Christian Constanda
Bardo E.J. Bodmann
Haroldo F. de Campos Velho
Editors

Integral Methods in Science and Engineering

Progress in Numerical and Analytic Techniques

Birkhäuser

Integral Methods in Science and Engineering

Christian Constanda • Bardo E.J. Bodmann
Haroldo F. de Campos Velho
Editors

Integral Methods in Science and Engineering

Progress in Numerical and Analytic Techniques

Birkhäuser

Editors
Christian Constanda
Department of Mathematics
The University of Tulsa
Tulsa, OK, USA

Bardo E.J. Bodmann
Mechanical Engineering
Federal University of
Rio Grande do Sul
Porto Alegre, RS, Brazil

Haroldo F. de Campos Velho
Associate Laboratory for Computing
and Applied Mathematics
National Institute for Space Research
São José dos Campos, SP, Brazil

ISBN 978-1-4614-7827-0 ISBN 978-1-4614-7828-7 (eBook)
DOI 10.1007/978-1-4614-7828-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013943828

Mathematics Subject Classification (2010): 00B25, 35-06, 41-06, 44-06, 45-06, 65-06, 76-06, 86-06, 86A10

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.birkhauser-science.com)

To Tom Grasso
for his professionalism and friendship,
on his retirement from the world of publishing
(CC)

Preface

The international conferences on Integral Methods in Science and Engineering (IMSE) are a forum where researchers in many theoretical and applied areas, whose working methodology includes integration, communicate their latest results and discoveries and create synergies based on their common interest in the use of a class of general—diverse but interconnected—mathematical procedures.

The first 11 IMSE conferences took place in a variety of venues all over the world:

1985, 1990: University of Texas–Arlington, USA;
1993: Tohoku University, Sendai, Japan;
1996: University of Oulu, Finland;
1998: Michigan Technological University, Houghton, MI, USA;
2000: Banff, AB, Canada (organized by the University of Alberta, Edmonton);
2002: University of Saint-Étienne, France;
2004: University of Central Florida, Orlando, FL, USA;
2006: Niagara Falls, ON, Canada (organized by the University of Waterloo);
2008: University of Cantabria, Santander, Spain;
2010: University of Brighton, UK.

The 2012 meeting, held in Bento Gonçalves, Rio Grande do Sul, Brazil, July 23–27, and attended by participants from 11 countries on 4 continents, enhanced even further the IMSE tradition as an important event on the international conference circuit, which makes it possible for scientists and engineers to talk about their research interests in a stimulating atmosphere of understanding and cooperation.

As in the past, the organization of IMSE 2012 was of a very high standard; by way of acknowledgement, the participants wish to thank CNPq, CAPES, and FAPERGS for their financial support, and Dall'Onder Grande Hotel for special conditions and discounts, which ensured that the daily proceedings of the conference took place in pleasant surroundings. Special thanks are due to the members of the Local Organizing Committee:

Bardo E.J. Bodmann (Federal University of Rio Grande do Sul), Chairman,
Claudio Pellegrini (Federal University of São João Del Rey),
Daniela Buske (Federal University of Pelotas),

Fernando Carvalho (Federal University of Rio de Janeiro),
Gervasio A. Degrazia (Federal University of Santa Maria),
Haroldo F. de Campos Velho (National Institute for Space Research),
Marco Túllio M.B. de Vilhena (Federal University of Rio Grande do Sul),
Renato M. Cotta (Federal University of Rio de Janeiro),
Ricardo C. Barros (Rio de Janeiro State University).

A distinguishing feature of IMSE 2012 was the increased number of young researchers who attended and presented their work. It was both reassuring and gratifying to see that the new generation is ready to join in and help our particular field of scientific interest move forward.

The next IMSE conference will be hosted by the Karlsruhe Institute of Technology, Germany, in July 2014. Further details will be posted in due course on the conference web site.

The peer-reviewed chapters of this volume, arranged alphabetically by first author's name, are an expansion of 26 papers from among those given in Bento Gonçalvez. The editors would like to thank the staff at Birkhäuser for their courteous and professional handling of the publication process.

Tulsa, OK, USA — Christian Constanda
Porto Alegre, RS, Brazil — Bardo E.J. Bodmann
São José dos Campos, SP, Brazil — Haroldo F. de Campos Velho

The International Steering Committee of IMSE:

C. Constanda (The University of Tulsa), *Chairman*
M. Ahues (University of Saint-Étienne)
B. Bodmann (Federal University of Rio Grande do Sul)
H. de Campos Velho (INPE, Saõ José dos Campos)
P. Harris (University of Brighton)
A. Kirsch (Karlsruhe Institute of Technology)
M. Lanza de Cristoforis (University of Padova)
S. Mikhailov (Brunel University)
D. Mitrea (University of Missouri-Columbia)
A. Nastase (RWTH Aachen University)
D. Natroshvili (Georgian Technical University)
M. Pérez (University of Cantabria)
K. Ruotsalainen (University of Oulu)
O. Shoham (The University of Tulsa)

Contents

Contributors

André S. de Aguiar Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Lourdes Alvarez The University of Tulsa, Tulsa, OK, USA

Antônio C.M. Alvim Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Alexandre S. Athayde Federal University of Pelotas, Pelotas, RS, Brazil

Carlos Avila Chevron ETC, Houston, TX, USA

Fabio S. Azevedo Institute for Mathematics, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Andrea Barbarino Energy Department, Politecnico di Torino, Torino, Italy

Ricardo C. Barros University of the State of Rio de Janeiro, Programa de Pós-graduação em Ciências Computacionais, Rio de Janeiro, RJ, Brazil

Bardo E.J. Bodmann Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Daniela Buske Federal University of Pelotas, Pelotas, Rio Grande do Sul, Brazil

Vincent Calvez École Normale Supérieure de Lyon, Lyon Cedex, France

Dayana Q. de Camargo Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Ke Chen University of Liverpool, Liverpool, UK

Christian Constanda The University of Tulsa, Tulsa, OK, USA

D.R. Doty The University of Tulsa, Tulsa, OK, USA

Sandra Dulla Energy Department, Politecnico di Torino, Torino, Italy

Julio C.L. Fernandes Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Barry D. Ganapol The University of Arizona, Tucson, AZ, USA

Izabel Gioveli Federal University of Fronteira Sul, Santo Ângelo, RS, Brazil

D. Gómez Dpto. Matemáticas, Estadísitica y Computación, Universidad de Cantabria, Santander, Spain

Luis Gomez–Morillo The University of Tulsa, Tulsa, OK, USA

Paul J. Harris University of Brighton, Brighton, UK

Eliete B. Hauser Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Hiroshi Hirayama Department of Vehicle System Engineering, Faculty of Creative Engineering, Kanagawa Institute of Technology, Kanagawa, Japan

Cibele A. Ladeia State University of Londrina, Londrina, Paraná, Brazil

Celso M.F. Lapa Institute of Nuclear Engineering, Rio de Janeiro, RJ, Brazil

Ralph S. Mansur University of the State of Rio de Janeiro, Programa de Pós-graduação em Ciências Computacionais, Rio de Janeiro, RJ, Brazil

Nicolas Meunier Université Paris Descartes, Paris, France

Ram S. Mohan The University of Tulsa, Tulsa, OK, USA

Nicolas Muller Université Paris Descartes, Paris, France

Luis P.L. de Oliveira University of Vale do Rio dos Sinos, São Leopoldo, RS, Brazil

Rubén P. Pazos University of Santa Cruz do Sul, Santa Cruz do Sul, RS, Brazil

Claudio Z. Petersen Federal University of Pelotas, Pelotas, RS, Brazil

M. Eugenia Pérez Dpto. Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander, Spain

Mauricio G. Prado The University of Tulsa, Tulsa, OK, USA

Regis Quadros Federal University of Pelotas, Pelotas, Rio Grande do Sul, Brazil

Piero Ravetto Energy Department, Politecnico di Torino, Torino, Italy

Neyva M.L. Romeiro State University of Londrina, Londrina, Paraná, Brazil

Esequia Sauter Institute for Mathematics, Federal Institute of Rio Grande do Sul, Porto Alegre, RS, Brazil

Marcelo Schramm Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Adalberto Schuck Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Cynthia F. Segatto Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Tatiana A. Shaposhnikova Department of Differential Equations, Moscow State University, Moscow, Russia

Ovadia Shoham The University of Tulsa, Tulsa, OK, USA

Jeronimo J.A. Silva Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Francisco F. Lamego Simões Filho Institute of Nuclear Engineering, Rio de Janeiro, RJ, Brazil

Abner D. Soares National Commission for Nuclear Energy, Rio de Janeiro, RJ, Brazil

Adelir J. Strieder Federal University of Pelotas, Pelotas, RS, Brazil

José A. Tenreiro Machado Institute of Engineering, Polytechnic of Porto, Department of Electrical Engineering, Porto, Portugal

Gavin R. Thomson A.C.C.A., Glasgow, UK

Mark Thompson Institute for Mathematics, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Fernanda K. Tomaschewski Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Rinaldo A.M. Vieira Petrobras, Rio de Janeiro, Brazil

Marco T.B.M. de Vilhena Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Raphaël Voituriez Université Pierre et Marie Curie, Paris Cedex, France

Maria A.V. Wasserman Institute of Nuclear Engineering, Rio de Janeiro, RJ, Brazil

Guilherme J. Weymar Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Jorge R.S. Zabadal Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Chapter 1
Multiphase Flow Splitting in Looped Pipelines

L. Alvarez, R.S. Mohan, O. Shoham, L. Gomez, and C. Avila

1.1 Introduction

The Petroleum Industry utilizes parallel and looped pipelines in order to decrease pressure drop and increase flow capacity. For the looped lines configuration, the flow splits at an impact tee into 2 lines, which are recombined downstream. In the parallel configuration, the flow splits into 2 lines, which are not recombined. The looped pipeline system design has been carried out in the past based on rule of thumb. For single-phase flow, the splitting of the flow and the corresponding pressure drop can be determined in a straightforward manner, based on first principles. However, the two-phase flow case is more complicated and no fundamental understanding of the splitting flow phenomena. This is the main reason for the lack of publications and studies in this area. Following is a brief summary of studies published on the parallel pipeline configurations.

Several studies were published on multiphase flow splitting in impacting tees. These include [HwSoLa89], [HoGr95], [AzPuGo88a] and [ElSoSi07]. Flow splitting behavior in steam flood distribution networks was studied in [ChRu92]. The splitting of gas–liquid two phases in 4 parallel pipelines with common inlet and outlet headers, capable of rotating between 0 and 15°, was investigated in [TaEtAl03], where it was found that for the horizontal case, the split is more or less even between all the pipes. However, for inclined flow, at low gas and liquid rates, the two-phase mixture prefers to flow into a single line, while stagnant liquid fills part of the other pipes. In the follow-up paper [TaEtAl06], a rigorous stability

L. Alvarez • R.S. Mohan • O. Shoham (✉) • L. Gomez
The University of Tulsa, Tulsa, OK 74104, USA
e-mail: lourdes-alvarez@utulsa.edu; ram-mohan@utulsa.edu; ovadia-shoham@utulsa.edu; luis-gomez-morillo@utulsa.edu

C. Avila
Chevron ETC, Houston, TX 77002, USA
e-mail: c.avila@chevron.com

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_1,

analysis was presented for the determination of the flow splitting in the lines, capable of predicting the number of pipes that are filled with stagnant liquid.

No studies have been published for two-phase flow splitting in looped lines, as operated by the Petroleum Industry. This is owing to the difficulty of measuring simultaneously the gas and liquid flow rates in each of the looped lines. This is the gap that the current study attempts to address.

1.2 Experimental Program

A unique experimental facility has been designed, constructed, and instrumented, which enables acquiring multiphase flow splitting data in both parallel and looped horizontal pipelines. Figure 1.1 shows a schematic of the splitting facility, which is divided into three sections: inlet, parallel/looped splitting lines, and separation/metering sections. The entire splitting facility is constructed of clear PVC. The inlet section and looped line section are 30 ft long each ($\frac{L}{d} = 180$), in order to ensure fully developed flow. The splitting section includes the impacting tee, which is located at the end of the inlet section.

The air and water mixture flows through the inlet pipe and into the impacting tee, which splits the flow into the 2 looped pipes (side 1 and side 2). Downstream, each of the looped lines is connected to a Gas-Liquid Cylindrical Cyclone (GLCC)©[1] compact separator, where the gas and liquid phases are separated and metered. The 2 GLCC©'s are identical, both instrumented with a gas rotameter to measure the gas flow rate, and a Micromotion® for measuring the liquid flow rate. As shown in Fig. 1.1, pressure transducers are installed at the inlet (upstream of the impacting

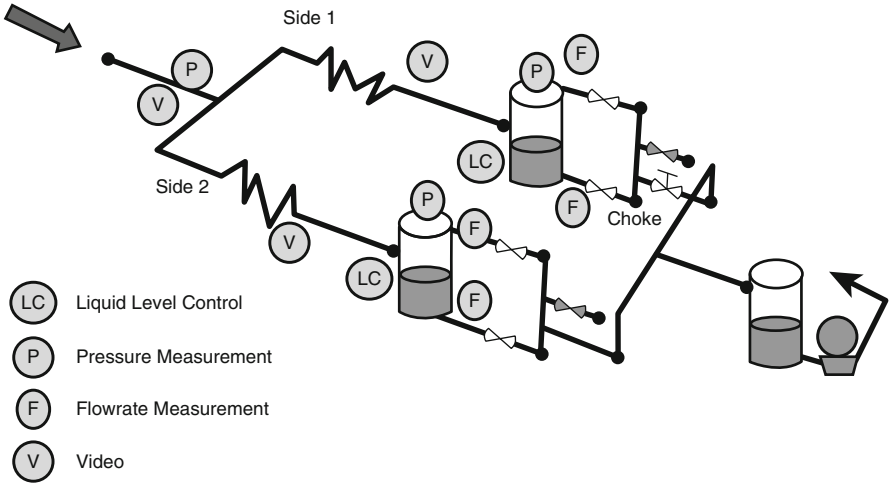

Fig. 1.1 Schematic of the parallel/looped splitting facility

[1]GLCC© - Gas-Liquid Cylindrical Cyclone - copyright, The University of Tulsa, 1994.

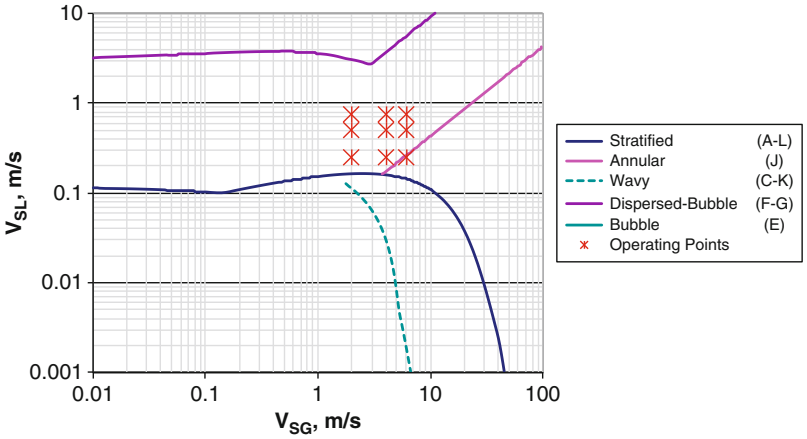

Fig. 1.2 Inlet operational conditions

tee) and on top of each GLCC© for measurement of the pressure at these locations. The outlet pressure of the system is kept at atmospheric conditions for all the test runs.

For the parallel lines configuration tests, the gas was vented off the GLCC© gas leg, and the liquid recirculated to the storage tank. For the looped configuration the gas and liquid were recombined downstream the GLCC©, and the recombined lines recombined again to form the looped lines configuration (see Fig. 1.1).

A total of 81 gas–liquid flow splitting experiments have been carried out in both the parallel and looped configurations. The test matrix is divided into three phases, as follows:

- ***Phase 1: Equal Split Conditions.*** In this phase, the parallel and looped configurations are symmetrical, with the same diameter pipes (2 in.) installed on one line (side 1) downstream the GLCC©. Experiments are conducted with valve.
- ***Phase 2: Uneven Split Conditions.*** Uneven splitting is achieved utilizing a globe valve settings of 100, 75, 50, 25, and 10 % open. These experiments were carried out only for the looped configuration (2-in. diameter pipes).
- ***Phase 3: Different Diameter Conditions.*** For this phase, one side (side 1) was 1-in. in diameter, while the other side (side 2) is 2-in. in diameter.

Air and water were used in this study, and the operational conditions are shown in Fig. 1.2 on a flow pattern map for the inlet conditions, based on the model in [Ba87].

Note that the legend includes the different transition boundaries (such as stratified and annular), and the code designated to each transition (such as A–L and J). As can be seen in the figure, the superficial gas velocities at inlet conditions for all 3 experimental phases are 2, 4, and 6 m/s, while the superficial liquid velocities are 0.25, 0.50, and 0.75 m/s, resulting in 9 different combinations of *Vsg* and *Vsl* splitting runs, under slug flow conditions.

1.3 Experimental Results

In this section, typical experimental results are presented for each of the 3 phases of the study (refer to [Al09] for more details). The results are presented in terms of the liquid fraction($F_{Liq\,1}$) and the gas fraction($F_{gas\,1}$) in the line with the valve (side 1). The fraction of a phase is the flow rate of the phase in line 1 divided by the total phase flow rate at the inlet, as given by

$$F_{gas\ 1} = \frac{q_{\text{gas 1}}}{q_{\ gas\ inlet}}, \quad F_{\ liq\ 1} = \frac{q_{\ liquid\ 1}}{q_{\ liquid\ \ inlet}},$$

where q (ft^3/sec) is the volumetric flow rate. Note that some of the test runs have been repeated, demonstrating the repeatability of the conducted tests.

Results Phase 1: Equal Split Conditions. Equal split conditions are obtained for both parallel and looped configurations. For this phase, all pipes have the same diameter and the GLCC© pressures on each side, P_1 and P_2, are the same. The splits of both the gas and liquid phases are close to 50–50. Figure 1.3 presents the results for both the parallel and looped configurations for Vsl = 0.75 m/s and the 3 superficial gas velocities tested.

The 45° line represents the equal-phase-splitting line, where the outlets GLRs are equal to the inlet one. As can be observed in the figure, for both the parallel and looped configurations, equal split conditions are reached with negligible discrepancy.

Results Phase 2: Uneven Split Conditions. Table 1.1 and Fig. 1.4 present the results for Vsg = 4 m/s and the different Vsl's tested. The results presented are for valve settings of 100, 75, 50, 25, and 10 % opening. For small valve opening, the

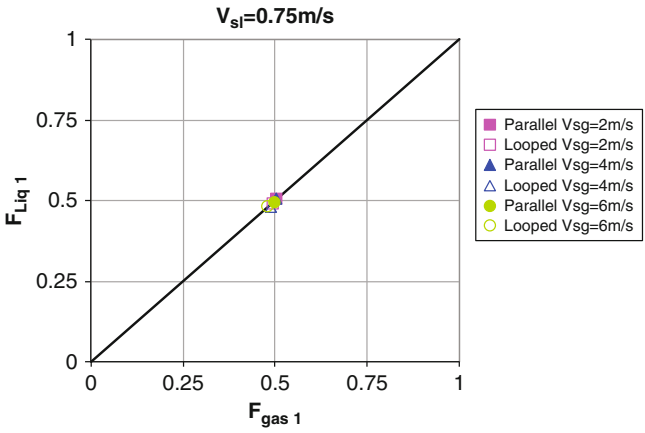

Fig. 1.3 Results for equal split conditions (phase 1, Vsg = 6 m/s)

Table 1.1 Sample results for uneven split condition (phase 2)

Superficial gas velocity V_{sg} = 4 m/s										
Vsl = 0.25 m/s	100 %		75 %		50 %		25 %		10 %	
	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2
Inlet PT (psi)	16.14 ± 0.15		16.14 ± 0.19		16.17 ± 0.12		16.20 ± 0.13		16.60 ± 0.11	
GLCC PT (psi)	15.92 ± 0.12	15.80 ± 0.12	15.88 ± 0.20	15.78 ± 0.12	16.00 ± 0.13	15.90 ± 0.14	16.01 ± 0.14	15.90 ± 0.13	16.46 ± 0.10	16.23 ± 0.08
F_{gas}	0.51 ± 0.02	0.49 ± 0.02	0.42 ± 0.01	0.58 ± 0.01	0.34 ± 0.01	0.66 ± 0.01	0.25 ± 0.01	0.63 ± 0.09	0	1
F_{liq}	0.50 ± 0.09	0.50 ± 0.09	0.46 ± 0.06	0.54 ± 0.06	0.43 ± 0.11	0.57 ± 0.11	0.37 ± 0.09	0.57 ± 0.04	0.13 ± 0.11	0.87 ± 0.11
Flow pattern (outlet)	Wavy	Wavy	Wavy	Wavy	Wavy	Wavy	Wavy	Wavy	Only liquid	Wavy
Vsl = 0.5 m/s	100 %		75 %		50 %		25 %		10 %	
	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2
Inlet PT (psi)	16.97 ± 0.25		16.99 ± 0.27		17.02 ± 0.12		17.11 ± 0.27		17.86 ± 0.20	
GLCC PT (psi)	16.45 ± 0.17	16.42 ± 0.18	16.55 ± 0.20	16.42 ± 0.17	16.58 ± 0.16	16.44 ± 0.14	16.73 ± 0.23	16.50 ± 0.19	17.66 ± 0.20	17.23 ± 0.08
F_{gas}	0.51 ± 0.02	0.49 ± 0.02	0.44 ± 0.02	0.56 ± 0.02	0.36 ± 0.02	0.66 ± 0.02	0.25 ± 0.01	0.75 ± 0.01	0	1
F_{liq}	0.49 ± 0.05	0.51 ± 0.05	0.46 ± 0.05	0.54 ± 0.05	0.44 ± 0.04	0.57 ± 0.04	0.39 ± 0.05	0.61 ± 0.05	0.13 ± 0.06	0.87 ± 0.06
Flow pattern (outlet)	Slug	Slug	Slug	Slug	Slug	Slug	Slug	Slug	Only liquid	Slug
Vsl = 0.75 m/s	100 %		75 %		50 %		25 %		10 %	
	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2	Side 1	Side 2
Inlet PT (psi)	17.93 ± 0.27		18.00 ± 0.28		18.12 ± 0.29		18.38 ± 0.30		19.82 ± 0.20	
GLCC PT (psi)	17.29 ± 0.13	17.10 ± 0.13	17.35 ± 0.15	17.16 ± 0.14	17.51 ± 0.15	17.27 ± 0.13	17.88 ± 0.21	17.45 ± 0.16	19.53 ± 0.17	18.19 ± 0.29
F_{gas}	0.49 ± 0.02	0.51 ± 0.02	0.42 ± 0.02	0.58 ± 0.02	0.35 ± 0.01	0.65 ± 0.01	0.25 ± 0.01	0.75 ± 0.01	0	1
F_{liq}	0.49 ± 0.04	0.51 ± 0.04	0.47 ± 0.04	0.53 ± 0.04	0.46 ± 0.04	0.54 ± 0.04	0.41 ± 0.06	0.59 ± 0.06	0.13 ± 0.08	0.87 ± 0.08
Flow pattern (outlet)	Slug	Slug	Slug	Slug	Slug	Slug	Slug	Slug	Only liquid	Slug

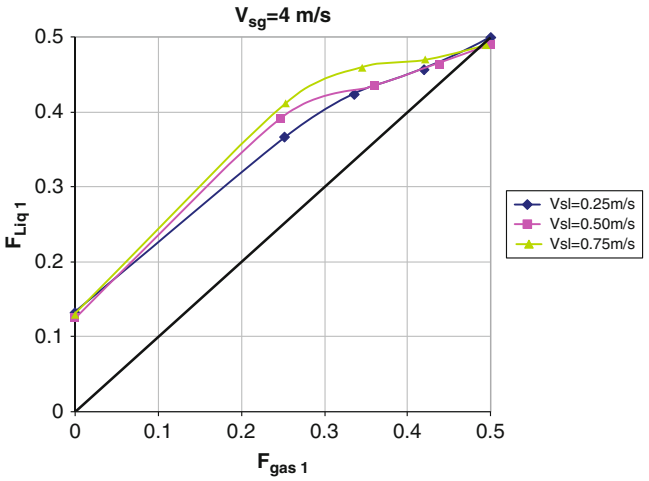

Fig. 1.4 Results for uneven split condition (phase 2, $Vsg = 4$ m/s)

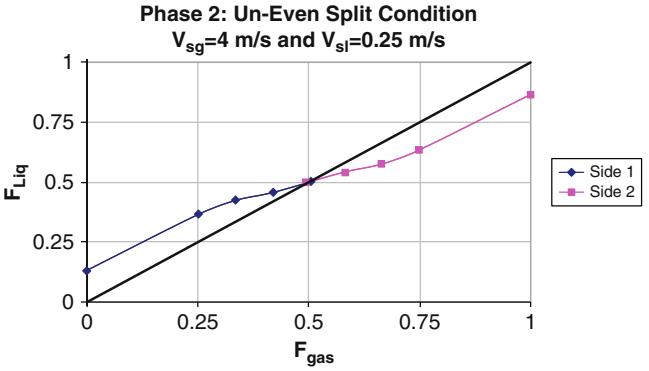

Fig. 1.5 Mass balance illustration for uneven split condition (phase 2)

gas and liquid flow preferentially through the other line (side 2) without the valve. A higher tendency for the gas to split preferentially into the line without the valve (side 2) is clearly observed for low valve settings. Also, the splitting gas fraction changes with valve setting are more significant, as compared to the liquid–phase fractions. At the valve setting of 10 % open, all the gas flows preferentially through side 2, whereby no more gas flows through the choked line. However, there is still some liquid flowing through this line. Note that for this condition, the liquid fraction in side 1, F_{liq1}, does not depend on the inlet liquid flow rate.

The mass balance is satisfied for all the uneven flow runs, namely, that the sum of the gas and liquid flow rates in the 2 looped lines equal to the respective total inlet flow rate. This is demonstrated in Fig. 1.5. Figure 1.6 shows the effect of increasing *Vsl* at a fixed value of *Vsg*, showing, respectively, an increasing deviation of the

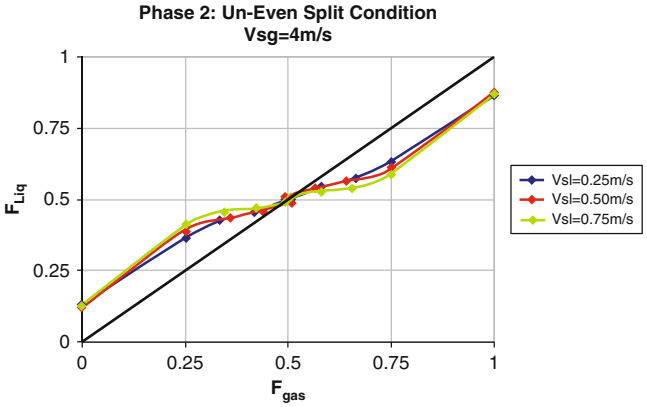

Fig. 1.6 Effect of liquid flow rate for uneven split condition (phase 2, $Vsg = 4$ m/s)

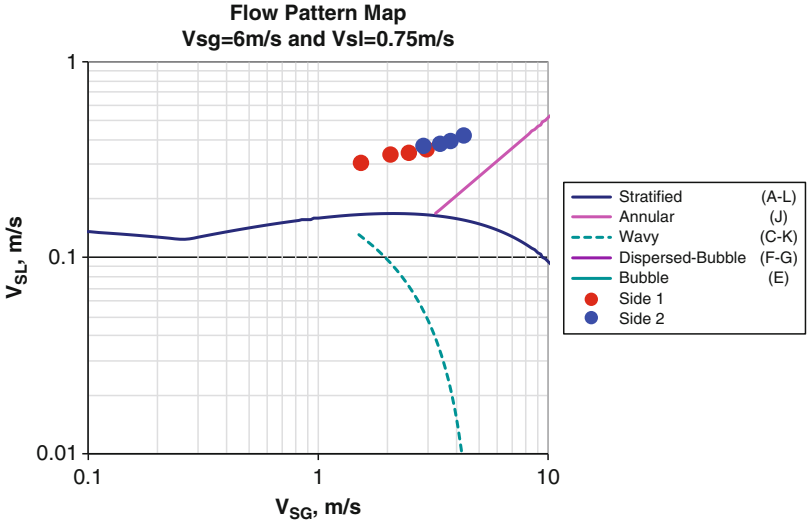

Fig. 1.7 Results for uneven split condition with slug flow in looped lines

results from the 45° equal splitting line. These results agree with the observation reported in [ChRu92], [HoGr95], and [FuEtAl95].

Slug flow occurs at the inlet for all the experimental test runs. For almost all the runs, slug flow also exists in the looped lines. Figure 1.7 shows the results for the $Vsg = 6$ m/s and $Vsl = 0.75$ m/s flow run. As can be seen, for these conditions, slug flow occurs not only at the inlet but also in both looped lines.

Results Phase 3: Different Diameter Conditions. For this case, side 2 line remained the same, 2-in. in diameter, while side 1 line was replaced with a

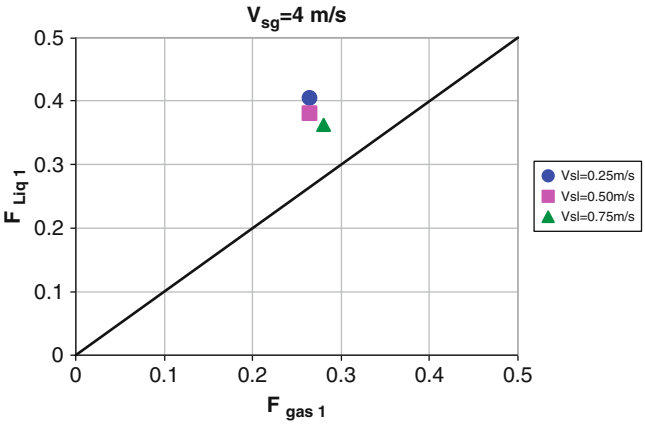

Fig. 1.8 Results for different diameter conditions (phase 3)

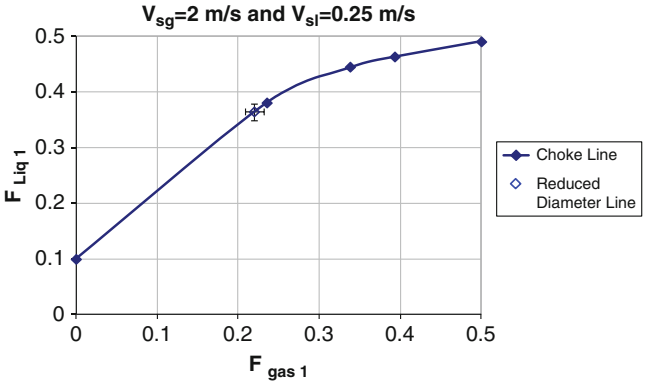

Fig. 1.9 Schematic of splitting model in parallel/looped pipelines

reduced 1-in. diameter pipe. Figure 1.8 presents the results for $Vsg = 4$ m/s and the 3 superficial liquid velocities tested. For a fixed superficial gas velocity, while increasing the superficial liquid velocity, more liquid tends to flow into the larger diameter line (side 2), reducing the fraction of liquid flowing in the reduced line (side 1).

An interesting comparison between the results of phase 2 and phase 3 is presented in Fig. 1.9, the $Vsg = 2$ m/s and $Vsl = 0.25$ m/s run. As can be seen, similar results are obtained for this case for phase 3 (different diameter lines) and phase 2 run with a valve setting of 25 % open. This implies that for this particular geometry the frictional pressure drop drives the flow split.

1.4 Model Development

This section describes the mechanistic model developed for predicting gas–liquid two-phase flow splitting in parallel and looped pipelines. The model is capable of predicting the gas and liquid splitting fractions in each of the lines, as well as the pressure drop across the system. Figure 1.10 shows a schematic of the model, which is presented next.

Pressure Drop Model. This sub-model is based on pressure equality in the looped lines, namely,

$$\Delta P_{total} = \Delta P_1 = \Delta P_2.$$

Side 1 line (ΔP_1) includes the valve, while side 2 (ΔP_2) has no flow restriction. The pressure drop in each line is given by

$$\Delta P_{side} = \Delta P_{inlet-side} + \Delta P_{side-outlet},$$

where $\Delta P_{inlet-side}$ represents the pressure drop between the inlet and GLCC© and $\Delta P_{side-outlet}$ is the pressure drop between each GLCC© and the outlet, including the side 1 valve (phase 2) and the pressure drop in the GLCC©. The horizontal inlet pressure drop ($\Delta P_{inlet-side}$) is determined using the model in [GoEtAl00]. The outlet pressure drop ($\Delta P_{side-outlet}$) includes two components, namely, frictional and gravitational. The frictional component integrates the valve, several elbows, instrumentation, and the GLCC© itself in one fitting, in terms of a resistance coefficient K defined as

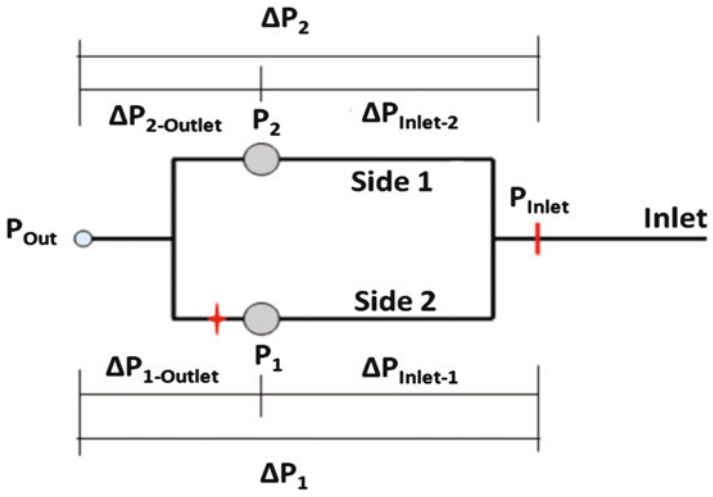

Fig. 1.10 Schematic of splitting model in parallel/looped pipelines

$$\Delta P_{frictional} = K_1 \left(\frac{\rho_{mix} \ v_{mix}^2}{2 g_c} \right),$$

where ρ_{mix} is the average mixture density (lbm/ft^3), v_{mix} is mixture velocity (ft/s), and $g_c = 32.17 \ \frac{lbm \ ft}{lbf \ s^2}$. The homogeneous no-slip model is utilized to determine the mixture velocity and density.

Flow Splitting Model. A flow splitting model and algorithm have been developed, utilizing the pressure drop model. The looped lines geometry, inlet and outlet pressures, and inlet gas and liquid flow rates are given as input data. The model predicts the flow split between the two looped lines. A trial and error procedure is utilized, whereby the split is guessed and the pressure drops in each of the looped line are calculated. Convergence is achieved when the pressure drop through the looped lines is equal. The same inlet GLR is assumed to occur in the looped lines. Note that no multiple solutions are obtained, namely, a unique solution is found for each of the flow conditions.

1.5 Results and Discussion

This section presents comparisons between the model predictions and experimental data for both the pressure drop and flow splitting, as presented next.

For phase 2, namely, the 54 unequal split runs, the average pressure drop error (across the looped configuration) between the model predictions and experimental data is $\pm 10\,\%$. The proposed flow splitting model predictions have also been compared with phase 2 experimental runs. Figure 1.11 shows a typical comparison for the $Vsg = 4$ m/s and $Vsl = 0.75$ m/s run, showing a good agreement. Similar agreement is observed for all 54 runs, comparing experimental data and model predictions for the liquid fraction $(F_{Liq\ 1})$ and the gas fraction $(F_{gas\ 1})$ in the line with the valve (side 1). For this comparison, the discrepancies between model predictions and experimental data are within $\pm 15\,\%$, as demonstrated in Fig. 1.12.

Field Case Example. A multiphase splitting in looped lines field case has been provided by Chevron. The looped lines are 6 in. and 8 in. diameter pipes, which have the same profile, as shown in Fig. 1.13. The actual flow conditions are given in Table 1.2, in terms of the total flow rates of the liquid and gas at the inlet (upstream the splitting into the 2 looped lines), the water cut, specific gravities of the phases and downstream separator pressure and temperature at the recombination location.

A field design code has been developed, based on the proposed model [Er10]. The developed code has been utilized to run the field case, for determining the splitting fractions of the gas and the liquid in each of the looped lines. Table 1.3 presents the developed computer code predictions of the gas and liquid splitting fractions, as well as the results predicted by OLGA. The developed computer code predicted that 31.5 % of the gas and liquid flows in the 6 in. pipe and 68.5 % of the

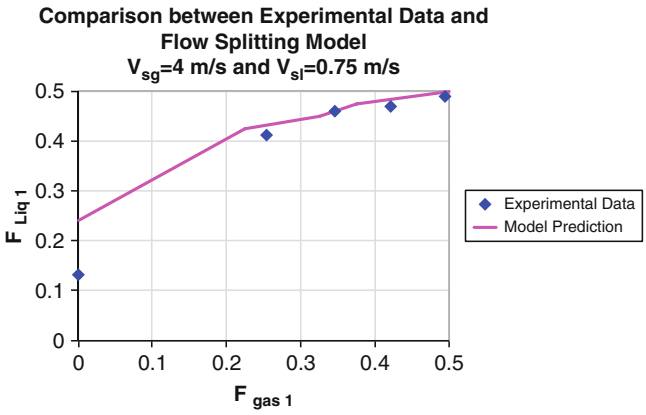

Fig. 1.11 Typical comparison between experimental data and model prediction for flow splitting

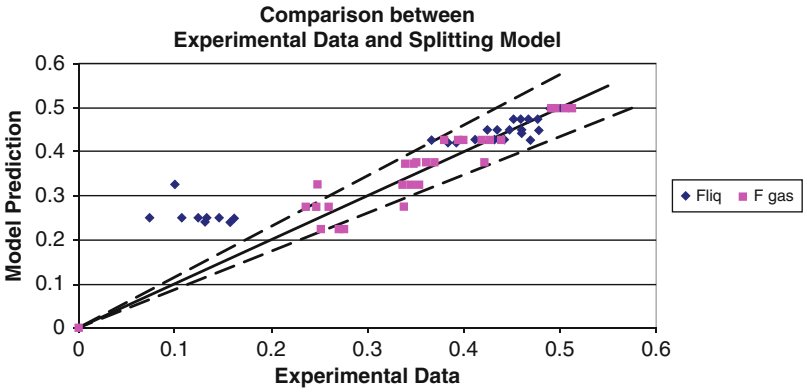

Fig. 1.12 Comparison between model predictions and experimental data for flow splitting fractions (phase 2)

Table 1.2 Field case flow conditions

Gas specific gravity	0.761
Water specific gravity	1.02
Oil specific gravity	32.2 API
Total gas flow rate	1.3 MMscf/d
Total liquid flow rate	25,581 STB/d
Water cut	46 %
Separator temperature	115 F°
Separator pressure	85 psig

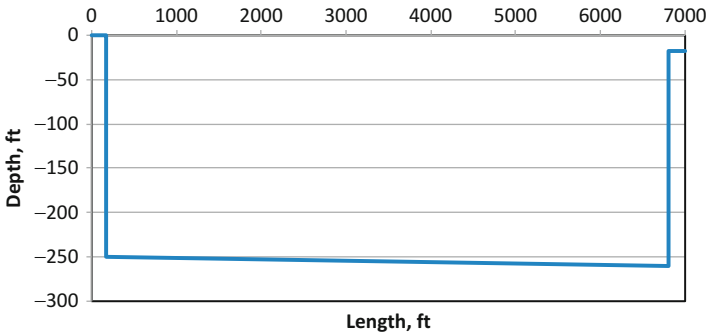

Fig. 1.13 Schematic of field case flow line profile

Table 1.3 Comparison between the OLGA and the computer code predictions

	OLGA		Current study	
Diameter (in.)	6"	8"	6"	8"
F_{liq} (%)	31	69	31.5	68.5
F_{gas} (%)	27	73	31.5	68.5

gas and liquid flows in 8 in. pipe. As can be seen in the table, the splitting liquid and gas fraction predictions show good agreement with the OLGA predictions. The small difference between the gas fraction predictions is due to the constant GOR assumption of the generated code.

1.6 Conclusion

The objective of this study is to investigate theoretically and experimentally multiphase flow splitting in both parallel and looped pipelines. Summary and conclusions of this study are given below:

1. A novel and unique experimental facility was designed, constructed, and instrumented suitable for acquiring pertinent slug flow splitting data. The facility can be operated in both parallel and looped configurations.
2. A total of 81 experimental runs were conducted. Nine different gas–liquid superficial velocity combinations under slug flow were employed. Uneven split condition was generated at 5 different settings (100, 75, 50, 25, and 10 % opening) of a globe valve installed on one of the looped lines and by utilizing different diameter looped lines.
3. For *Equal Split Conditions*, data analysis shows that the equal or even split condition is reached in both parallel and looped configurations for symmetric lines. For *Uneven Split Conditions*, the gas split is significantly more uneven

than liquid split. The gas–liquid ratio was found to be different in the two looped lines, and also different from the inlet.

4. For *Different Diameter Condition*s a similar behavior between the reduced diameter configuration and the same diameter configuration with valve setting at 25 % open was observed. This implies that for this particular geometry the frictional pressure drop drives the uneven split.
5. A mechanistic model is developed for the prediction of the pressure drop and the gas and liquid split in looped lines. The proposed model is compared with the experimental data sets acquired in this study. Good agreement was found between the proposed splitting model predictions and the experimental data with an average error of ± 15 %.
6. A field design code has been developed, based on the proposed model. The developed code has been utilized to run a field looped lines case. In the absence of flow split data, the predictions of the code were compared to the OLGA predictions, showing a good agreement.

Acknowledgments The authors wish to thank the financial support of the Tulsa University Separation Technology Projects (TUSTP), Chevron TU-CoRE (Tulsa University Center of Research Excellence), and National Science Foundation-Industry/University Cooperative Research Center on Multiphase Transport Phenomena (NSF-I/UCRC-MTP).

References

[Al09] Alvarez, L.: Multiphase Flow Splitting in Parallel/Looped Pipelines, MS thesis, The University of Tulsa (2009)

[AzPuGo88a] Azzopardi, B.J., Purvis, A., Govan, A.H.: Flow split of churn flow at a vertical impacting T. Int. J. Eng. Fluid Mech. **1**, 320–329 (1988)

[Ba87] Barnea, D.: A unified model for predicting flow-pattern transitions for the whole range of pipe inclinations. Int. J. Multiphas. Flow **13**, 1 (1987)

[ChRu92] Chien, S.F., Rubel, M.T.: Phase splitting of wet steam in annular flow through a horizontal impacting tee. SPE Prod. Eng. **7**, 368–374 (1992)

[ElSoSi07] El-Shaboury, A.M.F., Soliman, H.M., Sims, G.E.: Two-phase flow in a horizontal equal-sized impacting tee junction. Int. J. Multiphas. Flow **33**, 411–431 (2007)

[Er10] Er, M.O.: Onset of Water-Layer in Three-Phase Stratified Flow, MS thesis, The University of Tulsa (2010)

[FuEtAl95] Fujii, T., Takenaka, N., Nakazawa, T., Asano, H.: The phase separation characteristics of a gas–liquid two-phase flow in the impacting T-junction. In: Proceeding of the Second International Conference on Multiphase Flow, pp. 6.27–6.32 (1995)

[GoEtAl00] Gomez, L.E., Shoham, O., Schmidt, Z., Chokshi, R., Northug, T.: A unified mechanistic model for steady-state two-phase flow—horizontal to vertical upward flow. SPE J. **5**, 339–350 (2000)

[HoGr95] Hong, K.C., Griston, S.: Two-phase flow splitting at an impacting tee. SPE Prod. Facil. **10**, 184–190 (1995)

[HwSoLa89] Hwang, S.T., Soliman, H.M., Lahey, R.T. Jr.: Phase separation in impacting wyes and tees. Int. J. Multiphas. Flow **15**, 965–975 (1989)

[TaEtAl03] Taitel, Y., Pustylnik, L., Tshuva, M., Barnea, D.: Flow distribution of gas and liquid in parallel pipes. Int. J. Multiphas. Flow **29**, 1193–1202 (2003)

[TaEtAl06] Taitel, Y., Pustylnik, L., Tshuva, M., Barnea, D.: Prediction of two-phase flow distribution in parallel pipes using stability analysis. AIChE J. **52**, 3345–3352 (2006)

Chapter 2
Green's Function Decomposition Method for Transport Equation

F.S. Azevedo, E. Sauter, M. Thompson, and M.T. Vilhena

2.1 Introduction

The Green's Function Decomposition Method is a methodology to solve the transport equation in a slab with specular reflection at the boundaries. This method was initially derived to be applied to a non-typical problem arising from the asymptotic analysis of a radiative transport problem. In that problem, the equation to be solved takes the following form:

$$-\mu\frac{\partial}{\partial y}I(y,\mu)+\lambda I(y,\mu)=\int_{-1}^{1}\left(\frac{\sigma}{2}-\beta\mu'^{2}\right)I(y,\mu)d\mu',\ y>0, \tag{2.1a}$$

$$I(y,\mu)-\rho I(y,-\mu)=g_b(\mu),\quad \mu<0,\ y=0 \tag{2.1b}$$

$$\lim_{y\to\infty} I(y,\mu)=0. \tag{2.1c}$$

Here λ, σ and β are positive constants and $\lambda>\sigma$. The reflection coefficient ρ is a nonnegative mensurable function of the angular variable μ and is bounded above by the unit, i.e.

$$0\le\rho(\mu)\le 1.$$

We note that the domain of this equation is the half plane $y>0$ and there is no internal source term, the whole solution being determined by the boundary term $g_b(\mu)$ which is a known mensurable bounded function. For further details, the reader

F.S. Azevedo (✉) • M. Thompson • M.T. Vilhena • E. Sauter
Institute of Mathematics, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: fabio.azevedo@ufrgs.br; mark.thompson@ufrgs.br; vilhena@ufrgs.br; esequia@gmail.com

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_2,

is invited to read the paper [AzEtAl11a]. In that paper full details of the theory of existence and uniqueness of a solution for this problem is presented together with an important exponential decay estimate for the solution, i.e., the problem (2.1) has a unique solution $I(y,\mu)$ and satisfies the following inequality:

$$|I(y,\mu)| \leq Ce^{-\alpha y}, \tag{2.2}$$

where the exponent α must satisfy

$$\alpha < \sqrt{\lambda\left[\lambda - \int_{-1}^{1}\left(\frac{\sigma}{2} - \beta\mu^2\right)^{+} d\mu\right]};$$

here $\left(\frac{\sigma}{2} - \beta\mu^2\right)^{+}$ indicates the positive part of $\left(\frac{\sigma}{2} - \beta\mu^2\right)$. The constant C is a function of g_b and α.

