with Applications to Biology*, (S. Ruan, G. Wolkowicz, and J. Wu, eds.), Vol. 21, 1999, 217-224.
- [15] Giannakopoulos, F. and Zapp, A., Local and global Hopf bifurcation in a scalar delay differential equation, 1998, preprint.
- [16] Gilli, Strange attractors in delayed cellular neural networks, *IEEE Trans. Circuits Syst.*, 40, 1993, 849-853.
- [17] Gopalsamy, K. and He, X.-Z., Stability in asymmetric Hopfield nets with transmission delays, *Physica D*, 76, 1994, 344-358.
- [18] Hebb, D.O., *The Organization of Behavior*, John Wiley & Sons, New York, 1949.
- [19] Herz, A.V.M., Salzer, B., Kühn, R., and van Hemmen, J.L., Hebbian learning reconsidered: Representation of static and dynamic objects in associative neural nets, *Biol. Cybern.*, 60, 1989, 457-467.
- [20] Hirsch, M., Stability and convergence in strongly monotone dynamical systems, *J. Reine Angew. Math.*, 383, 1988, 1-53.
- [21] Hopfield, J.J., Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci.*, 81, 1984, 3088-3092.
- [22] Huang, L. and Wu, J., Joint effects of the threshold and synaptic delay on dynamics of artificial neural networks with McCulloch-Pitts nonlinearity, 1999a, preprint.

- [23] Huang, L. and Wu, J., The role of threshold in preventing delay-induced oscillations of frustrated neural networks with McCulloch-Pitts nonlinearity, 1999b, preprint.
- [24] Krisztin, T. and Walther, H.-O., Unique periodic orbits for delayed positive feedback and the global attractor, *J. Dynamics and Differential Equations*, 1999, to appear.
- [25] Krisztin, T., Walther, H.-O., and Wu, J., *Shape, Smoothness and Invariant Stratification of an Attracting Set for Delayed Monotone Positive Feedback*, the Fields Institute Monographs Series 11, Amer. Math. Soc., Rhode Island, 1999.
- [26] Levine, D.S., *Introduction to Neural and Cognitive Modelling*, Lawrence Erlbaum Associate, Inc., New Jersey, 1991.
- [27] Marcus, C.M. and Westervelt, R.M., Basins of attraction for electronic neural networks, in *Neural Information Processing Systems*, (D.Z. Anderson, ed.) American Institute of Physics, New York, 1988, 524-533.
- [28] Marcus, C.M. and Westervelt, R.M., Stability of analog neural networks with delay, *Physical Review A*, 39, 1989, 347-359.
- [29] May, R.M. and Leonard, W.J., Nonlinear aspects of competition between three species, *SIAM J. Appl. Math.*, 29, 1975, 243-253.
- [30] Müller, B. and Reinhardt, J., *Neural Networks, An Introduction*, Springer-Verlag, Berlin, 1991.
- [31] Olien, L. and Bélair, J., Bifurcation's, stability, and monotonically properties of a delayed neural network model, *Physica D*, 102, 1997, 349-363.
- [32] Pakdaman, K., Grotta-Ragazzo, C.P., Malta, K., and Vibert, J.-F., Delay-induced transient oscillations in a two-neuron network, *Resends*, 1997, 45-54.
- [33] Pakdaman, K., Grotta-Ragazzo, C.P., Malta, C.P., Rain, O., and Vibert, J.-F., Effect of delay on the boundary of the basin of attraction in a system of two neurons, *Neural Networks*, 11, 1998, 509-519.
- [34] Ruan, S. and Wei, J., Periodic solutions of planar systems with two delays, *Proc. Royal Soc. Edinburgh (A)*, 129, 1999, 1017-1032.
- [35] Smith, H.L., Monotone semiflows generated by functional differential equations, *J. Differential Equations*, 66, 1987, 420-442.
- [36] Smith, H.L., Convergent and oscillatory activation dynamics for cascades of neural nets with nearest neighbor competitive or cooperative interactions, *Neural Networks*, 4, 1991, 41-46.
- [37] van den Driessche, P. and Zou, X.F., Global attractivity in delayed Hopfield neural networks models, *SIAM J. Appl. Math.*, 58, 1998, 1878-1890.
- [38] Walther, H.-O., The singularities of an attractor of a delay differential equation, *Functional Differential Equations*, 5, 1998, 513-548.

- [39] Wu, J., Symmetric functional differential equations and neural networks with memory, *Trans. Amer. Math. Soc.*, 350, 1998, 4799-4838.
- [40] Wu, J., Introduction to Neural Dynamics and Signal Delays, 1999, preprint.

29 INVARIANT SETS AND GLOBAL ATTRACTOR OF A CLASS OF PARTIAL DIFFERENTIAL EQUATIONS

Daoyi Xu

Institute of Mathematics

Sichuan University

Chengdu 610064, China

and

Qingyi Guo

Kangding Teacher's College

Sichuan 626001, China

In recent years, the invariant and attractor of dynamical system have obtained considerable attention because in nonlinear system we don't know whether the equilibrium point exists. Various interesting results on this problem have been report [1-5]. In this paper, we will study invariant sets and attractor of reaction-diffusion equations, and give the particular method on determining invariant sets and attractor by virtue of the properties of operator semigroup[6], nonnegative matrices [7], and differential inequality technique [8].

1. PRELIMINARY

Let H is a metric space. A family of operator $T(t)$, $t \geq 0$ is called to be a operator semigroup if they map H into itself and enjoy the usual semigroup properties

$$T(t+s) = T(t) T(s), \quad \forall t, s \geq 0; \quad T(0) = I \text{ (Identity in } H).$$

We say that a set $X \subset H$ is positively invariant for the semigroup $T(t)$ if

$$T(x) X \subset X, \quad \forall t > 0$$

and the set X is invariant for the semigroup $T(t)$ if $T(t)X = X, \quad \forall t > 0$.