In view of solving numerically the problem (2.1) we find three difficulties: (1) the domain is not finite; (2) the scattering kernel, which is not a nonnegative function, is not bounded below by λ; (3) the reflecting coefficient $\rho(\mu)$ may not vary smoothly with μ, which typically happens with Fresnel's reflection.

The first difficulty is overcome by truncating the domain into a finite interval taking into account the estimate (2.2). The second difficulty means that well-known iterative methods will not converge easily outside the spectral radius, i.e., when $\sigma/2 - \beta < -\lambda$, which is the case of most interest in that work. The third difficulty implies one needs to use a large number of ordinate if one decides to employ any method involving the discretization of the angular variable. That said we conceived the Green's Function Decomposition Method (GFD) with the following features: (1) It is not iterative (2) It does not involve any discretization of the angular variable.

Here we will not focus on the solution of this very specific problem, showing how to use the GFD method to solve numerically the transport equation in a slab with anisotropic scattering kernel and specular reflection at the boundary.

In Sect. 2.2 we present the problem, solve, and reformulate it into an integral operator equation. In Sect. 2.3, we describe the discretization of the integral operators, resulting in a finite approximation of the problem, which we solve numerically. In Sect. 2.4, we present numerical results for a broad range of applications.

2.2 Reformulation as an Integral Equation

We consider the following transport equation with anisotropic scattering:

$$\mu\frac{\partial I}{\partial y} + \lambda I = \frac{1}{2}\int_{-1}^{1}\omega(\mu',\mu)I(y,\mu')d\mu' + S, \qquad y \in (0,L), t > 0 \tag{2.3a}$$

$$I(y,\mu) = \rho_0 I(y,-\mu) + (1-\rho_0)B_0(\mu), \qquad y = 0,\ \mu > 0, \tag{2.3b}$$

$$I(y,\mu) = \rho_L I(y,-\mu) + (1-\rho_L)B_L(\mu), \qquad y = L,\ \mu < 0, \tag{2.3c}$$

where $I = I(y,\mu)$ is a radiative intensity, $S = S(y,\mu)$ is a source, $B := (B_0, B_L)$ indicates the boundary condition, and $\omega(\mu,\mu')$ is the scattering kernel. In order to establish the existence theory, we assume the following hypothesis: exists $\omega_{max} < \lambda$ such that

$$|\omega(\mu,\mu')| \leq \omega_{max}. \tag{2.4}$$

and the function $B_0(\mu)$ and $B_L(\mu)$ are absolutely integrable, i.e.:

$$\int_0^1 |B_0(\mu)|d\mu < \infty \qquad \text{and} \qquad \int_{-1}^0 |B_L(\mu)|d\mu < \infty. \tag{2.5}$$

In order to establish the existence of unique solution for the problem (2.3a)–(2.3c) and derive from this analysis the operator formulation for the problem, we firstly show that $J(y,\mu)$, defined by

$$J(y,\mu) := \frac{1}{2}\int_{-1}^1 \omega(\mu,\mu')I(y,\mu')d\mu', \tag{2.6}$$

admits the representation

$$J(y,\mu) = S_g S(y,\mu) + S_b B, \tag{2.7}$$

where S_b e S_g are operators in $C^0([0,L],L^\infty[-1,1])$.

Theorem 1. *If $B_0(\mu)$ and $B_L(\mu)$ are integrable and the condition (2.4), is satisfied then (2.3a)–(2.3c) admits the representation (2.7), where S_g and S_b are operators in $C^0([0,L],L^\infty[-1,1])$.*

Proof. We consider the following auxiliary problem:

$$\begin{aligned}
&\mu\frac{\partial I(y,\mu)}{\partial y} + \lambda I(y,\mu) = q(y,\mu)\\
&I(y,\mu) = \rho_0 I(y,-\mu) + (1-\rho_0)B_0(\mu), \qquad y = 0,\ \mu > 0,\\
&I(y,\mu) = \rho_L I(y,-\mu) + (1-\rho_L)B_L(\mu), \qquad y = L,\ \mu < 0.
\end{aligned} \tag{2.8}$$

This problem can be solved by the method of ray tracing (see [Mo03], [AzEtAl11b], [Si95], [Si93], and [BeGl70]) which consists in integrating the transport equation along the ray direction. The equation

$$\frac{\partial I(y,\mu)}{\partial y} + \frac{\lambda}{\mu}I(y,\mu) = \frac{1}{\mu}q(y,\mu)$$

can be rewritten in the form

$$\frac{\partial}{\partial y}\left(e^{\frac{\lambda y}{\mu}} I(y,\mu)\right) = \frac{e^{\frac{\lambda y}{\mu}}}{\mu} q(y,\mu).$$

Now we integrate this equation from in the intervals $(0,y)$ and (y,L) in order to obtain the expressions

$$I(y,\mu) = I(0,\mu)e^{-\frac{\lambda y}{\mu}} + \frac{1}{\mu}\int_0^y q(s,\mu)e^{\frac{\lambda(s-y)}{\mu}} ds, \tag{2.9a}$$

$$I(y,\mu) = I(L,\mu)e^{\frac{\lambda(L-y)}{\mu}} - \frac{1}{\mu}\int_y^L q(s,\mu)e^{\frac{\lambda(s-y)}{\mu}} ds. \tag{2.9b}$$

We note that these expressions are valid for both $\mu > 0$ and $\mu < 0$. Nonetheless, we will favorite the solution constructed by integrating in the direction of the ray, that is to say, we will favorite (2.9a) for $\mu > 0$ and (2.9b) for $\mu < 0$. Applying the boundary conditions given by

$$\begin{aligned} I(0,\mu) &= \rho_0(\mu)I(0,-\mu) + (1-\rho_0(\mu))B_0(\mu), \qquad \mu > 0 \\ I(L,\mu) &= \rho_L(\mu)I(L,-\mu) + (1-\rho_L(\mu))B_L(\mu), \qquad \mu < 0 \end{aligned}$$

to (2.9a) and (2.9b), we obtain

$$\begin{aligned} I(0,\mu) = \rho_0(\mu)I(L,-\mu)e^{-\frac{\lambda L}{\mu}} &+ \frac{\rho_0(\mu)}{\mu}\int_0^L q(s,-\mu)e^{-\frac{\lambda s}{\mu}} ds \\ &+ (1-\rho_0(\mu))B_0(\mu), \qquad t > 0,\ \mu > 0 \end{aligned}$$

$$\begin{aligned} I(L,\mu) = \rho_L(\mu)I(0,-\mu)e^{\frac{\lambda L}{\mu}} &- \frac{\rho_L(\mu)}{\mu}\int_0^L q(s,-\mu)e^{-\frac{\lambda(s-L)}{\mu}} ds \\ &+ (1-\rho_L(\mu))B_L(\mu), \qquad t > 0,\ \mu < 0 \end{aligned}$$

We now substitute μ by $-\mu$ in the last expression in order to have a linear system in $I(0,\mu)$ and $I(L,-\mu)$ valid for $\mu > 0$, as follows:

$$\begin{bmatrix} 1 & -\rho_0(\mu)e^{-\frac{\lambda L}{\mu}} \\ -\rho_L(-\mu)e^{-\frac{\lambda L}{\mu}} & 1 \end{bmatrix} \begin{bmatrix} I(0,\mu) \\ I(L,-\mu) \end{bmatrix} = \begin{bmatrix} \frac{\rho_0(\mu)}{\mu}\int_0^L q(s,-\mu)e^{-\frac{\lambda s}{\mu}} ds \\ +(1-\rho_0(\mu))B_0(\mu) \\ \frac{\rho_L(-\mu)}{\mu}\int_0^L q(s,\mu)e^{\frac{\lambda(s-L)}{\mu}} ds \\ +(1-\rho_L(-\mu))B_L(-\mu) \end{bmatrix}. \tag{2.10}$$

This system has a unique solution since the determinant of the matrix involved is not zero due to the estimate

$$1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}} \geq 1-e^{-\frac{2\lambda L}{\mu}} > 0.$$

The solution of (2.10) is given by

$$I(0,\mu) = \frac{\frac{\rho_0(\mu)}{\mu}\int_0^L \left(q(s,\mu)\rho_L(-\mu)e^{\frac{\lambda(s-2L)}{\mu}} + q(s,-\mu)e^{-\frac{\lambda s}{\mu}}\right) ds}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}} \tag{2.11a}$$

$$+ \frac{(1-\rho_0(\mu))B_0(\mu) + e^{-\frac{\lambda L}{\mu}}\rho_0(\mu)(1-\rho_L(-\mu))B_L(-\mu)}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}}$$

$$I(L,-\mu) = \frac{\frac{\rho_L(-\mu)}{\mu}\int_0^L \left(q(s,\mu)e^{\frac{\lambda(s-L)}{\mu}} + q(s,-\mu)\rho_0(\mu)e^{-\frac{\lambda(s+L)}{\mu}}\right) ds}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}} \tag{2.11b}$$

$$+ \frac{(1-\rho_L(-\mu))B_L(-\mu) + e^{-\frac{\lambda L}{\mu}}\rho_L(-\mu)(1-\rho_0(\mu))B_0(\mu)}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}}.$$

We substitute (2.11a) and (2.11b) into (2.9a) and (2.9b), respectively, in order to write $I(y,\mu)$ as

$$I(y,-\mu) = \frac{\frac{\rho_L(-\mu)}{\mu}\int_0^L \left(q(s,-\mu)e^{\frac{\lambda(s-L)}{\mu}} + q(s,\mu)\rho_0(\mu)e^{-\frac{\lambda(s+L)}{\mu}}\right) ds}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}} e^{-\frac{\lambda(L-y)}{\mu}}$$

$$+ \frac{(1-\rho_L(-\mu))B_L(-\mu) + e^{-\frac{\lambda L}{\mu}}\rho_L(-\mu)(1-\rho_0(\mu))B_0(\mu)}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}} e^{-\frac{\lambda(L-y)}{\mu}}$$

$$+ \frac{1}{\mu}\int_y^L q(s,-\mu)e^{\frac{-\lambda(s-y)}{\mu}} ds, \quad \mu > 0, \tag{2.12a}$$

$$I(y,\mu) = \frac{\frac{\rho_0(\mu)}{\mu}\int_0^L \left(q(s,\mu)\rho_L(-\mu)e^{\frac{\lambda(s-2L)}{\mu}} + q(s,-\mu)e^{-\frac{\lambda s}{\mu}}\right) ds}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}} e^{-\frac{\lambda y}{\mu}} \tag{2.12b}$$

$$+ \frac{(1-\rho_0(\mu))B_0(\mu) + e^{-\frac{\lambda L}{\mu}}\rho_0(\mu)(1-\rho_L(-\mu))B_L(-\mu)}{1-\rho_0(\mu)\rho_L(-\mu)e^{-\frac{2\lambda L}{\mu}}} e^{-\frac{\lambda y}{\mu}}$$

$$+ \frac{1}{\mu}\int_0^y q(s,\mu)e^{\frac{\lambda(s-y)}{\mu}} ds, \quad \mu > 0.$$

We now interpret (2.12a) and (2.12b) as two integral operators acting on the functions $q(y,\mu)$ and $B(\mu)$

$$I(y,\mu) = L_g^{\mu} q(y,\mu) + L_b^{\mu} B(\mu). \tag{2.13}$$

The superscript μ in L_g^{μ} and L_b^{μ} explicits the dependence of the operators on the angular variable μ.

We now go back to our original problem (2.3) and observe that it takes the form of the auxiliary problem (2.8) if

$$q(y,\mu) = J(y,\mu) + S(y,\mu).$$

Taking into account the representation (2.13), the solution of (2.3) must satisfy

$$I(y,\mu) = L_g^{\mu}[J(y,\mu) + S(y,\mu)] + L_b^{\mu} B(\mu). \tag{2.14}$$

We recall that $J(y,\mu)$ was defined in (2.6) as

$$J(y,\mu) = \frac{1}{2}\int_{-1}^{1} \omega(\mu,\mu') I(y,\mu') d\mu'$$

We now substitute μ by μ' into (2.14), multiply it by $\omega(\mu,\mu')$, and integrate on $\mu' \in (-1,1)$:

$$\begin{aligned} J(y,\mu) &= \frac{1}{2}\int_{-1}^{1} \omega(\mu,\mu') \left\{ L_g^{\mu'}[J(y,\mu') + S(y,\mu')] + L_b^{\mu'} B(\mu') \right\} d\mu', \\ &= L_g[J(y,\mu) + S(y,\mu)] + L_b B(\mu), \end{aligned} \tag{2.15}$$

where the operators L_g and L_b are given by

$$\begin{aligned} L_g &= \frac{1}{2}\int_{-1}^{1} \omega(\mu,\mu') L_g^{\mu'} d\mu' \quad \text{and} \\ L_b &= \frac{1}{2}\int_{-1}^{1} \omega(\mu,\mu') L_b^{\mu'} d\mu' \end{aligned} \tag{2.16}$$

Now we write (2.15) as

$$J(y,\mu) - L_g J(y,\mu) = L_g S(y,\mu) + L_b B(\mu),$$

i.e.,

$$(1 - L_g) J(y,\mu) = L_g S(y,\mu) + L_b B(\mu). \tag{2.17}$$

This equation may be solved whenever the inverse of $(1 - L_g)$ exists and its solution is given by

$$J(y,\mu) = S_g S(y,\mu) + S_b B(\mu) \tag{2.18}$$

with

$$S_g := (1-L_g)^{-1} L_g \quad \text{and} \quad S_b := (1-L_g)^{-1} L_b. \tag{2.19}$$

We know that the inverse required here exists provided the norm of L_g is less than 1. That said, we will estimate the norm of this operator under conditions (2.4) and (2.5). The operator L_g is given explicitly by

$$\begin{aligned}
L_g q = &\frac{1}{2}\int_0^1 \omega(-\mu',\mu)\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\frac{\rho_L e^{-\frac{\lambda(L-y)}{\mu'}}}{\mu'}\times\right.\\
&\left.\times\int_0^L\left(q(s,-\mu')\rho_0 e^{-\frac{\lambda(s+L)}{\mu'}}+q(s,\mu')e^{\frac{\lambda(s-L)}{\mu'}}\right)ds+\frac{1}{\mu'}\int_y^L q(s,-\mu')e^{-\frac{\lambda(s-y)}{\mu'}}ds\right]d\mu'\\
&+\frac{1}{2}\int_0^1 \omega(\mu',\mu)\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\frac{\rho_0 e^{-\frac{\lambda y}{\mu'}}}{\mu'}\times\right.\\
&\left.\times\int_0^L\left(q(s,\mu')\rho_L e^{\frac{\lambda(s-2L)}{\mu'}}+q(s,-\mu')e^{-\frac{\lambda s}{\mu'}}\right)ds+\frac{1}{\mu'}\int_0^y q(s,\mu')e^{\frac{\lambda(s-y)}{\mu'}}ds\right]d\mu'\\
:= &A+B.
\end{aligned}\tag{2.20}$$

In order to abbreviate the notation we omitted the dependence on μ of $\rho_0=\rho_0(\mu)$ and $\rho_L=\rho_L(-\mu)$. We note that $|L_g|\le|A|+|B|$, where

$$\begin{aligned}
|A| \le &\frac{||q||_{C^0}\omega_{max}}{2}\int_0^1\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\frac{\rho_L e^{-\frac{\lambda(L-y)}{\mu'}}}{\mu'}\int_0^L\left(\rho_0 e^{-\frac{\lambda(s+L)}{\mu'}}+e^{\frac{\lambda(s-L)}{\mu'}}\right)ds\right.\\
&\left.+\frac{1}{\mu'}\int_y^L e^{-\frac{\lambda(s-y)}{\mu'}}ds\right]d\mu'\\
= &\frac{||q||_{C^0}\omega_{max}}{2}\int_0^1\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\frac{\rho_L e^{-\frac{\lambda(L-y)}{\mu'}}}{\mu'}\frac{\mu'}{\lambda}\left(-\rho_0 e^{-\frac{\lambda 2L}{\mu'}}+1+\right.\right.\\
&\left.\left.+\rho_0 e^{-\frac{\lambda L}{\mu'}}-e^{-\frac{\lambda L}{\mu'}}\right)+\frac{1}{\mu'}\frac{\mu'}{\lambda}\left(-e^{-\frac{\lambda(L-y)}{\mu'}}+1\right)\right]d\mu'\\
= &\frac{||q||_{C^0}\omega_{max}}{2\lambda}\int_0^1\left[\frac{\rho_L e^{-\frac{\lambda(L-y)}{\mu'}}\left(-\rho_0 e^{-\frac{\lambda 2L}{\mu'}}+1+\rho_0 e^{-\frac{\lambda L}{\mu'}}-e^{-\frac{\lambda L}{\mu'}}\right)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right.\\
&\left.+\left(-e^{-\frac{\lambda(L-y)}{\mu'}}+1\right)\right]d\mu'
\end{aligned}$$

$$
\begin{aligned}
&= \frac{||q||_{C^0}\omega_{max}}{2\lambda}\left(1+\int_0^1\left[\frac{-\rho_0\rho_L e^{-\frac{\lambda(3L-y)}{\mu'}}+\rho_0 e^{-\frac{\lambda(L-y)}{\mu'}}+\rho_0\rho_L e^{-\frac{\lambda(2L-y)}{\mu'}}-e^{-\frac{\lambda(2L-y)}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right.\right.\\
&\qquad\left.\left.+\frac{-e^{-\frac{\lambda(L-y)}{\mu'}}+\rho_0\rho_L e^{-\frac{\lambda(3L-y)}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right]d\mu'\right)\\
&= \frac{||q||_{C^0}\omega_{max}}{2\lambda}+\frac{||q||_{C^0}\omega_{max}}{2\lambda}\int_0^1\left[\frac{(\rho_0-1)e^{-\frac{\lambda(L-y)}{\mu'}}+\rho_0(\rho_L-1)e^{-\frac{\lambda(2L-y)}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right]d\mu'\\
&= \frac{||q||_{C^0}\omega_{max}}{2\lambda}-f_A, \qquad (2.21)
\end{aligned}
$$

with

$$
f_A := \frac{||q||_{C^0}\omega_{max}}{2\lambda}\int_0^1\left[\frac{(1-\rho_0)e^{-\frac{\lambda(L-y)}{\mu'}}+\rho_0(1-\rho_L)e^{-\frac{\lambda(2L-y)}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right]d\mu'
$$

and

$$
\begin{aligned}
|B| &\le \frac{||q||_{C^0}\omega_{max}}{2}\int_0^1\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\frac{\rho_0 e^{-\frac{\lambda y}{\mu'}}}{\mu'}\int_0^L\left(\rho_L e^{\frac{\lambda(s-2L)}{\mu'}}+e^{-\frac{\lambda s}{\mu'}}\right)ds\right.\\
&\qquad\left.+\frac{1}{\mu'}\int_0^y e^{\frac{\lambda(s-y)}{\mu'}}ds\right]d\mu'\\
&= \frac{||q||_{C^0}\omega_{max}}{2}\int_0^1\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\frac{\rho_0 e^{-\frac{\lambda y}{\mu'}}}{\mu'}\frac{\mu'}{\lambda}\left(\rho_L e^{-\frac{\lambda L}{\mu'}}-e^{-\frac{\lambda L}{\mu'}}-\right.\right.\\
&\qquad\left.\left.-\rho_L e^{-\frac{\lambda 2L}{\mu'}}+1\right)+\frac{1}{\mu'}\frac{\mu'}{\lambda}\left(1-e^{-\frac{\lambda y}{\mu'}}\right)\right]d\mu'\\
&= \frac{||q||_{C^0}\omega_{max}}{2\lambda}\left(\int_0^1 d\mu'+\int_0^1\left[\frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\rho_0 e^{-\frac{\lambda y}{\mu'}}\left(\rho_L e^{-\frac{\lambda L}{\mu'}}-e^{-\frac{\lambda L}{\mu'}}\right.\right.\right.\\
&\qquad\left.\left.\left.-\rho_L e^{-\frac{\lambda 2L}{\mu'}}+1\right)+\left(-e^{-\frac{\lambda y}{\mu'}}\right)\right]d\mu'\right)\\
&= \frac{||q||_{C^0}\omega_{max}}{2\lambda}\left(1+\int_0^1\left[\frac{\rho_0\rho_L e^{-\frac{\lambda(y+L)}{\mu'}}-\rho_L e^{-\frac{\lambda(y+L)}{\mu'}}-\rho_0\rho_L e^{-\frac{\lambda(y+2L)}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}+\right.\right.\\
&\qquad\left.\left.+\frac{\rho_L e^{-\frac{\lambda y}{\mu'}}-e^{-\frac{\lambda y}{\mu'}}+\rho_0\rho_L e^{-\frac{\lambda(y+2L)}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right]d\mu'\right)
\end{aligned}
$$

$$= \frac{||q||_{C^0}\omega_{max}}{2\lambda} + \frac{||q||_{C^0}\omega_{max}}{2\lambda}\int_0^1 \left[\frac{(\rho_0-1)\rho_L e^{-\frac{\lambda(y+L)}{\mu'}} + (\rho_L-1)e^{-\frac{\lambda y}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right] d\mu'$$

$$= \frac{||q||_{C^0}\omega_{max}}{2\lambda} - f_B, \tag{2.22}$$

with

$$f_B := \frac{||q||_{C^0}\omega_{max}}{2\lambda}\int_0^1 \left[\frac{(1-\rho_0)\rho_L e^{-\frac{\lambda(y+L)}{\mu'}} + (1-\rho_L)e^{-\frac{\lambda y}{\mu'}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu'}}}\right] d\mu'.$$

Therefore, using the condition (2.4), we obtain

$$||L_g q||_{C^0} \leq |A| + |B| \leq \frac{||q||_{C^0}\omega_{max}}{\lambda} - \inf(f_A + f_B) \leq \frac{||q||_{C^0}\omega_{max}}{\lambda} \leq ||q||_{C^0}. \tag{2.23}$$

We observe that $0 < \rho < 1$, $\inf(f_A + f_B) > 0$ and the strict inequality holds:

$$||L_g q||_{C^0} < \frac{||q||_{C^0}\omega_{max}}{\lambda} \leq ||q||_{C^0}.$$

2.3 Methodology

In this section we describe the discretization of the operators L_g and L_b. For better comprehension, we firstly deal with the isotropic case in Sect. 2.3.1. In Sect. 2.3.2 we explain the generalization for the anisotropic case.

2.3.1 The Isotropic Case

In this section we will consider the scattering kernel to be isotropic, i.e,

$$\omega(\mu, \mu') = \sigma,$$

where σ is a constant and $\sigma < \lambda$ so that L_g assumes the integral representation

$$(L_g q)(y) = \int_0^L K(s,y)q(s)ds \tag{2.24}$$

where

$$K(s,y) = \frac{1}{2}\int_0^1 \left[\frac{\rho_0}{\mu}q(s)\frac{\rho_L e^{-\frac{\lambda(s-y+2L)}{\mu}} + e^{\frac{\lambda(s+y-2L)}{\mu}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\right] d\mu$$
$$+ \frac{1}{2}\int_0^1 \left[\frac{\rho_L}{\mu}q(s)\frac{\rho_0 e^{\frac{\lambda(s-y-2L)}{\mu}} + e^{-\frac{\lambda(s+y)}{\mu}}}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}} + \frac{1}{\mu}q(s)e^{-\frac{\lambda|s-y|}{\mu}}\right] d\mu \quad (2.25)$$

In order to construct a discretized version of S_g and S_b, we first represent a continuous function $q:[0,L]\to\mathbb{R}$ as a continuous piecewise linear function given by

$$\tilde{q}(y) = \frac{s_{j+1}-y}{h_s}q(s_j) + \frac{y-s_j}{h_s}q(s_{j+1}),\ s_j \le y \le s_{j+1} \quad (2.26)$$

where $\{s_j\}_{j=1}^{N+1}$ is a uniform mesh consisting of N subintervals whose extreme points are given by $s_j=(j-1)h_s$ with $h_s=\frac{L}{N}$. We define the interpolation operator $I_N: C^0[0,L]\to C^0[0,L]$ mapping $q(s)$ to $\tilde{q}(s)$ in (2.26). The range of I_N on $C^0[0,L]$ is the $N+1$-dimensional space of first-order finite elements, which will denote $C_N^0([0,L]))$. Since $I_N q\to q$ for each $q\in C^0([0,L])$, i.e, I_N converges to the identity in the strong topology, it is natural to approximate L_g by $L_g^N := I_N L_g I_N$. Due to the natural isomorphism φ_N between $C_N^0([0,L]))$ and $\mathbb{R}^{N+1}$ given by

$$\varphi_N : C_N^0([0,L])) \longleftrightarrow \mathbb{R}^{N+1}$$
$$q(y) \longleftrightarrow [q(s_0),\ldots,q(s_N),q(s_{N+1})]^T,$$

we represent L_g^N by the matrix

$$W_0 := \varphi_N L_g^N \varphi_N^{-1} = \varphi_N I_N L_g I_N \varphi_N^{-1} =: (w_{ij})_{i,j=1}^{N+1},$$

where the entries w_{ij} are calculated using the integral representation (2.24) of L_g, as follows:

$$K_g^{l,k}(I_N q) = \int_0^L K^{l,k}(s,y)\tilde{q}(s)ds$$
$$= \sum_{j=1}^N \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)\tilde{q}(s)ds$$
$$= \sum_{j=1}^N \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)\left(q(s_j) + (s-s_j)\frac{q(s_{j+1})-q(s_j)}{h_s}\right)ds$$
$$= \sum_{j=1}^N q(s_j)\left(1+\frac{s_j}{h_s}\right)\int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds - \sum_{j=1}^N \frac{q(s_j)}{h_s}\int_{s_j}^{s_{j+1}} sK^{l,k}(s,y)ds$$

$$
\begin{aligned}
&- \sum_{j=1}^{N} \left(s_j \frac{q(s_{j+1})}{h_s} \right) \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds + \sum_{j=1}^{N} \frac{q(s_{j+1})}{h_s} \int_{s_j}^{s_{j+1}} sK^{l,k}(s,y)ds \\
&= \sum_{j=1}^{N} q(s_j) \left[\left(1 + \frac{s_j}{h_s} \right) \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds - \frac{1}{h_s} \int_{s_j}^{s_{j+1}} sK^{l,k}(s,y)ds \right] \\
&+ \sum_{j=1}^{N} q(s_{j+1}) \left[-\frac{s_j}{h_s} \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds + \frac{1}{h_s} \int_{s_j}^{s_{j+1}} sK^{l,k}(s,y)ds \right] \\
&= \sum_{j=1}^{N} q(s_j) \left[\left(1 + \frac{s_j}{h_s} \right) \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds - \frac{1}{h_s} \int_{s_j}^{s_{j+1}} sK^{l,k}(s,y)ds \right] \\
&+ \sum_{j=2}^{N+1} q(s_j) \left[-\frac{s_{j-1}}{h_s} \int_{s_{j-1}}^{s_j} K^{l,k}(s,y)ds + \frac{1}{h_s} \int_{s_{j-1}}^{s_j} sK^{l,k}(s,y)ds \right] \\
&= \sum_{j=1}^{N+1} w_j^{l,k}(y) q(s_j),
\end{aligned}
$$

where

$$
\begin{aligned}
w_1^{l,k} &= \left(1 + \frac{s_1}{h_s} \right) \int_{s_1}^{s_2} K^{l,k}(s,y)ds - \frac{1}{h_s} \int_{s_1}^{s_2} sK^{l,k}(s,y)ds \\
w_j^{l,k} &= \left(1 + \frac{s_j}{h_s} \right) \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds - \frac{1}{h_s} \int_{s_j}^{s_{j+1}} sK^{l,k}(s,y)ds \\
&\quad - \frac{s_{j-1}}{h_s} \int_{s_{j-1}}^{s_j} K^{l,k}(s,y)ds + \frac{1}{h_s} \int_{s_{j-1}}^{s_j} sK^{l,k}(s,y)ds \qquad \text{se } 2 \le j \le N \\
w_{N+1}^{l,k} &= -\frac{s_N}{h_s} \int_{s_N}^{s_{N+1}} K^{l,k}(s,y)ds + \frac{1}{h_s} \int_{s_N}^{s_{N+1}} sK^{l,k}(s,y)ds
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
w_1^{l,k} &= \int_{s_1}^{s_2} K^{l,k}(s,y)ds - \frac{1}{h_s} \int_{s_1}^{s_2} sK^{l,k}(s,y)ds \\
w_j^{l,k} &= \int_{s_j}^{s_{j+1}} K^{l,k}(s,y)ds - \frac{1}{h_s} \int_{s_j}^{s_{j+1}} (s - s_j) K^{l,k}(s,y)ds \\
&\quad + \frac{1}{h_s} \int_{s_{j-1}}^{s_j} (s - s_{j-1}) K^{l,k}(s,y)ds \qquad \text{se } 2 \le j \le N \qquad (2.27) \\
w_{N+1}^{l,k} &= \frac{1}{h_s} \int_{s_N}^{s_{N+1}} (s - s_N) K^{l,k}(s,y)ds
\end{aligned}
$$

The operator $L_b^N = I_N L_b I_N$ is obtained directly from L_b:

$$L_b B = \frac{1}{2}\int_0^1 \frac{e^{-\frac{\lambda L}{\mu}}\rho_L(1-\rho_0)B_0 + (1-\rho_L)B_L}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}} e^{-\frac{\lambda(L-y)}{\mu}} d\mu$$
$$+\frac{1}{2}\int_0^1 \frac{e^{-\frac{\lambda L}{\mu}}\rho_0(1-\rho_L)B_L + (1-\rho_0)B_0}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}} e^{-\frac{\lambda y}{\mu}} d\mu, \tag{2.28}$$

and its associated $(N+1)\times 2$ matrix $V_0 := (v_{ij})_{i=1,j=1}^{N+1,2} = \phi_N L_b^N \phi_N^{-1}$:

$$v_{i,1} = \frac{1}{2}\int_0^1 \frac{e^{-\frac{\lambda(2L-y_i)}{\mu}}\rho_L(1-\rho_0) + e^{-\frac{\lambda y_i}{\mu}}(1-\rho_0)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}} d\mu,$$
$$v_{i,2} = \frac{1}{2}\int_0^1 \frac{e^{-\frac{\lambda(y_i-L)}{\mu}}\rho_0(1-\rho_L) + e^{-\frac{\lambda(L-y_i)}{\mu}}(1-\rho_L)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}} d\mu.$$

The matrices W_σ and V_σ approximating the operators S_g and S_b are obtained, respectively, from (2.19):

$$W_\sigma := (1-\sigma W_0)^{-1} W_0 \quad \text{and} \quad V_\sigma := (1-\sigma W_0)^{-1} V_0.$$

Note that this notation is consistent when $\sigma = 0$.

Once the matrices W_σ and V_σ have been constructed, they are used as vector Green's functions for the transport equation, allowing us to solve the unknown I with a matrix vector multiplication.

2.3.2 The Anisotropic Case

Here we will consider the following anisotropic scattering kernel:

$$\omega(\mu', \mu) = \sum_{l=0}^{M} \beta_l P_l(\mu) P_l(\mu'), \tag{2.29}$$

where $P_l(\mu)$ is the lth Legendre polynomial and β_l are constants.

The kernel (2.29) allows us to define the moments $J_l(y)$ by

$$J_l(y) = \frac{1}{2}\int_{-1}^{1} P_l(\mu') I(y, \mu') d\mu'. \tag{2.30}$$

Using (2.6), the unknown variable $J(y, \mu)$ is written as

$$J(y, \mu) = \sum_{l=0}^{M} \beta_l J_l(y) P_l(\mu), \tag{2.31}$$

and $J_l(y)$ as

$$J_l(y) = \frac{1}{2}\sum_{k=0}^{M} \beta_k K_g^{l,k} J_k(y) + K_g^{l,0} S(y) + \frac{1}{2} K_b^l B \tag{2.32}$$

where the operators $K_g^{l,k} : C^0[0,L] \to C^0[0,L]$ and $K_b^l : (L^\infty[0,1] \times L^\infty[-1,0]) \to C^0[0,L]$ are given by

$$K_g^{l,k} q(y) = \frac{1}{2}\int_{-1}^{1} P_l(\mu') L_g^{\mu'} \left[q(y) P_k(\mu')\right] d\mu' \tag{2.33}$$

$$K_b^l B = \frac{1}{2}\int_{-1}^{1} P_l(\mu') L_b^{\mu'} B d\mu'. \tag{2.34}$$

The expression (2.32) can be represented by the system

$$\begin{bmatrix} 1-\beta_0 K_g^{0,0} & \cdots & -\beta_M K_g^{0,M} \\ -\beta_0 K_g^{1,0} & \cdots & -\beta_M K_g^{1,M} \\ \vdots & \ddots & \vdots \\ -\beta_0 K_g^{M,0} & \cdots & 1-\beta_M K_g^{M,M} \end{bmatrix} \begin{bmatrix} J_0(y) \\ J_1(y) \\ \vdots \\ J_M(y) \end{bmatrix} = \begin{bmatrix} K_g^{0,0} S(y) + K_b^0 B \\ K_g^{1,0} S(y) + K_b^1 B \\ \vdots \\ K_g^{M,0} S(y) + K_b^M B \end{bmatrix} \tag{2.35}$$

Now, each operator $K^{l,k}$ and K_b^l is discretized in a real matrix $(N+1)\times(N+1)$ and the system (2.35) is solved as a linear system of dimension $(M+1)(N+1)$. The approximation $W^{l,k}$ for $K_g^{l,k}$ and V^l for K_b^l are constructed similarly these of isotropic case.

2.3.3 The Calculation of the Coefficient of $W^{l,k}$ (and W_σ)

We observe that the kernels $K^{l,k}(s,y)$ and $sK^{l,k}(s,y)$ can be integrated analytically on s. Furthermore, since $K(s,y)$ can be decomposed in the form

$$\begin{aligned} K^{l,k}(s,y) = & \\ = & \int_0^1 P_l(-\mu) P_k(-\mu) h_0(y-s+L,\mu) d\mu + \int_0^1 P_l(\mu) P_k(\mu) h_0(s-y+L,\mu) d\mu \\ + & \int_0^1 P_l(-\mu) P_k(\mu) \rho_L h_1(s+y,\mu) d\mu + \int_0^1 P_l(\mu) P_k(-\mu) \rho_0 h_1(-y-s+2L,\mu) d\mu \\ + & \int_0^1 P_l(-Si(s-y)\ \mu) P_k(-Si(s-y)\ \mu) h_2(|s-y|,\mu) d\mu, \end{aligned} \tag{2.36}$$

where

$$h_0(y,\mu) = \frac{\rho_0\rho_L}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\frac{e^{\frac{\lambda(y-3L)}{\mu}}}{2\mu}, \qquad 0 \le y \le 2L \tag{2.37}$$

$$h_1(y,\mu) = \frac{1}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\frac{e^{\frac{\lambda(y-2L)}{\mu}}}{2\mu}, \qquad 0 \le y \le 2L \tag{2.38}$$

$$h_2(y,\mu) = \frac{e^{-\frac{\lambda y}{\mu}}}{2\mu}, 0 \le y \le L \tag{2.39}$$

the number of defined integrals to calculate the coefficients of W grows up with N. In order to explicit this fact, we combine the coefficients of the matrix W (2.27) with the expression (2.36) resulting in a list of double integrals. Here we omit the index k and l and use the following notation:

$$f_0^-(y,\mu) := P_l(-\mu)P_k(-\mu)h_0(y,\mu), \qquad f_0^+(y,\mu) = P_l(\mu)P_k(\mu)h_0(y,\mu)$$
$$f_1^-(y,\mu) := P_l(-\mu)P_k(\mu)\rho_L h_1(y,\mu), \qquad f_1^+(y,\mu) = P_l(\mu)P_k(-\mu)\rho_0 h_1(y,\mu)$$
$$f_2^-(y,\mu) := P_l(-\mu)P_k(-\mu)h_2(y,\mu), \qquad f_2^+(y,\mu) = P_l(\mu)P_k(\mu)h_2(y,\mu)$$

and we obtain the terms of the sum

$$\int_{s_j}^{s_{j+1}}\int_0^1 f_0^{\pm}(s_i - s + L,\mu)d\mu ds = \int_0^1\int_{(N+i-j-1)h_s}^{(N+i-j)h_s} f_0^{\pm}(y,\mu)dyd\mu$$
$$= \int_0^1\int_{(k-1)h_s}^{kh_s} f_0^{\pm}(y,\mu)dyd\mu, \quad k = N+i-j,$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_0^{\pm}(s_i - s + L,\mu)d\mu ds$$
$$= \int_0^1\int_{(N+i-j-1)h_s}^{(N+i-j)h_s}(s_i - y + L - s_j)f_0^{\pm}(y,\mu)dyd\mu$$
$$= \int_0^1\int_{(k-1)h_s}^{kh_s}(kh_s - y)f_0^{\pm}(y,\mu)dyd\mu, \quad k = N+i-j,$$

$$\int_{s_j}^{s_{j+1}}\int_0^1 f_0^{\pm}(s - s_i + L,\mu)d\mu ds = \int_0^1\int_{(N+j-i)h_s}^{(N+j-i+1)h_s} f_0^{\pm}(y,\mu)dyd\mu$$
$$= \int_0^1\int_{(k-1)h_s}^{kh_s} f_0^{\pm}(y,\mu)dyd\mu, \quad k = N+j-i+1,$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_0^{\pm}(s-s_i+L,\mu)d\mu ds$$
$$=\int_0^1\int_{(N+j-i)h_s}^{(N+j-i+1)h_s}(y+s_i-L-s_j)f_0^{\pm}(y,\mu)dyd\mu$$
$$=\int_0^1\int_{(k-1)h_s}^{kh_s}(y-(k-1)h_s)f_0^{\pm}(y,\mu)dyd\mu,\quad k=N+j-i+1,$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_0^{\pm}(s-s_i+L,\mu)d\mu ds$$
$$=-\int_0^1\int_{(k-1)h_s}^{kh_s}(kh_s-y)f_0^{\pm}(y,\mu)dyd\mu+h_s\int_0^1\int_{(k-1)h_s}^{kh_s}f_0^{\pm}(y,\mu)dyd\mu,$$
$$k=N+j-i+1,$$

$$\int_{s_j}^{s_{j+1}}\int_0^1 f_1^{\pm}(s+s_i,\mu)d\mu ds=\int_0^1\int_{(j+i-2)h_s}^{(j+i-1)h_s}f_1^{\pm}(y,\mu)dyd\mu$$
$$=\int_0^1\int_{(k-1)h_s}^{kh_s}f_1^{\pm}(y,\mu)dyd\mu,\quad k=j+i-1,$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_1^{\pm}(s+s_i,\mu)d\mu ds$$
$$=\int_0^1\int_{(j+i-2)h_s}^{(j+i-1)h_s}(y-(j+i-2)h_s)f_1^{\pm}(y,\mu)dyd\mu$$
$$=\int_0^1\int_{(k-1)h_s}^{kh_s}(y-(k-1)h_s)f_1^{\pm}(y,\mu)dyd\mu,\quad k=j+i-1,$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_1^{\pm}(s+s_i,\mu)d\mu ds$$
$$=-\int_0^1\int_{(k-1)h_s}^{kh_s}(kh_s-y)f_1^{\pm}(y,\mu)dyd\mu+h_s\int_0^1\int_{(k-1)h_s}^{kh_s}f_1^{\pm}(y,\mu)dyd\mu,$$
$$k=j+i-1,$$

$$\int_{s_j}^{s_{j+1}}\int_0^1 f_1^{\pm}(-s_i-s+2L,\mu)d\mu ds=\int_0^1\int_{(2N-i-j+1)h_s}^{(2N-i-j+2)h_s}f_1^{\pm}(y,\mu)dyd\mu$$
$$=\int_0^1\int_{(k-1)h_s}^{kh_s}f_1^{\pm}(y,\mu)dyd\mu,$$
$$k=2N-i-j+2,$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_1^{\pm}(-s_i-s+2L,\mu)d\mu ds$$

$$=\int_0^1\int_{(2N-i-j+1)h_s}^{(2N-i-j+2)h_s}((2N-j-i+2)h_s-y)f_1^{\pm}(y,\mu)dyd\mu$$

$$=\int_0^1\int_{(k-1)h_s}^{kh_s}(kh_s-y)f_1^{\pm}(y,\mu)dyd\mu,\quad k=2N-i-j+2,$$

$$\int_{s_j}^{s_{j+1}}\int_0^1 f_2^{\pm}(|s-s_i|,\mu)d\mu ds$$

$$=\int_0^1\int_{(j-i)h_s}^{(j-i+1)h_s} f_2^{\pm}(|y|,\mu)dyd\mu$$

$$=\begin{cases}\displaystyle\int_0^1\int_{(j-i)h_s}^{(j-i+1)h_s} f_2^{-}(y,\mu)dyd\mu,\ j-i\geq 0\\ \displaystyle\int_0^1\int_{(i-j-1)h_s}^{(i-j)h_s} f_2^{+}(y,\mu)dyd\mu,\ \ i-j\geq 1\end{cases}$$

$$=\begin{cases}\displaystyle\int_0^1\int_{(k-1)h_s}^{kh_s} f_2^{-}(y,\mu)dyd\mu,\quad k=j-i+1,\ k\geq 1\\ \displaystyle\int_0^1\int_{(k-1)h_s}^{kh_s} f_2^{+}(y,\mu)dyd\mu,\quad k=i-j,\ k\geq 1\end{cases}$$

$$\int_{s_j}^{s_{j+1}}(s-s_j)\int_0^1 f_2^{\pm}(|s-s_i|,\mu)d\mu ds$$

$$=\int_0^1\int_{(j-i)h_s}^{(j-i+1)h_s}(y+(i-j)h_s)f_2^{\pm}(|y|,\mu)dyd\mu$$

$$=\begin{cases}\displaystyle\int_0^1\int_{(j-i)h_s}^{(j-i+1)h_s}(y+(i-j)h_s)f_2^{-}(y,\mu)dyd\mu,\quad j-i\geq 0\\ \displaystyle\int_0^1\int_{(i-j-1)h_s}^{(i-j)h_s}(-y+(i-j)h_s)f_2^{+}(y,\mu)dyd\mu,\quad i-j\geq 1\end{cases}$$

$$=\begin{cases}\displaystyle\int_0^1\int_{(k-1)h_s}^{kh_s}(y-(k-1)h_s)f_2^{-}(y,\mu)dyd\mu,\quad k=j-i+1,\ k\geq 1\\ \displaystyle\int_0^1\int_{(k-1)h_s}^{kh_s}(-y+kh_s)f_2^{+}(y,\mu)dyd\mu,\quad k=i-j,\ k\geq 1\end{cases}$$

$$=\begin{cases}\displaystyle\int_0^1\int_{(k-1)h_s}^{kh_s}(y-(k-1)h_s)f_2^{-}(y,\mu)dyd\mu,\quad k=j-i+1,\ k\geq 1\\ \displaystyle-\int_0^1\int_{(k-1)h_s}^{kh_s}(y-(k-1)h_s)f_2^{+}(y,\mu)dyd\mu\\ \displaystyle\quad +h_s\int_0^1\int_{(k-1)h_s}^{kh_s} f_2^{\pm}(y,\mu)dyd\mu,\quad k=i-j,\ k\geq 1.\end{cases}$$