Let \mathcal{B} be a subset of H and \mathcal{U} an open set containing \mathcal{B} . We say that \mathcal{B} is absorbing in \mathcal{U} if the orbit of any bounded set of \mathcal{U} enters into \mathcal{B} after a certain time (which may depend on the set):

$$\begin{cases} \forall \mathcal{B}_0 \subset U, \quad \mathcal{B}_0 \text{ bounded} \\ \exists t_1(\mathcal{B}_0) \text{ such that } T(t)\mathcal{B}_0 \subset \mathcal{B}, \quad \forall t \geq t_1(\mathcal{B}_0) \end{cases} \quad (1.1)$$

Definition 1. A set $A \subset H$ is said to attract a set $U \subset H$ under $T(t)$ if for every $u_0 \in U$, $\text{dist}(T(t)u_0, A) \rightarrow 0$ as $t \rightarrow \infty$. The set A is said to attract uniformly the set U if for any given $\varepsilon > 0$, there is a $b = b(\varepsilon, U) > 0$ such that

$$\text{dist}(T(t)U, A) < \varepsilon, \quad \forall t \geq t_0 + b. \quad (1.2)$$

The set A is called to be global attracting set if it attracts each bounded set in H .

Definition 2. An invariant set A is said to be a global attractor if A is a maximal compact invariant and global attracting set.

Definition 3. The operators $T(t)$ are called to be uniformly compact for t large if for every bounded set \mathcal{B} there exists t_0 which may depend on \mathcal{B} such that

$$\bigcup_{t \geq t_0} T(t)\mathcal{B} \quad (1.3)$$

is relatively compact in H .

Lemma 1. [1] We assume that H is a metric space and the semigroup $T(t)$ is continuous and uniformly compact for t large. We also assume that there exists a bounded set \mathcal{B} such that \mathcal{B} is absorbing in H . Then there is a global attractor \mathcal{R} for the semigroup $\{T(t)\}_{t \geq 0}$. Furthermore, if H is a Banach space, then \mathcal{R} is connected too.

Let Ω be a bounded domain in R^n with smooth boundary $\partial\Omega$. $C(X, Y)$ denotes the class of continuous mapping from a Banach space X to a Banach space Y . $L^2(\Omega)$ is the space of real Lebesgue measurable functions on Ω . It is a Banach space for the norm

$$\|u\| = \left[\int_{\Omega} |u(x)|^2 dx \right]^{1/2},$$

where $|u|$ denotes the Euclid norm of a vector $u \in R^p$ for any integer p .

Let $H^1(\Omega) = \{u \in L^2(\Omega), \nabla u \in L^2(\Omega), \nabla \text{ is the gradient operator}\}$, H_0^1 is the closure of $C_0^\infty(\Omega)$ in $H^1(\Omega)$, where $C_0^\infty(\Omega)$ is the space of real C^∞ functions on Ω with a compact support in Ω . H_0^1 is also a Banach space for the norm

$$\|u\| = \left[\int_{\Omega} |\nabla u(x)|^2 dx \right]^{1/2}.$$

For $u \in R^m$, we define $[u]^+ = \text{Col}\{u^i\}$, and for $u \in L^m \triangleq \{L^2(\Omega)\}^m$ and for $u \in H^m \triangleq \{H_0^1(\Omega)\}^m$, we define $[u]^L = \text{Col}\{\|u^i\|\}$ and $[u]^V = \text{Col}\{\|u^i\|\}$, respectively.

$A \geq B$ ($A < B$) means that each pair of corresponding elements of A and B satisfies the inequality “ \geq ” (“ $<$ ”), especially, A is called a nonnegative matrix if $A \geq 0$.

Lemma 2. [7]

If $M \geq 0$ and $\rho(M) < 1$ then $(I - M)^{-1} \geq 0$.

The symbol $\rho(M)$ denotes the spectral radius of a square matrix M .

2. MAIN RESULTS

Consider semilinear reaction diffusion equations with delay, for $i = 1, \dots, m$,

$$\begin{cases} \frac{\partial u^i(t, x)}{\partial t} = a_i \Delta u^i(t, x) - b_i u^i(t, x) + f_i(u(x)), & t \geq t_0, \quad x \in \Omega, \\ u^i(t, x) = 0, & t > t_0, \quad x \in \partial \Omega, \\ u^i(x, t_0) = \phi_i(x), & x \in \Omega, \end{cases} \quad (2.1)$$

where $a_i > 0$, $b_i \geq 0$ are constants and if $b_i = 0$, we agree that no boundary condition applies to u^i , Δ is the Laplace operator, the initial data $\phi = \text{Col}\{\phi_i\} \in C$ is a given function. $f = \text{Col}\{f_i\} \in C(L^m, R^m)$ is globally Lipschitz uniformly in $x \in \Omega$.

Under the above conditions, the assumptions of existence and uniqueness results (2.1) are then satisfied. This is however sufficient to define the semigroup $T(t)$: we set

$$T(t) : \phi \in L^m \rightarrow u \in L^m.$$

In the following, we always suppose that:

$$(H1) \quad [f(u(x))]^+ \leq B[u]^+ + P, \quad B = (b_{ij})_{m \times m}, \quad P = (p_1, \dots, p_m)^T,$$

$$(H2) \quad \rho(M) < 1, \quad M = (m_{ij})_{m \times m}, \quad m_{ij} = \frac{b_{ij}}{\Pi_i}, \quad \frac{a_i}{\gamma_i} + b_i = \Pi_i,$$

where $\gamma_i = \gamma_i(\Omega)$ is a constant determined by Poincare inequality (see [7]). $\gamma_i = h/\sqrt{m}$ if $\Omega = \{x \in R^m \mid \|x_i\| < h\}$.

Theorem 1. Let $|\Omega|$ is the measure of Ω and $Q = [q_1, \dots, q_m]^T$, $q_i = p_i \sqrt{(\Omega)} / \eta_i$. If (H1) and (H2) hold, then the set $S_\alpha = \left\{ \phi \in L^m \mid [\phi]^L \leq \alpha K = \alpha(I - M)^{-1}Q, \alpha \geq 1 \right\}$, which is called a pseudo-rectangle, is a positively invariant for the semigroup $T(t)$ associated to the system (2.1).

Proof: After multiplying (1.1) by $u^i(t, x)$, by integration of (1.1), we find

$$\begin{aligned} \frac{d}{dt} \int_{\Omega} [u^i(t, x)]^2 dx &= a_i \int_{\Omega} u^i(t, x) \Delta u^i(t, x) dx - b_i \int_{\Omega} [u^i(t, x)]^2 dx \\ &\quad + \int_{\Omega} u^i(t, x) f_i(u(x)) dx, \quad t \geq t_0. \end{aligned} \quad (2.2)$$

By the Green formula

$$\int_{\Omega} u^i(t, x) \Delta u^i(t, x) dx = - \int_{\Omega} [\nabla u^i(t, x)]^2 dx. \quad (2.3)$$

We infer from the Poincare inequality the existence of a constant γ_i such that

$$\int_{\Omega} [u^i(t, x)]^2 dx \leq \gamma_i \int_{\Omega} [\nabla u^i(t, x)]^2 dx. \quad (2.4)$$

Then we deduce from (2.2), (2.3), condition (H1) and Hölder inequality

$$\frac{d}{dt} \|u^i\|^2 \leq -\frac{a_i}{\gamma_i} \|u^i\|^2 - b_i \|u^i\|^2 + \|u^i\| \left[\sum_{j=1}^m b_{ij} \|u^j\| + p_i \sqrt{(\Omega)} \right]. \quad (2.5)$$

From $\frac{a_i}{\gamma_i} + b_i = \eta_i$, we obtain

$$\|u^i\|^2 \leq e^{-\eta_i(t-t_0)} \|\phi_i\|^2 + \int_{t_0}^t e^{-\eta_i(t-s)} \eta_i \|u^i\| \left[\sum_{j=1}^m m_{ij} \|u^j\| + q_i \right] ds. \quad (2.6)$$

Without loss of generality, we assume that $P > 0$, i.e., $Q > 0$. Since $\rho(M) < 1$, $(I - M)^{-1} \geq 0$ and $K = (I - M)^{-1}Q > 0$. We now prove that, when $[\phi]^L < \alpha K$

$$[u(t)]^L < \alpha K \quad \text{for } t \geq t_0. \quad (2.7)$$

If (2.7) is not so, there must be some I , and $t_1 > t_0$, such that

$$\|u^i(t_1)\| = \alpha k_i, \quad \|u^i(t)\| < \alpha k_i, \quad \text{for } t < t_1, \quad (2.8)$$

and

$$[u(t)]^L \leq \alpha K, \quad \text{for } t_0 \leq t \leq t_1, \quad (2.9)$$

where k_i is the I -th component of vector K . Then

$$\|u^i(t_1)\|^2 < e^{-\eta_i(t_1-t_0)} \alpha^2 k_i^2 + \int_{t_0}^{t_1} e^{-\eta_i(t_1-s)} \eta_i \alpha k_i \left[\sum_{j=1}^m m_{ij} \alpha k_j + q_i \right] ds.$$

Let $\eta = [\eta_1, \dots, \eta_m]^T$ and $\text{diag}\{k_i\}$ (or $\text{diag}K$) is a diagonal matrix with diagonal entries k_i . Then

$$\begin{aligned} \text{Col}\left\{\|u^i(t_1)\|^2\right\} &\leq e^{-\eta(t_1-t_0)} \alpha^2 \text{diag}\{k_i\} K \\ &+ \int_{t_0}^{t_1} \text{diag}\left\{e^{-\eta(t_1-s)} \eta_i\right\} \alpha \text{diag}\{k_i\} [\alpha M K + Q] ds. \end{aligned}$$

Noting that $K = (I - M)^{-1}Q$, i.e., $MK + Q = K$, we have

$$\begin{aligned} \text{Col}\left\{\|u^i(t_1)\|^2\right\} &< e^{-\eta(t_1-t_0)} \alpha^2 \text{diag}\{k_i\} (MK + Q) \\ &+ (I - e^{-\eta(t_1-t_0)}) \alpha \text{diag}\{k_i\} [\alpha MK + Q] \\ &= e^{-\eta(t_1-t_0)} \alpha (\alpha - 1) \text{diag}\{k_i\} Q + \alpha \text{diag}\{k_i\} [\alpha MK + Q] \quad (2.10) \\ &\leq \alpha (\alpha - 1) \text{diag}\{k_i\} Q + \alpha \text{diag}\{k_i\} [\alpha MK + Q] \\ &= \alpha^2 \text{diag}\{k_i\} (MK + Q) \\ &= \alpha^2 \text{diag}\{k_i\} K = \alpha^2 \text{Col}\{k_i^2\}. \end{aligned}$$

This implies that $\|u^i(t_1)\| < \alpha k_i$, which contradicts the equality in (2.8), and so (2.7) holds.

For any given $\phi \in L^m$, there is a α such that $\phi \in S_\alpha$ and the solution of (1.1) satisfies $\|u_i(t)\| < \alpha k_i$. This leads to the following corollary.

Corollary. All solutions of (1.1) are uniformly bounded in $L^2(\Omega)$.

Theorem 2. If (H1) and (H2) hold, then the pseudo-rectangle $S = \{\phi \in L^m \mid [\phi]^L \leq K = (I - M)^{-1}Q\}$ is a global attracting set for the semigroup $T(t)$ associated to the system (2.1).