This list of integrals motivates us to define the vectors

$$F_k^{0\pm} = \int_0^1 \int_{(k-1)h_s}^{kh_s} f_0^{\pm}(y,\mu)dyd\mu, \qquad 1 \le k \le 2N \tag{2.40a}$$

$$G_k^{0\pm} = \int_0^1 \int_{(k-1)h_s}^{kh_s} (kh_s - y) f_0^{\pm}(y,\mu)dyd\mu, \qquad 1 \le k \le 2N \tag{2.40b}$$

$$F_k^{1\pm} = \int_0^1 \int_{(k-1)h_s}^{kh_s} f_1^{\pm}(y,\mu)dyd\mu, \qquad 1 \le k \le 2N \tag{2.40c}$$

$$G_k^{1\pm} = \int_0^1 \int_{(k-1)h_s}^{kh_s} (kh_s - y) f_1^{\pm}(y,\mu)dyd\mu, \qquad 1 \le k \le 2N \tag{2.40d}$$

$$F_k^{2\pm} = \int_0^1 \int_{(k-1)h_s}^{kh_s} f_2^{\pm}(y,\mu)dyd\mu, \qquad 1 \le k \le N \tag{2.40e}$$

$$G_k^{2\pm} = \int_0^1 \int_{(k-1)h_s}^{kh_s} (y - (k-1)h_s) f_2^{\pm}(y,\mu)dyd\mu, \qquad 1 \le k \le N \tag{2.40f}$$

Therefore, the coefficients of W given by (2.27) are calculated for $1 \le i \le N+1$ as

$$\begin{aligned}
w_{i1}^{l,k} &= F_{N+i-1}^{0-} - \frac{1}{h_s}G_{N+i-1}^{0-} + \frac{1}{h_s}G_{N-i+2}^{0+} + \frac{1}{h_s}G_i^{1-} + F_{2N-i+1}^{1+} \\
&\quad - \frac{1}{h_s}G_{2N-i+1}^{1+} + \mathscr{F}_{i1}^2 - \frac{1}{h_s}\mathscr{G}_{i1}^2 \\
w_{ij}^{l,k} &= F_{N+i-j}^{0-} - \frac{1}{h_s}G_{N+i-j}^{0-} + \frac{1}{h_s}G_{N+i-j+1}^{0-} + \frac{1}{h_s}G_{N+j-i+1}^{0+} - \frac{1}{h_s}G_{N+j-i}^{0+} + F_{N+j-i}^{0+} \\
&\quad + \frac{1}{h_s}G_{j+i-1}^{1-} - \frac{1}{h_s}G_{j+i-2}^{1-} + F_{j+i-2}^{1-} + F_{2N-j-i+2}^{1+} - \frac{1}{h_s}G_{2N-j-i+2}^{1+} + \frac{1}{h_s}G_{2N-j-i+3}^{1+} \\
&\quad + \mathscr{F}_{ij}^2 - \frac{1}{h_s}\mathscr{G}_{ij}^2 + \frac{1}{h_s}\mathscr{G}_{i,j-1}^2, \qquad 2 \le j \le N \\
w_{i,N+1}^{l,k} &= \frac{1}{h_s}G_i^{0-} - \frac{1}{h_s}G_{2N+1-i}^{0+} + F_{2N+1-i}^{0+} - \frac{1}{h_s}G_{N+i-1}^{1-} + F_{N+i-1}^{1-} \\
&\quad + \frac{1}{h_s}G_{N-i+2}^{1+} + \frac{1}{h_s}\mathscr{G}_{N,j}^2
\end{aligned} \tag{2.41}$$

where

$$\mathscr{F}_{ij}^2 = \begin{cases} F_{j-i+1}^{2-}, & \text{se} \quad i \le j \\ F_{i-j}^{2+}, & \text{se} \quad i > j. \end{cases} \qquad 1 \le j \le N, \;\; 1 \le i \le N+1, \tag{2.42}$$

and

$$\mathscr{G}_{ij}^2 = \begin{cases} G_{j-i+1}^{2-}, & \text{se} \quad i \le j \\ -G_{i-j}^{2+} + h_s F_{i-j}^{2+}, & \text{se} \quad i > j. \end{cases} \qquad 1 \le j \le N, \;\; 1 \le i \le N+1. \tag{2.43}$$

Each of the vectors (2.40a)–(2.40f) is integrated analytically in y:

$$F_k^{0\pm} = \int_0^1 f_0'^{\pm}(kh_s,\mu)d\mu - \int_0^1 f_0'^{\pm}((k-1)h_s,\mu)d\mu, \qquad 1 \le k \le 2N \tag{2.44a}$$

$$G_k^{0\pm} = \int_0^1 g_0'^{\pm}(kh_s,\mu)d\mu - \int_0^1 g_0'^{\pm}((k-1)h_s,\mu)d\mu, \qquad 1 \le k \le 2N \tag{2.44b}$$

$$F_k^{1\pm} = \int_0^1 f_1'^{\pm}(kh_s,\mu)d\mu - \int_0^1 f_1'^{\pm}((k-1)h_s,\mu)d\mu, \qquad 1 \le k \le 2N \tag{2.44c}$$

$$G_k^{1\pm} = \int_0^1 g_1'^{\pm}(kh_s,\mu)d\mu - \int_0^1 g_1'^{\pm}((k-1)h_s,\mu)d\mu, \qquad 1 \le k \le 2N \tag{2.44d}$$

$$F_k^{2\pm} = \int_0^1 f_2'^{\pm}(kh_s,\mu)d\mu - \int_0^1 f_2'^{\pm}((k-1)h_s,\mu)d\mu, \qquad 1 \le k \le N \tag{2.44e}$$

$$G_k^{2\pm} = \int_0^1 g_2'^{\pm}(kh_s,\mu)d\mu - \int_0^1 g_2'^{\pm}((k-1)h_s,\mu)d\mu, \qquad 1 \le k \le N \tag{2.44f}$$

where the functions $f_0'^{\pm}$, $g_0'^{\pm}$, $f_2'^{\pm}$, $g_1'^{\pm}$, $f_2'^{\pm}$, and $g_2'^{\pm}$ are obtained from (2.37)–(2.39):

$$f_0'^{\pm}(y,\mu) = \frac{P_l(\pm\mu)P_k(\pm\mu)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\frac{\rho_0\rho_L e^{\frac{\lambda(y-3L)}{\mu}}}{2\lambda}, \qquad 0 \le y \le 2L \tag{2.45a}$$

$$g_0'^{\pm}(y,\mu) = \frac{P_l(\pm\mu)P_k(\pm\mu)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\frac{\rho_0\rho_L(kh_s\lambda-\lambda y+\mu)}{2\lambda^2}e^{\frac{\lambda(y-3L)}{\mu}}, \qquad 0 \le y \le 2L \tag{2.45b}$$

$$f_1'^{\pm}(y,\mu) = \frac{P_l(\mp\mu)P_k(\pm\mu)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\frac{\rho_{L,0}e^{\frac{\lambda(y-2L)}{\mu}}}{2\lambda}, \qquad 0 \le y \le 2L \tag{2.45c}$$

$$g_1'^{\pm}(y,\mu) = \frac{P_l(\mp\mu)P_k(\pm\mu)}{1-\rho_0\rho_L e^{-\frac{2\lambda L}{\mu}}}\frac{\rho_{L,0}(kh_s\lambda-\lambda y+\mu)}{2\lambda^2}e^{\frac{\lambda(y-2L)}{\mu}}, \qquad 0 \le y \le 2L \tag{2.45d}$$

$$f_2'^{\pm}(y,\mu) = -\frac{P_l(\pm\mu)P_k(\pm\mu)e^{-\frac{\lambda y}{\mu}}}{2\lambda}, \qquad 0 \le y \le L \tag{2.45e}$$

$$g_2'^{\pm}(y,\mu) = -\frac{P_l(\pm\mu)P_k(\pm\mu)(\lambda y+\mu-kh_s\lambda+h_s\lambda)}{2\lambda^2}e^{-\frac{\lambda y}{\mu}}, \qquad 0 \le y \le L. \tag{2.45f}$$

The solution of the system

$$\begin{bmatrix} 1-\beta_0 W^{0,0} & -\beta_1 W^{0,1} & \cdots & -\beta_M W^{0,M} \\ -\beta_0 W^{1,0} & 1-\beta_1 W^{1,1} & \cdots & -\beta_M W^{1,M} \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_0 W^{M,0} & -\beta_1 W^{M,1} & \cdots & 1-\beta_M W^{M,M} \end{bmatrix} \begin{bmatrix} J_0(y) \\ J_1(y) \\ \vdots \\ J_M(y) \end{bmatrix} = \begin{bmatrix} W^{0,0}S(y)+V^0 B \\ W^{1,0}S(y)+V^1 B \\ \vdots \\ W^{M,0}S(y)+V^M B \end{bmatrix} \tag{2.46}$$

provides the vectors $J_l(y)$, $1 \leq l \leq M$, calculated at the points of the mesh.

This formulation allows us to calculate the coefficients of the matrix with the cost of order N. We observe that the numerical method does not require that the condition (2.29) be satisfied. The kernel may be of the form

$$\omega(\mu,\mu') = \sum_{l=1}^{M} \beta_l f_l(\mu) g_l(\mu'),$$

where f_l e g_l are continuous functions. The choice of the Legendre polynomials was done by physical motivations.

2.4 Numerical Results

The integrals involved in (2.27) were performed by an adaptive Gauss–Legendre quadrature scheme, which ran with a relative tolerance of 10^{-8}. Numerical experiments show that the round-off error introduced by this quadrature is not relevant compared to the truncation error.

The validation of our numerical scheme involves two different techniques: study of convergence when the number of points in mesh changes and comparison of our results with results previously published in the literature.

For an isotropic scattering kernel, we compared our solution with the work in [VaSeVi07] using the LTS_N method with $N = 300$, a well-established method in transport theory. The pointwise comparison is given in Tables 2.1 and 2.2. Looking at Table 2.2, we observe a very good agreement between the results obtained and the exact ones.

We also calculate the maximum value of the ratio λ/σ for which the isotropic transport equation admits a finite solution. This value is obtained from the spectral radius of the operator L_g. This value is easy to estimate numerically using the spectral radius of the matrix W_0. In Tables 2.3–2.6 we compare our results with those [NaLo08] and [At96].

We compared the quantity $2J_0$ obtained for the transport problem with those obtained by Vargas et al. [VaSeVi07], using the LTS_N method with $N = 300$, a well-established method in transport theory. Here, we not only calculate the quantities $2J_0$ but also calculate $2J_1$ and present pointwise comparisons in Tables 2.7 and 2.8. These comparisons were chosen in order to show the performance and limitation of

Table 2.1 Comparison between the values calculated for $\int_{-1}^{1} I(y,\mu)d\mu$ when $\lambda = \sigma = 1.0$, the source is given by $S(y) = -y^2 + y$, the boundary condition is $B_0 = B_L = 0$ and $\rho = 0$ with the results published by Vargas et al. [VaSeVi07]

y	LTS_{300}	$N = 100$	$N = 200$	$N = 400$	$N = 800$	$N = 1,600$
0.0	0.335952	0.335875	0.335928	0.335942	0.335946	0.335947
0.05	0.398459	0.398382	0.398441	0.398456	0.39846	0.398461
0.1	0.452925	0.452842	0.452905	0.452921	0.452925	0.452926
0.15	0.502992	0.502904	0.502971	0.502988	0.502992	0.502994
0.2	0.548165	0.548071	0.548142	0.54816	0.548165	0.548166
0.25	0.587762	0.587629	0.587702	0.587721	0.587726	0.587727
0.3	0.621014	0.620913	0.620989	0.621009	0.621014	0.621015
0.35	0.647477	0.647373	0.647452	0.647471	0.647477	0.647478
0.4	0.666684	0.666579	0.666659	0.666679	0.666684	0.666685
0.45	0.678336	0.678224	0.678305	0.678325	0.67833	0.678332
0.5	0.682233	0.682126	0.682207	0.682228	0.682233	0.682234

Table 2.2 Comparison between the values calculated for $\int_{-1}^{1} I(y,\mu)d\mu$ when $\lambda = \sigma = 1.0$, the source is given by $S(y) = 1/8$, the boundary condition is $B_0 = B_L = 1/8$ and $\rho = 0$ with the results and exact solution published in [VaSeVi07]

y	Exact	LTS_{300}	$N = 200$	$N = 400$	$N = 800$	$N = 1,600$
0.0	0.516842	0.516841	0.516829	0.516838	0.516841	0.516841
0.1	0.600637	0.600634	0.600626	0.600634	0.600636	0.600637
0.2	0.647999	0.647997	0.647988	0.647996	0.647998	0.647999
0.3	0.678718	0.678715	0.678707	0.678715	0.678717	0.678718
0.4	0.696308	0.696303	0.696297	0.696305	0.696307	0.696308
0.5	0.702056	0.702053	0.702045	0.702053	0.702055	0.702055

Table 2.3 Comparison between the values calculated for the critical value of σ/λ when $\rho = 0$ with the benchmark results published by Naz and Loyalka [NaLo08]

L	Naz and Loyalka	N = 100	N = 200	N = 400	N = 800	N = 1,600
1.0	1.615379	1.615471	1.615403	1.615385	1.615380	1.615379
2.0	1.277102	1.277187	1.277125	1.277108	1.277103	1.277102
4.0	1.108468	1.108551	1.108490	1.108474	1.108469	1.108468
6.0	1.058296	1.058377	1.058317	1.058301	1.058297	1.058296
8.0	1.036402	1.036483	1.036423	1.036407	1.036403	1.036402
10.0	1.024879	1.024959	1.024900	1.024885	1.024881	1.024880
12.0	1.018072	1.018152	1.018093	1.018078	1.018074	1.018073
16.0	1.010766	1.010845	1.010787	1.010772	1.010768	1.010767
20.0	1.007136	1.007214	1.007156	1.007141	1.007137	1.007136

our method. In Table 2.7, when τ_0 is a small number, the results from LTS_N and GFD coincide to four digits. In Table 2.8, when τ_0 is large, GFD does not perform well, an easy phenomenon to explain: GFD involves spatial discretization, requiring a refined mesh to work properly in large domains.

Table 2.4 Comparison between the values calculated for the critical value of σ/λ and the solution published by Atalay [At96] and Naz and Loyalka [NaLo08] when ρ ranges between 0 and 0.99 and $L = 0.2$

	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = .75$	$\rho = 0.99$
N = 400	3.83032	2.95952	2.23566	1.60333	1.02491
N = 800	3.83031	2.95951	2.23566	1.60333	1.02491
Atalay	3.81843	2.94902	2.23324	1.60373	1.02503

Table 2.5 Comparison between the values calculated for the critical value of σ/λ and the solution published by Atalay [At96] and Naz and Loyalka [NaLo08] when ρ ranges between 0 and 0.99 and $L = 2.0$

	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = .75$	$\rho = 0.99$
N = 400	1.27711	1.20373	1.13254	1.06444	1.00250
N = 800	1.27710	1.20373	1.13254	1.06444	1.00250
Naz and Loyalka	1.27710	–	–	–	–
Atalay	1.27704	1.20396	1.13287	1.06469	1.00252

Table 2.6 Comparison between the values calculated for the critical value of σ/λ and the solution published by Atalay [At96] and Naz and Loyalka [NaLo08] when ρ ranges between 0 and 0.99 and $L = 20.0$

	$\rho = 0$	$\rho = 0.25$	$\rho = 0.50$	$\rho = .75$	$\rho = 0.99$
N = 400	1.00714	1.00658	1.00568	1.00398	1.00025
N = 800	1.00713	1.00658	1.00567	1.00397	1.00025
Naz and Loyalka	1.00714	–	–	–	–
Atalay	1.00714	1.00658	1.00568	1.00398	1.00117

Table 2.7 Comparison between the values calculated for $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} \mu I(y,\mu)d\mu$ and $\frac{2J_1(y)}{2\pi} = \int_{-1}^{1} I(y,\mu')\mu' d\mu'$ when $\omega = 1.0$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = B_L = 1.0$, $L = 1.0$ and $\rho = 0$ with the results published in [VaSeVi07]

	LTS_{300}		GFD_{400}		GFD_{800}	
y	$\mathscr{J}$	$\frac{2J_1(y)}{2\pi}$	$\mathscr{J}$	$\frac{2J_1(y)}{2\pi}$	$\mathscr{J}$	$\frac{2J_1(y)}{2\pi}$
0.0	3.514736	−0.682658	3.514725	−0.682651	3.514742	−0.682656
0.2	4.193457	−0.320120	4.193456	−0.320119	4.193467	−0.320119
0.4	4.306992	−0.023298	4.306991	−0.023298	4.307001	−0.023298
0.6	4.162764	0.219718	4.162763	0.219718	4.162773	0.219719
0.8	3.820951	0.418684	3.820951	0.418683	3.820960	0.418684
1.0	3.196350	0.581583	3.196338	0.581579	3.196349	0.581582

That said, we restrict ourselves to small domains in all subsequent calculations and, in Tables 2.9 and 2.10, present a study of convergence as N varies in two problems with anisotropic scattering. Looking at these tables, we see that fixing $N = 400$ yields a suitable approximation.

In Tables 2.11 and 2.12 we report results for intensity $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ and flux $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when $\rho = 0.5$ where the source is given by

Table 2.8 Comparison between the values calculated for $\mathscr{J} = \int_{-1}^{1} \mu I(y,\mu)d\mu$ and $\frac{2J_1(y)}{2\pi} = \int_{-1}^{1} I(y,\mu')\mu' d\mu'$ when $\omega = 1.0$, the source is given by $S(y) = e^{-y^2/4}$, the boundary condition is $B_0 = 1.0$ and $B_L = 0.0$, $L = 100.0$ and $\rho = 0$ with the results published in [VaSeVi07]

	LTS_{300}		GFD_{1600}		GFD_{3200}	
y	$\mathscr{J}$	$\frac{2J_1(y)}{2\pi}$	$\mathscr{J}$	$\frac{2J_1(y)}{2\pi}$	$\mathscr{J}$	$\frac{2J_1(y)}{2\pi}$
0	8.42592	−3.47483	8.429517	−3.466102	8.443685	−3.472449
20	16.9490	0.0700450	16.929820	0.069921	16.945922	0.0699867
40	12.7473	0.0700153	12.734579	0.069921	12.746721	0.0699867
60	8.54702	0.0699940	8.539338	0.069921	8.547520	0.0699867
80	4.34782	0.0699811	4.344098	0.069921	4.348319	0.0699867
100	0.121203	0.0699766	0.120914	0.069845	0.121160	0.0699643

Table 2.9 Numerical results for $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$ and $\omega = 0.5 + 0.2\mu\mu'$, using GFD_N method for some values of N

N	0.0	0.2	0.4	0.6	0.8	1.0
50	2.175512	2.280829	2.221042	2.126824	2.040648	2.003121
100	2.175580	2.280847	2.221039	2.126810	2.040626	2.003093
200	2.175603	2.280852	2.221039	2.126807	2.040621	2.003087
400	2.175611	2.280854	2.221039	2.126806	2.040620	2.003085
800	2.175613	2.280855	2.221040	2.126806	2.040620	2.003084

Table 2.10 Numerical results for $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 0.25$, $L = 1.0$ and $\omega = 0.1 - 0.8P_2(\mu)P_2(\mu')$, using GFD_N method for some values of N

N	0.0	0.2	0.4	0.6	0.8	1.0
50	1.422216	1.420961	1.317678	1.196440	1.081076	0.978663
100	1.422195	1.420936	1.317653	1.196417	1.081055	0.978646
200	1.422189	1.420930	1.317647	1.196411	1.081050	0.978641
400	1.422188	1.420928	1.317645	1.196410	1.081049	0.978640
800	1.422188	1.420928	1.317645	1.196409	1.081048	0.978640

$Q(y) = 0$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, and $L = 1.0$ for several values of ω', using GFD_{400} method.

In Tables 2.13 and 2.14 we report results for intensity $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ and flux $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $Q(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$, and $\omega = 0.5 + \beta_1\mu\mu'$ for some values of β_1', using GFD_{400} method.

In Tables 2.15 and 2.16 we report results for intensity $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ and flux $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when ρ obey the Fresnel's Law, the source is given by $Q(y) = -y^2 + 1$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$, and $\omega = 0.5 + \beta_1\mu\mu'$ for some values of β_1', using GFD_{400} method.

Table 2.11 Numerical results for $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = 0$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, and $L = 1.0$ for some values of ω', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\omega = 0.1$	0.394085	0.308092	0.295597	0.322948	0.3981832	0.599197
$\omega = 0.3$	0.455255	0.376538	0.366975	0.397257	0.474942	0.670610
$\omega = 0.5$	0.547692	0.481334	0.476855	0.510642	0.589553	0.774877
$\omega = 0.7$	0.704241	0.660817	0.665823	0.703859	0.780781	0.945278
$\omega = 0.9$	1.027922	1.035246	1.061086	1.104359	1.169499	1.285451
$\omega = 1.0$	1.366484	1.428692	1.476816	1.523184	1.571308	1.633516

Table 2.12 Numerical results for $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = 0$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, and $L = 1.0$ for some values of ω', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\omega = 0.1$	0.091348	0.030373	−0.023318	−0.078390	−0.142388	−0.227966
$\omega = 0.3$	0.081589	0.024937	−0.026616	−0.079646	−0.140016	−0.217202
$\omega = 0.5$	0.066771	0.016286	−0.031290	−0.080343	−0.134896	−0.201215
$\omega = 0.7$	0.041587	0.001061	−0.038569	−0.079491	−0.123798	−0.174651
$\omega = 0.9$	−0.010599	−0.031191	−0.052127	−0.073750	−0.096445	−0.120810
$\omega = 1.0$	−0.065226	−0.065226	−0.065226	−0.065226	−0.065226	−0.065226

Table 2.13 Numerical results for $2J_0 := \frac{\mathscr{J}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$, and $\omega = 0.5 + \beta_1 \mu\mu'$ for some values of β_1', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\beta_1 = -0.4$	2.171404	2.287915	2.228408	2.129973	2.039060	1.999342
$\beta_1 = -0.2$	2.172810	2.285582	2.225975	2.128928	2.039572	2.000564
$\beta_1 = 0.0$	2.174212	2.283228	2.223519	2.127872	2.040092	2.001812
$\beta_1 = 0.2$	2.175611	2.280854	2.221039	2.126806	2.040620	2.003085
$\beta_1 = 0.4$	2.177005	2.278459	2.218536	2.125729	2.041158	2.004384

Table 2.14 Numerical results for $2J_1(y) = \int_{-1}^{1} I(y,\mu)\mu d\mu$ when $\rho = 0.5$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$, and $\omega = 0.5 + \beta_1 \mu\mu'$ for some values of β_1', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\beta_1 = -0.4$	−0.179493	−0.043131	0.027270	0.052272	0.042998	0.004747
$\beta_1 = -0.2$	−0.179721	−0.043268	0.027390	0.052573	0.043322	0.004973
$\beta_1 = 0.0$	−0.179948	−0.043404	0.027514	0.052879	0.043653	0.005203
$\beta_1 = 0.2$	−0.180173	−0.043537	0.027644	0.053192	0.043990	0.005439
$\beta_1 = 0.4$	−0.180398	−0.043668	0.027778	0.053511	0.044333	0.005679

Table 2.15 Numerical results for $2J_0 := \frac{\mathscr{I}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ when ρ obey the Fresnel's Law, the source is given by $S(y) = -y^2 + 1$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$, and $\omega = 0.5 + \beta_1 \mu\mu'$ for some values of β_1', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\eta_1 = 1.1$ e $\beta_1 = 0.0$	2.139750	2.162119	2.102876	2.020612	1.947920	1.916511
$\eta_1 = 1.3$ e $\beta_1 = 0.0$	2.538795	2.543428	2.435245	2.238801	2.007225	1.859391
$\eta_1 = 1.5$ e $\beta_1 = 0.0$	2.674822	2.658213	2.532869	2.323659	2.083518	1.931200
$\eta_1 = 1.1$ e $\beta_1 = 0.2$	2.283157	2.339636	2.274510	2.111902	1.904883	1.770508
$\eta_1 = 1.3$ e $\beta_1 = 0.2$	2.540122	2.539579	2.430887	2.236966	2.008923	1.862644
$\eta_1 = 1.5$ e $\beta_1 = 0.2$	2.673850	2.653722	2.528947	2.322986	2.086874	1.936458

Table 2.16 Numerical results for $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when ρ obey the Fresnel's Law, the source is given by $S(y) = -y^2 + 1$, the boundary condition is $B_0 = 0.5$, $B_L = 1.0$, $L = 1.0$, and $\omega' = 0.5 + \beta_1 \mu\mu'$ for some values of β_1', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\eta_1 = 1.1$ e $\beta_1 = 0.0$	−0.269024	−0.122701	−0.039495	−0.002692	−0.001957	−0.031744
$\eta_1 = 1.3$ e $\beta_1 = 0.0$	−0.207152	−0.067573	0.045285	0.109681	0.100084	−0.016932
$\eta_1 = 1.5$ e $\beta_1 = 0.0$	−0.151652	−0.024587	0.077688	0.132996	0.115376	−0.009009
$\eta_1 = 1.1$ e $\beta_1 = 0.2$	−0.299995	−0.137502	−0.006496	0.072206	0.073990	−0.033573
$\eta_1 = 1.3$ e $\beta_1 = 0.2$	−0.207165	−0.067414	0.045887	0.110610	0.101018	−0.016279
$\eta_1 = 1.5$ e $\beta_1 = 0.2$	−0.151257	−0.023878	0.078846	0.134400	0.116642	−0.008209

Table 2.17 Numerical results for $2J_0 := \frac{\mathscr{I}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 0.25$, $L = 1.0$, and $\omega = 0.1 + \beta_2 P_2(\mu)P_2(\mu')$ for some values of β_2', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\beta_2 = -0.8$	1.420313	1.4187334	1.316498	1.196469	1.082103	0.980052
$\beta_2 = -0.4$	1.422188	1.420928	1.317645	1.196410	1.081049	0.978640
$\beta_2 = 0.4$	1.426277	1.425790	1.320271	1.196412	1.078875	0.975655
$\beta_2 = 0.8$	1.428517	1.428410	1.321786	1.196498	1.077769	0.974080

Table 2.18 Numerical results for $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $S(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 0.25$, $L = 1.0$, and $\omega = 0.1 + \beta_2 P_2(\mu)P_2(\mu')$ for some values of β_2', using GFD_{400} method

y	0.0	0.2	0.4	0.6	0.8	1.0
$\beta_2 = -0.8$	−0.051574	0.052266	0.101442	0.117132	0.110456	0.087691
$\beta_2 = -0.4$	−0.052534	0.050794	0.100112	0.116247	0.110036	0.087532
$\beta_2 = 0.4$	−0.054730	0.047398	0.096981	0.114078	0.108919	0.087067
$\beta_2 = 0.8$	−0.055999	0.045420	0.095117	0.112738	0.108177	0.086735

In Tables 2.17 and 2.18 we report results for intensity $2J_0 := \frac{\mathscr{I}}{2\pi} = \int_{-1}^{1} I(y,\mu)d\mu$ and flux $2J_1(y) = \int_{-1}^{1} \mu I(y,\mu)d\mu$ when $\rho = 0.5$, the source is given by $Q(y) = e^{-y}$, the boundary condition is $B_0 = 0.5$, $B_L = 0.25$, $L = 1.0$, and $\omega = 0.1 + \beta_2 P_2(\mu)P_2(\mu')$ for some values of β_2', using GFD_{400} method.

References

[At96] Atalay, M.: The critical slab problem for reflecting boundary conditions in one-speed neutron transport theory. Ann. Nucl. Energ. **23**, 183–193 (1996)

[AzEtAl11b] Azevedo, F.S., Sauter, E., Thompson, M., Vilhena, M.T.: Existence theory and simulations for one-dimensional radiative flows. Ann. Nucl. Energ. **38**, 1115–1124 (2011)

[AzEtAl11a] de Azevedo, F.S., Thompson, M., Sauter, E., Vilhena, M.T.: Existence theory for a one-dimensional problem arising from the boundary layer analysis of radiative flows. Progr. Nucl. Energ. **53**(8), 1105–1113 (2011)

[BeGl70] Bel, M.G., Glasstone, S.: Nuclear Reactor Theory. Van Nostrand Reinhold, New York (1970)

[Mo03] Modest, M.F.: Radiative Heat Transfer, 2nd edn. Academic, San Diego (2003)

[NaLo08] Naz, S., Loyalka, S.: One speed criticality problems for a bare slab and sphere: some benchmark results, part ii. Ann. Nucl. Energ. **35**, 2426–2431 (2008)

[Si93] Siewert, C.E.: On intensity calculations in radiative transfer. J. Quant. Spectros. Radiat. Tran. **50**, 555–560 (1993)

[Si95] Siewert, C.E.: An improved iterative method for solving a class of coupled conductive-radiative heat transfer problems. J. Quant. Spectros. Radiat. Tran. **4**, 599–605 (1995)

[VaSeVi07] Vargas, R.F., Segatto, C.F., Vilhena, M.T.: Solution of the radiative heat transfer equation with internal energy sources in a slab by the LTS_N. J. Quant. Spectros. Radiat. Tran. **105**, 1–7 (2007)

Chapter 3
Integral Neutron Transport and New Computational Methods: A Review

A. Barbarino, S. Dulla, and P. Ravetto

3.1 Introduction

The neutron transport equation is the basis for the physics simulation of nuclear reactors and, in particular, for nuclear reactor core design. This equation is a linear version of the original Boltzmann equation and it is commonly used in its integro-differential form. Several numerical methods have been derived over the years to obtain solutions for realistic configurations [LeMi84] and efficient codes are available to carry out neutronic simulations of multiplying systems [CaNo10].

An alternative form of the transport equation can be obtained by spatial integration of the integro-differential form. A spatially integral equation is thus obtained [Da58], which has proved very useful to highlight some physical aspects of the transport phenomenon and also for practical applications. In the special case of isotropic emissions, by angular integration, the Peierls equation is obtained, which served as the starting point for the development of the first analytical attempts to solve the transport problem [CaDePl53], [CaZw67], and also for the derivation of numerical methods [Ca65]. The integral form of the transport equation yields also the foundation to the so-called method of characteristics [As72] that has made an important breakthrough into reactor analysis during the most recent years [SaSaMo08]. The main interest of the work presented and discussed herewith is the neutronics of nuclear reactors, although the approach is also extended to radiation problems that may be encountered in various fields of nuclear engineering and of other applied sciences.

This work is opened by reviewing the basics of integral neutron transport, deriving the integral form of the equation for completeness in its most general time-dependent form. Afterwards, restricting to the Peierls equation, the second-order

A. Barbarino • S. Dulla • P. Ravetto (✉)
Energy Department, Politecnico di Torino, 24, C.so Duca degli Abruzzi, I-10129 Torino, Italy
e-mail: andrea.barbarino@polito.it; sandra.dulla@polito.it; piero.ravetto@polito.it

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_3,

A_N method is derived, illustrating the advantageous features of the model. The presentation of two approaches for the numerical solution of the A_N equations concludes the work.

3.2 The Integral Transport Equation

The integral form of the transport equation can be derived starting from the integro-differential (Boltzmann form) transport equation and carrying out a spatial integration along the characteristic line of motion of the particles [PrLa10]. This procedure is particularly simple in the steady-state situation.

The general integro-differential neutron transport equation is written as

$$\begin{aligned}&\frac{1}{v}\frac{\partial \Phi(\mathbf{r},E,\Omega,t)}{\partial t}+\nabla\cdot(\Omega\Phi(\mathbf{r},E,\Omega,t))+\Sigma(\mathbf{r},E)\Phi(\mathbf{r},E,\Omega,t)\\&\quad=S(\mathbf{r},E,\Omega,t)\\&\qquad+\oint d\Omega'\int dE'\Sigma_s(\mathbf{r},E')\Phi(\mathbf{r},E',\Omega',t)f_s(\mathbf{r},E'\to E,\Omega'\to\Omega),\end{aligned}\tag{3.1}$$

where the phase space point is defined by the geometric position $\mathbf{r}$, the particle energy E, and its direction of motion Ω. The quantity Φ is the unknown of the problem and denotes the total distance traveled by particles per unit volume, energy and solid angle and per unit time. The material properties are assumed to be constant in time. The fission term is not written explicitly in (3.1), but it can be easily added in the r.h.s. of the equation and its structure is the same as that of the integral scattering term. The collision transfer function for neutrons appearing in the integral scattering term depends only on the angle between Ω and Ω', thus only on the inner product $\Omega\cdot\Omega'$, being the material properties isotropic for neutrons at energies of interest for nuclear reactor theory. The r.h.s. of the equation constitutes the particle emission density that depends on the unknown particle flux determining the scattering rate. This equation constitutes a particle balance in phase space and it can be derived using the approach as originally proposed by Boltzmann for gas kinetics [Bo02], where the emissions by collisions are treated statistically.

In a time-independent situation, the transport equation (3.1) can be written at any point $\mathbf{r}-s\Omega$ belonging to the line along which neutrons are moving. Clearly, the particle streaming term $\nabla\cdot(\Omega\Phi)=\Omega\cdot\nabla\Phi$ is simply the directional derivative along the s variable. Hence, we can write

$$\begin{aligned}&-\frac{d\Phi(\mathbf{r}-s\Omega,E,\Omega)}{ds}+\Sigma(\mathbf{r}-s\Omega,E)\Phi(\mathbf{r}-s\Omega,E,\Omega)\\&=S(\mathbf{r}-s\Omega,E,\Omega)\\&\quad+\oint d\Omega'\int dE'\Sigma_s(\mathbf{r}-s\Omega,E')\Phi(\mathbf{r}-s\Omega,E',\Omega')f_s(\mathbf{r}-s\Omega,E'\to E,\Omega'\to\Omega)\\&\equiv Q(\mathbf{r}-s\Omega,E,\Omega).\end{aligned}$$

This space first-order equation can be easily integrated along s, thus obtaining

$$\Phi(\mathbf{r},E,\Omega) = \Phi(\mathbf{r}-s\Omega,E,\Omega)\exp\left[-\int_0^s ds'\Sigma(\mathbf{r}-s'\Omega,E)\right] + \int ds' Q(\mathbf{r}-s'\Omega,E,\Omega)\exp\left[-\int_0^{s'} ds''\Sigma(\mathbf{r}-s''\Omega,E)\right].$$

The above equation can be physically interpreted by explicitly noticing that the term

$$\exp\left[-\int_0^s ds'\Sigma(\mathbf{r}-s'\Omega,E)\right]$$

is the probability for neutrons to travel between the point $\mathbf{r}-s\Omega$ and $\mathbf{r}$ without undergoing any collision event. The integral appearing in the argument of the exponent is also known as optical path length, being the distance s measured in terms of the local mean free path, and it is an anisotropic quantity, explicitly depending on both $\mathbf{r}-s\Omega$ and $\mathbf{r}$ and not only on the distance s. The first term on the right-hand side gives the contribution to the neutron flux at $\mathbf{r},E,\Omega$ due to particles that freely travel from point $\mathbf{r}-s\Omega$ to point $\mathbf{r}$ without suffering any collision. The second term collects the contributions of all neutrons emitted between $\mathbf{r}-s\Omega$ and $\mathbf{r}$ by scatterings and external sources that travel to point $\mathbf{r}$ without any further collisions.

To obtain an integral form for the time-dependent equation, the Laplace transformation is applied prior to the space integration. By indicating with p the Laplace transform variable, we get

$$\begin{aligned}
&-\frac{\tilde{\Phi}(\mathbf{r}-s\Omega,E,\Omega,p)}{ds} + \left[\Sigma(\mathbf{r}-s\Omega,E)+\frac{p}{v}\right]\tilde{\Phi}(\mathbf{r}-s\Omega,E,\Omega,p)\\
&= \oint d\Omega' \int dE'\Sigma_s(\mathbf{r}-s\Omega,E')\tilde{\Phi}(\mathbf{r}-s\Omega,E',\Omega',p) f_s(\mathbf{r}-s\Omega,E'\to E,\Omega'\to\Omega)\\
&\quad + \tilde{S}(\mathbf{r}-s\Omega,E,\Omega,p) + \frac{1}{v}\Phi(\mathbf{r}-s\Omega,E,\Omega,0)\\
&\equiv \tilde{Q}(\mathbf{r}-s\Omega,E,\Omega,p).
\end{aligned}$$

The above equation has the same structure as the steady-state equation, where the total cross section has been modified by the addition of the p/v term and the initial state contribution has been added in the generalized emission density. Therefore, the integration along the characteristic line can be carried out at each point p in the Laplace-transformed space, leading to the integral equation

$$\begin{aligned}
&\tilde{\Phi}(\mathbf{r},E,\Omega,p)\\
&\quad = \tilde{\Phi}(\mathbf{r}-s\Omega,E,\Omega,p)\exp\left[-\int_0^s ds'\Sigma(\mathbf{r}-s'\Omega,E)\right]\exp\left(-\frac{p}{v}s\right)
\end{aligned}$$

$$+\int_0^s ds'\tilde{Q}(\mathbf{r}-s'\Omega,E,\Omega,p)\exp\left[-\int_0^{s'} ds''\Sigma(\mathbf{r}-s''\Omega,E)\right]\exp\left(-\frac{p}{v}s'\right).$$

In order to obtain the integral form of the transport equation for the angular flux in the time domain, a Laplace inversion must be carried out. The exponentials of the form $\exp(-ps/v)$ introduce the translated Dirac delta distributions $\delta(t-s/v)$. Use must be made of the convolution theorem to obtain the final form

$$\begin{aligned}
&\Phi(\mathbf{r},E,\Omega,t)\\
&\quad=\Phi(\mathbf{r}-s\Omega,E,\Omega,t-s/v)\exp\left[-\int_0^s ds'\Sigma(\mathbf{r}-s'\Omega,E)\right]\vartheta(t-s/v)\\
&\qquad+\Phi(\mathbf{r}-vt\Omega,E,\Omega,0)\exp\left[-\int_0^{vt} ds'\Sigma(\mathbf{r}-s'\Omega,E)\right]\vartheta(s-vt)\\
&\qquad+\int_0^{\min(s,vt)} ds'Q(\mathbf{r}-s'\Omega,E,\Omega,t-s'/v)\exp\left[-\int_0^{s'} ds''\Sigma(\mathbf{r}-s''\Omega,E)\right],
\end{aligned}$$

where ϑ is the Heaviside step function. The first term in the r.h.s. gives the contribution to the neutron flux at $\mathbf{r},E,\Omega$ at time t due to particles that freely travel from point $\mathbf{r}-s\Omega$ to point $\mathbf{r}$ without suffering any collision. Of course these particles must depart from the point $\mathbf{r}-s\Omega$ at an instant prior to t of the flight time s/v and can appear only once this time has passed, thus justifying the appearance of the step function. The second term accounts for the initial state, whose contribution disappears once t becomes larger than the transit time s/v. The third term collects the contributions of all neutrons emitted between $\mathbf{r}-s\Omega$ and $\mathbf{r}$ by scatterings and external sources that travel to point $\mathbf{r}$ without any further collisions. Also in this case, the delay due to the finite particle velocity is correctly accounted for. The upper limit of the integration in the emission density term is the minimum between the distance that can be traveled by neutrons having velocity v (i.e., vt) and s itself, for times longer than s/v. If the distance s is taken to be the distance $s_B(\mathbf{r},\Omega)$ to the external boundary of a non reentrant body facing vacuum, we arrive at the integral equation

$$\begin{aligned}
&\Phi(\mathbf{r},E,\Omega,t)\\
&\quad=\Phi(\mathbf{r}-vt\Omega,E,\Omega,0)\exp\left[-\int_0^{vt} ds'\Sigma(\mathbf{r}-s'\Omega,E)\right]\vartheta(s-vt)\\
&\qquad+\int_0^{\min(s_B(\mathbf{r},\Omega),vt)} ds'Q(\mathbf{r}-s'\Omega,E,\Omega,t-s'/v))\exp\left[-\int_0^{s'} ds''\Sigma(\mathbf{r}-s''\Omega,E)\right].
\end{aligned}$$

After a sufficiently long time, so that the contribution of the initial population has died out everywhere, we obtain

$$\Phi(\mathbf{r},E,\Omega,t) = \int_0^{s_B(\mathbf{r},\Omega)} ds' Q(\mathbf{r}-s'\Omega,E,\Omega,t-s'/v) \times \exp\left[-\int_0^{s'} ds'' \Sigma(\mathbf{r}-s''\Omega,E)\right]. \quad (3.2)$$

A special form of the integral equation can be derived in the particular case when Q is assumed to be isotropic, namely

$$Q(\mathbf{r},E,\Omega,t) = \frac{1}{4\pi} Q(\mathbf{r},E,t).$$

In this case, (3.2) can be integrated over all directions, thus leading to an integral over the whole volume V and yielding the total flux integral equation

$$\Phi(\mathbf{r},E,t) = \frac{1}{4\pi} \int d\mathbf{r}' Q\left(\mathbf{r}',E,t-\frac{|\mathbf{r}-\mathbf{r}'|}{v(E)}\right) \frac{\exp\left[-\int_0^{|\mathbf{r}-\mathbf{r}'|} ds' \Sigma(\mathbf{r}-\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}s',E)\right]}{|\mathbf{r}-\mathbf{r}'|^2}.$$

This is known as Peierls equation, and in the time-independent case it takes the simpler form

$$\Phi(\mathbf{r},E) = \frac{1}{4\pi} \int d\mathbf{r}' Q(\mathbf{r}',E) \frac{\exp\left[-\int_0^{|\mathbf{r}-\mathbf{r}'|} ds' \Sigma(\mathbf{r}-\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}s',E)\right]}{|\mathbf{r}-\mathbf{r}'|^2}.$$

The kernel of the transport integral equation contains all the physical features of the transport phenomena. Any approximate model introduces some distortion of such a kernel. Discrete ordinates and spherical harmonics are the most popular approximate models [LeMi84]; these models are derived from the integro-differential form and appear in differential form, but they can also be easily given an integral form, thus clearly highlighting the distortion induced by the approximation of the kernel.

3.3 The A_N Model

A space second-order model taking a diffusive form can be derived from the integral transport equation by an approximation of the exact transport kernel by means of a superposition of diffusive kernels [CoRa82]. This procedure can be applied inside a medium having a constant total cross section. For media characterized by fully heterogeneous properties, proper continuity conditions must be introduced at the interfaces between regions. The starting point is thus the transport equation (that we

write in the one-velocity case, for simplicity)

$$\Phi(\mathbf{r}) = \frac{1}{4\pi} \int d\mathbf{r}' \left[\gamma(\mathbf{r}')\Phi(\mathbf{r}') + S\left(\mathbf{r}'\right)\right] \frac{e^{-\Sigma|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2}, \tag{3.3}$$

where $\gamma(\mathbf{r}) = \Sigma_s(\mathbf{r})/\Sigma$ is the number of secondaries emitted per collision. Inside a homogeneous domain, it is not restrictive to assume $\Sigma = 1$, thus measuring distances in terms of mean free paths. The kernel is now approximated according to a standard integration formula, as [StZw58]

$$\frac{e^{-r}}{4\pi r^2} = \int_0^1 \frac{e^{-r/\mu}}{4\pi r \mu^2} d\mu \simeq \sum_{\alpha=1}^{N} p_\alpha \frac{e^{-r/\mu_\alpha}}{4\pi r \mu_\alpha^2}. \tag{3.4}$$

If we consider the system of integral equations

$$f_\beta(\mathbf{r}) = \frac{1}{4\pi} \int d\mathbf{r}' \left[\gamma(\mathbf{r}') \sum_{\alpha=1}^{N} p_\alpha f_\alpha(\mathbf{r}') + S(\mathbf{r}')\right] \frac{e^{-r/\mu_\beta}}{\mu_\beta^2 |\mathbf{r}-\mathbf{r}'|}, \quad \beta = 1, 2, \ldots, N, \tag{3.5}$$

we immediately verify by direct summation that the weighted sum of the so-called pseudo-moments f_α is an approximation of the total flux; that is,

$$\Phi(\mathbf{r}) \simeq \sum_{\alpha=1}^{N} p_\alpha f_\alpha(\mathbf{r}),$$

since the sum of the kernels appearing in (3.5) is an approximation of the exact kernel of (3.3). Recalling the properties of the Green function for the diffusion equation, the application of the Laplace operator to (3.5) leads to a system of differential equations for f_α, namely

$$\mu_\beta^2 \nabla^2 f_\beta(\mathbf{r}) - f_\beta(\mathbf{r}) + \left[\gamma(\mathbf{r}) \sum_{\alpha=1}^{N} p_\alpha f_\alpha(\mathbf{r}) + S(\mathbf{r})\right] = 0, \quad \beta = 1, 2, \ldots, N. \tag{3.6}$$

This system of equations, that has become known as the A_N model, has the structure of a multigroup system of diffusion equations with a full coupling appearing among all groups. Hence, the pseudo-moments f_α play the role of pseudo-energy group fluxes, although they can be related to the angular even parity fluxes in planar geometry [CoRa82].

It is worth recalling the fact that the approach outlined above can be carried out on a rigorous basis without approximating the μ integral in (3.4), and thus obtaining a novel exact formulation of the transport model [CoRaSu85] in terms of an integro-differential equation. The procedure can even be extended to include the time dependence [CoEtAl08], obtaining a wave-like time second-order form of the model.

There is an alternative route to obtain the A_N model, through the simplified spherical harmonics (SP_N) approach. This technique is particularly interesting, because it leads to a more consistent foundation to the SP_N method, which was introduced in a somewhat arbitrary fashion by Gelbard [Ge61]. In this approach, the first-order simplified spherical harmonics equations are reduced to a second-order form by elimination of the odd-order moments and the resulting system is then diagonalized [CiEtAl02].

The diffusive nature of the A_N equations is certainly advantageous for numerical applications, as will be seen in the next sections. However, this model cannot be easily extended to treat scattering anisotropy, although some efforts have been made to include linear anisotropy effects [CoRaSu83]. Also, pseudo-moments cannot be given a physical meaning except for the special case of slab geometry.

3.4 The Boundary Element Approach

Second-order form equations are particularly suitable for response matrix formulations. Equations can be cast into a form involving only values of the unknowns at the boundary of the meshes in which the domain is subdivided, thus leading to a boundary element numerical scheme (BEM) [BrTeWr84]. The resulting scheme involves a reduction in the dimensionality of the problem with obvious computational advantages.

The A_N equations can be given a BEM formulation [CiEtAl02]. To that end, the Green functions of the constituent equations are needed in each subdomain having a constant total cross section. Therefore the following equations are preliminarily considered:

$$\mu_\alpha^2 \nabla^2 \varphi_{\alpha\beta}(\mathbf{r}) - \Sigma \varphi_{\alpha\beta}(\mathbf{r}) + \Sigma_s(\mathbf{r}) p_\alpha \sum_{\eta=1}^{N} \varphi_{\eta\beta}(\mathbf{r}) + \delta_{\alpha\beta}\delta(\mathbf{r}-\mathbf{r}') = 0,$$

to explicitly obtain the solution in a standard way, through a sum of exponentials, namely

$$\varphi_{\alpha\beta}(|\mathbf{r}-\mathbf{r}'|) = \sum_{\eta=1}^{N} g_{\eta\beta} C_{\alpha\eta} \frac{e^{-\kappa_\eta |\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}.$$

The second-order equations are then multiplied by the Green functions and integrated over the volume. By application of the Green identity, one obtains the following result:

$$c(\mathbf{r}) f_\beta(\mathbf{r}) + \sum_{\alpha=1}^{N} \left[\varphi_{\alpha\beta}(|\mathbf{r}-\mathbf{r}'_\sigma|) J_{n',\alpha}(\mathbf{r}'_\sigma) + J_{n',\alpha\beta}(\mathbf{r}, \mathbf{r}'_\sigma) f_\alpha(\mathbf{r}'_\sigma) \right] = \Psi_\beta(\mathbf{r}), \quad (3.7)$$

where σ denotes the surface of the volume on which the integration has been performed. The coefficient $c(\mathbf{r})$ takes the values 0, 1, or $1/2$ for values outside the volume considered, inside or on the boundary, respectively. The above integral equation shows that the solution at each point is connected to a source volume term, that is,

$$\Psi_\beta(\mathbf{r}) = \sum_{\alpha=1}^{N} \int d\mathbf{r}' \varphi_{\alpha\beta}(|\mathbf{r}-\mathbf{r}'|) S(\mathbf{r}'),$$

and to boundary terms involving the pseudo-fluxes and the pseudo-currents defined by

$$J_{n,\alpha}(\mathbf{r}_\sigma) = -\frac{\mu_\alpha^2}{\Sigma} \frac{\partial f_\alpha}{\partial \mathbf{n}}(\mathbf{r}_\sigma),$$

$$J_{n',\alpha}(\mathbf{r}, \mathbf{r}'_\sigma) = \frac{\mu_\alpha^2}{\Sigma} \frac{\partial \varphi_{\alpha\beta}}{\partial \mathbf{n}'}(|\mathbf{r}-\mathbf{r}'_\sigma|).$$

The equation (3.7) shows also that the solution at each point inside the volume can be reconstructed once the values of the solution at the boundary are known. By evaluating (3.7) on the boundary, an integral equation for the solution is readily obtained:

$$c(\mathbf{r}_\sigma) f_\beta(\mathbf{r}_\sigma) + \sum_{\alpha=1}^{N} \left[\varphi_{\alpha\beta}(|\mathbf{r}_\sigma - \mathbf{r}'_\sigma|) J_{n',\alpha}(\mathbf{r}'_\sigma) + J_{n',\alpha\beta}(\mathbf{r}_\sigma, \mathbf{r}'_\sigma) f_\alpha(\mathbf{r}'_\sigma) \right] = \Psi_\beta(\mathbf{r}_\sigma), \quad (3.8)$$

which constitutes the basis for the BEM. In fact such equation can be written in discrete form by applying any discretization scheme to the boundary itself.

At inner interfaces between different meshes, continuity is required for the physical quantities such as the neutron flux and the corresponding current that are given in A_N form as

$$\Phi(\mathbf{r}) = \sum_{\alpha=1}^{N} p_\alpha f_\alpha(\mathbf{r}),$$

$$\mathbf{J}(\mathbf{r}) = -\sum_{\alpha=1}^{N} p_\alpha \frac{\mu_\alpha^2}{\Sigma} \nabla f_\alpha(\mathbf{r}).$$

The continuity is guaranteed by requesting all the pseudo-fluxes and all the pseudo-currents to be continuous at any point at an inner interface [CoEtAl10]. At last, an external surface facing vacuum classical Mark boundary conditions [BeGl70] are imposed:

$$-\frac{\mu_\alpha}{\Sigma} \frac{\partial f_\alpha}{\partial \mathbf{n}}(\mathbf{r}_\sigma) = f_\alpha(\mathbf{r}_\sigma).$$

The above formulation leading to the integral equations (3.8) on mesh boundaries is particularly suitable to be coded in computational tools based on a response-matrix scheme. Each mesh is characterized by a response matrix that, once applied to incoming partial particle currents, can produce exiting partial currents, account taken for inner sources. This approach has been successfully employed for core neutronics simulations [CaEtAl08].

3.5 The Spectral Element Approach

The spectral element method (SEM) is a scheme for the spatial discretization based on a generalized Galerkin method with numerical quadratures formulae [DeFiMu08]. The method employs a Lagrangian interpolation formula and high degree polynomials can be used on any given mesh. The basis functions are chosen to be orthogonal and the Gauss–Lobatto–Legendre quadrature formula is adopted for the evaluation of integrals, as:

$$\int_{-1}^{+1} g(\xi)\, d\xi \sim \sum_{k=1}^{N+1} \rho_k g(\xi_k).$$

The method has been successfully used in fluid-flow applications. Recently it has been extended to the field of neutron transport [Mu11], [BaEtAl11].

For neutronic applications, the method is derived by multiplying the A_N equations (3.6) by a test function χ_ℓ and then integrating over the domain; thus,

$$-\frac{\mu_\alpha^2}{\Sigma}\int_{\mathscr{D}} \mathrm{d}\mathbf{r}\nabla f_\alpha \nabla\chi_\ell + \int_{\partial\mathscr{D}} \mathrm{d}\mathbf{s}\frac{\mathrm{d}f_\alpha}{\mathrm{d}\mathbf{n}}\chi_\ell + \int_{\mathscr{D}} \mathrm{d}\mathbf{r}\left(-\Sigma f_\alpha + \Sigma_s \sum_{\beta=1}^{N} w_\beta f_\beta + S\right)\chi_\ell = 0,$$

which constitutes the weak formulation of the transport problem.

Afterwards, expanding the unknown functions in the same basis functions, we have

$$f_\alpha(\mathbf{r}) = \sum_{p=1}^{L} \hat{f}_\alpha^p \chi_p(\mathbf{r}).$$

For the SEM method, the basis functions are products of one-dimensional Lagrange interpolation polynomials built on a Gauss–Lobatto–Legendre quadrature grid. The general multidimensional configuration can be treated by starting from the 1D situation and taking a tensorial product of the coefficient matrices. Finally, all equations are recast into the single matrix form [BaEtAl13]

$$\left(\mu_\alpha^2 \mathfrak{K} + \mu_\alpha^2 \mathfrak{B} + \Sigma\mathfrak{M}\right)\hat{\varphi}_\alpha + \Sigma_s \sum_{\beta=1}^{N} w_\beta \mathfrak{M}\hat{\varphi}_\beta = \mathfrak{M}\hat{\mathrm{S}}.$$

In two-dimensional Cartesian geometry,

$$\mathfrak{K} = \frac{L_y}{L_x}(\mathbf{M} \otimes \mathbf{K}) + \frac{L_x}{L_y}(\mathbf{K} \otimes \mathbf{M}),$$

where

$$\mathfrak{M} = \frac{L_x L_y}{4}(\mathbf{M} \otimes \mathbf{M}),$$

L_x and L_y being the edges of the Cartesian mesh along the x and y coordinates, respectively. The term $\mathfrak{B}$ includes the contributions of the boundary terms. The matrices $\mathbf{K}$ and $\mathbf{M}$ are known as the stiffness and mass matrices, respectively, and are calculated with the formulas

$$M_{ij} = \int_{-1}^{+1} \chi_i(\xi)\chi_j(\xi)\mathrm{d}\xi = \rho_i \rho_j,$$

$$K_{ij} = \sum_{k=1}^{K+1} D_{ki} D_{kj} \rho_k,$$

where

$$D_{ij} = \left.\frac{\mathrm{d}\chi_j}{\mathrm{d}\xi}\right|_{\xi=\xi_i} = \begin{cases} \dfrac{P_K(\xi_i)}{P_K(\xi_j)} \dfrac{1}{\xi_i - \xi_j}, & i \neq j, \\ -\dfrac{(K+1)K}{4}, & i = j = 1, \\ \dfrac{(K+1)K}{4}, & i = j = K+1, \\ 0 & \text{elsewhere.} \end{cases}$$

Both continuous and discontinuous Galerkin approaches can be used. When using the continuous approach, only structured meshes may be adopted and a local refinement shall propagate, thus increasing the number of points also elsewhere in the system. Discontinuous schemes are possible, provided some terms are properly added, allowing more flexibility in the refinement and opening the way for applications to unstructured meshes configurations.

In order to improve the flexibility of the scheme, it is possible to deform a Cartesian element by means of a spatial transformation, which determines the position of the new grid points and the value of the Jacobian associated to them, to take into account the change in surface and volume, while approximating the integrals of the weak form. For instance, the so-called *transfinite interpolation* technique [DeFiMu08] requires only to specify the parametric description of the border of the element, and it is able to give both coordinates and Jacobian in a closed analytical form. In this way the rectangles of the mesh can be fit also for complicated curved geometries.

At last, it is also possible to collapse one of the sides of a rectangle to obtain a triangle, useful for some specific applications in the nuclear field, by using transfinite interpolation in conjunction with a change of the polynomial basis in one of the directions. More specifically, the Radau polynomials and the corresponding quadrature scheme provide a set of grid points which include just one of the extremes of the reference domain $[-1,+1]$. After the tensor product operation with a Lobatto grid in the other direction, one side remains without degrees of freedom. This side can be collapsed to a single point using a suitable transformation. The Jacobian in the collapsed vertex vanishes as expected, but it does not enter the algorithm because no unknowns are defined on it. The numerical scheme is then perfectly identical to the previous case. Of course, discontinuos Galerkin can still be applied to deformed grids, with just a complication in the algebraic form of the interface terms.

3.6 Comparison of Numerical Results

The performance of the BEM applied to the A_N equations for neutron transport calculations has been extensively tested in recent times; the method has shown to yield accurate results and to be an effective approach for the solution of neutron transport problems for nuclear reactor simulations. In [CiEtAl02] and [CoEtAl10] the results of some test calculations are reported for benchmark configurations usually considered in reactor physics. The method has proved to perform very well in comparison with standard discrete ordinate techniques and also to be free of drawbacks such as ray effects associated with the angular discretization in multidimensional problems. Furthermore, comparisons with Monte Carlo calculations have shown the excellent level of accuracy that can be attained by A_N. The SEM is also being deeply investigated [BaEtAl13] and may prove to be a powerful tool to obtain accurate results and to be suitable for simulations in which a high fidelity is required.

Two classical benchmarks (see [Na71], [KaStSc79]) are now presented, where the A_N model is solved with different numerical schemes. Figure 3.1 shows the computational domain of the IAEA-EIR2 benchmark, whose cross sections and source terms are reported in Table 3.1. For the FEM calculations, also the mesh average size h is indicated. Results are given in Table 3.2, where also the results produced using the TWODANT S_N code [AlEtAl90] are included.

The second benchmark is the "modified" Natelson PWR problem, characterized by the domain illustrated in Fig. 3.2; the cross sections data and source are given in Table 3.3. In this case, the absorption rates are calculated and Table 3.4 gathers the numerical results.

As an example of the capabilities of the combined use of spectral elements and transfinite interpolation, we present also a test case reproducing in a schematic way a fuel pin in a square lattice configuration, divided into several concentric rings representing also the clad and the moderator region (see Fig. 3.3 and Table 3.5). Only one eighth of the domain is taken into account, using reflective boundary

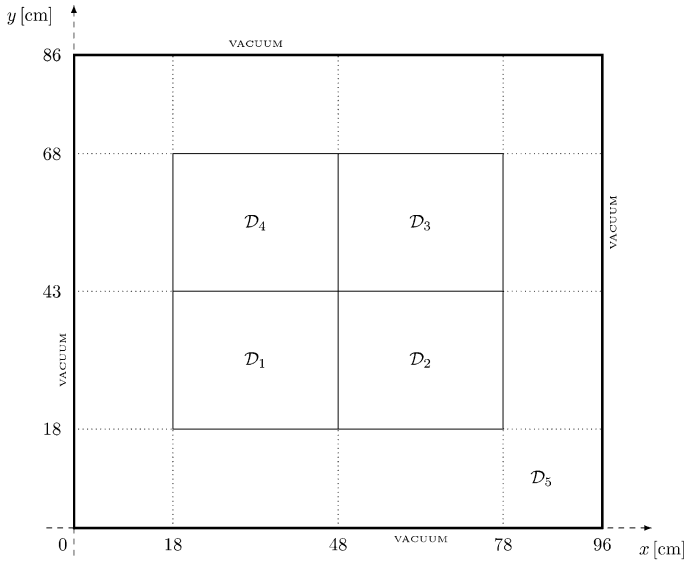

Fig. 3.1 IAEA EIR-2 problem domain

Table 3.1 Material properties for the IAEA EIR-2 benchmark problem

Region	Σ_t [cm^{-1}]	Σ_s [cm^{-1}]	Source strength [cm^{-3} s^{-1}]
$\mathcal{D}_1$	0.60	0.53	1.0
$\mathcal{D}_2$	0.48	0.20	0.0
$\mathcal{D}_3$	0.70	0.66	1.0
$\mathcal{D}_4$	0.65	0.50	0.0
$\mathcal{D}_5$	0.90	0.89	0.0

Table 3.2 Comparison of the average fluxes for the IAEA-EIR2 benchmark. All values are in n/(cm^2 s)

	FEM	FEM	FEM	SEM	BEM	TWODANT
	$h = 1.0$ cm	$h = 0.5$ cm	$h = 0.3$ cm			
$\overline{\Phi}_{\mathcal{D}_1}$	1.199869E+1	1.197052E+1	1.196234E+1	1.194591E+1	1.1973E+1	1.1960E+1
$\overline{\Phi}_{\mathcal{D}_2}$	6.018141E–1	5.703414E–1	5.608087E–1	5.438439E–1	5.3613E–1	5.3613E–1
$\overline{\Phi}_{\mathcal{D}_3}$	1.926494E+1	1.922154E+1	1.920514E+1	1.917657E+1	1.9222E+1	1.9202E+1
$\overline{\Phi}_{\mathcal{D}_4}$	9.095288E–1	8.715273E–1	8.599585E–1	8.384864E–1	8.2946E–1	8.3364E–1
$\overline{\Phi}_{\mathcal{D}_5}$	1.494612E+0	1.511750E+0	1.516965E+0	1.527069E+0	1.5318E+0	1.5263E+0

conditions on all sides. The central wedge is obtained by collapsing one side, which has no degrees of freedom because the basis in the radial direction is given by Radau polynomials.

Table 3.6 shows some results obtained comparing the SEM solution to a finite element solution, using elements of Courant P_1 and P_2 type.

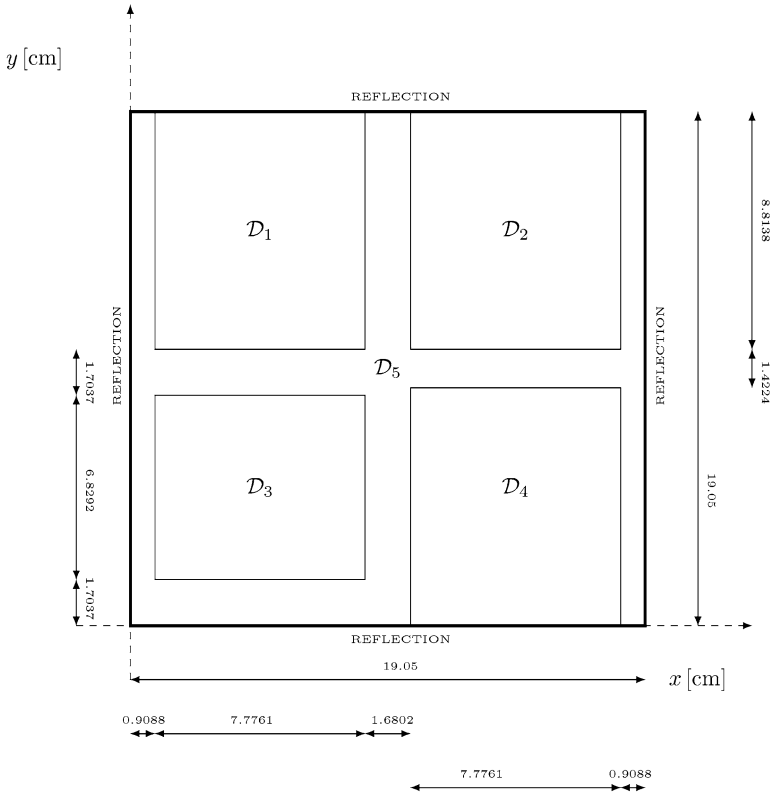

Fig. 3.2 Computational domain for the "modified" Natelson benchmark problem

Table 3.3 Material properties for the "modified" Natelson benchmark problem

Region	Σ_t [cm^{-1}]	Σ_s [cm^{-1}]	Source strength [cm^{-3} s^{-1}]
$\mathcal{D}_1$	0.200000	0.119230	0.00214230
$\mathcal{D}_2$	0.200000	0.119230	0.00215024
$\mathcal{D}_3$	0.200000	0.119230	0.00217729
$\mathcal{D}_4$	0.250000	0.147403	0.01048083
$\mathcal{D}_5$	0.200000	0.066703	0

3.7 Conclusions

In this contribution, the integral transport model for neutronic application is presented. The equation is derived in its most general form directly by spatial integration of the integro-differential form of the linear Boltzmann equation. For the time-dependent case a Laplace transform approach is used. The Peierls equation is then derived for the isotropic emission case.

Table 3.4 Results of the absorption rates for the modified Natelson PWR problem. All values are in n/(cm^2 s). The first five columns are calculated using the A_2 model

	FEM	FEM	FEM	SEM	BEM	S_{16}
	$h = 1.0$ cm	$h = 0.5$ cm	$h = 0.3$ cm			
$\overline{\Phi}_{\mathcal{D}_1}$	1.7173E–1	1.7722E–1	1.7891E–1	1.8212E–1	1.8052E–1	1.8132E–1
$\overline{\Phi}_{\mathcal{D}_2}$	1.4848E–1	1.5257E–1	1.5392E–1	1.5646E–1	1.5693E–1	1.5704E–1
$\overline{\Phi}_{\mathcal{D}_3}$	1.7101E–1	1.7688E–1	1.7879E–1	1.8217E–1	1.8101E–1	1.8178E–1
$\overline{\Phi}_{\mathcal{D}_4}$	3.1158E–1	3.2670E–1	3.3163E–1	3.4090E–1	3.4673E–1	3.4655E–1
$\overline{\Phi}_{\mathcal{D}_5}$	1.3251E–1	1.3559E–1	1.3657E–1	1.3835E–1	1.3481E–1	1.3329E–1

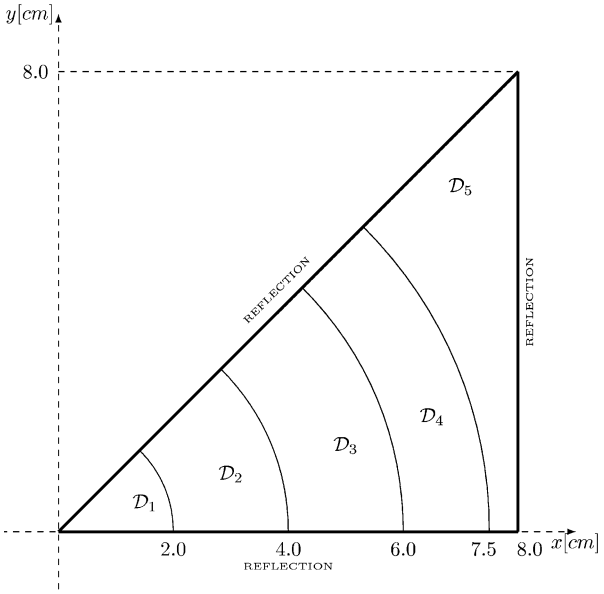

Fig. 3.3 Computational domain for the pin-cell benchmark problem

Table 3.5 Material properties for the pin-cell benchmark problem

Region	Σ_t [cm^{-1}]	Σ_s [cm^{-1}]	Source strength [cm^{-3} s^{-1}]
$\mathcal{D}_1$	1.0	0.5	1.0
$\mathcal{D}_2$	1.0	0.5	0.8
$\mathcal{D}_3$	1.0	0.5	0.3
$\mathcal{D}_4$	1.0	0.5	0.2
$\mathcal{D}_5$	1.0	0.5	0.1

The A_N approximation is derived by expanding the exact transport kernel as a superposition of diffusive kernels, leading to a system of second-order differential equations. Two numerical schemes that have been recently proposed are then reviewed, the boundary element method and the spectral element method.

Table 3.6 Relative distance in discrete L^2 norm between SEM (with discontinuous Galerkin) and FEM solution (at the same number of degrees of freedom) for the pin-cell benchmark problem

Comparison SEM–FEM Courant P_1				
	K = 9	K = 10	K = 11	K = 12
A_1	3.234924E–2	2.706710E–2	2.583746E–2	2.473618E–2
A_2	3.539272E–2	2.973357E–2	2.869604E–2	2.773729E–2
A_3	4.013103E–2	3.352912E–2	3.429756E–2	3.992359E–2
Comparison SEM–FEM Courant P_2				
	K = 9	K = 10	K = 11	K = 12
A_1	2.040621E–2	1.971461E–2	1.772286E–2	1.772286E–2
A_2	2.402395E–2	2.293065E–2	2.122532E–2	1.854835E–2
A_3	3.097324E–2	2.956092E–2	2.844892E–2	2.592764E–2

The results presented for benchmark configurations and for a system typical in nuclear reactor core simulations show that both numerical schemes are quite effective and flexible and they can yield accurate results as compared to finite element techniques. The development of these schemes could lead to computational tools appropriate for evaluations where a high fidelity is required.

Acknowledgements One of the authors (P.R.) is very grateful to the organizers of the IMSE-2012 Conference for the kind invitation and the generous support that enabled him to travel to Bento Gonçalves and to take part in the stimulating scientific sessions held in such a friendly atmosphere.

References

[AlEtAl90] Alcouffe, R.E., et al.: User's Guide for TWODANT - A Code Package for Two-Dimensional, Diffusion-Accelerated, Neutral-Particle, Transport, Report LA-10049-M, Los Alamos Scientific Laboratory, Los Alamos (1990) http://epubs.siam.org/doi/abs/10.1137/0902035

[As72] Askew, J.R.: A Characteristics Formulation of the Neutron Transport Equation in Complicated Geometries, Report AEEW-M 1108, United Kingdom Atomic Energy Establishment, Winfrith (1972)

[BaEtAl11] Barbarino, A., Dulla, S., Ravetto, P., Mund, E.: The spectral element approach for the solution of neutron transport problems. In: International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering, M&C2011, Rio de Janeiro, Brazil (2011)

[BaEtAl13] Barbarino, A., Dulla, S., Ravetto, P., Mund, E.: A spectral element method for neutron transport in A_N approximation; Part I. Ann. Nucl. Energ. **53**, 372–380 (2013)

[BeGl70] Bell, G.I., Glasstone, S.: Nuclear Reactor Theory. Van Nostrand Reinhold, New York (1970)

[Bo02] Boltzmann, L.E.: Leçons sur la Théorie des Gaz. Gauthier-Villars, Paris (1902)

[BrTeWr84] Brebbia, C.A., Telles, J.C.F., Wrobel, L.C.: Boundary Element Technique. Springer, Berlin (1984)

[CaNo10] Calvin, C., Nowak, D.: High performance computing in nuclear engineering. In: Cacuci, D.G. (ed.) Handbook of Nuclear Engineering, vol. II(12). Springer, New York (2010)

[CaEtAl08] Canepa, S., Van Geemert, R., Porsch, D., Dulla, S., Ravetto, P.: A response matrix formulation of multidimensional transport problems. In: International Conference on the Physics of Reactors, PHYSOR'08, Interlaken, Switzerland (2008)

[Ca65] Carlvik, I.: A method for calculating collision probabilities in general cylindrical geometry and applications to flux distributions and Dancoff factors. In: Proceedings of the United Nations International Conference on Peaceful Uses of Atomic Energy, vol. 2, p. 255, Geneva, Switzerland (1965)

[CaZw67] Case, K.M., Zweifel, P.F.: Linear Transport Theory. Addison-Wesley, Reading (1967)

[CaDePl53] Case, K.M., De Hoffmann, F., Placzek, G.: Introduction to the Theory of Neutron Diffusion. Los Alamos Scientific Laboratory, Los Alamos (1953)

[CiEtAl02] Ciolini, R., Coppa, G.G.M., Montagnini, B., Ravetto, P.: Simplified P_N and A_N methods in neutron transport. Progr. Nucl. Energ. **40**, 245–272 (2002)

[CoRa82] Coppa, G., Ravetto, P.: An approximate method to study the one-velocity neutron integral transport equation. Ann. Nucl. Energ. **9**, 169–174 (1982)

[CoRaSu83] Coppa, G., Ravetto, P., Sumini M.: Approximate solution to neutron transport equation with linear anisotropic scattering. J. Nucl. Sci. Tech. **20**, 822–831 (1983)

[CoRaSu85] Coppa, G., Ravetto, P., Sumini, M.: An alternative formulation of the monokinetic transport equation. Transport Theor. Stat. Phys. **14**, 83–102 (1985)

[CoEtAl08] Coppa, G.G.M., Dulla, S., Peano, F., Ravetto, P.: Alternative forms of the time-dependent neutron transport equation. Progr. Nucl. Energ. **50**, 934–938 (2008)

[CoEtAl10] Coppa, G.G.M., Giusti, V., Montagnini, B., Ravetto, P.: On the relation between spherical harmonics and simplified spherical harmonics methods. Transport Theor. Stat. Phys. **39**(2), 164–191 (2010)

[Da58] Davison, B.: Neutron Transport Theory. Clarendon Press, Oxford (1958)

[DeFiMu08] Deville, M.O., Fisher, P.F., Mund, E.: High-Order Methods for Incompressible Fluid Flow. Cambridge University Press, Cambridge (2008)

[Ge61] Gelbard, E.M.: Simplified Spherical Harmonics Equations and Their Use in Shielding Problems. Technical Report WAPD-T-1182, Westinghouse Electric Corp. Bettis Atomic Power Laboratory, Pittsburgh, Pennsylvania (1961)

[KaStSc79] Kavenoky, A., Stepanek, J., Schmidt, F.: Benchmark Problems. Transport Theory and Advanced Reactor Simulations. IAEA-TECDOC-254, International Atomic Energy Agency, Vienna, Austria (1979)

[LeMi84] Lewis, E.E., Miller, W.F. Jr.: Computational Methods of Neutron Transport. Wiley, New York (1984)

[Mu11] Mund, E.: Spectral element solutions for the P_N neutron transport equations. Comput. Fluid **43**(1), 102–106 (2011)

[Na71] Natelson, M.: Variational derivation of discrete ordinate-like approximations. Nucl. Sci. Eng. **43**, 131–144 (1971)

[PrLa10] Prinja, A.K., Larsen, E.W.: General principles of neutron transport. In: Cacuci, D.G. (ed.) Handbook of Nuclear Engineering, vol. II(12). Springer, New York (2010)

[SaSaMo08] Santandrea, S., Sanchez, R., Mosca, P.: A linear surface characteristics approximation for neutron transport in unstructured meshes. Nucl. Sci. Eng. **160**, 23–40 (2008)

[StZw58] Stewart, J.C., Zweifel, P.F.: A review of self-shielding effects in the absorption of neutron. In: Second International Conference on the Peaceful Uses of Atomic Energy, vol. 16, pp. 650–662, Geneva, Switzerland (1958)

Chapter 4
Scale Invariance and Some Limits in Transport Phenomenology: Existence of a Spontaneous Scale

B.E.J. Bodmann, M.T. Vilhena, J.R.S. Zabadal, L.P. Luna de Oliveira, and A. Schuck

4.1 Introduction

In transport phenomenology it is a common practice to express equations for continuous quantities such as fluxes, current densities among others, in a dimensionless fashion, i.e. independent of scales. This may be understood from the fact that transport phenomena in fluids are the continuum limit of scalable multi-particle distributions and their respective flows [Kr97], [LeLa12], [Po94]. If from the physical point of view one respects the microscopic origin of fluids, then these equations, when scaled to the microscopic or particle level such as the mean free path or the mean inter-particle distance, should break scale invariance or invariance under dilatation transformation. Nevertheless, physical parameters that are typically present in the equations establish a connection of the macroscopic with the microscopic world by their relations to distributions. For instance, the diffusion parameter is linked with particle distributions manifest in Avogadro's number together with the multi-particle system's equation of state. The microscopic or macroscopic cross sections reflect particle interaction probabilities typical for the physical forces that drive the dynamics of the particle ensemble in consideration. One could continue this reasoning with many other examples.

While for multi-particle systems the continuum limit seems adequate and is sufficient as long as mean(-field) values are sufficient and effects due to fluctuations may be neglected. Theoretically, if one starts with the complete physics of the many-particle system, mean values and all higher significant moments can be

B.E.J. Bodmann (✉) • M.T. Vilhena • J.R.S. Zabadal • A. Schuck
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: bardo.bodmann@ufrgs.br; vilhena@mat.ufrgs.br; jorge.zabadal@ufrgs.br
schuck@iee.ufrgs.br

L.P. Luna de Oliveira
Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, Brazil
e-mail: lpluna@unisinos.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_4,

determined; however, this is not possible in practice. Hence, there seems to be no smooth transition between a distributional continuous and a particle picture without resorting to additional techniques such as stochastic models that translate distributions into ensemble descriptions. In the distributional picture one assumes in principle an uncountable set of constituents, whereas the latter (particle picture) is based on a countable set. Moreover, if there were a natural transition between the continuous (macro) and the discrete (micro) scale, there would be need for a hybrid description below a certain micro-scale [GrPi07]. Such a "natural" transition was not found until now and thus is a supporting argument in favor of our reasoning, to look for a transition by means of a spontaneous symmetry breaking, that as the present discussion will show has the broken scale invariance as a consequence. In other words, what to look for is whether it is in principle possible to consider the discrete limit, starting from the continuous description together with a spontaneously broken invariance.

Since it is not obvious at all, how to get a mechanism that transforms a symmetric case into a non-symmetric one, we recall that the fact to break a symmetry is nothing else than obtaining an asymmetry, which in turn may be interpreted as a reference quantity, i.e. a normalization. In order to show how transformations, their invariants, asymmetry, and normalization are related, we should start from a transport equation, determine the Lie invariants and determine the generator for symmetry breaking from some of these operators, we adopt a simpler procedure based on geometry arguments, that nevertheless have its replica in differential geometry. Although we show by means of hyperspace arguments and geometric properties of that space how to identify the generator for symmetry breaking, the analogue way should in principle work for differential geometry-based arguments, but that are certainly very much more complicated to identify and handle as compared to the procedure that we present in the following.

4.2 A Geometric Invariant

As a next step we introduce a geometric space–time invariant for hydrodynamical quantities. To this end, consider a hydrodynamical flux $\mathbf{j}$ (momentum transport for instance) and associated (energy) density ρ that in a static limit reduce to the thermodynamic density ω (inner energy), that may be determined from the thermodynamic density of a sufficiently small control volume in motion with the flux contribution subtracted. The geometric relation for the respective densities and hydrodynamical flux shall obey the first fundamental form of Gauss for the differential quantities [SaToBa06]

$$d\omega^2 = d\rho^2 - d\mathbf{j}^2 = g_{\mu\nu} dj^\mu dj^\nu. \tag{4.1}$$

Here, in the right-hand side of the equation, we have made use of the sum convention that implies in summation over double appearing indices and $g_{\mu\nu}$ is the metric tensor. In this equation, if $d\omega^2$ is an invariant, then it could well serve as a local

reference scale and is defined by invariance under a set of some transformations, that have to be determined. Note that, at this point, the existence of an invariant will lead to the most general form of local transformations.

Any transformation in momentum transport is then generically given by

$$j^\mu \to j^\mu + \varepsilon k^\mu(j)\,.$$

Inserting the changes into the shell equation (4.1) yields the infinitesimal change in the metric tensor:

$$\begin{aligned}\delta(d\omega^2) &= \underbrace{\delta(g^{\mu\nu})}_{\equiv 0} dj_\mu dj_\nu + g^{\mu\nu}\delta(dj_\mu)dj_\nu + g^{\mu\nu}dj_\mu\delta(dj_\nu)\\ &= \varepsilon\left(\frac{\partial k_\mu}{\partial j_\nu} + \frac{\partial k_\nu}{\partial j_\mu}\right)dj_\mu dj_\nu = \varepsilon G^{\mu\nu}dj_\mu dj_\nu.\end{aligned}$$

Taking into account the causality constraint one determines the modified metric $G^{\mu\nu}$ in terms of $\sigma g^{\mu\nu} = G^{\mu\nu}$ with $\sigma = \frac{1}{4}g_{\mu\nu}G^{\mu\nu}$, where σ represents a local scale factor, which in turn defines the constraints for the most general flux dependence of the infinitesimal transformation by $k^\mu(j)$:

$$G^{\mu\nu} - \sigma g^{\mu\nu} = \left(\frac{\partial k^\mu}{\partial j_\nu} + \frac{\partial k^\nu}{\partial j_\mu}\right) - \frac{1}{2}g^{\mu\nu}\frac{\partial k^\lambda}{\partial j^\lambda} = 0. \tag{4.2}$$

The specific form of the transformation may be determined using a power expansion of $k^\mu(j)$; that is,

$$k^\mu = {}_0a^\mu + {}_1^1a^\mu_\nu j^\nu + {}_1^2a^\nu_\nu j^\mu + {}_2^1a^\mu_{\nu\lambda}j^\nu j^\lambda + {}_2^2a^\nu_{\nu\lambda}j^\lambda j^\mu + {}_2^2a^\lambda_{\nu\lambda}j^\nu j^\mu + \mathcal{O}(j^3).$$

Here the coefficients ${}_2^i a_{\nu\lambda} = {}_2^i a_{\lambda\nu}$ are symmetric under interchange of the lower indices. The symmetry conditions (4.2) then read for the respective terms that go with a specific power in $\mathcal{O}(j^n)$:

1. Equality (4.2) puts no restriction except for causality on $\mathcal{O}(j^0)$ and represents the Poincaré translation.
2. For $\mathcal{O}(j^1)$ the scalar coefficient ${}_1^2a$ is an arbitrary factor, reflecting the dilatation transformation. In addition, one gets

$$0 = {}_1^1a^{\mu\nu} + {}_1^1a^{\nu\mu} - \tfrac{1}{2}\,g^{\mu\nu}\,{}_1^1a^\lambda_\lambda,$$

   which may be identified with the Lorentz transformation.
3. From the terms that go with $\mathcal{O}(j^2)$ one obtains

$$\begin{aligned}0 = 2\,{}_2^1a^{\mu\nu}_\lambda j^\lambda + 2\,{}_2^1a^{\nu\mu}_\lambda j^\lambda - {}_2^1a^\kappa{}_{\lambda\kappa}g^{\mu\nu}j^\lambda\\ + 2\left({}_2^2a^\nu j^\mu + {}_2^2a_\lambda g^{\mu\nu}j^\lambda + {}_2^2a^\mu j^\nu\right).\end{aligned} \tag{4.3}$$

Contracting (4.3) by $g_{\mu\nu}$ eliminates all terms except for the one in parentheses and, thus, ${}^{2}_{2}a_{\lambda} \equiv 0$. For the remaining coefficients ${}^{1}_{2}a^{\mu\nu\lambda}$ one observes symmetry under exchange of the second and third index, ${}^{1}_{2}a^{\mu\nu\lambda} = {}^{1}_{2}a^{\mu\lambda\nu}$, which permits one to rewrite the coefficient in terms of an arbitrary vector c^{μ} and the metric.

$$ {}^{1}_{2}a^{\mu\nu\lambda} = g^{\mu\nu}c^{\lambda} + g^{\mu\lambda}c^{\nu} - g^{\nu\lambda}c^{\mu}. $$

Note that this contribution has got the characteristics of a conformal translation.

4. All terms with higher powers in $\mathcal{O}(j^n)$, for all $n > 2$ vanish identically, because of symmetry under interchange of indices except for the first one.

Thus, the most general admissible form of the infinitesimal transformation is

$$ k^{\mu} = \underbrace{b^{\mu}}_{\text{Poincaré}} + \underbrace{\Lambda^{\mu}_{\nu} j^{\nu}}_{\text{Lorentz}} + \underbrace{\lambda\, j^{\mu}}_{\text{Dilatation}} + \underbrace{2c_{\lambda} j^{\lambda} j^{\mu} - c^{\mu} j^{\lambda} j_{\lambda}}_{\text{Conformal}} . $$

Successive application of the infinitesimal conformal translation yields

$$ j^{\mu} \to \frac{j^{\mu} - j_{\nu} j^{\nu} c^{\mu}}{1 - 2 j^{\lambda} j_{\lambda} + j^{\lambda} j_{\lambda} c^{\kappa} c_{\kappa}} . $$

From the finite form of the conformal translation one recognizes that these transformations may turn singular in a sub-manifold, where the denominator vanishes. Therefore the transformation has to be restricted in c^{μ} such as to define a diffeomorphism in the physically relevant region of momentum transport space.

4.3 The Hyperspace Hypothesis

The aforementioned transformation analysis made use of the usual $1 \oplus 3$ time–space dimensions, but no link to an asymmetry and normalization was established yet. Recalling that the invariant was based on geometrical arguments it seems plausible to extend geometry by adding two extra dimensions, where the asymmetry may be defined by a difference and the normalization by a sum, respectively, of the components of these two extra dimensions. Note that one could have chosen another way introducing curvature into the $1 \oplus 3$ dimensional space and probably come to a similar result; however, the advantage of using a hyper-space lies in the fact that the symmetry group may be represented by linear transformations of the pseudo-orthogonal group $SO(4,2)$ with the hypercone defined by $\mathscr{S}_6 = \left\{ j | g_{\alpha\beta} j^{\alpha} j^{\beta} = 0 \right\}$. The representation of the pseudo-orthogonal transformation $\boldsymbol{\Omega}$, which transform the six-flux $j^{\alpha} \to \Omega^{\alpha}_{\beta} j^{\beta}$, shall maintain the hyper-cone invariant, i.e. $g_{\alpha\beta}\Omega^{\alpha}_{\gamma}\Omega^{\beta}_{\delta} = g_{\gamma\delta}$, where $||\boldsymbol{\Omega}|| = 1$ holds. Together with the restrictions in the parameter space $\{c^{\mu}\}$ of the conformal translations, the conditions (4.2) in the spirit of the first fundamental form of Gauss are necessary and sufficient to permit a self-consistent implementation of a scale.

One may now use the fact that it is the second fundamental form of Gauss that contains all curvature properties of a given space [Da94], [Fr97] and interpret the normal vector on an oriented four-dimensional flux hypersurface as a reciprocal normalization N^{-1}, which is fixed but may be arbitrarily chosen, and an asymmetry A, which is then a function of four-flux. One possibility is to define the normalization and asymmetry by $N^{-1} = j^4 + j^5$ and $A = j^4 - j^5$, and the shell equation is then

$$\omega^2 = N^{-1}A = (j^5 + j^4)(j^5 - j^4) = j^\mu j_\mu \,.$$

For convenience and since we have the freedom to define a scale, i.e. fix the normalization, we define unitless momentum transport $\upsilon^\mu = N j^\mu$ with the scale invariant shell equation and NA dimensionless.

$$\varpi^2 = N^{-2}\omega^2 = \upsilon^\mu \upsilon_\mu = NA \,.$$

An analysis of transformation properties on A and N constitute the next step in the procedure.

4.4 *SO(4,2)* Symmetry Breaking

In the following the effect of the subgroups on normalization and asymmetry are shown. Inspection shall indicate the relevant transformations for the construction of the generator capable of spontaneously breaking a symmetry.

1. *Poincaré translation*: The subgroup which leaves the normalization invariant defines the translation in energy–momentum space:

$$\upsilon^{\mu\prime} = \upsilon^\mu + N b^\mu, \quad N' = N, \quad A' = A + 2\upsilon^\mu b_\mu + N b^\mu b_\mu . \tag{4.4}$$

2. *Lorentz transformation*: Maintaining the normalization and the asymmetry constant, the transformation reduces to the Lorentz one, namely

$$\upsilon^{\mu\prime} = \Lambda^\mu_\nu \upsilon^\nu, \quad N' = N, \quad A' = A.$$

3. *Dilatation*: The one parameter subgroup defines the dilatation which leaves invariant the reduced flux υ^μ but changes the normalization as well as the asymmetry:

$$\upsilon^{\mu\prime} = \upsilon^\mu, \quad A' = \lambda A, \quad N' = \lambda^{-1} N.$$

4. *Conformal translation*: The subgroup with four parameters represents conformal translations and leaves the asymmetry invariant:

$$\upsilon^{\mu\prime} = \upsilon^\mu - A c^\mu, \quad A' = A, \quad N' = N - 2\upsilon^\mu c_\mu + A c^\mu c_\mu . \tag{4.5}$$

From (4.4) and (4.5) one may identify the Poincaré as well as the conformal translation as the candidates because they change either the normalization or the asymmetry. It is remarkable that in a specific system with $A = 0$, upon transformation, the asymmetry may turn nonzero. One may verify this by an example, suppose, that initially equation $\upsilon_\mu \upsilon^\mu = 0$ holds. Assuming that flux is displaced on the cone with $b_\mu b^\mu = 0$, then there is still the possibility of getting an asymmetry according to

$$A' = \underbrace{A}_{=0} + 2b_\mu \upsilon^\mu + N \underbrace{b_\mu b^\mu}_{=0} ,$$

where $b_\mu \upsilon^\mu \neq 0$ might play the role of a momentum transfer, which is a typical interaction feature.

In order to show that from (4.4) and (4.5) one may construct an operator, which transforms a scale invariant description with

$$\upsilon_\mu \upsilon^\mu = 0 ,$$

i.e. $A = 0$ into a nonvanishing one, one may some sort of "transport" the flux υ^μ first by a Poincaré displacement $\mathscr{P}$ followed by a conformal translation $\mathscr{C}$ and then return by the inverse sequence. Thus the change after "transport" to the original system is

$$[\mathscr{C}^\mu_\nu, \mathscr{P}^\nu_\lambda]\upsilon^\lambda = Nb^\mu - Ac^\mu ,$$

where $[\cdot, \cdot]$ is the usual commutator, which plays the role of a generator and transforms a specific symmetric description into another equivalent description.

The change from a singular to a finite scale is then

$$\upsilon^\nu [\mathscr{C}^\lambda_\nu, \mathscr{P}^\kappa_\lambda] g_{\kappa\mu} \upsilon^\mu = Nb_\mu \upsilon^\mu - Ac_\mu \upsilon^\mu .$$

Even in the limit of a vanishing asymmetry $A \to 0$, there remains the scale invariant term $Nb_\mu \upsilon^\mu$, which may be nonzero; however, j_4 and j_5 shall be finite. The fact that we have found a generator, which transforms a scale invariant description into one with a scale may be understood as an implementation of a spontaneous symmetry breaking.