Proof. For any $\phi \in L^m$, there is a pseudo-rectangle S_α such that $\phi \in S_\alpha$. From Theorem 1, the solution $u(t, x)$ satisfies

$$[u(t)]^L \leq \alpha K. \quad (2.11)$$

We will prove $\lim_{t \rightarrow +\infty} \sup[u(t)]^L \leq K$. If it is not true, there exist a constant vector $\sigma \geq 0$ and $\sigma \neq 0$ such that

$$\lim_{t \rightarrow +\infty} \sup[u(t)]^L = \sigma. \quad (2.12)$$

According to definition of superior limit and (2.12), for sufficient small positive constant $\varepsilon < \sigma$, there is $t_2 > t_0$, such that, for any $t \geq t_2$,

$$[u]^L \leq (\sigma + \varepsilon) < 2\sigma, \quad E = [1, \dots, 1]^T. \quad (2.13)$$

Since $\eta > 0$, for the above ε and K , there must be $T > 0$ such that for $t \geq T$,

$$\alpha^2 e^{-\eta(t-t_0)} \operatorname{diag} KK + \int_T^\infty \operatorname{diag} \left\{ e^{-\eta s} \eta_i \right\} \operatorname{diag} \{ \alpha K \} (M(\alpha K) + Q) ds \leq E\varepsilon.$$

So, when $t \geq t_2 + T$,

$$\begin{aligned} \operatorname{Col} \left\{ \|u^i\|^2 \right\} &\leq e^{-\eta(t-t_0)} \operatorname{Col} \left\{ \|\phi_i\|^2 \right\} \\ &\quad + \int_{t_0}^t \operatorname{diag} \left\{ e^{-\eta_i(t-s)} \eta_i \right\} \operatorname{diag} \left\{ \|u^i\| \right\} (M[u(s)]_r^L + Q) ds \\ &= e^{-\eta(t-t_0)} \operatorname{Col} \left\{ \|\phi_i\|^2 \right\} + \left\{ \int_{t_0}^{t-T} \right. \\ &\quad \left. + \int_{t-T}^t \right\} \operatorname{diag} \left\{ e^{-\eta_i(t-s)} \eta_i \right\} \operatorname{diag} \left\{ \|u^i\| \right\} (M[u(s)]_r^L + Q) ds \\ &\leq e^{-\eta(t-t_0)} \operatorname{Col} \left\{ \alpha^2 k_i^2 \right\} \\ &\quad + \int_T^\infty \operatorname{diag} \left\{ e^{-\eta s} \eta_i \right\} \operatorname{diag} \{ \alpha K \} (M(\alpha K) + Q) ds \\ &\quad + \int_{t-T}^t \operatorname{diag} \left\{ e^{-\eta_i(t-s)} \eta_i \right\} \operatorname{diag} \{ \sigma \} \\ &\quad + E\varepsilon \} (M(\sigma + E\varepsilon) + Q) ds \\ &\leq E\varepsilon + (I - e^{-\eta T}) \operatorname{diag} \{ \sigma + E\varepsilon \} (M(\sigma + E\varepsilon) + Q) \\ &\leq E\varepsilon + \operatorname{diag} \{ \sigma + E\varepsilon \} (M(\sigma + E\varepsilon) + Q). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$, we have

$$\operatorname{Col} \left\{ \|u^i\|^2 \right\} \leq \operatorname{diag} \sigma (Q + M\sigma).$$

Combining (2.12), we obtain

$$\operatorname{diag} \sigma \sigma \leq \operatorname{diag} \sigma (M\sigma + Q),$$

that is, $\sigma \leq M\sigma + Q$ or $\sigma \leq (I - M)^{-1}Q$. Hence, (1.2) holds, and the proof is completed.

We now prove the existence of an absorbing set in $H_0^1(\Omega)$. For that purpose, we give the following condition

$$(H3) \quad \rho(N) < 1, \quad N = (n_{ij})_{m \times m}, \quad n_{ij} = \frac{m\gamma_i^2 b_{ij}^2}{2a_i b_i},$$

γ_i is given in (22). Especially, $n_{ij} = \frac{h^2 b_{ij}^2}{2a_i b_i}$ if $\Omega = \{x \in R^m \mid \|x_i\| < h\}$.

Theorem 3. Let $\hat{P} = [\hat{p}_1, \dots, \hat{p}_m]^T$, $\hat{p}_i = \frac{p_i^2 \mathcal{M}(\Omega)}{2a_i b_i}$. If (H1) and (H3) hold, we have

the pseudo-rectangle $\Gamma_\alpha = \left\{ \phi \in H^m \mid \text{Col} \left\{ \|\phi_i\|^2 \right\} \leq \alpha \hat{K} = \alpha (I - N)^{-1} \hat{P}, \alpha \geq 1 \right\}$, is a positively invariant for the semigroup $S(t)$ associated to the system (2.1).

Proof. In order to prove this theorem, we multiply equation (2.1) by $-\Delta u^i$ and integrate over Ω .

We have, using the boundary condition of (2.1) and the Green formula,

$$-\int_{\Omega} \Delta u^i \frac{\partial u^i}{\partial t} dx = \sum_{j=1}^n \int_{\Omega} \frac{\partial u^i}{\partial x_j} \frac{\partial (u^i)^2}{\partial x_j \partial t} dx = \frac{1}{2} \frac{d}{dt} \|\|u^i\|^2\|, \quad (2.14)$$

$$-\int_{\Omega} u^i \Delta u^i dx = \int_{\Omega} |\nabla u^i|^2 dx = \|\|u^i\|^2\|. \quad (2.15)$$

We then obtain

$$\frac{1}{2} \frac{d}{dt} \|\|u^i\|^2\| = -a_i \|\Delta u^i\|^2 + \int_{\Omega} b_i u^i \Delta u^i dx - \int_{\Omega} \Delta u^i f_i(u) dx. \quad (2.16)$$

Then we deduce from (2.15) and (2.16)

$$\frac{d}{dt} \|\|u^i\|^2\| \leq -2a_i \|\Delta u^i\|^2 - 2b_i \|\|u^i\|^2\| + 2 \int_{\Omega} \|\Delta u^i\| f_i(u(x)) dx. \quad (2.17)$$