4.5 Conclusions

In the present work we showed the possibility using a dimensionally extended space, where the extra dimensions allow to define an asymmetry together with a normalization by a closed line integral defined by commutators of Poincaré and diffeomorphic conformal displacements applied to a density current, respectively.

The asymmetry plays the role of an indicator of (spontaneous) symmetry breaking. We show how starting from a model without a reference scale that a scale emerges through spontaneous $SO(4,2)$-symmetry breaking.

The symmetry transformations of the differential shell equation, which transform any physically meaningful four-flux into a feasible new one (allowed by dynamics) are diffeomorphisms on the Poincaré group, the dilatation, and the conformal momentum translation. In order to prepare the playground for Hydrodynamics Field theory with its usually linear operators we have chosen the group representation by linear transformations of the pseudo-orthogonal group $SO(4,2)$. We found indeed a generator defined by the commutation of the Poincaré with the conformal translation which allows one to change the dimensionless description into one that contains a scale. Since in a scale invariant theory a scale term breaks dilatational symmetry one may understand those as a consequence of the breaking of diffeomorphic $SO(4,2)$ symmetry.

Further, from the ordinary space–time point of view, there is no necessity for splitting into separate points or any other structural change of space–time [BoMc73], [So98]. It is the curved flux hypersurface which puts a symmetry condition on the allowed solutions of the transport equations and belongs in this sense to these equations.

We are completely aware of the fact that our discussion at the present status is restricted to the question how spontaneous symmetry breaking in a curved flux space may explain a spontaneous scale in an *ab initio* scale invariant model.

In the literature [BuVe95], [FoGrSt11] one finds discussions of space–time transformations embedded in a higher dimensional (>4) flat hyperspace. In these approaches the algebra is setup by 15 generators (the Poincaré group, dilatation and conformal translations). If one applies the before mentioned transformations on vectors in four-flux space, symmetry considerations are likely to reflect dynamical properties of the system in consideration.

Although this discussion appears to have a rather academic than practical character, this reasoning may well be applied for convergence problems, where the discretization represents a length scale and the continuous limit has to be recovered. In this case the question would mean that symmetry restoration is considered. It is noteworthy that in practice the continuous limit in discrete approaches does not exist, there is a maximum precision that may be achieved. Since the symmetry argument is independent of numerical specifications, the convergence could be analyzed by symmetry restoration arguments.

References

[BoMc73] Bose, S.K., McGlinn, W.D.: Space–time symmetries and the spontaneous breakdown of dilation invariance. Phys. Rev. D **7**, 1949–1949 (1973)

[BuVe95] Buchholz, D., Verch, R.: Scaling algebras and renormalization group in algebraic quantum field theory. Rev. Math. Phys. **7**, 1195–1239 (1995)

[Da94] Darling, R.W.R.: Differential Forms and Connections. Cambridge University Press, Cambridge (1994)

[FoGrSt11] Fortin, J.-F., Grinstein, B., Stergiou, A.: Scale without conformal invariance: an example. Phys. Lett. B **704**, 74–80 (2011)

[Fr97] Frankel, T.: The Geometry of Physics. Cambridge University Press, Cambridge (1997)

[GrPi07] Graña, M., Pinasco, J.P.: Discrete scale invariance in scale free graphs. Phys. A **380**, 601–610 (2007)

[Kr97] Krug, J.: Origins of scale invariance in growth processes. Adv. Phys. **46**, 139–282 (1997)

[LeLa12] Lesne, A., Lagües, M.: Scale Invariance: From Phase Transitions to Turbulence. Belin, Paris (2012)

[Po94] Pocheau, A.: Scale invariance in turbulent front propagation. Phys. Rev. E **49**, 1109–1122 (1994)

[SaToBa06] Sadeghi, J., Tofighi, A., Banijamali, A.: Scale and conformal invariance and variation of the metric. J. Mod. Phys. A **21**, 3641–3647 (2006)

[So98] Sornette, D.: Discrete scale invariance and complex dimensions. Phys. Rep. **297**, 239–270 (1998)

Chapter 5
On Coherent Structures from a Diffusion-Type Model

B.E.J. Bodmann, J.R.S. Zabadal, A. Schuck, M.T. Vilhena, and R. Quadros

5.1 Introduction

Turbulent structures are quite common in nature, sometimes not only directly visible as in fluid flows, meteorological phenomena in the upper atmosphere but also indirectly observable through measurements usually based on correlation techniques. These aforementioned picturesque structures are not an effect of mere fluctuations that are also present in purely dissipative flows [SaCo09], though are of stochastic origin, because of a clear time ordering. Moreover, if a fluid is described as an ensemble of atoms or molecules that obeys microscopic laws and collectively constitutes a stochastic system with laws provided by statistical thermodynamics and hydrodynamics, then the continuous macroscopic system shall have manifestations with origin in microscopic properties [HoHo03]. Such a reasoning implies two essential questions: "Why do particles move in an orchestrated way in turbulent phenomena?", and more specifically, "How do particles sense their partially phase-locked position and movement that give rise to vortices and thus turbulence?" As the following discussion will show, this quest may find some explanation if one considers the phenomenon as a result of the formation of coherent structures. However, only an adequate dynamical model that produces such coherent structures together with a proof that excludes the possibility of a collection of statistical fluctuations will shed further light on the subject.

Contrary to the particle-based approach, a fluid is considered a continuous entity and has its associated velocity or momentum field or a related kinetic energy

B.E.J. Bodmann (✉) • J.R.S. Zabadal • A. Schuck • M.T. Vilhena
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: bardo.bodmann@ufrgs.br; jorge.zabadal@ufrgs.br; schuck@iee.ufrgs.br; mtmbvilhena@gmail.com

R. Quadros
Federal University of Pelotas, Pelotas, Rio Grande do sul, Brazil
e-mail: regis.quadros@ufpel.edu.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_5,

distribution [JiBo05]. Thus, at first sight one might accept that there is no need for the microscopic multi-particle structure of such a field. However, axial and shear stresses are not sufficient to explain complex vortex structures that are known from wind tunnel and similar experiments with controlled conditions, unless the stress tensor itself is based on a specific field model [GuEtAl05], [LuTsKi05]. Hence, we reason that some missing arguments shall result from the fact that any fluid has last but not least particles as its constituents.

5.2 Motivation from "Arm-Waving Arguments"

A common starting point for turbulent phenomena are transport equations based on continuous conservation laws or symmetries (scale, translational invariance among others) that describe a continuous medium fluid. Further, it is convenient to cast these transport equations in a dimensionless form, so that in principle one may scale down the solution up to molecular or atomic dimensions, where clearly such a scale invariance should break down. Moreover, these transport equations have continuous translational invariance. Thus, if one simplifies the microscopic particle structure of the fluid to a pseudo-periodic arrangement of atoms, molecules, or any microscopic particles that constitute the fluid, then on the average a discrete translational invariance shall hold, only. This is due to the fact that atoms, molecules are subject to pseudo periodical perturbations caused by particle–particle collisions along the direction of motion, in other words translation invariance down to length scales below the mean free path is in contradiction to the well-established microscopic structure of matter.

In order to map out our further reasoning, we make use of an "arm-waving argument" that motivates our discussion that follows. Two thin sheets of a fluid with an average particle density $1/a^3$ shall be in relative motion with Δv. In such a pseudo-crystal model on the average discrete translational invariance holds. Due to "synchronous" particle–particle collisions slip produces a pseudo-periodic perturbation with frequency $\nu_S = \Delta v/a$. However, pseudo-periodic perturbations could give origin to a radiation field with coherent content because of the periodic "phase-locked" perturbation. The radiation field could be characterized by a pitch frequency $\nu_R \sim \nu_S$ and a correlated wavelength $\lambda \sim ac/\Delta v$, where c is the propagation speed of the radiation field. Numerical values for λ may range depending on the assumptions as far as hundreds of meters or even more and thus could be responsible for coherent structures of the same size.

At this point, two observations are in order, first the radiation field that interacts with other constituents of the fluid does not follow the translational direction but introduces effects with components perpendicular to the original flow direction, and second, a model system to be constructed in the further shall represent properties of fluid constituents as well as interaction mediators. Mediators propagate in general faster as its origin the constituents movement and thus may well be the reason

for formation of large (macroscopic) partially coherent structures compared to the microscopic inter-particle distance scale that gave rise to a continuous approach in the first place.

5.3 A Coherent Constituent–Mediator Model

5.3.1 The Concept of Coherent States

So far, there is no consensus as to which definition is the most significant one for coherent structures. In turbulent transport, there is lack of a universal definition for coherence in an Eulerian frame. Furthermore, quantitative Eulerian measures of coherence are frame dependent and therefore fail to reflect intrinsic properties of turbulent flows. Also in a Lagrangian frame [BoEtAl06], [MaEtAl07], [ToBo09], [Ye02], coherent structures are rather designed to capture properties of chaotic advection than measuring genuine coherence.

In mathematical physics coherence was introduced as an idealized property of waves that allows for interference, which is also the fundamental concept of quantum theory. Historically, coherence was related to a constant phase difference, manifest in constructive or destructive interference, that may be of spatial, temporal, spectral, or of more general origin. The introduced measure for coherence in mathematical physics though makes use of correlation functions between different waves, based on formal definitions sketched out next.

One assumes that the phenomenon in consideration may be formulated on a separable complex Hilbert space, which is locally compact, and there exists a well-defined local measure $d\mu(\mathbf{x})$ (a generalized volume in the most simple cases). Any observable $\mathbf{x}$ in this space has a vector representation, and there exists a complex field $\phi(\mathbf{x})$ (usually the wave function) that represents distributional properties of that observable. Separability guarantees that measurable quantities $\boldsymbol{\Omega}$ are independent of representation, shall be integrable, and have bilinear form $\int \phi(\mathbf{x})\boldsymbol{\Omega}\psi(\mathbf{x})\, d\mu(\mathbf{x})$, including unity (a normalization). Any complete set of vectors that satisfies these properties defines a manifold of general coherent states.

Following this concept of coherence, it seems that kibitzing Quantum Theory might be a promising direction. Hence, in the next section we thoroughly deduce and construct a simple model starting from Quantum equations that shall comply with some characteristics of our "arm-waving argument" toy model, i.e. the fluid constituent–interaction mediator model. To this end, we interpret constituents in terms of Fermions and identify, as usual in Quantum Field Theory, interaction mediators in terms of Bosons, where the fact that we use field equations, that are also employed in Quantum Theory naturally entails coherence properties.

5.3.2 Modeling Coherent Fluid Constituents

A quantum equation, that is exclusively compatible with fermionic degrees of freedom is the Dirac equation, which we consider in two-component form in the Weyl representation. Note that although this equation is associated with fermions it further includes the coupling of fermions to bosons:

$$\begin{aligned}\imath\hbar\partial_t\Phi_+ - eA_0\Phi_+ + \boldsymbol{\sigma}(\imath\hbar c\boldsymbol{\nabla} + e\mathbf{A})\Phi_+ &= mc^2\Phi_-,\\ \imath\hbar\partial_t\Phi_- - eA_0\Phi_- - \boldsymbol{\sigma}(\imath\hbar c\boldsymbol{\nabla} + e\mathbf{A})\Phi_- &= mc^2\Phi_+.\end{aligned}$$

Since, the Dirac–Weyl equation is a relativistic equation, and with our fluid considerations we are far from a relativistic regime, we reduce the equation to a one component equation similar to a Schrödinger–Pauli equation. This may be attained relating the time component of the four-vector potential to the mass of the constituent particle $eA_0 \approx mc^2$, and upon applying the Lorenz gauge condition $\boldsymbol{\partial}\mathbf{A} = 0$. Further we make use of the usual definitions for the electric field $\mathbf{E}$, the magnetic induction $\mathbf{B}$ and the Bohr magneton $\mu_B = \frac{e\hbar}{2mc}$, respectively:

$$\begin{aligned}0 = -\frac{\hbar^2}{2mc^2}\partial_t^2\Phi_+ - \imath\hbar\partial_t\Phi_+ + \frac{\hbar^2}{2m}\Delta\Phi_+ + \mu_B\boldsymbol{\sigma}\left(\mathbf{B} - \frac{\imath}{c}\mathbf{E}\right)\Phi_+ \\ -2\imath\mu_B\mathbf{A}\boldsymbol{\nabla}\Phi_+ - \frac{e^2}{2mc^2}\mathbf{A}^2\Phi_+.\end{aligned}$$

The closest one can get to a classical model is considering only leading order terms and assuming weak vector fields, that results in a diffusion alike equation. It is noteworthy that approaching a classical model does not mean here to use the limit $\hbar \to 0$ where one ends up with triviality, but keep the minimum terms that still maintain coherent properties:

$$-\imath\hbar\partial_t\Phi_+ + \frac{\hbar^2}{2m}\Delta\Phi_+ = 0.$$

The essence of this finding is that the diffusion alike equation has an imaginary diffusion constant $|D| = \hbar/(2m)$. Even using typical values for quantum systems yields orders of magnitudes that are comparable to typical values in classical diffusion scenarios ($\sim 10^{-3}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$). Hence, it seems that apart from its possible numerical values for $|D|$, the factor $\sqrt{-1}$ provides coherence properties.

5.3.3 Modeling a Coherent Interaction Mediator

Following an analogue strategy as for the constituents, the starting point is the equation for boson dynamics, i.e. the time-dependent Maxwell equation.

For convenience we use the vector potential representation for the electric field **E** and the magnetic induction **B**, respectively:

$$\nabla^2\phi + \frac{\partial}{\partial t}(\nabla \mathbf{A}) = -\frac{\rho}{\varepsilon_0},$$
$$\left(\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}\right) - \nabla\left(\nabla\mathbf{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}\right) = \mu_0\mathbf{j}.$$

Again, we reduce the equation to leading order terms only and for simplicity neglect also source terms so that the simplified equation system reads

$$\nabla^2\phi + \frac{\partial}{\partial t}(\nabla \mathbf{A}) = 0, \quad \nabla \times \nabla \times \mathbf{A} = \mathbf{0}. \tag{5.1}$$

There is again the possibility to make this equation system coincide with a diffusion-like equation if the following condition holds:

$$\nabla\mathbf{A} = \frac{\imath 2m}{\hbar}\phi. \tag{5.2}$$

It is noteworthy that in semi-classical approaches that involve the Schrödinger equation, the time component of the four-vector potential yields a static and classical interaction potential, whereas the transverse components of the vector field remain perpendicular to the propagation direction, which is also known as the Coulomb gauge. The present condition represents an inhomogeneous Coulomb gauge condition that relates the transverse with the longitudinal field and may be considered a closure if interpreted in analogy with the classical closure that one generally needs to reduce a transport equation to a diffusion equation, by relating a density gradient to a current density. Upon inserting (5.2) into the second equation (5.1) one verifies consistency, so that we can use the result as a simple model for coherent interaction mediator contributions.

5.4 A Simple Model with Coherence Content

The result of the previous two sections is not surprising since the Schrödinger equation is known to work for solutions that arise either from commutator or from anti-commutator relations, in other words, it does not distinguish between Fermions or Bosons. From the derivation of the coherent constituent and mediator model one observes that the most simple approach needs only one dynamical equation, since both results have diffusion alike form.

However, fluids in reality are more likely to show a mixture of coherent and noncoherent contributions so that a simplified model with more realistic properties shall be represented by a diffusion alike equation with complex diffusion coefficient:

$$\frac{\partial \phi(\mathbf{r},t)}{\partial t} = D\boldsymbol{\nabla}^2 \phi(\mathbf{r},t), \quad D \in \mathbb{C}. \tag{5.3}$$

The difference to the classical diffusion equation comes from the fact that the concept for coherence demands for a flux description in bilinear form of amplitudes $\rho(\mathbf{r},t) = \phi^*(\mathbf{r},t)\phi(\mathbf{r},t)$. It is worth mentioning that for a real diffusion coefficient the amplitudes ϕ and the bilinear distribution ρ obey the same diffusion equation, that is not the case for a purely imaginary diffusion coefficient. Thus a more realistic solution for the amplitude is a superposition of coherent (wave like) and incoherent (dissipative) contributions that enter into the distribution in a bilinear form that furthermore guarantees semi-positiveness of the distribution.

Equation (5.3) describes the dynamics of the model and ϕ represents the amplitude field for any space–time point of interest. However, flow phenomena are typically open systems, i.e. depend on a source term. In our discussion we do not add a source term to the equation and solve it, but introduce the source contribution in form of a "continuous initial condition" and without any proof, we assume that this approach should be reasonable if the source strength is sufficiently weak. In the present discussion such a limitation is not crucial, since the main focus is put on coherence that shall represent the turbulent character in a qualitative fashion, rather than determine a general solution for a complex diffusion equation with an arbitrary source term.

Let Γ denote the domain of the source which may be continuous or discrete depending whether there are point, line, surface, or volumetric sources and $\mathbf{s} \in \Gamma$ a vector pointing to one of the countable or uncountable set of point sources of that domain, then the flow field amplitude ϕ is a superposition of all contributions from all point sources of Γ that add up coherently and determine $\phi(\mathbf{r},t)$:

$$\phi(\mathbf{r},t) = \int_\Gamma \psi(\mathbf{r},\mathbf{s},t)\, d\mu(\mathbf{s}). \tag{5.4}$$

Here, $d\mu(\mathbf{s}) = Q\, d^\delta s$ and Q is the distribution of source strengths with δ the dimension of the considered space (usually $\delta = 2$ or 3).

If the distribution that gives rise to the amplitudes is a discrete distribution, then $d\mu(\mathbf{s}) = \sum_i \delta(\mathbf{s} - \mathbf{q}_i) d^{\delta_\Gamma} s$, where $\mathbf{q}_i$ are the locations of the point sources. For example if there is a superposition of flow amplitude contributions originating from two point sources, then the description reminds one on the Young experiment. Alternatively, if the amplitude origin is a line, then $d\mu(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{q}(\kappa))\, d^\delta s$, where the parameter κ sweeps through all source points. If the source domain Γ is an effective volume (area, volume or any hyper-volume), then $d\mu(\mathbf{s}) = \delta(\mathbf{s} - \mathbf{q}(\{\kappa_i\}))\, d^\delta s$ with $i \in \{1,\ldots,\delta_\Gamma\}$ and δ_Γ is the dimension of the source domain which does not necessarily coincide with the dimension of the integration domain. In this representation the parametrization by κ_i yields an ordinate scheme and sweeps the whole of the sources' domain. Thus, independent of the specific form of the source term, the observed contribution that corresponds to a fluid flow in a bilinear form is

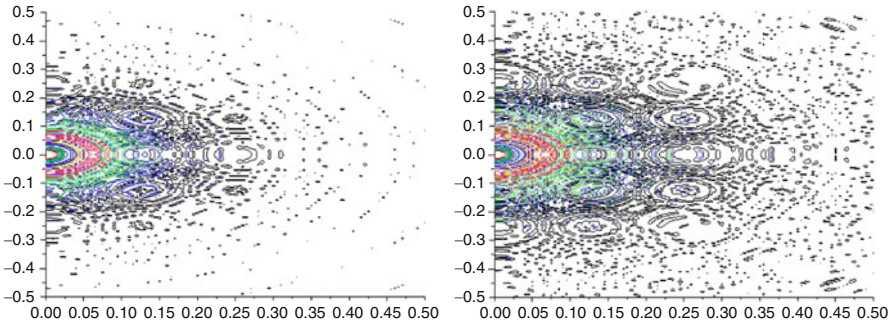

Fig. 5.1 Contour plot in the $x-y$-plane for two subsequent instants from the superposition of amplitude contributions and resulting density in bilinear form that gives rise to a turbulent character

$$\rho = \phi^* \phi = \int_\Gamma \psi^*(\mathbf{r}, \mathbf{s}', t)\psi(\mathbf{r}, \mathbf{s}, t)\, d\mu(\mathbf{s}')\, d\mu(\mathbf{s}).$$

Besides spatial variations the source may also have a specific characteristics with time. The way we absorbed a possible source term into the initial condition ($\lim_{t\to 0}$ of (5.4)) represents an instantaneous source term only. Another simple source term not compatible with this condition is a continuous one, which may be assembled from the solutions for instantaneous synchronous source rates that are shifted by an infinitesimal delay and summed up. The procedure of such a superposition is compatible with a continuous source but cast into a "continuous initial condition." This type of approach should be reasonable if the source strength is weak, since weak sources allow to assume that contributions are linear so that the solution found for an "instantaneous initial condition" may be used multiply in order to construct the final distribution, which is also consistent with the fact that the diffusion equation is invariant under time translations:

$$\rho = \int_0^t \int_0^t \dot{\phi}^*(\mathbf{r}, t-\tau')\dot{\phi}(\mathbf{r}, t-\tau)\, d\tau'\, d\tau.$$

For convenience we assume the initial condition in form of its spectral composition with a Gaussian shape, consider a two-dimensional complex diffusion problem with the source aligned along the y axis and consider the flow in the direction of a semi-open plane with $x \geq 0$. The following Fig. 5.1 shows two subsequent snapshots of contour lines that represent the resulting rugged distribution for $t > 0$. Differently than the purely diffusive case where the smooth distribution flattens out with progressive distance, the present case shows fluctuations that remind on turbulent structures. The fact that we used an approach that was based on the presence of coherent structures classifies them as turbulent and not mere uncorrelated fluctuations.

From our reasoning lined out in Sect. 5.3.1 we introduced formally coherence as the phenomenon that has in a limited space–time domain regions with constant

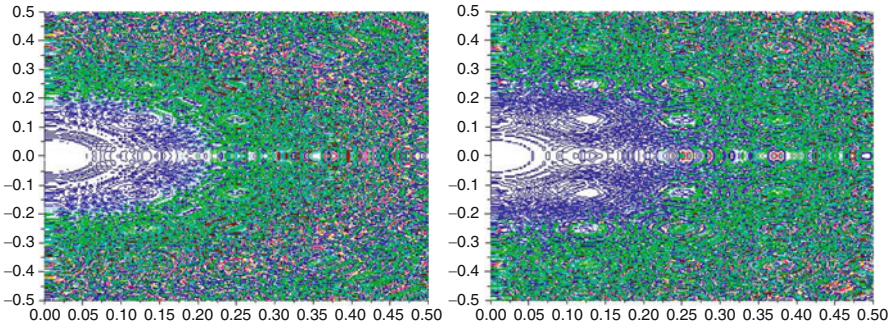

Fig. 5.2 Contour plot in the $x-y$-plane of phase distributions for two subsequent instants

phase relation, resulting from space–time interference integrals. Although the phase argument with coherent states is a clear statement, a representation for phase-locking is not that obvious. One possible reproduction for the phase relation is contour plots with equal phase isolines for ϕ:

$$\tau(\mathbf{r},t) = \tan^{-1}\left(\frac{\Im m\left[\phi(\mathbf{r},t)\right]}{\Re e\left[\phi(\mathbf{r},t)\right]}\right).$$

Figure 5.2 shows these phase relations, which reflect a remarkable similarity to the density plots in Fig. 5.1. One incontestable interpretation is that these phase distributions are responsible for the turbulent fluctuations and thus are a manifestation of coherence for this type of phenomenon. It is noteworthy that although the superposition of amplitudes is linear, the observed density is not, because of its bilinear implementation.

5.5 Conclusions

In the present discussion we focused on the question, how one may make plausible the phenomenon that individual particles that constitute a fluid, that macroscopically appears as a continuum, organize their motion in a way to form coherent structures like eddies. The key to such an understanding may not be found in a purely macroscopic approach, but needs the microscopic reality of the existence of particles and their capability to interact with each other by mediators. A typical system that has those properties is a fermion–boson system, which we called generically the constituent–interaction mediator model.

Thus one may recognize that probably one of the crucial reasons that stalled the process of a more profound understanding of the turbulent character in fluid flows may be due to the missing of a unique and significant definition for coherent structures. Nevertheless, in a realm that is different than the one of transport

phenomena, i.e. quantum theory, coherence is one of the key issues and one may build a meaningful definition for coherence on the formal properties for coherent states. Moreover, the formal properties that are essential and necessary for coherent states, no reference to quantum properties is necessary, so that these may be considered useful also for classical considerations.

A reduction of equations for Fermions and Bosons to a quasi-classical level showed that the most simple dynamics with coherence properties is supplied by a diffusion-like equation with complex diffusion coefficient. Our simple model generated results that are compatible with turbulent appearances manifest in coherent structures. As the discussion showed, the microscopic particle origin of matter cannot be ignored because of its breaking of continuous symmetries such as scaling and translation. Although the resulting solutions for the fluid are continuous, their origin trace back to some discrete particular structure.

In a previous work [BoEtAl10] a similar reasoning was applied and resulted in a Reynolds number definition, that compared microscopic to macroscopic scales ($Re \sim \Lambda v/(\lambda c_{therm})$, more specifically a vortex correlation length Λ, that might be related to a coherence length, and the mean free path (λ, the inverse macroscopic scattering cross section) that reflects the fact that there exist particle collisions, moreover the effective displacement speed of the fluid flow (v) and the thermal velocity (c_{therm}) that reflects the random walk of the microscopic constituents of the fluid. The new definition solved two critical points of the traditional formulation. For ideal fluids the Reynolds number approaches zero (instead of infinity) and the vortex correlation length that sets up a length scale allows for vorticity from a potential model, which is incompatible with the traditional expression for vorticity and makes use of the rotational operation, only.

A considerable number of works literally discard the possibility to understand turbulence in the end from first principles and focus on numerical analysis of existent continuous models [BuEa05], [CaEtAl04], [PeHu06], [RuSmHu10], [vaClWi08], [ZhPr05], without challenging the breaking with sedated paradigms. Although we presented only a simple model but with the desired properties, we believe to have opened a doorway for a new theoretical and progressive understanding of turbulent phenomena by another step into a new but promising direction.

References

[BoEtAl10] Bodmann, B.E.J., Vilhena, M.T., Zabadal, J.R., Beck, D.: On a new definition of the Reynolds number from the interplay of macroscopic and microscopic phenomenology. In: Constanda, C., Harris, P.J. (eds.) Integral Methods in Science and Engineering: Computational and Analytic Aspects, pp. 7–14. Birkhäuser, Boston (2011)

[BoEtAl06] Bonetto, F., Gallavotti, G., Giuliani, A., Zamponi, F.: Chaotic hypothesis, fluctuation theorem and singularities. J. Stat. Phys. **123**, 39–54 (2006)

[BuEa05] Burton, T.M., Eaton, J.K.: Fully resolved simulations of particle-turbulence interaction. J. Fluid Mech. **545**, 67–111 (2005)

[CaEtAl04] Cate, A.T., Derksen, J.J., Portela, L.M., Van den Akker, H.E.A.: Fully resolved simulations of colliding mono-disperse spheres in forced isotropic turbulence. J. Fluid Mech. **519**, 233–271 (2004)

[GuEtAl05] Guala, M., Luthi, B., Liberzon, A., Tsinober, A., Kinzelbach, W.: On the evolution of material lines and vorticity in homogeneous turbulence. J. Fluid Mech. **533**, 339–359 (2005)

[HoHo03] Hoover, W.G., Hoover, C.G.: Links between microscopic and macroscopic fluid mechanics. Mol. Phys. **101**, 1559–1573 (2003)

[JiBo05] Jirkovsky, L., Bo-ot, L.: Momentum transport equation for the fluids using projection perturbation formalism and onset of turbulence. Physica A **352**, 241–251 (2005)

[LuTsKi05] Lüthi, B., Tsinober, A., Kinzelbach, W.: Lagrangian measurement of vorticity dynamics in turbulent flow. J. Fluid Mech. **528**, 87–118 (2005)

[MaEtAl07] Mathur, M., Haller, G., Peacock, Th., Ruppert-Felsot, J.E., Swinney, H.L.: Uncovering the Lagrangian skeleton of turbulence. Phys. Rev. Lett. **98**, 144502 (2007)

[PeHu06] Perrin, A., Hu, H.H.: An explicit finite-difference scheme for simulation of moving particles. J. Comput. Phys. **212**, 166–187 (2006)

[RuSmHu10] Rumsey, C.L., Smith, B.R., Huang, G.P.: Consistency, verification, and validation of turbulence models for Reynolds-averaged Navier–Stokes applications. J. Aerospace Eng. **224**, 1211–1218 (2010)

[SaCo09] Salazar, J.P.L.C., Collins, L.R.: Two-particle dispersion in isotropic turbulent flows. Ann. Rev. Fluid Mech. **41**, 405–432 (2009)

[ToBo09] Toschi, F., Bodenschatz, E.: Lagrangian properties of particles in turbulence. Ann. Rev. Fluid Mech. **41**, 375–404 (2009)

[vaClWi08] van Aartrijk, M., Clercx, H.J.H., Winters, K.B.: Single-particle, particle-pair, and multi-particle dispersion of fluid particles in forced stably stratified turbulence. Phys. Fluids **20**, 025104–025116 (2008)

[Ye02] Yeung, P.K.: Lagrangian investigations of turbulence. Ann. Rev. Fluid Mech. **34**, 115–142 (2002)

[ZhPr05] Zhang, Z., Prosperetti, A.: A second-order method for three-dimensional particle simulation. J. Comput. Phys. **210**, 292–324 (2005)

Chapter 6
Numerical Simulation of the Dynamics of Molecular Markers Involved in Cell Polarization

V. Calvez, N. Meunier, N. Muller, and R. Voituriez

6.1 Introduction

Cell polarization is a major step involved in several important cellular processes such as directional migration, growth, oriented secretion, cell division, mating, or morphogenesis. When a cell is not polarized molecular markers (proteins CDC42) are uniformly distributed on the membrane while polarization is characterized by the concentration of molecular markers in a small area of the cell membrane. In [WeAlLi03], it has been observed that if the external pheromone concentration is above a critical concentration, polarization can occur spontaneously. It has also been observed that cell asymmetry can be driven by an external asymmetric stimulus.

Cell polarization in yeast cells has been intensively studied during the past decade. Recently, many models describing cell polarization have been developed. The majority of these models are based on reaction–diffusion systems where polarization appears as a type of Turing instability [IgDe08], [OnRa07], [LeKeRa06], or due to stochastic fluctuations [AlEtAl08], other models include cytoskeleton proteins as a regulatory factor [EuEtAl07], [WeAlLi03]. Many biological studies have shown that the cytoskeleton plays an important role in polarization. It has been suggested that the cytoskeleton has a positive feedback on molecular markers density. Indeed, disruption of transport along the cytoskeleton greatly reduces the

V. Calvez
École Normale Supérieure de Lyon, Lyon Cedex, France
e-mail: vincent.calvez@umpa.ens-lyon.fr

N. Meunier (✉) • N. Muller
Université Paris Descartes, Paris, France
e-mail: nicolas.meunier@parisdescartes.fr; nicolas.muller@parisdescartes.fr

R. Voituriez
Université Pierre et Marie Curie, Paris Cedex, France
e-mail: voiturie@lptmc.jussieu.fr

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_6,

stability of polar cap [WeAlLi03]. The cell cytoskeleton is a network of long semi-flexible filaments made up of protein subunits [PhKoTh09]. These filaments (mainly actin or microtubules) act as roads along which motor proteins are able to perform a biased ballistic motion and carry various molecules. Molecular markers play a key role in the formation of these filaments.

Following [HaEtAl09], [CaMeVo10], and [CaEtAl12], in this work we study models that describe the dynamics of cell polarization. In these models, molecular markers, such as proteins, diffuse in the cytoplasm and are actively transported along the cytoskeleton. The resulting motion is a biased diffusion regulated by the markers themselves. Using numerical simulations and mathematical heuristics, we observe that the coupling on the velocity field achieves an inhomogeneous distribution of molecular markers without any external asymmetric field. Such an inhomogeneous distribution is only due to interaction between molecular markers.

Throughout this paper, the density of molecular markers (resp. advection field) is denoted by $\rho(t,\mathbf{x})$ (resp. $\mathbf{u}(t,\mathbf{x})$). The advection is obtained though a coupling with the membrane concentration of markers. The cell is figured by the domain $\Omega \subset \mathbb{R}^n$ with $n = 1,2$ and a part of the boundary of the domain will be the active membrane denoted by Γ. The time evolution of the molecular markers satisfies the following advection–diffusion equation, see [HaEtAl09] and [CaEtAl12]:

$$\begin{cases} \partial_t \rho(t,\mathbf{x}) = D\Delta\rho(t,\mathbf{x}) - \chi\nabla.\left(\rho(t,\mathbf{x})\,\mathbf{u}(t,\mathbf{x})\right), \quad t>0, \quad \mathbf{x}\in\Omega, \\ \rho(0,\mathbf{x}) = \rho_0(\mathbf{x}). \end{cases} \tag{6.1}$$

There is neither creation nor degradation of molecular markers in the cell, so the quantity of molecular markers remains constant in time:

$$M = \int_{x\in\Omega} \rho_0(\mathbf{x})d\mathbf{x} = \int_{x\in\Omega} \rho(t,\mathbf{x})d\mathbf{x}. \tag{6.2}$$

This condition is ensured by a zero flux boundary condition on the boundary. A first simplified step is to assume that the cell is essentially bidimensional and to neglect curvature effects. The membrane boundary is then a 1D line along the y-axis and the cytoplasm is parametrized by $\mathbf{x} = (x,y) \in \mathbb{R}_+ \times \mathbb{R}$.

The plan of this work is the following. First, we recall the main mathematical results of the simplified model in 1D for $\Omega = (0,\infty)$ and $\Gamma = \{x = 0\}$, see [CaMeVo10], [CaEtAl12] for more details. Then we study a more realistic model that includes dynamical exchange of markers on the boundary for a general Ω. This model was introduced in [HaEtAl09] and studied in [CaEtAl12] in the one-dimensional case. Here, we will perform a first numerical analysis of this model in the two-dimensional case, for periodic (in one direction) and bounded (in the other direction) domain. Finally, we provide a methodology for parameter estimation by using mathematical heuristics and biological literature.

6.1.1 One-Dimensional Case

In this section, we study the one-dimensional case on the half line for $\Omega = (0,\infty)$. The membrane is then the point $\Gamma = \{x = 0\}$. For the first model, the advection field towards the membrane is equal to the density of molecular markers on the boundary $\rho(t,0)$. Then we improve this model by considering that only the trapped molecular markers on the membrane contribute to the advection field.

6.1.1.1 Simplified Model Set on the Half Line

In [CaEtAl12] the first mathematical study has been done on this model. We define an advection field $\mathbf{u}(t,x)$ for (6.1)

$$\mathbf{u}(t,x) = -\rho(t,0),$$

in such a case (6.1) reads as (with $D = 1$ and $\chi = 1$):

$$\partial_t \rho(t,x) = \partial_{xx}\rho(t,x) + \rho(t,0)\,\partial_x \rho(t,x), \quad t > 0, \quad x > 0, \tag{6.3}$$

with the following zero flux condition on the boundary $\{x = 0\}$, that ensures the mass conversation (6.2),

$$\partial_x \rho(t,0) + \rho(t,0)^2 = 0.$$

In [CaEtAl12], it has been proved that solutions of (6.3) blow-up in finite time if their masses are above a certain critical mass, $M > 1$, and exist globally in time if $M \leq 1$. Let us first recall the definition of weak solutions of (6.3).

Definition 1. We say that $\rho(t,x)$ is a weak solution of (6.3) on $(0,T)$ if it satisfies

$$\rho \in L^\infty(0,T;L^1_+(\mathbb{R}_+)), \quad \partial_x \rho \in L^1((0,T)\times\mathbb{R}_+),$$

and $\rho(t,x)$ is a solution of (6.3) in the sense of distributions in $\mathscr{D}'(\mathbb{R}_+)$.

Let us now recall the main results for weak solutions of (6.3).

Theorem 1 (Global existence: $M \leq 1$). *Assume that the initial data ρ_0 satisfies both $\rho_0 \in L^1((1+x)\,\mathrm{d}x)$ and $\int_{x>0} \rho_0(x)(\log \rho_0(x))_+ \,\mathrm{d}x < +\infty$. Assume in addition that $M \leq 1$, then there exists a global weak solution of (6.3).*

Theorem 2 (Blow-up: $M > 1$). *Assume $M > 1$. Any weak solution of (6.3) with non-increasing initial data ρ_0 blows-up in finite time.*

Remark 1. It would tempting to interpret blow-up of solutions of the one-dimensional model as cell polarization. But it is to be noticed that concentration

of markers on the boundary doesn't mean polarization. Indeed, consider a radially symmetric 2D cell case. Equation then reduces to the one-dimensional one. Above a threshold on the total mass, the convection wins and markers concentrate on the boundary. In some situations, these markers may be homogeneously distributed on the boundary and in such a case there is no symmetry breaking.

6.1.1.2 The Model with Dynamical Exchange of Markers at the Boundary

Such a direct activation of transport on the boundary seems to be unrealistic. Indeed possible occurrence of blow-up in finite time suggests this claim. We improve the previous model by distinguishing between cytoplasmic content $\rho(t,x)$ and the concentration of trapped molecules on the boundary that will be denoted by $\mu(t)$. The dynamical exchange of markers at the boundary is done with an attachment rate k_{on} and a detachment rate k_{off}, hence the time evolution of $\mu(t)$ is

$$\frac{d}{dt}\mu(t) = k_{on}\,\rho(t,0) - k_{off}\,\mu(t). \tag{6.4}$$

The advection field $\mathbf{u}(t,x)$ in (6.1) is now defined by

$$\mathbf{u}(t,x) = -\mu(t),$$

hence (6.1) (with $D=1$ and $\chi=1$) reads as

$$\partial_t\rho(t,x) = \partial_{xx}\rho(t,x) + \mu(t)\,\partial_x\rho(t,x), \quad t>0, \quad x>0,$$

with a modified boundary condition

$$\partial_x\rho(t,0) + \rho(t,0)\,\mu(t) = \frac{d}{dt}\mu(t).$$

This ensures the following mass conservation shared among $\rho(t,x)$ and $\mu(t)$:

$$M = \int_{\mathbb{R}_+}\rho_0(x)dx + \mu_0 = \int_{\mathbb{R}_+}\rho(t,x)dx + \mu(t).$$

With (6.4), the self-activation of transport by $\rho(t,0)$ is then delayed in time. Since the transport speed is bounded $\mu(t)\le M$, the solution of the model with dynamical exchange on the boundary exists globally in time. More precisely it is possible (see [CaEtAl12]) to prove that it converges towards a nontrivial stationary state.

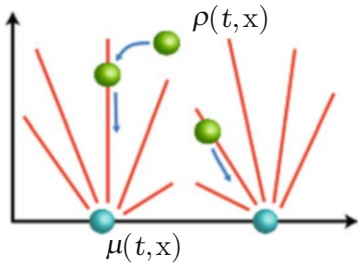

Fig. 6.1 Advection field orientation due to actin networks

6.1.2 *Two-Dimensional Case: The Model with Dynamical Exchange of Markers at the Boundary*

Let $\Omega \subset \mathbb{R}^2$ be the cytoplasm domain, as in the one-dimensional case (6.4) we consider dynamical exchange of markers at the boundary, so for $\mathbf{x} \in \Gamma$ we have the evolution in time of $\mu(t,\mathbf{x})$

$$\partial_t \mu(t,\mathbf{x}) = k_{on}\,\rho(t,\mathbf{x}) - k_{off}\,\mu(t,\mathbf{x}).$$

with a modified boundary condition for $\rho(t,\mathbf{x})$ at point $\mathbf{x} \in \Gamma$

$$(D\nabla\rho(t,\mathbf{x}) - \chi\,\rho(t,\mathbf{x})\,\mathbf{u}(t,\mathbf{x})).\vec{n}_{\mathbf{x}} = -\partial_t \mu(t,\mathbf{x}),$$

where $\vec{n}_{\mathbf{x}}$ is the outward normal to Γ. This ensures the following mass conservation sharing by $\rho(t,\mathbf{x})$ and $\mu(t,\mathbf{x})$:

$$M = \int_\Omega \rho_0(\mathbf{x})d\mathbf{x} + \int_\Gamma \mu_0(\mathbf{x})d\mathbf{x} = \int_\Omega \rho(t,\mathbf{x})d\mathbf{x} + \int_\Gamma \mu(t,\mathbf{x})d\mathbf{x}.$$

We consider the advection field deriving from a harmonic potential modeling the transport by actin filaments (cytoskeleton), namely

$$\mathbf{u}(t,\mathbf{x}) = \nabla c(t,\mathbf{x}),\ \text{where}\ \begin{cases} -\Delta c(t,\mathbf{x}) = 0, & \text{if } \mathbf{x} \in \Omega, \\ \nabla c(t,\mathbf{x}).\vec{n}_{\mathbf{x}} = S(\mathbf{x})\mu(t,\mathbf{x}), & \text{if } \mathbf{x} \in \Gamma. \end{cases} \tag{6.5}$$

This advection field orientation is due to the actin networks (see Fig. 6.1).

Actin filaments are attached on the membrane and randomly distributed, there orientations are mixed up. We also add the external pheromone concentration at $\mathbf{x} \in \Gamma$ which acts by the mating-pheromone MAPK cascade on the actin transport.

In dimension 2, we have global existence for the model without exchange on the boundary (replacing (6.5) by $\nabla c(t,\mathbf{x}).\vec{n}_{\mathbf{x}} = S(\mathbf{x})\rho(t,\mathbf{x})$ if $\mathbf{x} \in \Gamma$) with $\Omega = (0,+\infty) \times \mathbb{R}$ and $\Gamma = \{0\} \times \mathbb{R}$. For clarity, we recall this result, see [CaEtAl12] for more details.

Theorem 3 (Global existence in dimension 2). *Assume that the advection field satisfies the two following conditions:* $\nabla \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u}(t,0,y) \cdot \mathbf{e}_e = \rho(t,0,y)$. *Assume that the initial data* ρ_0 *satisfies both* $\rho_0 \in L^1((1+|\mathbf{x}|^2)\,\mathrm{d}\mathbf{x})$ *and* $\|\rho_0\|_{L^2}$ *is smaller than some constant c. Then there exists a global weak solution to* (6.1), (6.2).

In the two-dimensional case, for the model with exchange on the boundary, blow-up or global existence has not been proved yet. In this work, we make a first step in this direction by using a mathematical heuristic and numerical simulations.

6.1.3 Heuristics

The mathematical analysis performed in [CaEtAl12] has demonstrated that a class of models exhibit pattern formation (either blow-up or convergence towards a non-homogeneous steady state) under some conditions. However the main question still remains unanswered: do these models describe cell polarization or not? Thus in order to provide a first answer to this question, we will perform numerical simulations. Our aim is to see if, under some conditions, the model leads to a concentration of markers, not only on the boundary but as on a small region of the boundary. In such a case polarization occurs. In order to obtain more information on the critical value distinguishing the polarized case and the stable case, in the two-dimensional case we will use a mathematical heuristics that we describe now.

Let $\mathbf{x} = (x,y)$ be in $\Omega = \mathbb{R}_+ \times \mathbb{R}$, and let $\Gamma = \{0\} \times \mathbb{R}$ be the boundary, we have

$$\mathbf{u}(t,\mathbf{x}) = \nabla c(t,\mathbf{x}), \text{ where } \begin{cases} -\Delta c(t,\mathbf{x}) = 0, & \text{if } \mathbf{x} \in \mathbb{R}_+ \times \mathbb{R}, \\ -\partial_x c(t,0,y) = S(y)\mu(t,y), & \text{if } y \in \mathbb{R}, \end{cases}$$

hence (see, e.g., [Ev98]), it is well known that

$$c(x,y) = -\frac{1}{\pi} \int_{y' \in \mathbb{R}} \log\left(\sqrt{(y-y')^2 + x^2}\right)(S\mu)(y')dy'.$$

The tangential component at the boundary is then given by

$$\mathbf{u}(t,0,y) \cdot \vec{\mathbf{e}}_y = -\mathscr{H}(S\mu)(y), \quad y \in \mathbb{R},$$

where $\mathscr{H}$ denotes the one-dimensional Hilbert transform that we recall now (see, e.g., [CaPeTa07]) with respect to the y variable:

$$\mathscr{H}(f\mu)(y) = \frac{1}{\pi} \text{p.v.} \int_{\mathbb{R}} \frac{1}{y-y'} f(y')\,dy'.$$

Integrating the main equation (6.1) with respect to x with zero flux condition on $\Gamma = \{x = 0\}$, we obtain:

$$\partial_t \int_{x>0} \rho(t,x,y)\,dx = D\,\partial_{yy}\left(\int_{x>0} \rho(t,x,y)\,dx\right) - \chi\,\partial_y\left(\int_{x>0} \rho(t,x,y)(\mathbf{u}(t,x,y)\cdot\vec{\mathbf{e}}_y)dx\right).$$

In the super-critical case, numerical simulations, see [Mu13], suggest that the density $\rho(t,\mathbf{x})$ concentrates on the boundary $\{x = 0\}$. Assuming $\rho(t,x,y) = \nu(t,y)\delta(x=0)$, we can formally write the dynamics of $\nu(t,y)$ as

$$\partial_t \nu(t,y) = D\,\partial_{yy}\nu(t,y) + \chi\,\partial_y\left(\nu(t,y)\mathscr{H}(S\mu)(y)\right).$$

Assuming S constant on $\mathbb{R}$ and $\mu(t,y) = \frac{k_{on}}{k_{off}}\nu(t,y)$ for $y \in \mathbb{R}$, it reads as

$$\partial_t \nu(t,y) = D\,\partial_{yy}\nu(t,y) + \chi S\frac{k_{on}}{k_{off}}\,\partial_y\left(\nu(t,y)\mathscr{H}(\nu)(y)\right).$$

The Hilbert transform has a critical singularity to offset the diffusion on this equation [CaPeTa07]. We have a blow-up if $\int_{\mathbb{R}} \nu(t,y)\,dy = M$ is above $\frac{2\pi D k_{off}}{S\chi k_{on}}$. This is the first step to observe a critical mass phenomenon and this may lead to blow-up if the mass is large enough. In this way, we define an order of magnitude for some parameters.

It is to be noticed that this latter criterion is valid for an infinite domain, namely $y \in \mathbb{R}$. In the case of a cell, the domain will be finite and the existence of such a dichotomy has not been proved yet. In order to see if such a dichotomy holds true we will perform numerical simulations. This is the object of the following section.

6.2 Numerical Analysis

We first give a discretization of the convection–diffusion model set on a 1D periodic domain. This first step allows us introducing the discretization of this model on a 2D domain which is periodic in one direction and bounded on the other direction.

6.2.1 One-Dimensional Case

Let $u(t,x)$ be a given function. We consider the following advection–diffusion equation on the periodic domain $\Omega = \mathbb{R}/\mathbb{Z}$

$$\partial_t \rho(t,x) = \partial_x(\partial_x \rho(t,x) - u(t,x)\,\rho(t,x)), \quad t>0, \quad x \in \Omega. \tag{6.6}$$

Let $t^n = n\,dt$ be the time discretization and $\{x_j = j\,dx, j \in \{1,\dots,N_x\}\}$ be the space discretization of the periodic interval $\mathbb{R}/\mathbb{Z}$. Since the equations of the model are written in a conservative form, the natural framework to be used for the spatial discretization is the finite volume framework. We hence introduce the control volume defined for $j \in \{1,\dots,N_x\}$

$$V_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}). \tag{6.7}$$

Let ρ_j^n (resp. $u^n_{j+\frac{1}{2}}$) be the approximated value of the exact solution $\rho(t^n,x_j)$ (respectively, $u(t^n,x_{j+\frac{1}{2}})$), the classical upwind scheme for (6.6) is

$$\frac{\rho_j^{n+1}-\rho_j^n}{dt} = \frac{\mathscr{F}_{j+\frac{1}{2}} - \mathscr{F}_{j-\frac{1}{2}}}{dx}, \quad j \in \{1,\dots,N_x\},$$

where the numerical flux $\mathscr{F}_{j+\frac{1}{2}}$ and $\mathscr{F}_{j-\frac{1}{2}}$ are defined by

$$\mathscr{F}_{j+\frac{1}{2}} = \frac{\rho_{j+1}^{n+1}-\rho_j^{n+1}}{dx} - A^{up}(u^n_{j+\frac{1}{2}}, \rho_j^n, \rho_{j+1}^n),$$

$$\mathscr{F}_{j-\frac{1}{2}} = \frac{\rho_j^{n+1}-\rho_{j-1}^{n+1}}{dx} - A^{up}(u^n_{j-\frac{1}{2}}, \rho_{j-1}^n, \rho_j^n),$$

with the advection numerical flux given by

$$A^{up}(u,x_-,x_+) = \begin{cases} u x_-, & \text{si } u>0, \\ u x_+, & \text{si } u<0. \end{cases} \tag{6.8}$$

The periodic flux condition on boundary reads as $\mathscr{F}_{\frac{1}{2}} = \mathscr{F}_{N_x+\frac{1}{2}}$ and we set the value $u^n_{\frac{1}{2}} = u^n_{N_x+\frac{1}{2}}$. The diffusion part is treated implicitly and it is then unconditionally stable, while the advection term is treated explicitly. The CFL condition of the scheme is

$$\left|\left|\left(u^n_{j+\frac{1}{2}}\right)_{j\in\{1,\dots,N_x\}}\right|\right|_\infty < \frac{dx}{dt}.$$

We define the column vector $\rho^n = \left(\rho_1^n\ \rho_2^n\ \dots\ \rho_{N_x}^n\right)^T$. As usual (see, e.g., [Al07]), the discrete heat matrix $A \in M_{N_x}(\mathbb{R})$ with periodic flux condition on the boundary is defined by

$$A = \begin{pmatrix} 2+\frac{dx^2}{dt} & -1 & & & -1 \\ -1 & 2+\frac{dx^2}{dt} & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 2+\frac{dx^2}{dt} & -1 \\ -1 & & & -1 & 2+\frac{dx^2}{dt} \end{pmatrix}. \tag{6.9}$$

Periodic flux condition adds the top-right term and the bottom-left term. Next, in order to use A^{up} defined by (6.8), we define

$$(u)^+ = \max(u,0), \quad (u)^- = \min(u,0).$$

The discrete advection matrix $B \in M_{N_x}(\mathbb{R})$ with periodic flux condition on the boundary is then defined as in [Al07]

$$B = \frac{dx^2}{dt} I_{N_x} - dx \begin{pmatrix} \left(u^n_{\frac{3}{2}}\right)^+ & \left(u^n_{\frac{3}{2}}\right)^- & & & \\ & \ddots & \ddots & & \\ & & \left(u^n_{j+\frac{1}{2}}\right)^+ & \left((u^n_{j+\frac{1}{2}}\right)^- & \\ & & & \ddots & \left(u^n_{N_x-\frac{1}{2}}\right)^- \\ \left(u^n_{N_x+\frac{1}{2}}\right)^- & & & & \left(u^n_{N_x+\frac{1}{2}}\right)^+ \end{pmatrix}$$

$$+ dx \begin{pmatrix} \left(u^n_{\frac{1}{2}}\right)^- & & & & \left(u^n_{\frac{1}{2}}\right)^+ \\ \left(u^n_{\frac{3}{2}}\right)^+ & \ddots & & & \\ & \left(u^n_{j-\frac{1}{2}}\right)^+ & \left(u^n_{j-\frac{1}{2}}\right)^- & & \\ & & \ddots & \ddots & \\ & & & \left(u^n_{N_x-\frac{1}{2}}\right)^+ & \left(u^n_{N_x-\frac{1}{2}}\right)^- \end{pmatrix}. \tag{6.10}$$

We use a standard numerical method to invert the symmetric positive definite matrix A. Finally, at each time step we resolve

$$\rho^{n+1} = A^{-1} B \rho^n.$$

6.2.2 Two-Dimensional Case

We perform numerical simulations on the model with dynamical exchange of markers at the boundary. In this work, we assume that the cell occupies a disk of radius $r > 0$. Furthermore for simplicity, we consider a bounded-periodic domain $\Omega = [0,r] \times \mathbb{R}/2\pi r\mathbb{Z}$ with $\Gamma = \{r\} \times \mathbb{R}/2\pi r\mathbb{Z}$. This simplifies our numerical approach by using finite difference schemes on Cartesian grid. We start with the numerical study of the equation on ρ by assuming that the advection field $u(t,\mathbf{x}) = \nabla c(t,\mathbf{x})$ is known. Then we perform the discretization of c.

In this section, for simplicity we fix all the parameter values to 1 except M. Let us first recall the model with dynamical exchange of markers at the boundary on $\Omega = [0,r] \times \mathbb{R}/2\pi r\mathbb{Z}$:

$$\partial_t \rho = \nabla.(\nabla\rho - \rho\,\nabla c) \text{ in } (0,r) \times \mathbb{R}/2\pi r\mathbb{Z}, \tag{6.11}$$

$$\partial_x \rho - \rho\,\partial_x c = -\partial_t \mu \text{ on } \{r\} \times \mathbb{R}/2\pi r\mathbb{Z}, \tag{6.12}$$

$$\partial_x \rho - \rho\,\partial_x c = 0 \text{ on } \{0\} \times \mathbb{R}/2\pi r\mathbb{Z}. \tag{6.13}$$

Dynamical exchange markers on active boundary $\{r\} \times \mathbb{R}/2\pi r\mathbb{Z}$ is given by

$$\partial_t \mu = \rho - \mu, \text{ on } \{r\} \times \mathbb{R}/2\pi r\mathbb{Z}. \tag{6.14}$$

Laplace equation on c with inappropriate Neumann conditions on a bounded domain is ill-posed (see, e.g., [Al07]). In order to handle this problem, we add the degradation term

$$-\Delta c + \alpha\, c = 0 \text{ in } (0,r) \times \mathbb{R}/2\pi r\mathbb{Z}, \tag{6.15}$$

$$-\partial_x c = \mu \text{ on } \{r\} \times \mathbb{R}/2\pi r\mathbb{Z}, \tag{6.16}$$

$$-\partial_x c = 0 \text{ on } \{0\} \times \mathbb{R}/2\pi r\mathbb{Z}. \tag{6.17}$$

We take random initial conditions c, μ_0 and ρ_0 satisfying the following mass conservation

$$\int_\Omega \rho_0 + \int_\Gamma \mu_0 = M. \tag{6.18}$$

Let $t^n = n\,dt$ be the time discretization and $\{x_j = j\,dx, j \in \{1,\dots,N_x\}\}$ (respectively, $\{y_k = k\,dy, k \in \{1,\dots,N_y\}\}$) be the space discretization of the bounded interval $[0,r)$ (respectively, periodic interval $\mathbb{R}/2\pi r\mathbb{Z}$). We define the control volume $W_{(j,k)} \subset \mathbb{R}^2$ by

$$W_{(j,k)} = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}) \times (y_{k-\frac{1}{2}}, y_{k+\frac{1}{2}}).$$

Let $P^n_{(j,k)}$ (resp. μ^n_k) be the approximate value of the exact solution $\rho(t^n, x_j, y_k)$ of (6.11)–(6.13) and (6.18) (resp. $\mu(t^n, y_k)$ of (6.14) and (6.18)). Let $c_{(j,k)}$ be the approximated value of the exact solution $c(x_j, y_k)$ of (6.15)–(6.17).

6.2.2.1 Equation for μ

We can resolve at each time step for $k \in \{1, \ldots, N_y\}$

$$\mu^{n+1}_k = \mu^n_k + dt\,(\rho^n_k - \mu^n_k).$$

6.2.2.2 Equation for c

For simplicity, we call $\mathscr{F}$ the numerical flux as in the 1D case, we can write the following scheme: for $(j,k) \in \{1, \ldots, N_x\} \times \{1, \ldots, N_y\}$

$$\frac{\mathscr{F}_{(j+\frac{1}{2},k)} - \mathscr{F}_{(j-\frac{1}{2},k)}}{dx} + \frac{\mathscr{F}_{(j,k+\frac{1}{2})} - \mathscr{F}_{(j,k-\frac{1}{2})}}{dy} - \alpha c_{(j,k)} = 0.$$

with numerical flux defined by

$$\mathscr{F}_{(j+\frac{1}{2},k)} = \frac{c_{(j+1,k)} - c_{(j,k)}}{dx}, \quad \mathscr{F}_{(j-\frac{1}{2},k)} = \frac{c_{(j,k)} - c_{(j-1,k)}}{dx}.$$

The zero flux boundary condition (6.17) means that $\mathscr{F}_{(\frac{1}{2},k)} = 0$ and the boundary condition (6.16) $\mathscr{F}_{(N_x+\frac{1}{2},k)} = -\mu^n_k$ for $k \in \{1, \ldots, N_y.\}$ Similarly, the periodic conditions lead to $\mathscr{F}_{(j,N_y+\frac{1}{2})} = \mathscr{F}_{(j,\frac{1}{2})}$ for $j \in \{1, \ldots, N_x.\}$ We define the column vector $\mathscr{C}$ by $\mathscr{C}(k + (j-1)N_y) = C_{(j,k)}$ with $(j,k) \in \{1, \ldots, N_x\} \times \{1, \ldots, N_y.\}$ As before, the rigidity matrix $A_{2D,\alpha}$ is defined by

$$A_{2D,\alpha} = \begin{pmatrix} A_\alpha + Id & -Id & & & & & \\ -Id & A_\alpha + 2Id & -Id & & & & \\ & -Id & A_\alpha + 2Id & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & A_\alpha + 2Id & -Id & \\ & & & & -Id & A_\alpha + 2Id & -Id \\ & & & & & -Id & A_\alpha + Id \end{pmatrix},$$

where the discrete Poisson matrix $A_\alpha \in M_{N_y}(\mathbb{R})$ in 1D with periodic flux condition on boundary is defined by

$$A_\alpha = \begin{pmatrix} 2+\alpha\,dx^2 & -1 & & & -1 \\ -1 & 2+\alpha\,dx^2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 2+\alpha\,dx^2 & -1 \\ -1 & & & -1 & 2+\alpha\,dx^2 \end{pmatrix}.$$

The flux boundary condition $\{r\} \times \mathbb{R}/2\pi r\mathbb{Z}$ generates the right-hand side column vector of length $N_x N_y$

$$R_c = -dx\left((\mu_k^n)_k\ 0\ \dots\ 0\right).$$

We use a standard numerical method to invert the symmetric positive definite matrix $A_{2D,\alpha}$ and then solve at each time step; that is,

$$\mathscr{C} = A_{2D,\alpha}^{-1} R_c.$$

6.2.2.3 Equation for ρ

For simplicity, we call $\mathscr{F}$ the numerical flux as in the previous cases, we can write the upwind scheme as follows:

$$\frac{P^{n+1}_{(j,k)} - P^n_{(j,k)}}{dt} = \frac{\mathscr{F}_{(j+\frac{1}{2},k)} - \mathscr{F}_{(j-\frac{1}{2},k)}}{dx} + \frac{\mathscr{F}_{(j,k+\frac{1}{2})} - \mathscr{F}_{(j,k-\frac{1}{2})}}{dy},$$

where

$$u^n_{(j+\frac{1}{2},k)} = \frac{c^n_{(j+1,k)} - c^n_{(j,k)}}{dx}, \quad u^n_{(j-\frac{1}{2},k)} = \frac{c^n_{(j,k)} - c^n_{(j-1,k)}}{dx},$$

and the numerical flux is defined by

$$\mathscr{F}_{(j+\frac{1}{2},k)} = \frac{\rho^{n+1}_{(j+1,k)} - \rho^{n+1}_{(j,k)}}{dx} - A^{up}\left(u^n_{(j+\frac{1}{2},k)}, P^n_{(j,k)}, P^n_{(j+1,k)}\right),$$

$$\mathscr{F}_{(j-\frac{1}{2},k)} = \frac{\rho^{n+1}_{(j,k)} - \rho^{n+1}_{(j-1,k)}}{dx} - A^{up}\left(u^n_{(j-\frac{1}{2},k)}, P^n_{(j-1,k)}, P^n_{(j,k)}\right).$$

The zero flux boundary conditions (6.13) leads to $\mathscr{F}_{(\frac{1}{2},k)} = 0$, while the flux boundary conditions (6.12) mean that

$$\mathscr{F}_{(N_x+\frac{1}{2},k)} = -\frac{\mu_k^{n+1} - \mu_k^n}{dt}$$

for $k \in \{1,\dots,N_y\}$. Similarly, the periodic conditions generate the equality

$$\mathscr{F}_{(j,N_y+\frac{1}{2})} = \mathscr{F}_{(j,\frac{1}{2})}$$

for $j \in \{1,\dots,N_x\}$. We define the column vector $\mathscr{P}^n$ by

$$\mathscr{P}^n(k+(j-1)N_y) = P^n_{(j,k)}$$

with $(j,k) \in \{1,\dots,N_x\} \times \{1,\dots,N_y\}$. For simplicity, in what follows we consider that $dx = dy$. We define the rigidity matrix $A_{2D} \in M_{N_xN_y}(\mathbb{R})$ with $A \in M_{N_y}(\mathbb{R})$ defined by (6.9):

$$A_{2D} = \begin{pmatrix} A+Id & -Id & & & \\ -Id & A+2Id & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & A+2Id & -Id \\ & & & -Id & A+Id \end{pmatrix}.$$

We define the following diagonal matrices for $j \in \{1,\dots,N_x\}$, $U^+_{j+\frac{1}{2}} \in M_{N_y}(\mathbb{R})$ and $U^-_{j+\frac{1}{2}} \in M_{N_y}(\mathbb{R})$:

$$U^+_{j+\frac{1}{2}} = \begin{pmatrix} \ddots & & & & \\ & (u^n_{(j+\frac{1}{2},k-1)})^+ & & & \\ & & (u^n_{(j+\frac{1}{2},k)})^+ & & \\ & & & (u^n_{(j+\frac{1}{2},k+1)})^+ & \\ & & & & \ddots \end{pmatrix},$$

$$U^-_{j+\frac{1}{2}} = \begin{pmatrix} \ddots & & & & \\ & (u^n_{(j+\frac{1}{2},k-1)})^- & & & \\ & & (u^n_{(j+\frac{1}{2},k)})^- & & \\ & & & (u^n_{(j+\frac{1}{2},k+1)})^- & \\ & & & & \ddots \end{pmatrix}.$$

With $B \in M_{N_y}(\mathbb{R})$ defined by (6.10), the discrete advection matrix $B_{2D} \in M_{N_xN_y}(\mathbb{R})$ with zero flux boundary condition in the x-axis direction and periodic flux boundary condition in the y-axis direction is defined by

$$B_{2D} = \begin{pmatrix} B & & & & \\ & B & & & \\ & & \ddots & & \\ & & & B & \\ & & & & B \end{pmatrix} - dx \begin{pmatrix} U^+_{\frac{3}{2}} & U^-_{\frac{3}{2}} & & & & \\ & \ddots & \ddots & & & \\ & & U^+_{j+\frac{1}{2}} & U^-_{j+\frac{1}{2}} & & \\ & & & \ddots & U^-_{N_x-\frac{1}{2}} & \\ & & & & U^+_{N_x+\frac{1}{2}} & \end{pmatrix}$$

$$+ dx \begin{pmatrix} U^-_{\frac{1}{2}} & & & & & \\ U^+_{\frac{3}{2}} & \ddots & & & & \\ & U^+_{j-\frac{1}{2}} & U^-_{j-\frac{1}{2}} & & & \\ & & \ddots & \ddots & & \\ & & & U^+_{N_x-\frac{1}{2}} & U^-_{N_x-\frac{1}{2}} & \end{pmatrix}.$$

The flux boundary condition $\{r\} \times \mathbb{R}/2\pi r\mathbb{Z}$ imposes this right-hand side column vector of length $N_x N_y$:

$$R_\rho = -dx \left(\left(\frac{\mu_k^{n+1} - \mu_k^n}{dt} \right)_k 0 \dots 0 \right).$$

We use a standard numerical method to inverse the symmetric positive definite matrix A_{2D} and then resolve at each time step

$$\mathscr{P}^{n+1} = A_{2D}^{-1} (B_{2D} \mathscr{P}^n + R_\rho).$$

6.2.3 Graphics

With the previous numerical analysis, we implement all numerical simulations using MATLAB. The results of testing different values of M are shown in Fig. 6.2.

6.3 Conclusion

In this work we have provided a first answer to the following question: do the nonlinear convection–diffusion models given in [HaEtAl09] and [CaEtAl12] describe cell polarization or not? To do so, we have used both a mathematical heuristic and numerical simulations. Numerical simulations were necessary because the heuristic is only valid for an infinite geometry while the cell is obviously finite. The numerical simulations ensure that solutions develop symmetry breaking over

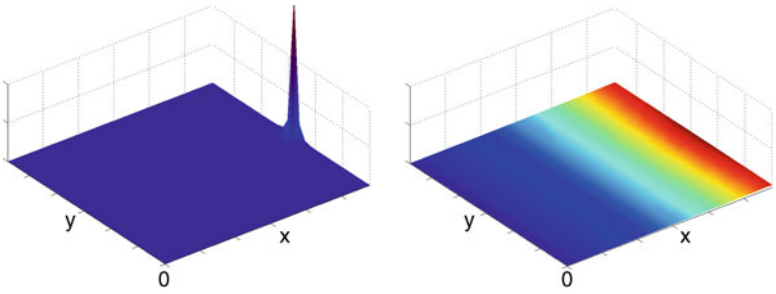

Fig. 6.2 Numerical simulations on $\Omega = [0,1] \times \mathbb{R}/2\pi\mathbb{Z}$ with $\Gamma = \{r\} \times \mathbb{R}/2\pi r\mathbb{Z}$ and all parameters equal to 1. *Left*: for $M = 20$ large enough, symmetry breaking appears. Molecular markers are concentrated on one point of the membrane in finite time. *Right*: for $M = 0.01$ small, steady state is homogeneous in the y-axis

a critical value M^* given us a first justification of the mathematical heuristic. In a further work, we will estimate an approximate value of this critical mass.

References

[Al07] Allaire, G.: Numerical Analysis and Optimization. An Introduction to Mathematical Modelling and Numerical Simulation. Oxford University Press, Oxford (2007)

[AlEtAl08] Altschuler, S., Angenent, S., Wang, Y., Wu, L.: On the spontaneous emergence of cell polarity. Nature **454**, 886–890 (2008)

[CaPeTa07] Calvez, V., Perthame, B., Tabar, M.S.: Modified Keller-Segel system and critical mass for the log interaction kernel. Stochastic analysis and partial differential equations, Contemp. Math., Amer. Math. Soc., Providence, RI, **429**, 45–62 (2007)

[CaMeVo10] Calvez, V., Meunier, N., Voituriez, R.: A one-dimensional Keller-Segel equation with a drift issued from the boundary. C. R. Acad. Sci. Paris, Ser. 1 **348**, 629–634 (2010)

[CaEtAl12] Calvez, V., Hawkins, R., Meunier, N., Voituriez, R.: Analysis of a nonlocal model for spontaneous cell polarization. SIAM J. Appl. Math. **72**, 594–622 (2012)

[EuEtAl07] Eugenio, M., Wedlich-Soldner, R., Li, R., Altschuler, S.J., Wu, L.F.: Principles for the dynamic maintenance of cortical polarity. Cell **129**, 411–422 (2007)

[Ev98] Evans, L.C.: Partial Differential Equations. American Mathematical Society, Providence (1998)

[HaEtAl09] Hawkins, R.J., Benichou, O., Piel, M., Voituriez, R.: Rebuilding cytoskeleton roads: active transport induced polarisation of cells. Phys. Rev. E **80**, 040903 (2009)

[IgDe08] Iglesias, P.A., Devreotes, P.N.: Navigating through models of chemotaxis. Cell Biol. **20**, 35 (2008)

[LeKeRa06] Levine, H., Kessler, D.A., Rappel, W.-J.: Directional sensing in eukaryotic chemotaxis: a balanced inactivation model. Proc. Natl. Acad. Sci. **103**, 9761 (2006)

[Mu13] Muller, N.: Mathematical and numerical studies of nonlinear and nonlocal models involved in biology, Doctoral dissertation, Paris Descartes (2013)

[OnRa07] Onsum, M., Rao, C.V.: A mathematical model for neutrophil gradient sensing and polarisation. PLoS Comput. Biol. **3**, e36 (2007)

[PhKoTh09] Phillips, R., Kondev, J.,Theriot, J.: Physical Biology of the Cell. Garland Science, New York, (2008)

[WeAlLi03] Wedlich-Soldner, L.W.R., Altschuler, S., Li, R.: Spontaneous cell polarization through actomyosin-based delivery of the cdc42 GTPase. Science **299**, 1231 (2003)

Chapter 7
Analytical Study of Computational Radiative Fluxes in a Heterogeneous Medium

D.Q. de Camargo, B.E.J. Bodmann, M.T. Vilhena, and C.F. Segatto

7.1 Introduction

During the last decades, an increasing number of authors developed a diversity of approaches concerning radiative transfer for a variety of applications. The common of these problems is the physical phenomenon of energy transfer by radiation and conduction in a medium. This phenomenon occurs in a lot of areas, including optics [LiEtAl06], astrophysics [PiEtAl09], atmospheric sciences [ThSt02], remote sensing [ShEtAl07], and engineering applications, such as the transport of heat by radiation [Br92], e.g., laser applications or radiative transfer in cooling processes [KyGu04] among others.

The propagation of radiation through a homogeneous or heterogeneous medium suffers changes by isotropic or non-isotropic processes like absorption, emission, and scattering that enter the mathematical approach in form of a nonlinear radiative transfer equation. The nonlinearity of the equation originates from a local thermal description using the Stefan–Boltzmann law, which relates heat transport to the radiation intensity and thus renders the radiative transfer a radiative-conductive problem, which we discuss in this work [Oz73], [Po05].

The radiative transfer equation is an integro-differential equation and its complexity derives from the fact that it is described in a phase space that consists of seven independent variables (three positions, two directions, a frequency and time). Several methods have been proposed to solve this time-dependent equation in a plane parallel geometry. In 1981, Levermore and Pomraning [LePo81] deduced the diffusion theory based on the equation of radiative transfer; in 1986 Ganapol [Ga86] obtained a numerical solution for the time-dependent transport equation

D.Q. de Camargo • B.E.J. Bodmann (✉) • M.T. Vilhena • C.F. Segatto
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: dayanadecamargo@gmail.com; bardo.bodmann@ufrgs.br; vilhena@mat.ufrgs.br; cynthia.segatto@ufrgs.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_7,

using an expansion by Legendre polynomials; Larsen and Pomraning [LaPo91] showed in 1991 that P_N equations are an asymptotic limit of the time-dependent transport equation.

In general, the equation of radiative-conductive transfer is difficult to solve without the introduction of some approximations, such as linearization or a reduction to a diffusion equation, which facilitates the construction of a solution to an approximate problem. The approach used in this study is not different in the sense that approximations shall be introduced; nevertheless, the nonlinearity that represents the crucial ingredient in the problem is solved without resorting to linearization or perturbation like procedures and to the best of our knowledge is the first analytical approach of this kind. The solution of the modified or approximate problem can be given in closed analytical form that permits to calculate numerical results in principle to any desired accuracy. Moreover, the influence of the nonlinearity can be analyzed in an analytical fashion directly from the formal solution.

Solutions found in the literature are typically linearized and of numerical nature (see, for instance, [As01], [AsEtAl02], [At00], [KrNaDu01], [MeVi83], [MuEtAl04], [SiTh91], [SpSi96], and references therein). To the best of our knowledge no analytical approach for a heterogeneous medium as well as considering the nonlinearity exists so far. A possible reason for considering a simplified problem (homogeneous and linearized) is that such a procedure makes feasible the determination of a convergent solution. It is worth mentioning that a general solution from an analytical approach for this type of problems exists only in the discrete ordinate approximation for a homogeneous medium as reported in [SeEtAl10].

Various applications allow to segment the medium in plane parallel sheets, where the radiation field is invariant under translation in directions parallel to that sheet. In other words the only spatial coordinate of interest is the one perpendicular to the sheet that indicates the penetration depth of the radiation in the medium. Frequently, it is justified to assume the medium to have an isotropic structure which reduces the angular degrees of freedom of the radiation intensity to the azimuthal angle θ or equivalently to its cosine μ. Further simplifications may be applied which are coherent with measurement procedures. On the one hand measurements are conducted in finite time intervals where the problem may be considered (quasi-)stationary, which implies that explicit time dependence may be neglected in the transfer equation. On the other hand, detectors have a finite dimension (extension) with a specific acceptance angle for measuring radiation and thus set some angular resolution for experimental data. Such an uncertainty justifies to segment the continuous angle into a set of discrete angles (or their cosines), which renders the original equation with angular degrees of freedom a set of equations known as the S_N approximation.

In the work [BoViSe11] an analytical approach was discussed and compared to the nonlinear S_N problem of radiative-conductive transfer in a heterogeneous medium of plane-parallel geometry using a composite method by Laplace transform and Adomian decomposition [Ad88], here called the $D_M LTS_N$ method. In the same reference it was shown that the heterogeneous problem can be expressed

in a set of homogeneous problems, so that the general solution can be obtained through a hierarchical algorithm. The Laplace technique opens way to use classical procedures for linear problems, while the decomposition method allows to separate the nonlinear contribution of the problem, which is then solved in a recursive fashion. In the afore cited work, the decomposition of a heterogeneous problem into a set of homogeneous problems was discussed on a theoretical basis only; however, a solution was not presented. In the present work this particular procedure is applied to a specific problem, where the partial solutions are matched at the interfaces between neighboring sheets due to different physical parameter of each sheet. Thus one may construct the solution for the heterogeneous problem of radiative-conductive transfer.

7.2 Radiative-Conductive Transfer

In problems of radiative transfer in plane parallel media it is convenient to measure linear distances normal to the plane of stratification using the concept of optical thickness τ which is measured from the boundary inward and is related to the density ρ, an attenuation coefficient κ and the geometrical projection on the direction perpendicular to that plane, say along the z-axis, so that $d\tau = -\kappa\rho dz$. Further, the temperature is measured in multiples of an arbitrary reference temperature $T(\tau) = \Theta(\tau)T_r$, typically taken at $\tau = 0$.

Based on the photon number balance and expressed as a Boltzmann-type equation one arrives at the radiative transfer equation in a volume that shall be chosen in a way so that no boundaries, that separate media with different physical properties, cross the control volume. To this end, five photon number changing contributions shall be taken into account which may be condensed into the four terms that follow. The first term describes the net rate of streaming of photons through the bounding surface of an infinitesimal control volume, the second term combines absorption and out-scattering from μ to all possible directions μ' in the control volume. The third term contemplates in-scattering from all directions μ' into the direction μ, and last not least a black-body like emission term according to the temperature dependence of Stefan–Boltzmann's law for the control volume.

$$\frac{dI(\tau,\mu)}{d\tau} + \frac{1}{\mu}I(\tau,\mu) = \frac{\omega(\tau)}{2\mu}\int_{-1}^{1}\mathscr{P}(\mu,\mu')I(\tau,\mu')d\mu' + \frac{1-\omega(\tau)}{\mu}\Theta^4(\tau). \quad (7.1)$$

Here, I is the radiation intensity, ω is the single scattering albedo, and $\mathscr{P}(\mu)$ signifies the differential scattering coefficient or also called the phase function, that accounts for the rate at which photons are scattered into an angle $d\mu'$ and with inclination μ with respect to the normal vector of the sheet. Note that the phase function is normalized; that is,

$$\frac{1}{2}\int \mathscr{P}(\mu)\,d\mu = 1.$$

Upon simplifying the phase function in plane geometry one may expand the angular dependence in Legendre polynomials $P_n(\mu)$,

$$\mathscr{P}(\mu,\mu') = \sum_{\ell=0}^{\infty} \beta_\ell P_\ell(\mu'-\mu),$$

with β_n the expansion coefficients that follow from orthogonality. Further one may employ the identity for Legendre polynomials using azimuthal symmetry (hence the zero integral)

$$P_\ell(\mu'-\mu) = P_\ell(\mu)P_\ell(\mu') + 2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}P_n^m(\mu)P_n^m(\mu') \times \underbrace{\int_0^{2\pi}\cos(m(\phi-\phi'))d\phi'}_{=0},$$

and write the integral on the right-hand side of (7.1) as

$$\int_{-1}^{1}\mathscr{P}(\mu,\mu')I(\tau,\mu')d\mu' = \sum_{\ell=0}^{\infty}\beta_\ell\int_{-1}^{1}P_\ell(\mu)P_\ell(\mu')I(\tau,\mu')d\mu',$$

where the summation index refers to the degree of anisotropy. For practical applications only a limited number of terms indexed with ℓ have to be taken into account in order to characterize qualitatively and quantitatively the anisotropic contributions to the problem. Also higher ℓ terms oscillate more significantly and thus suppress the integral's significance in the solution. The degree of anisotropy may be indicated truncating the sum by an upper limit L. The integro-differential equation (7.1) together with the aforementioned manipulations may be cast into an approximation known as the S_N equation upon reducing the continuous angle cosine to a discrete set of N angles. This procedure opens a pathway to apply standard vector algebra techniques to obtain a solution from the equation system.

In order to define boundary conditions we have to specify in more detail the scenario in consideration. Furthermore we analyze nonlinear radiative-conductive transfer in a gray plane-parallel participating medium with opaque walls, where specular (mirror like) as well as diffuse reflections occur besides thermal photon emission according to the Stefan–Boltzmann law (see [El09] and references therein). If one subdivides the medium into sheets of thickness $\Delta\tau$ with sufficiently small depth so that for each sheet a homogeneous medium applies, then for each face or interface the condition for the upper sheet boundary (at $\tau = \tau_i$) is

$$I(\tau,\mu)=\varepsilon(\tau)\Theta^4(\tau)+\rho^s(\tau)I(\tau,-\mu)+2\rho^d(\tau)\int_0^1 I(\tau,-\mu')\mu' d\mu', \tag{7.2}$$

with ρ^s and ρ^d the specular and diffuse reflections at the boundary, which are related to the emissivity ε by

$$\varepsilon+\rho^s+\rho^d=1.$$

For the lower boundary (at $\tau=\tau_i+\Delta\tau$) equation (7.2) applies but with reflected angles $\mu\to-\mu$ and $\mu'\to-\mu'$ in the argument of $I(\tau,\mu^{(\prime)})$. Suppose we have N_S sheets and N_S+1 boundaries, one might think that for a first order differential equation (7.1) in τ the supply of N_S+1 boundary conditions results in an ill-posed problem with no solutions at all. However, we still have to set up an equation system that uniquely defines the nonlinearity in terms of the radiation intensity.

The relation may be established in two steps, first recognizing that the dimensionless radiative flux is expressed in terms of the intensity by

$$q_r^*=2\pi\int_{-1}^1 I(\tau,\mu)\mu d\mu,$$

and the energy equation for the temperature that connects the radiative flux to a temperature gradient is

$$\frac{d^2}{d\tau^2}\Theta^4=\frac{1}{4\pi N_c}\frac{d}{d\tau}q_r^*(\tau)=\frac{1}{4\pi N_c}\frac{d}{d\tau}\left(2\pi\int_{-1}^1 I(\tau,\mu)\mu d\mu\right), \tag{7.3}$$

Here N_c is the conduction–radiation parameter, defined as

$$N_c=\frac{k\beta_{ext}}{4\sigma n^2 T_r^3},$$

with k the thermal conductivity, β_{ext} the extinction coefficient, σ the Stefan–Boltzmann constant, and $\mathfrak{n}$ the refractive index. Note that the radiative flux results from the integration of the intensity over angular variables, so that the thermal conductivity is considered here isotropic. Equation (7.3) is subject to prescribed temperatures at the top- and bottommost boundary, respectively:

$$\Theta(0)=\Theta_T \quad \text{and} \quad \Theta(\tau_0)=\Theta_B. \tag{7.4}$$

7.3 Solution by Decomposition Method

The set of equations (7.1) and (7.3), which are continuous in the angle cosine, may be simplified using an enumerable set of discrete angles following the collocation method that defines the radiative-convective transfer problem in the S_N approximation

$$\frac{dI_n(\tau)}{d\tau}+\frac{1}{\mu_n}I_n(\tau)=\frac{\omega(\tau)}{2\mu_n}\sum_{\ell=0}^{L}\beta_\ell P_\ell(\mu_n)\sum_{k=1}^{N}\omega_k P_\ell(\mu_k)I_k(\tau)+\frac{1-\omega(\tau)}{\mu_n}\Theta^4(\tau), \tag{7.5}$$

$$\frac{d\Theta(\tau)}{d\tau}-\frac{d\Theta(\tau)}{d\tau}\bigg|_{\tau=0}=\frac{1}{2N_c}\sum_{k=1}^{N}\omega_k(I_k(\tau)-I_k(0))\mu_k, \tag{7.6}$$

for $n=1,\ldots,N$, and are subject to the following boundary conditions:

$$I_n(0)=\varepsilon(0)\Theta^4(0)+\rho^s(0)I_{N-n+1}(0)+2\rho^d(0)\sum_{k=1}^{N/2}\omega_k I_{N-k+1}(0)\mu_k,$$

$$I_{N-n+1}(\tau_0)=\varepsilon(\tau_0)\Theta^4(\tau_0)+\rho^s(\tau_0)I_n(\tau_0)+2\rho^d(\tau_0)\sum_{k=1}^{N/2}\omega_k I_k(\tau_0)\mu_k.$$

Note that the integrals over the angular variables are replaced by a Gaussian quadrature scheme with weight factors w_k, where k refers to one of the discrete directions μ_k.

For convenience we introduce a shorthand notation in matrix operator form, where the column vector

$$\boldsymbol{\Phi}(\tau)=(\mathbf{I},\Theta(\tau))^T=(I_1(\tau),\ldots,I_N(\tau),\Theta(\tau))^T,$$

combines the anisotropic intensities and the isotropic temperature function, the nonlinear terms and boundary terms from integration (i.e., the temperature gradient and the conduction radiation intensity at $\tau=0$) are absorbed in an inhomogeneity:

$$\boldsymbol{\Psi}=\left(\frac{1-\omega(\tau)}{\mu_1}\Theta^4(\tau),\ldots,\frac{1-\omega(\tau)}{\mu_n}\Theta^4(\tau),\frac{d\Theta}{d\tau}(0)-\frac{1}{2N_c}\sum_{k=1}^{N}\omega_k I_k(0)\mu_k\right)^T.$$

This procedure allows to cast the equation system (7.5) and (7.6) in compact form of a first order matrix equation

$$\frac{d}{d\tau}\boldsymbol{\Phi}-\mathscr{L}_M\boldsymbol{\Phi}=\boldsymbol{\Psi}, \tag{7.7}$$

where $\mathscr{L}_M$ has elements

$$(\mathscr{L}_M)_{nk}=\delta_{nk}(1-\delta_{n,N+1})\frac{1}{\mu_n}+f_{nk}\qquad\text{for}\qquad n,k=1,\ldots,N+1.$$

Here, δ_{ij} is the Kronecker delta and θ_H the usual Heaviside function; that is,

$$\delta_{ij}=\begin{cases}1 & \text{if } i=j,\\ 0 & \text{otherwise,}\end{cases}\qquad \theta_H(x)=\begin{cases}1 & \text{if } x>0,\\ 0 & \text{otherwise,}\end{cases}$$

and the factors f_{nk} are

$$f_{nk} = \theta_H(N-n+1/2)\theta_H(N-k+1/2)\frac{\omega(\tau)}{2\mu_n}\sum_{\ell=0}^{L}\beta_\ell P_\ell(\mu_n)\omega_k P_\ell(\mu_k) + (1-\delta_{k,N+1})\delta_{n,N+1}\frac{\mu_k}{2N_c}.$$

Note that the increment $1/2$ in the Heaviside functional was introduced merely to make the argument positive definite in the range of interest which otherwise could lead to conflicts with possible definitions for $\theta_H(x)$ at $x = 0$.

The boundary conditions are combined accordingly, except for the limiting temperatures (7.4) that are kept separately for simplicity because they would add only an additional diagonal block leading to a reducible representation and does not bring any advantage from an algorithmic point of view.

$$\mathscr{B}_D\mathbf{I} - \mathscr{B}_M\mathbf{I} = \Gamma. \tag{7.8}$$

Equation (7.8) has a block form where one block represents forward angle contributions $\mu > 0$ and the other one backward terms $\mu < 0$ originating from the top and bottom boundary, respectively. Here, $\mathscr{B}_D$ is an $N \times N$ diagonal matrix, and

$$\mathscr{B}_M = \begin{pmatrix} 0 & \rho^s\mathscr{C}_{N/2} + 2\rho^d\mathscr{G}^-_{N/2} \\ \rho^s\mathscr{C}_{N/2} + 2\rho^d\mathscr{G}^+_{N/2} & 0 \end{pmatrix},$$

with $\mathscr{C}_{N/2}$ an $N/2 \times N/2$ matrix, which results from column reversion in the unit matrix, i.e. after mapping column position k to position $N/2-k+1$. The remaining matrices that control the diffuse forward and backward reflection ($\mathscr{G}^{\pm}_{N/2}$), respectively, have elements

$$(\mathscr{G}^+_{N/2})_{nk} = \theta_H(N/2-n+1/2)\theta_H(k-N/2-1/2)\mu_{N-k+1}\omega_{N-k+1}$$
$$(\mathscr{G}^-_{N/2})_{nk} = \theta_H(n-N/2-n-1/2)\theta_H(N/2-k+1/2)\mu_k\omega_k.$$

In these expressions the Heaviside functions restrict the nonzero elements to the off-diagonal blocks with row indices $n \in \{1,\ldots,N/2\}$ and column indices $k \in \{N/2+1,\ldots,N\}$ and with row indices $n \in \{N/2+1,\ldots,N\}$ and column indices $k \in \{1,\ldots,N/2\}$, respectively. The vector representation for the intensity is $\mathbf{I} = (\mathbf{I}_+,\mathbf{I}_-)^T$ with

$$\mathbf{I}_+ = (I_1(\tau),\ldots,I_{N/2}(\tau)) \quad \text{and} \quad \mathbf{I}_- = (I_{N/2+1}(\tau),\ldots,I_N(\tau)).$$

The inhomogeneity Γ has the same emission term in each component.

$$\Gamma_n = \varepsilon(\tau)\Theta^4(\tau) \quad \forall n.$$

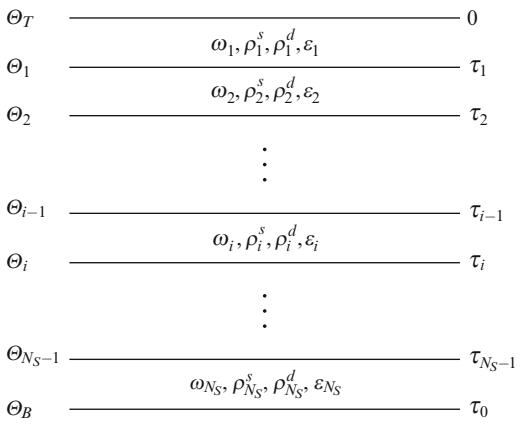

Fig. 7.1 Schematic illustration of a heterogeneous medium in form of a multi-layer slab

The principal difficulty in constructing a solution for the radiative-conductive transfer problem in the S_N approximation (7.7) subject to the boundary conditions (7.8) and (7.4) is due to the fact that the single scattering albedo $\omega(\tau)$, the emissivity $\varepsilon(\tau)$, and the specular and diffuse reflection ($\rho^s(\tau)$ and $\rho^d(\tau)$) have an explicit dependence on the optical depth τ, that is the heterogeneity of the medium in consideration. It is worth mentioning that the proposed methodology is quite general in the sense that it can be applied to other approximations of (7.1) that make use of spectral methods, as, for instance, the spherical harmonic P_N-, the Chebyshev Ch_N- and the Walsi W_N-approximation [ViSe99], [ViEtAl99], among others.

In the sequel we report on two approaches to solve the heterogeneous problem ((7.7), (7.8), and (7.4)). The principal idea of this techniques relies on the reduction of the radiative-conductive transfer problem in heterogeneous media to a set of problems in domains of homogeneous media. In the first approach we consider the standard approximation of the heterogeneous medium in form of a multi-layer slab (see Fig. 7.1). For each of the layers the problem reduces to a homogeneous problem but with the same number of boundary conditions as the original problem. In order to solve the unknown boundary values of the intensities and the temperatures at the interfaces between the slabs, matching these quantities using the bottom boundary values of the upper slab and the top boundary values of the lower slab eliminates these unknowns.

In the second approach we introduce a new procedure to work the heterogeneity. To begin with, we take the averaged value for the albedo coefficient $\omega(\tau)$,

$$\bar{\omega} = \frac{1}{\tau_0} \int_0^{\tau_0} \omega(\tau) d\tau,$$

and rewrite the problem as a homogeneous problem plus an inhomogeneous correction. Note that $\mathscr{L}_M$ as well as Ψ depend on the local albedo coefficient $\omega(\tau)$.