By using Hölder inequality and condition (H1), we have

$$\begin{aligned} \frac{d}{dt} \|\|u^i\|^2\| &\leq -2a_i \|\Delta u^i\|^2 - 2b_i \|\|u^i\|^2\| + 2 \|\Delta u^i\| \left[\sum_{j=1}^m b_{ij} \|\|u^j\|\| + p_i \sqrt{\mathcal{M}(\Omega)} \right] \\ &= -2b_i \|\|u^i\|^2\| + \sum_{j=1}^m \left(-\frac{a_i}{m} \|\Delta u^i\|^2 + 2b_{ij} \|\Delta u^i\| \|\|u^j\|\| \right) \\ &\quad + \left(-a_i \|\Delta u^i\|^2 + 2 \|\Delta u^i\| p_i \right) \\ &\leq -2b_i \|\|u^i\|^2\| + \sum_{j=1}^m \frac{m}{a_i} b_{ij}^2 \|\|u^j\|\|^2 + \frac{p_i^2 \mathcal{M}(\Omega)}{a_i}. \end{aligned} \quad (2.18)$$

We infer from the Poincaré inequality (2.4), that is,

$$\|\|u^i\|\| \leq \gamma_i \|\|u^i\|\| \quad \forall u^i \in H_0^1(\Omega). \quad (2.19)$$

Thus,

$$\frac{d}{dt} \|\|u^i\|^2\| \leq -2b_i \|\|u^i\|^2\| + \sum_{j=1}^m \frac{m \gamma_i^2}{a_i} b_{ij}^2 \|\|u^j\|\|^2 + \frac{p_i^2 \mathcal{M}(\Omega)}{a_i}. \quad (2.20)$$

We now prove that, when $\text{Col} \left\{ \|\phi_i\|^2 \right\} \leq \alpha \hat{K}$

$$\text{Col} \left\{ \|\|u^i(t)\|\| \right\} \leq \alpha \hat{K} \quad \text{for } t \geq t_0. \quad (2.21)$$

If (2.21) is not so, there must be some i and $t_1 > t_0$, such that

$$\|u^i(t_1)\|^2 = \alpha \hat{k}_i, \quad \|u^i(t)\|^2 < \alpha \hat{k}_i, \quad \text{for } t < t_1, \quad (2.22)$$

and

$$\text{Col}\left\{\|u^i(t)\|^2\right\} \leq \alpha \hat{K}, \quad \text{for } t_0 \leq t \leq t_1, \quad (2.23)$$

where k_i is the i th component of vector \hat{K} . From (2.20),

$$\|u^i(t)\|^2 \leq e^{-2b_i(t-t_0)} \|\phi_i\|^2 + \int_{t_0}^t e^{-2b_i(t-s)} 2b_i \sum_{j=1}^m [n_{ij} \|u^j\|^2 + \hat{p}_i] ds.$$

Letting $b = \text{diag}\{\phi_i\}$ and noting that $\hat{K} = (I - N)^{-1} \hat{P}$, i.e., $N \hat{K} + \hat{P} = \hat{K}$, we have

$$\begin{aligned} \text{Col}\left\{\|u^i\|^2\right\} &< e^{-2b(t-t_0)} \alpha \hat{K} + (I - e^{-2b(t-t_0)}) [N \alpha \hat{K} + \hat{P}] \\ &= e^{-2b(t-t_0)} (\alpha - 1) \hat{P} + N \alpha \hat{K} + \hat{P} \\ &\leq (\alpha - 1) \hat{P} + N \alpha \hat{K} + \hat{P} = \alpha (N \hat{K} + \hat{P}) = \alpha \hat{K}. \end{aligned} \quad (2.24)$$

This implies that $\|u^i(t)\|^2 < \alpha \hat{k}_i$, which contradicts the equality in (2.22), and so (2.21) holds.

Theorem 4. If (H1) and (H3) hold, then the pseudo-rectangle $\Gamma = \left\{ \phi \in H^m \mid \text{Col}\left\{\|\phi_i\|^2\right\} \leq \hat{K} = (I - N)^{-1} \hat{P} \right\}$, uniformly attracts each bounded set $\mathcal{B} \subset H^m$ under $T(t)$. The semigroup $T(t)$ is associated to the system (2.1).

Proof. For any bounded set $\mathcal{B} \subset H^m$, there is an $\alpha \geq 1$ such that $\mathcal{B} \subset \Gamma_\alpha$. From Theorem 3, for any $\phi \in \mathcal{B}$, the solution $u(t, x)$ satisfies

$$\|u^i(t)\|^2 \leq \alpha \hat{k}_i.$$

For any given $\varepsilon > 0$, we choose T_1 such that for $t \geq T_1$

$$e^{-2b(t-t_0)} \text{Col}\left\{\|\phi_i\|^2\right\} + e^{-2bT_1} \alpha N \hat{K} + \hat{P} \leq \beta \text{Col}\{1\}, \quad (2.25)$$

where $\beta = \varepsilon/4 \|(I - N)^{-1} \text{Col}\{1\}\|$. Then, from (2.20), we have

$$\begin{aligned} \|u^i(t)\|^2 &\leq e^{-2b_i(t-t_0)} \|\phi_i\|^2 \\ &\quad + \left\{ \int_{t_0}^{t-T_1} + \int_{t-T_1}^t \right\} e^{-2b_i(t-s)} 2b_i \left[\sum_{j=1}^m n_{ij} \|u^j\|^2 + \hat{p}_i \right] ds \\ &\leq \int_{t-T_1}^t e^{-2b_i(t-s)} 2b_i \left[\sum_{j=1}^m n_{ij} \|u^j\|^2 + \hat{p}_i \right] ds + \beta. \end{aligned} \quad (2.26)$$

Let $T_2 = \max\{r, T_1\}$ and

$$t_k = t_0 + k T_2, \quad k = 0, 1, \dots$$

Since $u(t, x)$ is uniformly bounded in H^1 , we choose $\pi_{ik} = \sup_{t_k < \theta < \infty} \left\{ \|u^i(t)\|^2 \right\}$.