Since the terms containing the coefficient are linear in ω permits to separate an average factor $\bar{\omega}$ and the difference $\omega(\tau)-\bar{\omega}$:

$$\frac{d}{d\tau}\Phi-\mathscr{L}_M(\bar{\omega})\Phi=\Psi(\bar{\omega})+\mathscr{L}_M(\omega(\tau)-\bar{\omega})\Phi+\Psi(\omega(\tau)-\bar{\omega}). \tag{7.9}$$

Now, following the idea of the decomposition method proposed originally by Adomian [Ad88], to solve nonlinear problems without linearization, we handle equation (7.9), constructing the following recursive system of equations. Here, $\Psi-\sum_{m=0}^{\infty}\Psi_m$ is a formal decomposition and the nonlinearity is written in terms of the so-called Adomian polynomials $\Theta^4(\tau)=\sum_{m=0}^{\infty}\hat{A}_m(\tau)$. The first equation of the recursive system is the same as in a homogeneous slab, and the influence of the heterogeneity is governed by the source term. The homogeneous problem is explicitly solved in [BoViSe11], so that we concentrate here on the inhomogeneity:

$$\frac{d}{d\tau}\Phi_0-\mathscr{L}_M(\bar{\omega})\Phi_0=\Psi_0(\bar{\omega}),$$
$$\frac{d}{d\tau}\Phi_i-\mathscr{L}_M(\bar{\omega})\Phi_i=\Psi_i(\bar{\omega})+\mathscr{L}_M(\omega(\tau)-\bar{\omega})\Phi_{i-1}+\Psi_{i-1}(\omega(\tau)-\bar{\omega})$$

for $i\geq 1$, and

$$\Psi_{i-1}(\omega(\tau)-\bar{\omega})=(\bar{\omega}-\omega(\tau))A_m(\tau)(\mu_1^{-1},\ldots,\mu_1^{-N},0)^T. \tag{7.10}$$

Note that the $(N+1)$th component of $\Psi_0(\bar{\omega})$ contains the inhomogeneous term of the temperature equation:

$$(\Psi_0(\bar{\omega}))_{N+1}=\Psi_{N+1}=\frac{d\Theta}{d\tau}(0)-\frac{1}{2N_c}\sum_{k=1}^{N}\omega_k I_k(0)\mu_k.$$

To complete our analysis considering the boundary conditions, the first equation of the recursive system satisfies the boundary condition, whereas the remaining equations satisfy homogeneous boundary conditions. By this procedure we guarantee that the solution Φ determined from the recursive scheme and truncated at a convenient limit $\mathscr{M}$ satisfies the boundary conditions of the problem (7.8) and (7.4). Therefore, we are now in a position to construct a solution with a prescribed accuracy by controlling the number of terms in the series solution given by (7.10). From the previous discussion it becomes apparent that it is possible by the proposed procedure to obtain a solution of the heterogeneous problem by a reduction to a set of homogeneous problems. To complete the construction of a solution for the heterogeneous problem, in the next section we present the derivation of the solution of the S_N radiative-conductive transfer problem in a sheet-like homogeneous slab.

7.4 Problem Parameter and Numerical Results

To check if the proposed method is appropriate for radiative-conductive transfer problems in heterogeneous media, we evaluate the behavior of the temperature and radiative, conductive, and total heat fluxes:

$$Q_r(\tau) = \frac{1}{4\pi N_c} q_r^* \quad Q_c(\tau) = -\frac{d}{d\tau}\Theta(\tau) \quad \text{and} \quad Q(\tau) = Q_r(\tau) + Q_c(\tau).$$

The coefficient β_ℓ is defined considering a binomial scattering law which is given by

$$\beta_\ell = \left(\frac{2\ell+1}{2\ell-1}\right)\left(\frac{L+1-\ell}{L+1+\ell}\right)\beta_{\ell-1}, \quad 0 \le \ell \le L, \quad \text{and} \quad \beta_0 = 1.$$

The medium considered for this consistency test of the method is composed by two different materials, and each half is considered homogeneous, as illustrated in Fig. 7.2. The numerical values of the parameters of the problem may be found in Table 7.1. In Table 7.1, Θ_1 and Θ_3 are the prescribed temperatures at the top- and bottommost boundary of the medium (7.4).

Following the idea proposed, the heterogeneous domain was divided into homogeneous sheets, and for each of these sheets the solution was determined by the $D_M LTS_N$ method. Finally, applying the continuity and boundary conditions we couple the individual homogeneous problems and thus define the last term of (7.9). The temperature Θ_2 shown in Table 7.1 is a result of matching the solution at the interface.

The numerical results for the temperature profile (Θ), for radiative ($Q_r(\tau)$), convective ($Q_c(\tau)$), and total ($Q(\tau)$) heat fluxes along the optical thickness are shown in Figs. 7.3, 7.4, 7.5, and 7.6, respectively, for τ ranging from 0 to 1. In this

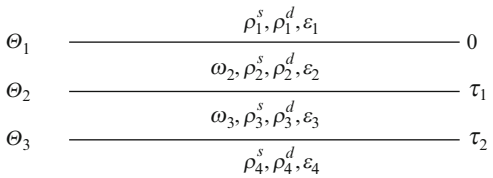

Fig. 7.2 Definition of a heterogeneous medium by sheet-wise homogeneous domains

Table 7.1 Parameter and properties of the medium, boundary and interfaces

ω_I	ω_{II}	N_{c_I}	$N_{c_{II}}$	L_I	L_{II}	Θ_1	Θ_2	Θ_3
0.95	0.95	0.05	0.04	0	0	1.5	1.21533	0.0

ε_1	ε_2	ε_3	ε_4	ρ_1^s	ρ_2^s	ρ_3^s	ρ_4^s	ρ_1^d	ρ_2^d	ρ_3^d	ρ_4^d
0.4	0.6	0.3	0.5	0.2	0.1	0.3	0.2	0.4	0.3	0.4	0.3

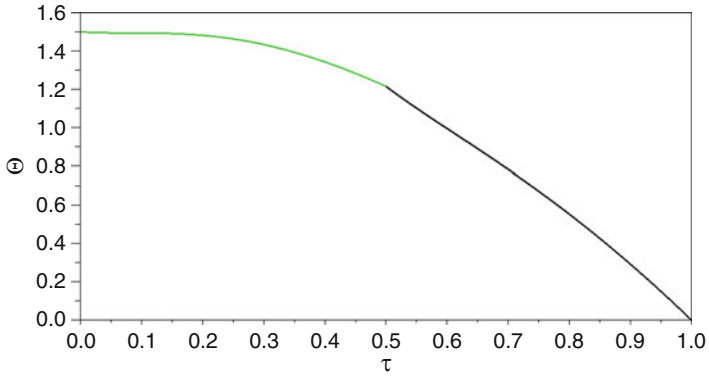

Fig. 7.3 Temperature profile along the optical thickness

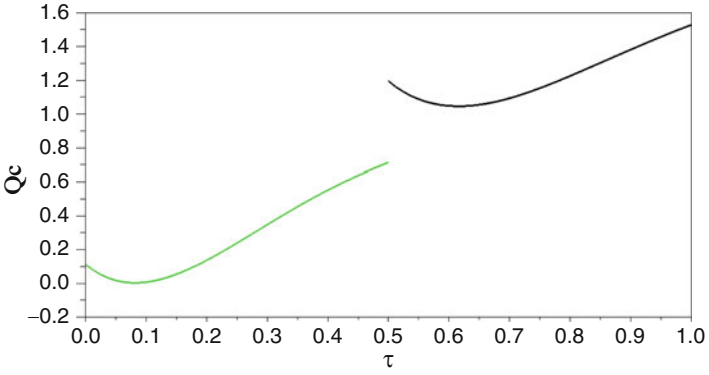

Fig. 7.4 Conductive heat flux along the optical thickness

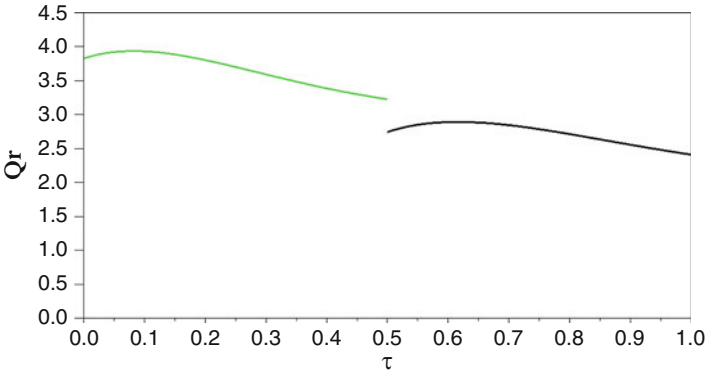

Fig. 7.5 Radiative heat flux along the optical thickness

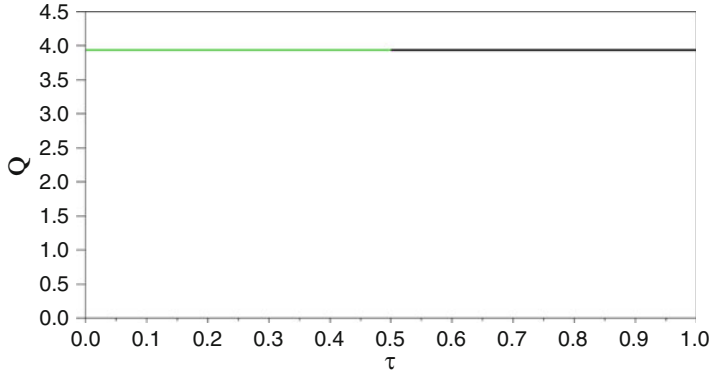

Fig. 7.6 Total heat flux along the optical thickness

analysis we used $N = 200$ as quadrature order and the truncation in the Adomian polynomial series was $M = 10$.

Since the prescribed temperatures Θ_1 and Θ_3, on the top and bottom boundary of the medium are constant, one expects a constant heat flux through the medium as a manifestation of a steady state, shown in Fig. 7.6.

7.5 Conclusions

The proposed methodology, which is composed by the Laplace transform and the Adomian decomposition [Ad88], reproduces the exact analytical solution for the approximate S_N problem in the limit $M \to \infty$. The present study demonstrated that the idea of reducing the radiative-conductive heat transfer problem in heterogeneous medium into a set of problems with homogeneous media is valid, so that the general solution can be obtained through a hierarchical algorithm. The Laplace technique opens the path to make use of well-established classical methods for linear problems, while the decomposition method allows to separate the nonlinear contribution of the problem, and thus allows to solve the equations in a closed form and recursive fashion.

The solution of the equation system involves computational operations among scalars, vectors, and matrices. There exist several programming libraries that implement the necessary functionality for manipulating this set, however with more or less reliability. More specifically, they usually work with simple problems but fail to yield numerically trustworthy results for more realistic problems. One objective of this work was to develop a program of the proposed problem in the programming language C++, which works for a wider range of parameter sets and allowing for scalability and optimization of the solution process. A variety of experiments with several existing libraries showed that all of them lacked satisfactory results in some of the parameter combinations. The first library we tested was the CLAPACK (Linear Algebra PACKage for C) [La12], this library does not provide correct

results, so that we resorted to other libraries some of them based on the codes available in the book *Numerical Recipes in C* [PrEtAl02]. Again we verified that some of the algorithms do not provide valid results for large order matrices. Among the public domain libraries the GNU Scientific Library (GSL) [Gs12] showed the best results for this type of problem. Currently we are working on a new open access library that is suitable for huge matrix systems, where tests so far are promising.

References

[Ad88] Adomian, G.: A review of the decomposition method in applied mathematics. J. Math. Anal. Appl. **135**, 501–544 (1988)

[As01] Asllanaj, F., Jeandel, G., Roche, J.R.: Numerical solution of radiative transfer equation coupled with non-linear heat conduction equation. Int. J. Numer. Meth. Heat Fluid Flow **11**, 449–473 (2001)

[AsEtAl02] Asllanaj, F., Milandri, A., Jeandel, G., Roche, J.R.: A finite difference solution of non-linear systems of radiative-conductive heat transfer equations. Int. J. Numer. Meth. Eng. **54**, 1649–1668 (2002)

[At00] Attia, M.T.: On the exact solution of a generalized equation of radiative transfer in a two-region inhomogeneous slab. J. Quant. Spectros. Radiative Tran. **66**, 529–538 (2000)

[Br92] Brewster, M.Q.: Thermal Radiative Transfer and Properties. Wiley, New York (1992)

[El09] Elghazaly, A.: Conductive-radiative heat transfer in a scattering medium with angle-dependent reflective boundaries. J. Nucl. Radiation Phys. **4**, 31–41 (2009)

[Ga86] Ganapol, B.D.: Solution of the one-group time-dependent neutron tranport equation in an infinite medium by polynomial reconstruction. Nucl. Sci. Eng. **92**, 272–279 (1986)

[Gs12] GSL—GNU Scientific Library, http://www.gnu.org/software/gsl (2012)

[KrNaDu01] Krishnapraka, C.K., Narayana, K.B., Dutta, P.: Combined conduction and radiation heat transfer in a gray anisotropically scattering medium with diffuse-specular boundaries. Int. Comm. Heat Mass Tran. **28**, 77–86 (2001)

[KyGu04] Kyunghan Kim, K., Guo, Z.: Ultrafast radiation heat transfer in laser tissue welding and soldering. Numer. Heat Tran. A **46**, 23–40 (2004)

[La12] LAPACK—Linear Algebra PACKage, http://www.netlib.org/lapack (2012)

[LaPo91] Larsen, E.W., Pomraning, G.C.: The P_N theory as an asymptotic limit of transport theory in planar geometry. Nucl. Sci. Eng. **109**, 49–75 (1991)

[LePo81] Levermore, C.D., Pomraning, G.C.: A flux-limited diffusion theory. Astrophys. J. **248**, 321–334 (1981)

[LiEtAl06] Liu, X., Smith, W.L., Zhou, D.K., Larar, A.: Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. Appl. Optic. **45**, 201–209 (2006)

[MeVi83] Mengüç, M.P., Viskanta, R.: Comparison of radiative transfer approximations for a highly forward scattering planar medium. J. Quant. Spectros. Radiative Tran. **29**, 381–394 (1983)

[MuEtAl04] Muresan, C., Vaillon, R., Menezo, Ch., Morlot, R.: Discrete ordinates solution of coupled conductive radiative heat transfer in a two-layer slab with Fresnel interfaces subject to diffuse and obliquely collimated irradiation. J. Quant. Spectros. Radiative Tran. **84**, 551–562 (2004)

[Oz73] Ozisik, M.N.: Radiative Transfer and Interaction with Conduction and Convection. Wiley, New York (1973)

[PiEtAl09] Pinte, C., Harries, T.J., Min, M., Watson, A.M., Dullemond, C.P., Woitke, P., Ménard, F., Durán-Rojas, M.C.: Benchmark problems for continuum radiative transfer. High optical depths, anisotropic scattering, and polarisation. Astron. Astrophys. **498**, 967–980 (2009)

[Po05] Pomraning, G.C.: The Equations of Radiation Hydrodynamics. Dover, Mineola (2005)

[PrEtAl02] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipies in C. Cambridge University Press, New York (2002)

[SeEtAl10] Segatto, C.F., Vargas, R.F., Vilhena, M.T., Bodmann, B.E.J.: A solution for the non-linear S_N radiative conductive problem in a grey plane-parallel participating medium. Int. J. Therm. Sci. **49**, 1493–1499 (2010)

[ShEtAl07] Shabanov, N.V., Huang, D., Knjazikhina, Y., Dickinson, R.E., Mynenia, R.B.: Stochastic radiative transfer model for mixture of discontinuous vegetation canopies. J. Quant. Spectros. Radiative Tran. **107**, 236–262 (2007)

[SiTh91] Siewert, C.E., Thomas, J.R.: A computational method for solving a class of coupled conductive-radiative heat transfer problems. J. Quant. Spectros. Radiative Tran. **45**, 273–281 (1991)

[SpSi96] Spuckler, C.M., Siegel, R.: Two-flux and diffusion methods for radiative transfer in composite layers. J. Heat Tran. **118**, 218–222 (1996)

[ThSt02] Thomas, G.E., Stamnes, K.: Radiative Transfer in the Atmosphere and Ocean. Cambridge University Press, Cambridge (2002)

[ViSe99] Vilhena, M.T., Segatto, C.F.: The state of art of the LTS_n method. In: Aregones, M., Ahnert, C., Cabellos, D. (eds.) Mathematica and Computational, Reactor Physics and Environmental Analysis in Nuclear Applications. M & C '1999 – Madrid: Mathematics and Computation, Reactor Physics and Environmental Analysis in Nuclear Applications: International Conference, Madrid, Spain, Senda Editorial, Madrid, **2**, pp. 1618–1631, (1999). ISBN, 8469909436, 9788469909430

[ViEtAl99] Vilhena, M.T., Barichelo, L.B., Zabadal, J.R., Segatto, C.F., Cardona, A.V., Pazos, R.P.: Solution to the multidimensional linear transport equation by the spectral method. Progr. Nucl. Energ. **35**, 275–291 (1999)

[BoViSe11] de Vilhena, M.T.M.B., Bodmann, B.E.J., Segatto, C.F.: Non-linear radiative-conductive heat transfer in a heterogeneous gray plane-parallel participating medium. In: Ahsan, A. (ed.) Convection and Conduction Heat Transfer. InTech, New York (2011) ISBN: 978-953-307-582-2. DOI: 10.5772/22736

Chapter 8
A Novel Approach to the Hankel Transform Inversion of the Neutron Diffusion Problem Using the Parseval Identity

J.C.L. Fernandes, M.T. Vilhena, and B.E.J. Bodmann

8.1 Introduction

The neutron diffusion equation is still one of the most frequently employed equations for nuclear reactor neutronics calculations, although its limitations are well known [GoLeVi09, ViSeGo04]. The equation is obtained under the assumptions that scattering is isotropic in the laboratory coordinate system and the region of interest is considered piecewise homogeneous, so that the diffusion coefficients are invariant under spatial transforms like translation and others. It is well known that such a derivation of diffusion theory rests on certain assumptions, i.e. the flux being sufficiently smooth especially by virtue of neutron absorption or production, which is reasonable since the mean free path is typically larger than the dimensions of the fuel cell and moderator space geometry. The solution of the diffusion equation system is thus an average description of a large number of neutrons, where fluctuations (higher moments) are neglected [LeEtAl08]. Further, the continuous energy distribution of neutrons is reduced by the use of energy groups (in the present case two).

8.2 Multi-group Steady State Neutron Diffusion

Our starting point is the steady state multi-energy group neutron diffusion equation, with the usual diffusion, removal, out-scattering, fission, and in-scattering terms. Here D_g is the diffusion coefficient for energy group g, $\Delta_r = r^{-1}\partial_r(r\partial_r)$ represents the radial part of the Laplace operator in cylindrical coordinates. Note that we assume translational symmetry of the neutron flux ϕ_g along the cylinder axis and

J.C.L. Fernandes • M.T. Vilhena • B.E.J. Bodmann (✉)
Federal University of Rio Grande do Sul, RS, Porto Alegre, Brazil
e-mail: julio.lombaldo@ufrgs.br; vilhena@mat.ufrgs.br; bardo.bodmann@ufrgs.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_8,

thus $\partial_{zz}\phi = 0$. Σ_{Rg} are the respective removal cross section, $\Sigma_{g\to g'}, \Sigma_{g'\to g}$ $(g \neq g')$ the out- and in-scattering cross sections, $\nu_g\Sigma_{fg}$ the fission cross section times the average neutron yield per fission, χ_g the spectral weight of energy group $g \in [1, G]$, k_{eff} the effective multiplication factor, and S_g a generic source term per energy group:

$$-D_g\Delta_r\phi_g + \left(\Sigma_{Rg} + \sum_{g'=1}^{G}\Sigma^s_{g\to g'}\right)\phi_g = \chi_g\sum_{g'=1}^{G}\nu_{g'}\Sigma_{fg'}\phi_{g'} + \sum_{g'=1}^{G}\Sigma_{g'\to g}\phi_{g'} + S_g. \tag{8.1}$$

The diffusion problem is subject to the boundary conditions of zero current density at the center of the cylinder $D_g(\partial\phi_g/\partial r)(0) = 0$ and zero flux at the boundary; that is, $\phi_g(R) = 0$.

8.3 The Hankel-Transformed Problem

The diffusion problem (8.1) previously introduced may be solved by the use of the zero order Hankel transform

$$\bar{f}(\xi) = H_n[f(r); r\to\xi] = \int_0^\infty rf(r)J_n(r\xi)\,dr$$

(here $n = 0$) that renders (8.1) a nonhomogeneous problem and may be cast into matrix form. As an example we show the equation for two energy groups:

$$\begin{pmatrix} D_1\xi^2 + \Sigma_{R1} & -(\chi_1\nu\Sigma_{f2} + \Sigma_{12}) \\ -(\chi_2\nu\Sigma_{f1} + \Sigma_{21}) & D_2\xi^2 + \Sigma_{R2} \end{pmatrix}\begin{pmatrix}\bar{\phi}_1 \\ \bar{\phi}_2\end{pmatrix} = \begin{pmatrix}\bar{S}_1 \\ \bar{S}_2\end{pmatrix}.$$

In shorthand notation, the equation reads $\mathbf{M}(\xi)\bar{\boldsymbol{\Phi}} = \bar{\mathbf{S}}$. In general $\mathbf{M}(\xi)$ is invertible, so that

$$\det(\mathbf{M}(\xi)) = A(\xi)B(\xi) - \mu_{12}\mu_{21} \neq 0,$$

with $A(\xi) = D_1\xi^2 + \Sigma_{R1}$, $B(\xi) = D_2\xi^2 + \Sigma_{R2}$, $\mu_{12} = \chi_1\nu\Sigma_{f2} + \Sigma_{12}$ and $\mu_{21} = \chi_2\nu\Sigma_{f1} + \Sigma_{21}$. The solution for the system in transformed variables is

$$\bar{\boldsymbol{\Phi}} = (\det(\mathbf{M}(\xi)))^{-1}\begin{pmatrix} B(\xi)\bar{S}_1 + \mu_{12}\bar{S}_2 \\ \mu_{21}\bar{S}_1 + A(\xi)\bar{S}_2 \end{pmatrix}.$$

In what follows, we introduce a simplification, that does not compromise the generality of the procedure, and consider a source term for group $g = 1$, only. Then

$$\bar{\phi}_1 = B(\xi)\frac{\bar{S}_1}{\det(M(\xi))}, \quad \bar{\phi}_2 = \mu_{21}\frac{\bar{S}_1}{\det(M(\xi))},$$

and upon applying the inverse Hankel transformation one may determine the analytical solution of the problem [Fe11].

8.3.1 Fast Flux Solution

Application of the inversion formula yields

$$\phi_1 = \int_0^\infty \xi \frac{B(\xi)J_0(r\xi)}{\det(M(\xi))}\bar{S}_1\, d\xi\ ,$$

which together with the Hankel inversion theorem and Parseval's identity allows us to derive the desired result.

Theorem 1 (The Hankel inversion theorem). *If $\sqrt{r'}f(r')$ is piecewise continuous and absolutely integrable on the positive half of the real line, and if $\gamma \geq -\frac{1}{2}$, then $\bar{f}_\gamma(\xi) = H_\gamma[f(r'); r\prime \to \xi]$ exists and*

$$\int_0^\infty \xi \bar{f}_\gamma(\xi)J_\gamma(\xi r')\, d\xi = \frac{1}{2}[f(r'_+) + f(r'_-)].$$

Theorem 2 (Parsenval's relation). *If the functions $f(r')$ and $g(r')$ satisfy the conditions of Theorem 1 and if $\bar{f}_\gamma(\xi)$ and $\bar{g}_\gamma(\xi)$ denote the Hankel transforms of order $\gamma \geq -\frac{1}{2}$, then*

$$\int_0^\infty r' f(r')g(r')\, dr' = \int_0^\infty \xi \bar{f}_\gamma(\xi)\bar{g}_\gamma(\xi)\, d\xi.$$

Making use of the theorem with $\bar{f}_0(\xi) = \frac{B(\xi)J_0(r\xi)}{\det(M(\xi))}$ and $\bar{g}_0(\xi) = \bar{S}_1$, establishes that

$$\int_0^\infty \xi \frac{B(\xi)J_0(r\xi)}{\det(M(\xi))}\bar{S}_1\, d\xi = \int_0^\infty r' H_0^{-1}\left\{\frac{B(\xi)J_0(r\xi)}{\det(M(\xi))}\right\} S_1(r')\, dr'\ ,$$

and by definition the following identity holds:

$$H_0^{-1}\left\{\frac{B(\xi)J_0(r\xi)}{\det(M(\xi))}\right\} = \int_0^\infty \xi \frac{B(\xi)J_0(r\xi)}{\det(M(\xi))}J_0(r'\xi)\, d\xi.$$

The physically meaningful range of nuclear parameters implies $0 < \frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)} < 1$, so that

$$\frac{B(\xi)}{A(\xi)B(\xi)-\mu_{12}\mu_{21}} = \frac{1}{A(\xi)} + \frac{1}{A(\xi)} O\left(\left(\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)}\right)^2\right),$$

which by virtue of the fact that $\left(\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)}\right) \ll 1$ allows to safely neglect higher-order terms. The integral may be evaluated [Ba54] as

$$\int_0^\infty \xi \frac{J_0(r\xi)}{A(\xi)} J_0(r'\xi)\, d\xi = \begin{cases} \frac{1}{D_1} I_0(\sqrt{\alpha_1} r') K_0(\sqrt{\alpha_1} r) & \text{for} \quad 0 < r' < r \\ \frac{1}{D_1} I_0(\sqrt{\alpha_1} r) K_0(\sqrt{\alpha_1} r') & \text{for} \quad r < r' < \infty, \end{cases}$$

where $\alpha_g = \Sigma_{Rg}/D_g$. Here, I_0 and K_0 are the modified Bessel functions and outside of the cylinder the source term is identically zero. The solution for the fast flux is then

$$\phi_1 = \frac{K_0(\sqrt{\alpha_1} r)}{D_1} \int_0^r r' I_0(\sqrt{\alpha_1} r') S_1(r')\, dr' + \frac{I_0(\sqrt{\alpha_1} r)}{D_1} \int_r^R r' K_0(\sqrt{\alpha_1} r') S_1(r')\, dr'.$$

8.3.2 The Thermal Flux Solution

The procedure for the thermal flux follows similar steps to the ones introduced in the solution scheme for the fast flux. Using the inversion formula

$$\phi_2 = \mu_{21} \int_0^\infty \xi \frac{J_0(r\xi)}{\det(M(\xi))} \bar{S}_1\, d\xi$$

together with Theorem 2,

$$\int_0^\infty \xi \frac{J_0(r\xi)}{\det(M(\xi))} \bar{S}_1\, d\xi = \int_0^\infty r' H_0^{-1} \left\{ \frac{J_0(r\xi)}{\det(M(\xi))} \right\} S_1(r')\, dr'$$

and, by definition,

$$H_0^{-1} \left\{ \frac{J_0(r\xi)}{\det(M(\xi))} \right\} = \int_0^\infty \xi \frac{J_0(r\xi)}{\det(M(\xi))} J_0(r'\xi)\, d\xi.$$

Using arguments analogous to those for the fast flux, we arrive at

$$H_0^{-1} \left\{ \frac{J_0(r\xi)}{\det(M(\xi))} \right\} = \frac{1}{(\Sigma_{R2} D_1 - \Sigma_{R1} D_2)} \int_0^\infty \xi \frac{J_0(r\xi)}{\xi^2 + (\sqrt{\alpha_1})^2} J_0(r'\xi)\, d\xi$$

$$
\begin{aligned}
&- \frac{1}{(\Sigma_{R2}D_1 - \Sigma_{R1}D_2)} \int_0^\infty \xi \frac{J_0(r\xi)}{\xi^2 + (\sqrt{\alpha_2})^2} J_0(r'\xi)\, d\xi \\
&= \begin{cases} \frac{I_0(\sqrt{\alpha_1}r')K_0(\sqrt{\alpha_1}r) - I_0(\sqrt{\alpha_2}r')K_0(\sqrt{\alpha_2}r)}{(\Sigma_{R2}D_1 - \Sigma_{R1}D_2)} & \text{for} \quad 0 < r' < r, \\ \frac{I_0(\sqrt{\alpha_1}r)K_0(\sqrt{\alpha_1}r') - I_0(\sqrt{\alpha_2}r)K_0(\sqrt{\alpha_2}r')}{(\Sigma_{R2}D_1 - \Sigma_{R1}D_2)} & \text{for} \quad r < r' < \infty, \end{cases}
\end{aligned}
$$

so that the thermal flux is

$$
\begin{aligned}
\phi_2 = c_1 \Bigg(& K_0(\sqrt{\alpha_1}r) \int_0^r r' I_0(\sqrt{\alpha_1}r') S_1(r')\, dr' \\
& + I_0(\sqrt{\alpha_1}r) \int_r^R r' K_0(\sqrt{\alpha_1}r') S_1(r')\, dr' \\
& - K_0(\sqrt{\alpha_2}r) \int_0^r r' I_0(\sqrt{\alpha_2}r') S_1(r')\, dr' \\
& - I_0(\sqrt{\alpha_2}r) \int_r^R r' K_0(\sqrt{\alpha_2}r') S_1(r')\, dr' \Bigg),
\end{aligned}
$$

where $c_1 = \frac{\mu_{21}}{(\Sigma_{R2}D_1 - \Sigma_{R1}D_2)}$.

Because of the similarity of the solutions the integral expressions may be used to formulate both solutions as

$$
\phi_1 = T_1[S_1](r) \quad \text{and} \quad \phi_2 = c_1(D_1 T_1[S_1](r) - D_2 T_2[S_1](r)),
$$

where

$$
\begin{aligned}
T_g[f](r) = \frac{K_0(\sqrt{\alpha_g}r)}{D_g} & \int_0^r r' I_0(\sqrt{\alpha_g}r') f(r')\, dr' \\
& + \frac{I_0(\sqrt{\alpha_g}r)}{D_g} \int_r^R r' K_0(\sqrt{\alpha_g}r') f(r')\, dr'.
\end{aligned}
$$

8.4 Multi-regions

In this section we present the first approximation for a solution in a piecewise homogeneous medium, where each region (with index κ) has its specific and in general distinct parameter set [BoEtAl10]. In order to simplify the problem, we ignore the energy group mixing terms (coupling between different energy groups) and consider as an approximation the diffusion equation for each group separately. A more general solution for a coupled system is beyond the scope of the present work but will be the issue in a future discussion:

$$-D_g^{(\kappa)}\Delta_r\phi_g^{(\kappa)}(r)+\sigma^{(\kappa)}\phi_g^{(\kappa)}=0,\quad \text{with}\quad \sigma_g^{(\kappa)}=\Sigma_{Rg}^{(\kappa)}-\nu\Sigma_{fg}^{(\kappa)}.$$

Basically two approaches may be used to solve the multi-region problem, the usual one determines the solution for each region separately and the integration constants are determined from the matching of the fluxes and current densities at the boundaries and interfaces, respectively [BoEtAl10]. In the further we follow a different reasoning, here the solution of the first region is extended to the whole domain of interest across all N regions with increasing boundaries at $R_1,\ldots,R_N$ and the modification of the solution for the change in the parameter set of the second region is determined by a correction to the already obtained solution. All corrections for the parameter changes of the successive regions are treated this way, so that the general solution gets a progressive character. If the solution for region κ is given by $\phi_g^{(\kappa)}$, then

$$\phi_g^{(\kappa)}=\sum_{i=1}^{\kappa}\phi_{gi}=\phi_{g\kappa}+\phi_g^{(\kappa-1)},$$

where $\kappa\in[1,\ldots,N.]$

The progressive solution is then determined by a recursive scheme with a finite recursion depth. The initialization is given by

$$-\Delta_r\phi_{g1}+\frac{\sigma_g^{(1)}}{D_g^{(1)}}\phi_{g1}=0\,,$$

and the generic recursion steps are

$$-D_g^{(\kappa)}\Delta_r\phi_{g\kappa}+\sigma_g^{(\kappa)}\phi_{g\kappa}=\underbrace{\left(\frac{D_g^{(\kappa)}}{D_g^{(\kappa-1)}}\sigma_g^{(\kappa-1)}-\sigma_g^{(\kappa)}\right)}_{\gamma_g^{(\kappa)}}\phi_g^{(\kappa-1)}\,. \tag{8.2}$$

Thus, once the solution for the preceding region is known it enters as a source term in the subsequent equation, which may be solved. The solution for the first region is the solution for a homogeneous problem:

$$\phi_g^{(1)}(r)=A_1J_0(\lambda_1 r)+B_1Y_0(\lambda_1 r). \tag{8.3}$$

Here A_i and B_i are constants, J_0, Y_0 are the Bessel and Neumann functions and $\lambda_i=(\sigma_g^{(i)})^{1/2}(D_g^{(i)})^{-1/2}$, in our case $B_1=0$ in order to render the solution regular at the origin. The solution for the recursion steps is composed of the aforementioned homogeneous solution (8.3) plus a particular solution that we will determine in the following. To this end, the Hankel transform is applied to (8.2), yielding

$$D_g^{(\kappa)}\xi^2\bar{\phi}_{g\kappa}+\sigma_g^{(\kappa)}\bar{\phi}_{g\kappa}=\gamma_g^{(\kappa)}\bar{\phi}_g^{(\kappa-1)}.$$

The solutions of the transformed problem are then

$$\bar{\phi}_{g\kappa}=\left(\frac{\gamma_g^{(\kappa)}}{D_g^{(\kappa)}\xi^2+\sigma_g^{(\kappa)}}\right)\bar{\phi}_g^{(\kappa-1)}\,.$$

From the inversion formula of the Hankel transform we get

$$H_0^{-1}\{\bar{\phi}_{g\kappa}\}=\phi_{g\kappa}=\int_0^\infty \xi\,\frac{\gamma_g^{(\kappa)}}{D_g^{(\kappa)}\xi^2+\sigma^{(\kappa)}}\bar{\phi}_g^{(\kappa-1)}J_0(r\xi)\,d\xi\,.$$

As already practised in the previous sections the inversion is done using Theorems 1 and 2, with $\bar{f}_0(\xi)=\frac{J_0(r\xi)}{D_g^{(\kappa)}\xi^2+\sigma_g^{(\kappa)}}$ and $\bar{g}_0(\xi)=\bar{\phi}_g^{(\kappa-1)}$, respectively:

$$\begin{aligned}\phi_{g\kappa}&=\gamma_g^{(\kappa)}\int_0^\infty \xi\left(\frac{J_0(r\xi)}{D_g^{(\kappa)}\xi^2+\sigma^{(\kappa)}}\right)\bar{\phi}_g^{(\kappa-1)}\,d\xi\\&=\gamma_g^{(\kappa)}\int_0^\infty r' H_0^{-1}\left\{\frac{J_0(r\xi)}{D_g^{(\kappa)}\xi^2+\sigma^{(\kappa)}}\right\}\phi_g^{(\kappa-1)}(r')\,dr'.\end{aligned}$$

Further, the integral may be solved analytically [Ba54] as

$$\begin{aligned}H_0^{-1}\left\{\frac{J_0(r\xi)}{D_g^{(\kappa)}\xi^2+\sigma_g^{(\kappa)}}\right\}&=\frac{1}{D_g^{(\kappa)}}\int_0^\infty \xi\,\frac{J_0(r\xi)}{\xi^2+(\sqrt{\alpha_\kappa})^2}J_0(r'\xi)\,d\xi,\\&=\begin{cases}\frac{1}{D^{(\kappa)}}I_0(\sqrt{\alpha_\kappa}r')K_0(\sqrt{\alpha_\kappa}r) & \text{for}\quad 0<r'<r,\\ \frac{1}{D^{(\kappa)}}I_0(\sqrt{\alpha_\kappa}r)K_0(\sqrt{\alpha_\kappa}r') & \text{for}\quad r<r'<R,\end{cases}\end{aligned}$$

with $\alpha_\kappa=\sigma_g^{(\kappa)}/D_g^{(\kappa)}$. The particular solution may be combined with the homogeneous solution in order to compose the general solution by the components $\phi_{g\kappa}$,

$$\begin{aligned}\phi_{g\kappa}=\frac{\gamma_g^{\kappa}}{D_g^{(\kappa)}}K_0(\sqrt{\alpha_\kappa}r)\int_0^r r' I_0(\sqrt{\alpha_\kappa}r')\phi_g^{(\kappa-1)}(r')\,dr'\\+\frac{\gamma_g^{\kappa}}{D_g^{(\kappa)}}I_0(\sqrt{\alpha_\kappa}r)\int_r^R r' K_0(\sqrt{\alpha_\kappa}r')\phi_g^{(\kappa-1)}(r')\,dr'\\+A_\kappa J_0(\lambda_\kappa r)+B_\kappa Y_0(\lambda_\kappa r).\end{aligned}$$

8.5 Error Estimates

The error of the solution comes merely from the expansion of the integrand

$$\frac{B(\xi)}{A(\xi)B(\xi)-\mu_{12}\mu_{21}} = \frac{1}{A(\xi)}\frac{1}{1-\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)}}.$$

For any choice of meaningful nuclear parameter the aforementioned relation

$$\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)} < 1$$

holds and the integral may be approximated by the leading order term of the integrand's expansion:

$$\begin{aligned} T &= \int_0^\infty \xi \frac{B(\xi)}{A(\xi)B(\xi)-\mu_{12}\mu_{21}} J_0(\xi r)J_0(\xi r')\,d\xi \\ &= \int_0^\infty \xi \frac{1}{A(\xi)} J_0(\xi r)J_0(\xi r')\,d\xi \\ &\quad + \int_0^\infty \xi \frac{\mu_{12}\mu_{21}}{A^2(\xi)B(\xi)} J_0(\xi r)J_0(\xi r')\,d\xi \\ &\quad + \int_0^\infty \xi \frac{(\mu_{12}\mu_{21})^2}{A^3(\xi)B^2(\xi)} J_0(\xi r)J_0(\xi r')\,d\xi + \cdots. \end{aligned}$$

The error of the integral is then given by

$$\delta T = \sum_{n=1}^{\infty} \left\{ \int_0^\infty \xi \frac{1}{A(\xi)} \left(\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)} \right)^n J_0(\xi r)J_0(\xi r')\,d\xi \right\}.$$

The final expression for flux is

$$\Phi = \int_0^\infty r' T S(r')\,dr'$$

and, consequently, the expression for the error is

$$\begin{aligned} \delta\Phi &= \int_0^\infty r' \delta T S(r')\,dr' \\ &= S_0 \int_0^R r' \left[\sum_{n=1}^{\infty} \left\{ \int_0^\infty \xi \frac{1}{A(\xi)} \left(\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)} \right)^n J_0(\xi r)J_0(\xi r')\,d\xi \right\} \right] S(r')\,dr', \end{aligned}$$

where $\frac{1}{A(\xi)}\left(\frac{\mu_{12}\mu_{21}}{A(\xi)B(\xi)}\right)^n$ is a strong monotone decreasing sequence. The dominating term of the error $E_\phi(r)$ is

$$\begin{aligned}\delta\phi^1(r) &= S_0\int_0^R r'\delta T^{(1)}(r')\,dr' \\ &= S_0\mu_{12}\mu_{21}\int_0^R r'\int_0^\infty \xi\frac{1}{A^2(\xi)B(\xi)}J_0(\xi r)J_0(\xi r')\,d\xi dr'.\end{aligned}$$

One may introduce an estimate for the explicit expression A^2B, namely

$$A^2(\xi)B(\xi) = D_1^2D_2\xi^6+\ldots+\Sigma_{R1}^2\Sigma_{R2} > (2D_1\Sigma_{R1}\Sigma_{R2}+D_2\Sigma_{R1})\xi^2+\Sigma_{R1}^2\Sigma_{R2}.$$

For convenience, we introduce the abbreviations

$$a = 2D_1\Sigma_{R1}\Sigma_{R2}+D_2\Sigma_{R1},\quad b=\Sigma_{R1}^2\Sigma_{R2}$$

and estimate the dominant error contribution by

$$\begin{aligned}\delta^{(1)}\phi(r') &< \frac{1}{a}\int_0^\infty \xi\frac{1}{\xi^2+\sqrt{\frac{b}{a}}^2}J_0(\xi r)J_0(\xi r')\,d\xi \\ &= \frac{1}{a}\int_0^\infty \xi\frac{1}{\xi^2+\sqrt{\frac{b}{a}}^2}J_0(\xi r)J_0(\xi r')\,d\xi \\ &= \begin{cases}\frac{1}{a}I_0(\sqrt{\frac{b}{a}}r')K_0(\sqrt{\frac{b}{a}}r) \text{ for } 0<r'<r \\ \frac{1}{a}I_0(\sqrt{\frac{b}{a}}r)K_0(\sqrt{\frac{b}{a}}r') \text{ for } r<r'<R\end{cases} \\ &= \frac{S_0\mu_{12}\mu_{21}}{a}\left(K_0(\sqrt{\frac{b}{a}}r)\int_0^r r'I_0(\sqrt{\frac{b}{a}}r')\,dr'\right. \\ &\qquad\left.+I_0(\sqrt{\frac{b}{a}}r)\int_r^R r'K_0(\sqrt{\frac{b}{a}}r')\,dr'\right).\end{aligned}$$

By a numerical test, one verifies that the error at the center is an order of magnitude larger than the error at the outer radius R, and both are several orders smaller than unity. Thus,

$$\{E_\phi^n\}_{n=1}^\infty = \{E_\phi^1, E_\phi^2, E_\phi^3, \ldots\}$$

is a monotonically decreasing sequence of functions inside the domain $[0,R]$.

8.6 Conclusions

In this work a novel approach to solve neutron diffusion problems in cylindrical geometry [DaKhOd11] was developed. The analytical expression found represents an accurate solution of an approximate problem for the multi-group steady state and multi-region diffusion equation in cylinder coordinates. An immediate conclusion that may be drawn from this work is that for neutron diffusion problems the Parseval identity is a considerably efficient technique to solve this type of problem. As can be seen from the formulation, the present method provides an analytical final expression without making use of simplifications. It is noteworthy that from Parseval's identity one obtains contributions by Bessel functions that are the eigenfunctions of the radial Sturm-Liouville problem. If an analytical solution was obtained by a spectral theory approach, the solution would have been expressed as an expansion of orthogonal functions with an associated functional basis. Parseval's identity indicates a natural basis that should be used by a spectral method approach and allows to truncate the basis to a small dimension still maintaining an acceptable precision in the numerical results. It is noteworthy that the eigenvalue spectrum that may be determined from the set of eigenfunctions seems to be independent of the geometry considered, which was also indicated in [GoViBo10], where it was called eigenvalue universality. Concluding, this method in cylindrical geometry can be considered a reliable tool for solving more general problems in neutron diffusion, for example, with more energy groups. We further plan to investigate results for a variety of situations of interest, where we hope to support this new method in the future.

References

[Ba54] Bateman, H.: Tables of Integral Transforms. McGraw-Hill, New York (1954)

[BoEtAl10] Bodmann, B.E.J., Vilhena, M.T., Ferreira, L.S., Bardaji, J.B.: An analytical solver for the multi-group two dimensional neutron-diffusion equation by integral transform techniques. Nuovo Cimento **C 33**, 1–10 (2010)

[DaKhOd11] Dababneh, S., Khasawneh, K., Odibat, Z.: An alternative solution of the neutron diffusion equation in cylindrical symmetry. Ann. Nucl. Energ. **38**, 1140–1143 (2011)

[Fe11] Fernandes, J.C.L.: Solução Analítica da equaçõ de difusão de nêutrons multi-grupo em cilindro infinito homogêneo através da transformada de Hankel, PhD dissertation, UFRGS, Porto Alegre, Brazil (2011)

[GoLeVi09] Goncalves, G.A., Leite, S.B., Vilhena, M.T.: Solution of the neutron transport equation problem with anisotropic scattering. Ann. Nucl. Energ. **36**, 98–102 (2009)

[GoViBo10] Goncalves, G.A., Vilhena, M.T., Bodmann, B.E.J.: Heuristic geometric "eigenvalue universality" in a one-dimensional neutron transport problem with anisotropic scattering. Kertechnik **75**, 50–52 (2010)

[LeEtAl08] Lemos, R., Vilhena, M.T., da Silva, F.C., Wortmann, S.: Analytic solution for two-group diffusion equations in a multi-layered slab using Laplace transform technique. Progr. Nucl. Energ. **50**, 747–756 (2008)

[ViSeGo04] Vilhena, M.T., Segatto, C.F., Goncalves, G.A.: Analytical solution of the one-dimensional discrete ordinates equation by the Laplace and Hankel integral transform. In: Integral Methods in Science and Engineering, pp. 267–272. Birkhäuser, Boston (2004)

Chapter 9
What Is Convergence Acceleration Anyway?

B.D. Ganapol

9.1 Introduction

One of the primary objectives of the twelfth in the series of international conferences on Integral Methods in Science and Engineering (IMSE) is

> *"… to promote new research tools, methods and procedures beyond the specific realms of mathematics, …"*

an objective within which convergence acceleration, to be described, clearly falls. More appropriately, convergence acceleration can be considered "experimental (applied) mathematics" in contrast to "theoretical mathematics." More aptly stated by Feynman [FeLeSa09],

> *Mathematics is not a science…, in the sense that it is not a natural science. The test of its validity is not experiment.*

Also, as has been stated by the author in the past,

> *Applied (numerical) mathematics is a science, since the test of its validity is experiment.*

This presentation will provide the essence of convergence acceleration by example. Two applications are considered—one in biophysics and the other in reactor physics. The first is the simulation of self-assembly of proteins responsible for infectious disease via misfolding. The second is for the characterization of nuclear reactor transients by the point kinetics equations (PKEs). The connection between these two applications is dynamics as represented by ordinary differential equations.