Then $\pi_{ik} \leq \pi_{ik-1} (\forall k = 1, 2, \dots)$ and there are $\bar{t}_{ik} \geq t_k$ (or $\bar{t}_{ik} = \infty$) such that $\pi_{ik} = \|u^i(\bar{t}_{ik})\|$. Thus (2.26) leads to

$$\begin{aligned}\pi_{ik} &\leq \int_{\bar{t}_{ik}-T_1}^{\bar{t}_{ik}} e^{-2b_i(\bar{t}_{ik}-s)} 2b_i \sum_{j=1}^m [n_{ij}\pi_{ik-2} + \hat{p}_i] ds + \beta \\ &\leq \sum_{j=1}^m [n_{ij}\pi_{ik-2} + \hat{p}_i] ds + \beta.\end{aligned}$$

Let $\Pi_k = \text{Col}\{\pi_{ik}\}$, we obtain

$$\begin{aligned}\Pi_{2k} &\leq N\Pi_{2k-2} + \beta\text{Col}\{1\} + \hat{P} \\ &\leq N^k\Pi_0 + (N^{k-1} + N^{k-2} + \dots + I)(\beta\text{Col}\{1\} + \hat{P}) \\ &= N^k\Pi_0 + (I - N)^{-1}(I - N^k)(\beta\text{Col}\{1\} + \hat{P}).\end{aligned}\quad (2.27)$$

By $\rho(N) < 1$, we have $N^k \rightarrow 0$ as $k \rightarrow \infty$ so that for the above ε there is a \bar{k} such that

$$|N^k| \leq \min \left\{ 1, \varepsilon/4 |\Pi_0|, \beta/\|\hat{P}\|_E \right\},$$

where $\|\hat{P}\|_E$ denotes the Euclid norm of matrix \hat{P} . Combining (2.27), we obtain

$$|\Pi_{2k} - (I - N)^{-1}\hat{P}| \leq \varepsilon/4 + 2|(I - N)^{-1}|\|\text{Col}\{1\}\|\beta + |(I - N)^{-1}|\beta \leq \varepsilon, \quad \forall k \geq \bar{k}.$$

Taking $b = t_0 + 2\bar{k}T_2$, we obtain

$$[u(t)]^\top < \hat{K} + E\varepsilon, \quad \forall t \geq t_0 + b.$$

This implies (1.2) and the proof is completed.

Theorem 5. We assume that the hypotheses (H1), (H2), and (H3) are satisfied. Then the semigroup $T(t)$ associated to the system (2.1) possesses a global attractor A which is bounded in H^m , compact and connected in L^m ; \mathcal{A} attracts the bounded sets of $L^2(\Omega)$.

Proof. We shall prove that the operators $S(t)$ are uniformly compact.

For any given $\phi \in L^m$, by Theorem 1, there is an α such that $\phi \in S_\alpha$, i.e., $\|u^i(t)\| \leq \alpha^2 k_i^2$ for all $t \geq t_0$. Then from (2.3) and (2.4), we obtain

$$\begin{aligned}\frac{d}{dt} \|u^i(t)\|^2 &\leq -a_i \|u^i\|^2 - b_i \|u^i\|^2 + \|u^i\| \sum_{j=1}^m [b_{ij} \|u^j\| + p_i \sqrt{|\Omega|}] \\ &\leq -a_i \|u^i\|^2 + \beta \quad \forall t \geq t_0,\end{aligned}\quad (2.28)$$

where $\beta = \beta(\alpha K) = b_i \alpha^2 k_i^2 + \alpha k_i \sum_{j=1}^m [b_{ij} \alpha k_j + p_i \sqrt{|\Omega|}]$.

For $h > 0$ fixed, we integrate (2.28) between t and $t+h$ and obtain

$$\int_t^{t+h} \|u^i(s)\|^2 ds \leq \frac{h\beta}{a_i} + \frac{1}{a_i} \|u^i(t)\|^2 \leq \frac{h\beta}{a_i} + \frac{\alpha^2 k_i^2}{a_i}. \quad (2.29)$$

On the other hand, from Theorem 1 and (2.18), we have

$$\frac{d}{dt} \|u^i(t)\|^2 \leq \sum_{j=1}^m \frac{m}{a_i} b_{ij}^2 \alpha^2 k_j^2 + \frac{p_i^2 |\Omega|}{a_i} \triangleq \xi_i. \quad (2.30)$$

We multiply (2.30) by t and obtain

$$\frac{d}{dt} \left(t \|u^i\|^2 \right) \leq \xi_i t + \|u^i\|^2. \quad (2.31)$$

By integration between 0 and $h > r$ and using (2.29)

$$\|u^i(h)\|^2 \leq \frac{1}{2} \xi_i h + \frac{1}{h} \int_0^h \|u^i\|^2 ds \leq \frac{1}{2} \xi_i h + \frac{1}{h} \left[\frac{h\beta}{a_i} + \frac{\alpha^2 k_i^2}{a_i} \right] \triangleq \zeta_i. \quad (2.32)$$

For the above $\zeta = \text{Col}\{\zeta_i\}$, there must be α' such that $\zeta \leq \alpha' K$. If $\phi \in S_\alpha \subset L^m$, then $u_h = T(h)\phi \in \Gamma_{\alpha'} \subset H^m$. It is easy to deduce from Theorem 3 and 4 that the pseudo-rectangle $\Gamma_{\alpha'} = \left\{ \phi \in H^m \mid \text{Col}\{\|\phi_i\|^2\} \leq \hat{K}, \alpha' \geq 1 \right\}$ is positively invariant and uniformly absorbing in H^m . If \mathcal{B} is any bounded set of L^m included in S_α , after a certain time $t_1 = t_1(\mathcal{B}, \alpha')$, we find that $u(t)$ belongs to the absorbing set $\Gamma_{\alpha'}$. This shows that

$$T(t)\mathcal{B} \subset \Gamma_{\alpha'}, \quad \forall t \geq t_1.$$

The embedding $H^1(\Omega) \subset L^2(\Omega)$ is compact [9], so is $\Gamma_{\alpha'} \subset L^m$. Therefore, (1.3) is proved.

From the above proof and Theorem 1-4, all the assumptions of Lemma 1 are now satisfied with $H = L^m$ and the proof of Theorem 5 is completed.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 19831030).

REFERENCES

- [1] Teman, R., *Infinity-Dimensional Dynamical Systems in Mechanics and Physics*, Springer-Verlag, New York, 1988.
- [2] Hale, J.K., *Asymptotic Behavior of Dissipative Systems*, American Mathematical Society, Providence, 1988.

- [3] Hale, J.K., Large diffusivity and asymptotic behavior in parabolic systems, *J. Math. Anal. Appl.*, 1986, 118, 455-466.
- [4] Marion, M., Attractor for reaction-diffusion equation; Existence and estimate of their dimension, *Appl. Anal.*, 1987, 25, 101-147.
- [5] Min, Q., Global behavior in the dynamical equation of J-J type, *J. Differential Equations*, 1988, 77, 315-333.
- [6] Pazy, A., *Semigroups of Linear Operator and Applications to Partial Differential Equations*, Springer-Verlag, New York, 1983.
- [7] Horn, R.A. and Johnson, C.R., *Matrix Analysis*, Cambridge University Press, London, 1990.
- [8] Xu, D., Integro-differential equations and delay integral inequalities, *Tohoku Math. J.*, 1992, 44, 365-378.
- [9] Adams, R.S., *Sobolev Space*, Academic Press, New York, 1975.

INDEX

- α -decomposition, 186
Almost periodic, 136
Almost periodic function, 141
Almost periodic solution, 139
Alternating sequence, 235
Amount of dissipation, 288
Analytic continuation, 273
Analytical methods, 235
Artificial intelligence, 183
Asymptotic analysis, 32
Asymptotic equilibrium, 175
Asymptotic reduction, 52
Asymptotic solution, 36
Asymptotic time limit, 169
Asymptotically stable, 3
Autonomous differential equation, 2
Auxiliary results, 12
- Banach space, 187, 420
Banach spaces, 24
Bandwidth processes, 294
Bifurcation, 156
Biological modeling, 401
Blow-up point, 28
Boltzmann constant, 336
Boundary condition, 97
Boundary-value problem, 30, 90, 91
Boundary-layer solution, 36
Boundary-layer theory, 236
Bounded domain, 107, 339
Bounded functions, 137
Boundedness, 19
Boussinesq equations, 57
Boussinesq Paradigm, 50
Boussinesq program, 53
Boussinesq simplications, 57
- Capillary action, 128
Casorati's determinant, 274
Characteristic equations, 157
Classical differentiability, 16
Classical solution, 28, 30
Closed cavities, 309
Cognitive simulation, 401
Coherent structure, 164, 176
Compact subset, 3
Compact support, 420
Comparison theorem, 184, 391
Complete solution, 39
Computer simulation, 113
Conditional total stability, 8
Conduction, 309
Containment device, 27
Continuous function, 136
Continuum mechanics, 262
Contour lines, 90
Convection, 309
Convergence theorem, 401
Convergent solutions, 71, 313
Convex functions, 14
Coronal streamers, 31
Creeping flow, 90
Cylindrical outlets, 355
- Delamination problems, 12
Delay equations, 153
Diffusion coefficient, 27
Dimensionless variables, 51
Directional derivative, 14
Dirichlet boundary conditions, 167
Dirichlet set, 53
Discontinuous nonlinearities, 20
Discrete parameter games, 306

- Discretization schemes, 167
- Dispersion parameter, 51
- Dispersive system, 49
- Dissipative phenomena, 289
- Double radio sources, 32
- Double-layer potential, 106
- Dynamic condition, 54
- Dynamical equation, 132
- Dynamical system, 9, 419

- Eigenfunction profile, 42
- Eigenvalue, 27
- Elliptic differential operator, 13
- Elliptic operator, 11, 392
- Empirical phenomena, 183
- Energy flux, 33
- Equi-asymptotic stability, 195
- Equilibrium solutions, 155
- Ergodic optimal, 296
- Ergodic payoff criterion, 297
- Ergodic perturbation, 136
- Ergodic solutions, 135
- Euclid norm, 427
- Eulerian representation, 285
- Eulerian-Lagrangian passage, 54
- Evolutionary models, 224
- Existence and nonexistence theorems, 30
- Existence and uniqueness, 28
- Exponential attractor, 224
- Exponential trichotomy, 139
- Exterior domain, 350
- Extremal solutions, 12, 15, 17

- Finite dimensional matrix, 301
- Flat plate, 90
- Fractal dimension, 224
- Fundamental matrix, 139
- Fundamental solutions, 265

- Fuzzy control, 188
- Fuzzy differential equations, 183, 193
- Fuzzy mapping, 188
- Fuzzy sets, 183
- Fuzzy topological spaces, 186

- Galilean invariance, 49, 50
- Gamma functions, 272
- Gaussian elimination, 61
- Gaussian random variables, 172
- Generalized directional derivative, 11
- Generalized gradient, 21
- Generalized quasilinear scheme, 235
- Generalized subdifferential calculus, 14

- Geometrical properties, 183
- Global attractor, 420
- Global problem, 265
- Global properties, 213
- Global solution, 32
- Greatest fixed point, 17
- Greatest solution, 16
- Green's function, 101

- Hamiltonian, 164
- Hamiltonian form, 165
- Hamiltonian system, 166, 168
- Heat flux, 313
- Heat transfer, 310
- Helmholtz equation, 102, 107
- Hemivariational inequalities, 11
- Heyman's theorem, 274
- Hilbert spaces, 109
- Hill equation, 140
- Hölder continuity, 223
- Homogenous deformation, 285
- Hooke's law, 375
- Hyperasymptotic estimates, 42

- Image processing, 183, 249
Implicit scheme, 204
Inhibitory inputs, 405
Integral equation, 106
Invariant measure, 305
Invariant set, 5
Inverse acoustics, 101
Inverse problem, 109
Iterative scheme, 209
- Kelvin-Helmholtz instability, 48
Kinematic viscosity, 336
Kinetic energy, 170
Kronecker product, 249
- Lagrangian density, 54
Laplace equation, 51
Laplace operator, 95
Large-scale structure, 166
Lebesgue measurable, 140, 149, 420
Lebesgue space, 227
Linear momentum, 129
Linear perturbation, 33
Linearized system, 158
Liouville property, 168
Lipschitz condition, 144, 238
Lipschitz constant, 144
Locally Lipschitzian, 18
Long-time saturation, 179
Long-wave solutions, 59
Lorentz transformation, 133
Lower solution, 15
Lower-semicontinuous, 19
Łukasiewicz logic, 186
Lyapunov function, 194
- Magnetic field, 27
Mappings, 17
- Markov strategies, 294
Martingale problem, 299
Maximal monotonicity, 23
Mechanical systems, 9
Minimal surfaces, 128
Molecular potential, 114
Molecular system, 118
Monotone sequences, 202
Monotonicity, 17
- N*-body problems, 114
 n -dimensional space, 258
Navier-Stokes equations, 89, 115
Nemytskij operator, 16
Nervous system, 402
Nested family, 3
Neural network, 153, 402
Neural signals, 402
Neurotransmitters, 402
Neutral difference equations, 69
Newton's method, 129
NLS equation, 164
Non-differentiable functionals, 24
Non-Newtonian fluids, 339
Nonsingular, 388
Numerical algorithm, 209
Numerical calculations, 321
Numerical computation, 401
Numerical examples, 235
Numerical scheme, 201
Numerical solution, 90
Nusselt number, 311
- Oldroyd type, 347
Open cavities, 309
Optimal equilibrium, 293
Ordered pair, 15
Orthogonal transformations, 258
Oscillating oblateness, 116

- Oscillatory solution, 75
 Overrelaxation algorithm, 95
 Parabolic systems, 201
 Parameter space, 32
 Partial ordering, 13, 185
 Pattern recognition, 183
 Periodic component, 136
 Periodic sequences, 148
 Periodic solutions, 145
 Peripheral velocity, 90
 Permutation tensor, 257
 Phase shift, 62
 Phase speeds, 66
 Point-source, 107
 Potential energy, 119, 168
 Pressure function, 351
 Propagation angle, 32
 Pseudomomentum, 53
 Pseudomonotone, 19
 Quasilinear elliptic operator, 14
 Quasimonotone, 209
 Quasimonotone methods, 367
 Quasimonotone property, 202
 Radial velocity, 91
 Radiation condition, 106
 Radiative wave, 31
 Radii of stability, 1
 Radius of stability, 1
 Random variables, 301
 Rayleigh number, 336
 Reaction diffusion equations, 391
 Reaction rate, 367
 Reaction-diffusion problem, 368
 Region of stability, 1
 Regularity method, 219
 Relaxed control, 293
 Resonance layer, 40
 Reynolds number, 89
 Riccati algorithm, 373
 Rotating machinery, 373
 Rotor configurations, 373
 Rotor vibration, 373
 Seamount problem, 101
 Second order tensor, 250
 Self-adjoint systems, 385
 Semigroup, 419
 Semigroup properties, 419
 Separation, 115
 Shaded drop, 116
 Shear layer, 32, 35, 39
 Short-wavelength region, 43
 Singular behavior, 103
 Sobolev space, 11
 Solar system, 114
 Solitary wave, 49, 57, 164, 179
 Solitary wave solution, 172
 Spectral approximation, 167
 Square-integrable functions, 109
 Stabilization methods, 361
 State-dependent subdifferentials, 24
 Statistical equilibrium, 163, 178
 Statistical mechanics, 168
 Statistical theory, 179
 Steady problem, 221
 Stochastic differential game, 297
 Stokes discontinuity, 40
 Stokes problem, 345
 Stream function, 90
 Strictly differentiable, 14
 Strongly nonlinear, 50, 66
 Subcritical waves, 57
 Subdifferentials, 12
 Subsonic velocity, 32
 Sufficient condition, 70

- Superasymptotic level, 44
Superasymptotic solution, 42
Superasymptotic truncation, 42
Symmetric velocity gradient, 221
Symmetry classes, 249
Symmetry group, 289
Synaptic connection, 404
Synaptic excitation, 404
Synchronous equilibria, 153
Synchronous equilibrium, 153
Synchronous solutions, 154
System of particles, 113
- Theory of elasticity, 285
Thermal balance, 309
Thermal diffusivity, 336
Third order tensor, 257
Tikhonov regularization, 111
Time discretization, 204
Total stability, 1
Transient algorithms, 361
Transition function, 188
Transport equation, 345
Traveling wave, 31
Truncation method, 219, 220
Truth-value, 185
Turbine-generator, 373
Turbulent fluctuations, 164
Turning point, 41
- Unbounded domains, 340
Uniform attractor, 5
Uniformly absorbing, 428
Uniformly continuous, 388
Unique solution, 69, 150
- Universal appropriateness, 185
Unstrained specimen, 287
Upper and lower sequences, 202
Upper and lower solution, 20, 201,
391, 392
Upper solution, 15
- Variation of parameters, 135
Variational formulation, 11
Viscoelastic fluid, 347
Volterra integral inequalities, 235
Vortex motion, 128
- Wave equations, 50
Wave models, 50
Wave momentum, 53
Wave number, 34
Wave turbulence, 163
Wave-fluid resonance, 35
Weak convergence, 300
Weak dispersion, 50
Weak solution, 15, 27, 222
Weaker nonlinearities, 175
Weal dispersion, 50
Wedge domain, 102
Weissenberg number, 348
Wide-band system, 300
- Zero solution, 2