For each application, the underlying (bio) physics will briefly be presented, followed by the mathematical description. Next, the standard numerical approach is

B.D. Ganapol (✉)
The University of Arizona, Tucson, AZ, USA
e-mail: ganapol@cowboy.ame.arizona.edu

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_9,

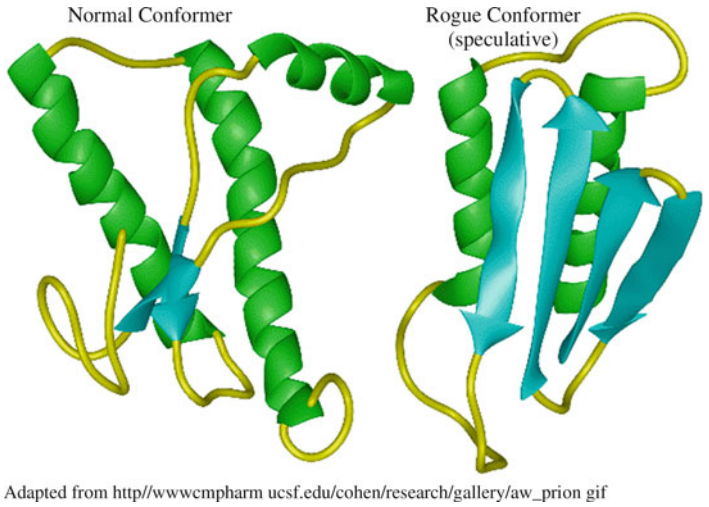

Fig. 9.1 A simulation of protein misfolding

outlined followed by application of convergence acceleration to enhance numerical accuracy. Finally, each application is demonstrated by several illustrative examples.

9.2 Simulation of Abnormal Protein Growth

9.2.1 Biophysical Setting

A prion is an infectious agent promoting protein misfolding thought responsible for a host of neurodegenerative diseases in mammals. Nobelist Stanley Prusiner coined the word from the two words "proteinaceous" and "infectious" adding "on" by analogy to "virion." Misfolding of proteins is known to be responsible for Creutzfeldt–Jakob disease and Kuru in humans and Bovine Spongiform Encephalopathy (BSE) in animals. Initially, misfolded proteins become a template for further protein folding as shown in Fig. 9.1 leading to the corruption of cell protein and disease onset. Misfolding induces fine fibers call fibrils, which grow from their ends and replicate upon breakage. It is this dynamic behavior that we intend to simulate.

Recently, in [KnEtAl09] there appeared an article entitled "An Analytical Solution to the Kinetics of Breakable Filament Self-Assembly." The title was particularly novel in that only infrequently will SCIENCE publish an article concerned with analytical solutions. Realizing that self-assembly must be a nonlinear process only heightens the mystery of the title and is the primary motivation for the analysis to follow.

We begin with the master equation describing the balance of filament growth from monomers

$$\frac{\partial f(t,j)}{\partial t} = 2k_+ m(t) f(t,j-1) - 2k_+ m(t) f(t,j) - k_-(j-1) f(t,j) + 2k_- \sum_{i=j+1}^{\infty} f(t,i) + k_n m(t)^{n_c} \delta_{j,n_c}. \quad (9.1)$$

Here, $f(t,j)$ is the concentration of filaments of length j and $m(t)$ is the concentration of free monomers. The terms on the RHS of (9.1) represent, in order from the equality sign,

- increase in filament length to j by addition of a monomer;
- loss of filaments j by growth to length $j+1$;
- loss from breakage at any of the $j-1$ internal links;
- two possible lengths for $i > j$ where breakage results in filaments of length j;
- spontaneous formation of a filament of length n_c.

In addition, we use the notation

$$\text{number density: } P(t) = \sum_{j=n_c}^{\infty} f(t,j),$$

$$\text{mass density: } M(t) = \sum_{j=n_c}^{\infty} j f(t,j).$$

We can show that these moments obey the following moments equations exactly:

$$\begin{aligned} \frac{dP(t)}{dt} &= k_- \left[m(t) + (2n_c - 1)P(t)\right] + k_n m(t)^{n_c} + k_- m_{tot}, \\ \frac{dm(t)}{dt} &= -2\left[k_+ m(t) - n_c(n_c-1)k_-/2\right] P(t) - n_c k_n m(t)^{n_c}, \end{aligned} \quad (9.2)$$

where

$$M(t) = m_{tot} - m(t),$$

with initial conditions

$$P(0) = P_0, \quad m(0) = m_{tot} - M_0.$$

Our goal is to solve these equations to a high numerical (extreme) accuracy. In [KnEtAl09], the authors claim an analytical solution and write

$$P(t) = \frac{m_{tot}}{2n_c - 1} - \frac{m_{tot} k_-}{\kappa} e^{-(2n_c-1)k_- t} Ei(-C_+ e^{\kappa t}) + B_2 e^{-(2n_c-1)k_- t},$$

$$M(t) = m_{tot}\left[1 - \exp\left(-C_+ e^{\kappa t} + C_- e^{-\kappa t} + \frac{k_n m_{tot}^{n_c-1}}{k_-}\right)\right], \tag{9.3}$$

where

$$\begin{aligned}
\kappa &\equiv \sqrt{2m_{tot}k_+k_-},\\
C_\pm &\equiv \frac{P_0}{m_{tot}}\sqrt{2m_{tot}\frac{k_+}{k_-}} \pm \frac{1}{2}\frac{M_0}{m_{tot}} \pm k_n\frac{m_{tot}^{n_c-1}}{2k_-},\\
B_2 &\equiv \frac{P_0}{m_{tot}} - \frac{1}{3} + \frac{1}{\sqrt{2m_{tot}\frac{k_+}{k_-}}}Ei(-C_+),
\end{aligned}$$

and Ei is the exponential integral. Their solution derives from a fixed-point iteration of one turn in M and two turns in P and neglects $O(k_n)$ terms. Clearly, this is not an analytical solution but an analytical approximation. Nevertheless, a significant result of their analysis is the establishment of scaling laws, for example the dimensionless parameter kappa. However, they go on to assert that only by having an analytical form at hand can such scaling laws be found. Here, we shall show that this assertion is not true.

9.2.2 Numerical Formulation

A numerical solution is facilitated by recasting (9.2) in vector form by letting

$$\mathbf{y}(t) \equiv \begin{bmatrix} P(t) \\ m(t) \end{bmatrix},$$

to give

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{A}(\mathbf{y}(t))\mathbf{y}(t) + \mathbf{S}(t),$$

where

$$\mathbf{A}(\mathbf{y}(t)) \equiv \begin{bmatrix} -k_-(2n_c-1) & -k_- - k_n m(t)^{n_c-1}, \\ n_c(n_c-1)k_- & -2k_+P(t) - n_c k_n m(t)^{n_c-1} \end{bmatrix}, \quad \mathbf{S}(t) \equiv \begin{bmatrix} k_- m_{tot} \\ 0 \end{bmatrix}.$$

The initial conditions are

$$\mathbf{y}(0) \equiv \begin{bmatrix} P_0 \\ m_{tot} - M_0 \end{bmatrix}.$$

A numerical algorithm is found by first discretizing time into uniform steps of $h = [t_j, t_{(j+1)}]$ and integrating over each step to give the algorithm

$$\begin{aligned} \mathbf{y}_{j+1}^0 &\equiv \mathbf{y}_j, \\ \mathbf{y}_{j+1}^l &= \left[\mathbf{I} - \frac{h}{2}\mathbf{A}_{j+1}(\mathbf{y}_{j+1}^{l-1})\right]^{-1} \left\{ \left[\mathbf{I} - \frac{h}{2}\mathbf{A}_j\right] \mathbf{y}_j + \frac{h}{2}\left[\mathbf{S}_{j+1} + \mathbf{S}_j\right] \right\}, \end{aligned} \tag{9.4}$$

where $\mathbf{y}_j$ is the approximation to $\mathbf{y}(t_j)$. This is a standard Runge–Kutta second-order scheme [ShGlTh03], where one can show that

$$\mathbf{y}(t_j) = \mathbf{y}_j(h) + \sum_{l=1}^{\infty} a_{l,j} h^{2l} \tag{9.5}$$

with the dependency of the approximation on h now indicated. Also, note that (9.4) is nonlinear as $\mathbf{y}_j$ is in the $\mathbf{A}$- matrix, which requires an additional iteration during the time step.

To this point, we have derived the classical numerical treatment, which certainly gives high accuracy for small time steps. Our goal here, however, is to achieve as high an accuracy as possible in the most straightforward way possible. This will be through convergence acceleration [Si03].

9.2.2.1 Convergence Acceleration

Consistency, the transition from a purely mathematical to a numerical algorithm and back, requires taking a limit. This is the case for fixed-point iteration, finite differences, finite elements, numerical quadrature, summation, and essentially all numerical processes known. We clearly see this in the above formulation, where the numerical method depends upon the time discretization (from mathematical formulation to numerical approximation) and the accuracy of the numerical approximation (from approximation back to the analytical (mathematical) solution) depends upon the smallness of h, or, in other words, its limit to zero. Thus

$$\mathbf{y}(t_j) = \lim_{h \to 0} [\mathbf{y}_j(h)].$$

The limit can also be viewed as the limit of the sequence of discretizations

$$\mathbf{y}(t_j) = \lim_{k \to \infty} [\mathbf{y}_j(h_k)],$$

where the steps h_k are essentially "free" to choose restricted only to give $t_j = jh_k$. It is important to emphasize that the numerical approximation, as a limit, is no longer a single approximation since it is replaced by a sequence of approximations tending toward their limit

$$\mathbf{y}_j(h_1), \mathbf{y}_j(h_2), \ldots, \mathbf{y}_j(h_k) \to \mathbf{y}(t_j).$$

The role of convergence acceleration is to simply force convergence to the limit more rapidly in the sense that one can construct a new sequence $\mathbf{y}_{j,k}$ such that (by component r)

$$\lim_{k\to\infty}\left[\frac{\mathbf{y}_{j,k}-\mathbf{y}(t_j)}{\mathbf{y}_j(h_k)-\mathbf{y}(t_j)}\right]_r = 0.$$

Now, the numerator approaches the limit faster than the original sequence in the denominator. In this way, we achieve a higher degree of accuracy with fewer terms. The key to this procedure, however, is the construction of the faster converging sequence, which we now consider.

9.2.2.2 Application of Richardsons and Wynn-Epsilon Accelerations

Richardsons deferred limit [Si03] comes directly from the relation between the exact and approximate solutions (see (9.5))

$$\mathbf{y}(t_j) = \mathbf{y}_{j,0}(h) + \sum_{l=1}^{\infty} \mathbf{a}_{j,l,0} h^{2l}, \tag{9.6}$$

where the initial approximation, based on h, is now called $\mathbf{y}_{j,0}(h)$. By continually refining the time grid by a factor of 2 and sequentially eliminating h^{2l} in (9.6), one arrives at a recurrence relation for higher-order approximations

$$\begin{aligned}
&\mathbf{y}_{j,0}(h) \equiv \mathbf{y}_j(h),\\
&\mathbf{y}_{j,k}(h) \equiv \left[\frac{2^{2k}\mathbf{y}_{j,k-1}(h/2)-\mathbf{y}_{j,k-1}(h)}{2^{2k}-1}\right], \quad k=1,2,\ldots.
\end{aligned}$$

Now, the higher-order approximation is

$$\mathbf{y}(t_j) = \mathbf{y}_{j,k}(h) + \sum_{l=k+1}^{\infty} \mathbf{a}_{l,j,k} h^{2l}.$$

$\mathbf{y}_{j,k}$ is the accelerating sequence. The power of this sequence to accelerate is observed by forming (by component r)

$$\left[\frac{\mathbf{y}_{j,k}-\mathbf{y}(t_j)}{\mathbf{y}_j(h)-\mathbf{y}(t_j)}\right]_r = \frac{\sum_{l=k+1}^{\infty} a_{r,l,j,k} h^{2l}}{\sum_{l=1}^{\infty} a_{r,l,j} h^{2l}} h^{2k},$$

which tends to zero with h and gives an approximation order of $2k$.

To decide when converged, one interrogates the sequences

$$\mathbf{y}_{j,0}(h/2^k), \mathbf{y}_{j,k}(h); \quad k = 0, 1, 2, \ldots$$

for convergence by component; that is,

$$\begin{cases} e_0 \equiv \max\limits_{r=1,2} \left| \dfrac{y_{j,0,r}(h/2^k) - y_{j,0,r}(h/2^{k-1})}{y_{j,0,r}(h/2^k)} \right| < \varepsilon, \\ e_R \equiv \max\limits_{r=1,2} \left| \dfrac{y_{j,k,r}(h) - y_{j,k-1,r}(h)}{y_{j,k,r}(h)} \right| < \varepsilon. \end{cases}$$

The limitation of Richardsons deferred approach is the requirement that the time step be continually reduced. To offset this disadvantage, a sometimes more quickly convergent sequence can be constructed from the nonlinear Wynn-epsilon (W-e) acceleration [Si03], namely

$$\begin{aligned} \varepsilon_{-1}^{(l)} &\equiv 0 \\ \varepsilon_0^{(l)} &\equiv P_j^l \ \ or \ \ m_j^l, \ \ l = 0, \ldots, L \\ \varepsilon_{k+1}^{(l)} &= \varepsilon_{k-1}^{(l+1)} + \left[\varepsilon_k^{(l+1)} - \varepsilon_k^{(l)} \right]^{-1}, \ k = 0, \ldots, L; \ l = 0, \ldots, L-k-1. \end{aligned}$$

Here, each subsequent sequence $\varepsilon_{k+1}^{(l)}$ formed by starting from the original sequence represents an accelerated sequence that can be arranged in the array

$$\begin{array}{ccccccc} \varepsilon_0^{(0)} & \varepsilon_1^{(0)} & \varepsilon_2^{(0)} & \ldots & \varepsilon_{L-1}^{(0)} & \varepsilon_L^{(0)} \\ \varepsilon_0^{(1)} & \varepsilon_1^{(1)} & \varepsilon_2^{(1)} & \ldots & \varepsilon_{L-1}^{(1)} & \\ \varepsilon_0^{(2)} & \ldots & & \ldots & & \\ \ldots & & \varepsilon_2^{(L-2)} & & & \\ & \varepsilon_1^{(L-1)} & & & & \\ \varepsilon_0^{(L)} & & & & & \end{array}$$

Only the entries in the even columns, starting from column zero, are relevant and the most accurate approximation is usually the last entry in these columns; hence, we interrogate

$$\varepsilon_{We} \equiv \left| \frac{\varepsilon_i^{(L-i)} - \varepsilon_i^{(L-i-2)}}{\varepsilon_i^{(L-i)}} \right| < \varepsilon, \ i = 2, \ldots, 2[L/2]$$

for convergence.

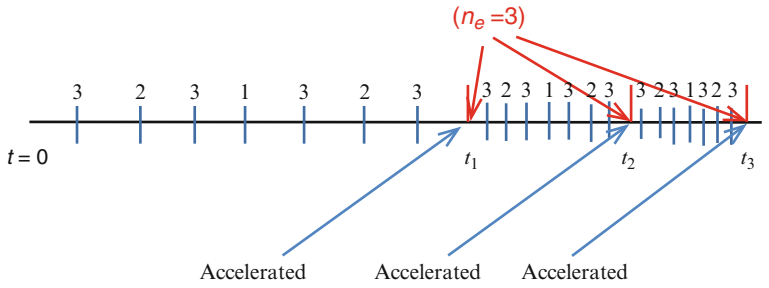

Fig. 9.2 Application of convergence acceleration

Figure 9.2 shows how convergence acceleration is implemented in the algorithm of (9.4). Say, we are interested in 3 edits, t_j, $j = 1,2,3$. A calculation is now performed with (9.4) to find the moments at t_1 with $h = t_1$. Next, the initial interval $[0,t_1]$ is partitioned by 2, where the added center point is indicated by 1 in Fig. 9.2. A second calculation for the moments at t_1 is performed, now with $h = t_1/2$, thus using the added edit. The interval is again halved, where the added edits are indicated now by 2 and the moments at t_1 are again determined with the added edits included. In this way, we are building a sequence of approximations for the moments at t_1. Both Richardsons and W-e accelerations are then applied to this sequence to accelerate convergence over the sequential grids to determine the high order approximation at t_1. When the moment approximations at t_1 are within a desired limit, we move to the second interval $[t_1,t_2]$ and repeat the process for the moments at t_2 with the newly converged moments at the end of the first interval as the initial condition. The calculation is particularly resistant against propagation error since each interval begins with a highly converged initial condition.

It is possible to confirm the accuracy of convergence acceleration through a manufactured solution. To do this, we assume a solution which, in this case, is the starting iterate, of [KnEtAl09], for the fixed-point iteration leading to (9.3)

$$\begin{aligned} P_0(t) &= D_+e^{\kappa t} + D_-e^{-\kappa t} - \frac{n_c k_n m_{tot}^{n_c-1}}{2k_+}, \\ M_0(t) &= \frac{2k_+m_{tot}D_+}{\kappa}e^{\kappa t} - \frac{2k_+m_{tot}D_-}{\kappa}e^{-\kappa t} - \frac{k_n m_{tot}^{n_c}}{k_-}, \end{aligned} \tag{9.7}$$

with

$$\begin{aligned} D_\pm &\equiv \frac{n_c k_n m_{tot}^{n_c-1}}{4k_+} \pm \frac{k_n m_{tot}^{n_c}\kappa}{4m_{tot}k_+k_-}, \\ P_0(0) &= M_0(0) = 0. \end{aligned}$$

When (9.7) are introduced into (9.2) now including sources S_p and S_m, the sources can be found. Thus, we are to solve

$$\frac{dP(t)}{dt} = -k_-[m(t) + (2n_c - 1)P(t)] + k_n m(t)^{n_c} + k_- m_{tot} + S_p(t),$$
$$\frac{dm(t)}{dt} = -2[k_+ m(t) - n_c(n_c - 1)k_-/2]P(t) - n_c k_n m(t)^{n_c} + S_m(t).$$

Since we know the solution for these sources, the numerical method should return the analytical solution of (9.7), which is the diagnostic value of a manufactured solution giving a precise error assessment.

We now consider an example with parameters

$$\begin{aligned} k_+ &\equiv 5 \times 10^4 M^{-1} s^{-1}, \\ k_- &\equiv 2 \times 10^{-8} s^{-1}, \\ n_c &= 2, \\ M_0 &= 0\mu M, \\ P_0 &= 0\mu M, \\ k_n &= 2 \times 10^{-5} s^{-1} M^{-1}, \\ m_{tot} &= 5\mu M. \end{aligned}$$

Figure 9.3a shows the moment time evolution. Figure 9.3b illustrates convergence behavior of the original and three additional sequences accelerations. The latter three accelerations are offset in time to better view the contrast between them. Each horizontal line represents the $M(t)$ approximation for a specific grid discretization, which in this case is $h_k = 30/2^k$, $k = 0, 1, \ldots, 7$. As observed, the third grid of Richardsons acceleration nearly outperforms 8 grids of the original calculation. The W-e is not nearly as effective; while, a W-e/Richardsons acceleration is no better than the original Richardson acceleration. Since only the original finite difference sequence is used in Richardsons extrapolation, from this confirmation, we concluded that convergence acceleration is indeed an effective way to use limited accuracy data to achieve extreme accuracy results.

Figure 9.4 shows a comparison of the accelerated solution to the analytical approximation of (9.3). The rapid growth of the prions is clearly evident resulting in a sigmoid curve. The initial rise comes directly from elongation through monomer adhesion as represented by the last term in (9.1) and is characterized by (9.7). As secondary filament breakage and elongation represented by the remaining terms on the RHS of (9.1) become increasingly important, elongation is arrested and the sigmoid nature of the evolution emerges.

In addition, the converged accelerated solution seems to indicate a less rapid rise than the approximate solution and therefore gives a slightly decreased phase lag time, which is the measure of the onset of a spongiform disease. The phase lag time is determined by the time from growth initiation to the time found by a tangent from the inflection point to zero moment as noted in Fig. 9.4.

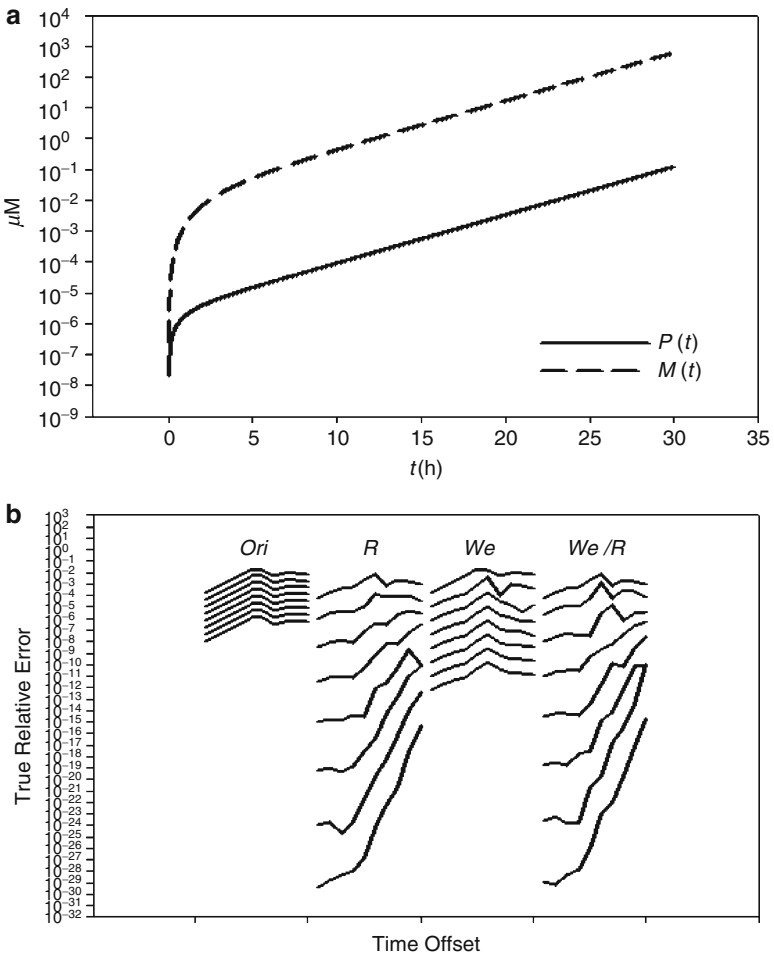

Fig. 9.3 (**a**) Moments evolution and (**b**) Error by sequence acceleration for a manufactured solution

To be entirely transparent, it must be remarked that the highly accurate converged accelerated solution seems to add little to the analytical approximation. It does, however, provide a confidence in the solution that was lacking. The above formulation in the larger sense is a demonstration of a new solution paradigm associated with convergence acceleration, where near analytical accuracies are achievable from the simplest of finite difference approximations.

Regardless of how the equations are solved, the real impact of their solution is that they allow the consolidation of experimental results through scaling laws. In particular, Fig. 9.5 shows that for in vitro experiments varying total insulin content m_{tot}, the analytical approximation and converged accelerated solution well

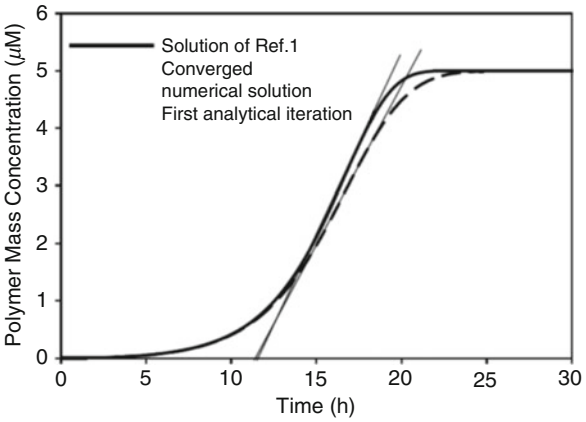

Fig. 9.4 Mass evolution of the filaments indicating the onset of a spongiform disease

represent the data. This has also been shown for other experiments, like Formin-Binding protein FBP28 [FeEtAl03] in Fig. 9.6. Here, the kinetics parameters have been derived from the last curve in Fig. 9.6b to predict the first three curves. Both the converged accelerated and analytical approximation capture the biophysics as shown in Fig. 9.6a.

While the ability to predict a relatively large number of experiments is quite remarkable, what about the requirement of an analytical solution or approximation to be able to derive the scaling laws?

9.2.2.3 Dimensional Analysis

Apparently, the authors of [KnEtAl09] are unaware of dimensional analysis. It is easy to see that the moments solution depends upon a number of parameters, namely

$$\mathbf{y}(t) = \mathbf{F}[k_-, k_+, k_n, m_{tot}; t, P(t), M(t)],$$

where

$$\begin{aligned}
[k_+] &= T^{-1}, \\
[k_-] &= T^{-1}M^{-1}, \\
[k_n] &= T^{-1}M^{-1}, \\
[m_{tot}] &= M, \\
[t] &= T, \\
[P(t)] &= M, \\
[M(t)] &= M.
\end{aligned}$$

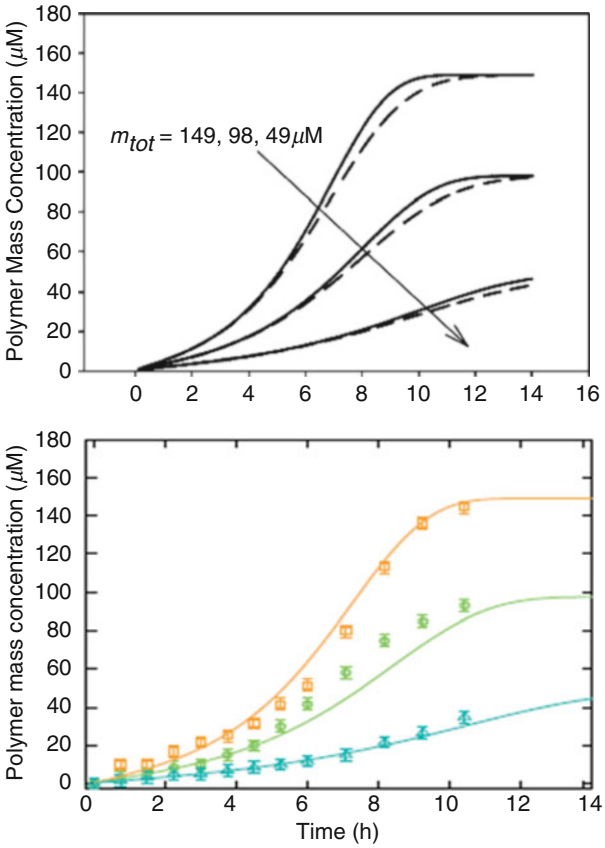

Fig. 9.5 Scaling of data for in vitro insulin experiments

Thus, two fundamental units moles (M) and time (T) characterize the moments solution. Therefore, from a dimensional analysis even without a solution, the following two of six or so independent dimensionless quantities emerge

$$\pi_1 \equiv k_- t, \quad \pi_2 \equiv m_{tot} \frac{k_+}{k_-},$$

which, when combined, give the additional dimensionless quantity

$$\pi_7(t) = \pi_1^2 \pi_2 = k_- k_+ m_{tot} t^2 = \kappa^2 t^2/2.$$

Here, κ represents the rate of multiplication of the filament secondary nucleation. Since the analytical approximation (9.3) can be recast entirely in terms of these three dimensionless numbers and the additional quantities

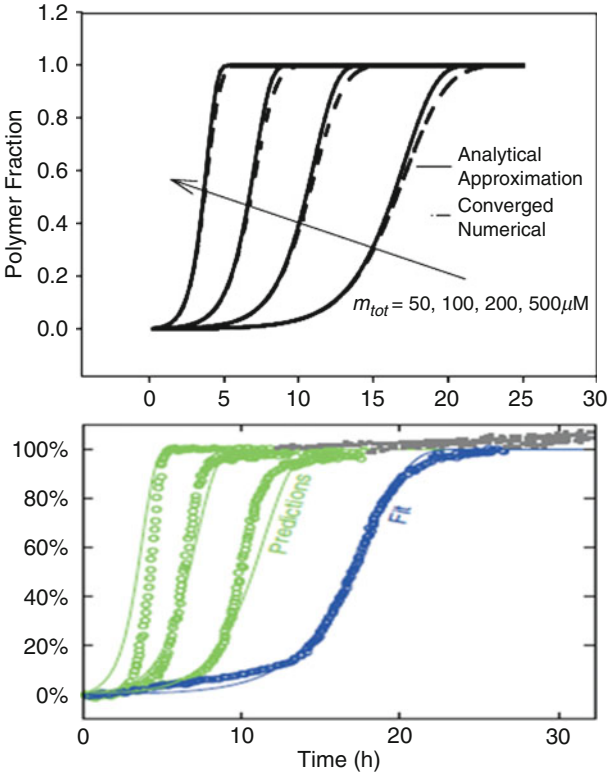

Fig. 9.6 Evolution of Formin-binding protein FBP28

$$\pi_P(t) \equiv \frac{P(t)}{m_{tot}}, \quad \pi_M(t) \equiv \frac{M(t)}{m_{tot}},$$

they indeed govern the time evolution of the solution.

If the moment equations are solved with convergence acceleration for 2,400 random sets of kinetic data and for several experiments with predetermined kinetic data, then a plot of dimensionless quantities

$$\pi_5 \equiv m_{tot} \frac{k_n}{k_-}$$

and π_7, remarkably gives the straight line of Fig. 9.7. The equation of this line is

$$\ln \pi_5 = -1.132 - 1.414\sqrt{\pi_7(\tau_{lag})}$$

which, in the original variables, is

$$\tau_{lag} = [\ln(1/C_+) - 1.825]/\kappa. \tag{9.8}$$

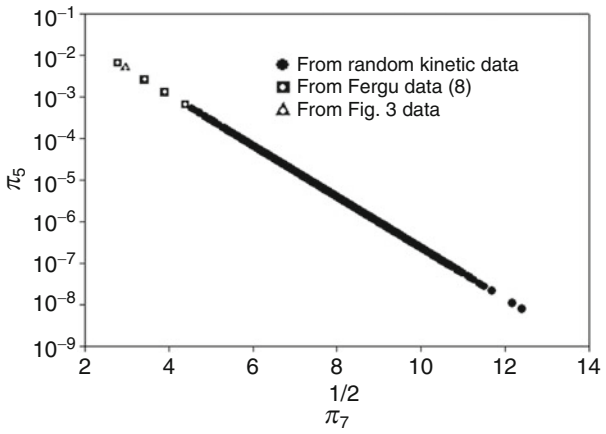

Fig. 9.7 Primary scaling law for phase lag

To interpret the physics in (9.8) for zero initial conditions, we write

$$C_+ \equiv k_n \frac{m_{tot}^{n_c-1}}{2k_-},$$

where k_n is the growth from primary nucleation. Hence, the phase lag time has weak dependence on primary nucleation and strong dependence on κ, which controls secondary nucleation through k_+ and k_-. Thus, it is important to include secondary nucleation in any protein growth studies.

It is important to remark that the constant 1.825 in (9.8) is apparently more accurate than the 1.718 found in [KnEtAl09] (see also (Knowles, Private communication)) and makes the phase lag time more accurate, presumably because of the more accurate converged accelerated solution.

In summary, we began with one of the most standard and fundamental numerical algorithms for solving a nonlinear ODE and wrapped it in a convergence acceleration. It was then demonstrated that, if indeed, high accurate solutions are desired, then convergence acceleration is a simple and reliable way to accomplish this. Next, we consider a second example, which reinforces this conclusion.

9.3 Nuclear Reactor Kinetics

The second example to which convergence acceleration is applied is to nuclear reactor transients. Here, as we shall see, convergence acceleration can be beneficial in obtaining solutions which otherwise could not be found. The basic approach is different from the first example, but is similar in implementation.

Currently, there are a host of numerical methods applied to simulate reactor transient behavior, including

- Runga–Kutta (RK) methods;
- exponential transform;
- better basis functions;
- imposed stiffness integrating factors;
- Padé approximants;
- Laplace transform inversion;
- piecewise constant reactivity approximations (PCA);
- converged accelerated finite differences;
- Taylor series (TS) solution.

Until recently, it seemed quite apparent that there was no one universal numerical algorithm that was suited for both prescribed and nonlinear reactivity insertions. Fortunately, the landscape has changed with the re-invention of the TS solution, as will be shown. The TS solution has had a long history and was one of the first methods to be considered (see [Vi67] and [Vi71]). The method has been modified many times thereafter (see, for example, [Mi77], [Le95], [AbHa02], and [Ke65]). We will again consider the TS solution, but now with convergence acceleration.

There are advantages and disadvantages to the TS solution. Two advantages are that a piecewise continuous reactivity can be considered and both prescribed and nonlinear reactivities can be treated. In addition, an analytical solution, not in closed form however, is found. Disadvantages include, lack of closed form solution, and the evaluation of an infinite series which, in the past, has proven quite difficult. This will not be the case here.

In the solution to be featured, continuous analytical continuation (CAC) will be coupled to convergence acceleration to provide a highly accurate numerical evaluation of the PKEs.

9.3.1 Reactor Transients

9.3.1.1 Reactor Kinetics Equations

The PKEs describing a nuclear reactor transient are

$$\begin{aligned} \frac{dN(t)}{dt} &= \left[\frac{\rho(t,N)-\beta}{\Lambda}\right] N(t) + \sum_{i=1}^{m} \lambda_i C_i(t), \\ \frac{dC_i(t)}{dt} &= \frac{\beta_i}{\Lambda} N(t) - \lambda_i C_i(t), \quad i = 1,\ldots,m. \end{aligned} \tag{9.9}$$

The PKE equations relate the reactor neutron density, N, to changes in the fissioning and absorbing properties of a reactor, such as from control rod motion.

These changes are characterized by the reactivity ρ, which can be imposed or depend upon N. As the fission product inventory changes, some elements, called precursors classified into m-groups i, will produce additional delayed neutrons, from decay with decay constant λ_i. The precursors are also being created directly from fission with a yield of β_i, where β is the total yield over all delayed precursors groups. The neutron generation time from birth to absorption is Λ.

When the Taylor series

$$\mathbf{y}(t) \equiv \begin{bmatrix} N(t) \\ C_i(t) \\ \rho(t,N) - \beta \end{bmatrix} = \sum_{k=0}^{\infty} \begin{bmatrix} N_{k,j-1} \\ C_{i,k,j-1} \\ \rho_{k,j-1} - \beta\,\delta_{k0} \end{bmatrix} (t - t_{j-1})^k \tag{9.10}$$

is placed in (9.9), we find that the Taylor coefficients satisfy the recurrence relation

$$\begin{aligned} (k+1)N_{k+1,j} &= \frac{1}{\Lambda}\sum_{l=0}^{k}(\rho_{k-l,j} - \beta\,\delta_{k-l,0})N_{l,j} + \sum_{i=0}^{m}\lambda_i C_{i,k,j}, \\ (k+1)C_{i,k+1,j} &= \frac{\beta_i}{\Lambda}N_{k,j} - \lambda_i C_{i,k,j}. \end{aligned} \tag{9.11}$$

This procedure is called CAC [Vi71] and helps ensure convergence of the TS by adjusting the time step appropriately. The initial conditions begin from a critical reactor, so for the initial interval,

$$N_{0,0} = N(0), \quad C_{i,0,0} = \frac{\beta_i}{\lambda_i \Lambda} N_{0,0}, \tag{9.12}$$

and for all subsequent intervals,

$$\begin{aligned} N_{0,j} &= N(t_j) = \sum_{k=0}^{\infty} N_{k,j-1}(t_j - t_{j-1})^k, \\ C_{i,0,j} &= C_{i,0}(t_j) = \sum_{k=0}^{\infty} C_{i,k,j-1}(t_j - t_{j-1})^k. \end{aligned} \tag{9.13}$$

9.3.2 Numerical Implementation

9.3.2.1 Application of Richardsons and Wynn-Epsilon Accelerations

Richardsons and the W-e accelerations are applied exactly as in the first example. Every interval begins with a converged accelerated initial condition through either Richardsons or the W-e acceleration, where now all evaluations are through the

Table 9.1 1$ step insertion

t	N(t)	
0.1	7.59084115305179414332394 38	E+07
0.2	3.38871492981176968523922 18	E+13
0.3	1.51279525957596757971839 71	E+19
0.4	6.75344354658506400992449 04	E+24
0.5	3.01488251289837640834027 78	E+30

TS of (9.10)–(9.13). A difference from Example 1 is that if any interval fails to converge in 12 sub-grids, the calculation restarts that interval with additional edits introduced. We allow 2^{12} additions to the original interval. In addition, a passive time step control is introduced. If the Taylor series fails to converge in 25 terms, the current interval is subdivided and the calculation restarted. In this way, the Taylor series will nearly always converge. In addition, we apply the W-e acceleration to the Taylor series partial sums for acceleration. The entire algorithm is called the convergence accelerated Taylor series (CATS).

A very convenient verification for CATS, or any other algorithm, exists for step reactivity insertion. For this first case, reactivity is a step at time zero, i.e., the instantaneous removal of a control rod. Then (9.9) written in vector form is

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{A}\mathbf{y}(t),$$

with initial conditions

$$\mathbf{y}(0) \equiv \left[1 \quad \frac{\beta_1}{\lambda_1 \Lambda} \quad \cdots \quad \frac{\beta_m}{\lambda_m \Lambda} \right]^T,$$

gives the solution

$$\mathbf{y}(t) = e^{\mathbf{A}t}\mathbf{y}(0).$$

If $\mathbf{A}$ is diagonalizable such that $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{U}^{-1}$, where $\mathbf{W}$ is a diagonal matrix of eigenvalues and $\mathbf{U}$ is the matrix of eigenvectors, then

$$\mathbf{y}(t) = \mathbf{U}e^{\mathbf{W}t}\mathbf{U}^{-1}\mathbf{y}(0). \tag{9.14}$$

The results for $N(t)$ from the CATS algorithm for \$1 ($\rho = \beta$) reactivity insertion in a thermal reactor are shown in Table 9.1. The results agree to all 25 places with the analytical solution (9.14), thus demonstrating true extreme accuracy.

A second test is through a manufactured solution as in Example 1. If (9.9) is solved for reactivity, we find that

$$\rho(t) = \beta + \frac{\Lambda}{N(t)}\frac{dN(t)}{dt} - \frac{1}{N(t)}\sum_{i=1}^{m}\left[\beta_i e^{-\lambda_i t} + \lambda_i \int_0^t dt' e^{-\lambda_i (t-t')} N(t')\right]. \tag{9.15}$$

One can now specify $N(t)$ and find the reactivity for that specified neutron density trace. For example, if

$$N(t) \equiv 1 + f(1 - e^{-\alpha t}), \tag{9.16}$$

then the reactivity from (9.15) is found from

$$N(t)\rho(t) = \beta + f\Lambda\alpha e^{-\alpha t} - \sum_{i=1}^{m} \beta_i \left\{ e^{-\lambda_i t} + (1+f)(1-e^{-\alpha t}) - f\frac{\lambda_i}{\lambda_i - \alpha}(e^{-\alpha t} - e^{-\lambda_i t}) \right\}.$$

To put (9.16) into an appropriated Taylor series, we write

$$N(t) \equiv 1 + f\left(1 - e^{-\alpha t_{j-1}} e^{-\alpha(t-t_{j-1})}\right) = \sum_{k=0}^{\infty} N_{k,j}(t - t_{j-1})^k$$

and

$$\rho(t)N(t) = H(t - t_{j-1}) = \sum_{k=0}^{\infty} H_{k,j}(t - t_{j-1})^k,$$

which implies

$$N_{0,j}\rho_{k,j} = H_{k,j} - \sum_{l=0}^{k-1} N_{k-l,j}\rho_{l,j}.$$

The density and reactivity traces are shown in Fig. 9.8. The number of correct digits is given in Fig. 9.9, where one observes a minimum of 24-digit agreement.

This seems quite remarkable for an algorithm that was previously thought to perform so poorly.

9.3.2.2 Demonstration

Table 9.2 gives the neutron density for a prescribed ramp reactivity [$\rho = at$] of $a = 1\$/\text{s}$ for a neutron generation time of $\Lambda = 10^{-7}$ s representative of a fast reactor. The novelty of this solution is that it is able to achieve extremely high neutron densities. This is an extreme test of any discrete numerical method.

This case has been verified by an entirely different numerical method and again provides confidence in the CATS algorithm.

As a final application, we consider the Doppler shutdown of a reactor reactivity transient. For this case, the reactivity behaves as

$$\rho(t) = at - B\int_0^t dt' N(t'), \tag{9.17}$$

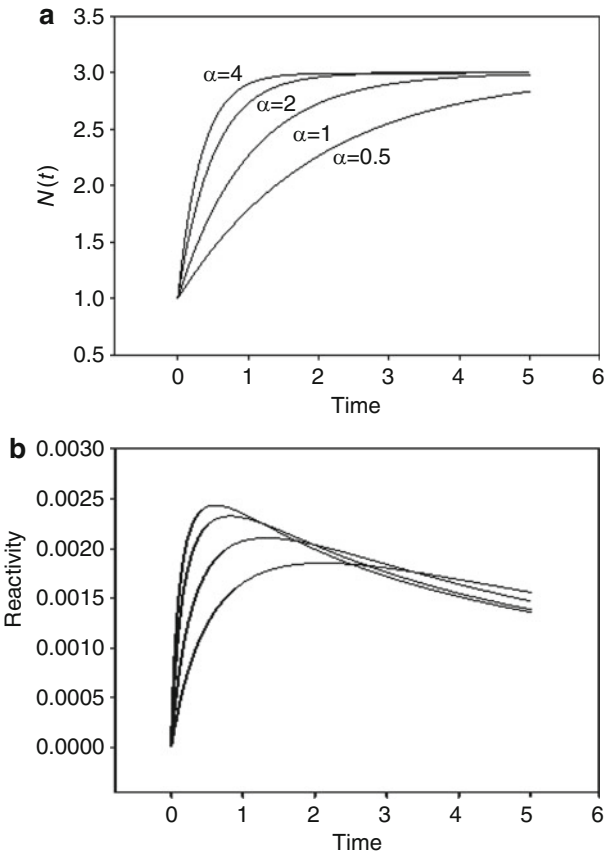

Fig. 9.8 (**a**) Neutron density and (**b**) reactivity for a manufactured solution

Table 9.2 Ramp \$1/s

t	N
0.01	$1.010097111E+00$
0.10	$1.113320112E+00$
0.20	$1.260559925E+00$
0.50	$2.136409107E+00$
1.00	$1.207814197E+03$
1.10	$3.257593355E+99$
1.15	$1.028975360+219$

where B is the Doppler shutdown reactivity coefficient. The TS coefficients for this reactivity are

$$\begin{aligned} \rho_{1,j} &= a, \\ \rho_{k,j} &= -BN_{k-1,j}, \quad k = 2, \ldots . \end{aligned} \tag{9.18}$$

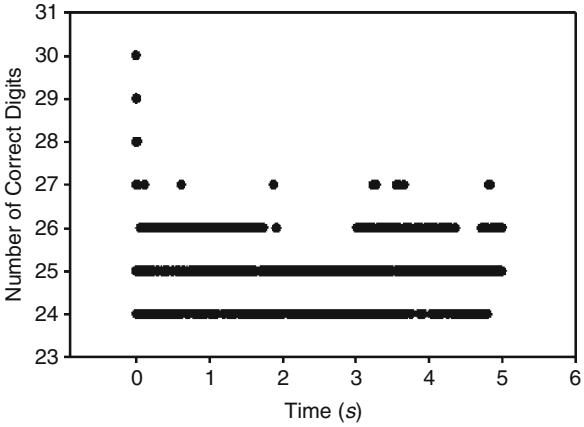

Fig. 9.9 Number of correct digits in agreement with the manufactured solution

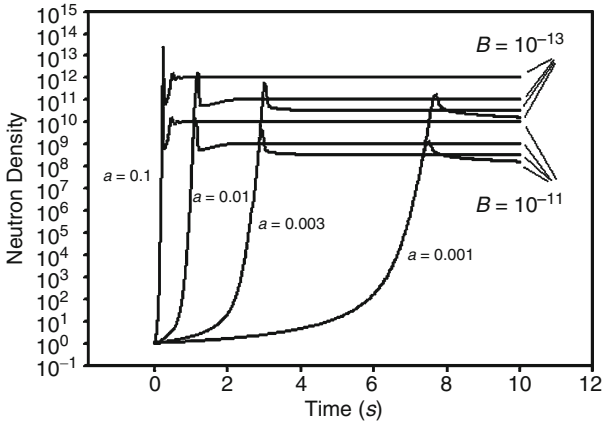

Fig. 9.10 Reactor shutdown transient

The neutron density is shown in Fig. 9.10, which shows how the initial transient sends the density to high values after which the temperature rises steeply and the Doppler shutdown reactivity begins to dominate to reduce the density. The algorithm seems to works flawlessly. Table 9.3 gives the time to the first peak and peak densities by interrogating the time derivative for a sign change. These are the first accurately published values for the peak times and densities and can serve as a valuable benchmark.

In summary, the CATS method enables the Taylor series solution. This is done by convergence acceleration of the initial points to each interval and CAC to ensure convergence of the TS. A demonstration indicates that the method can deliver extreme accuracy by wrapping the TS solution in convergence acceleration.

Table 9.3 Time to peak and peak neutron density for compensated transients

a	B	t_{peak}	N_{peak}
0.1	10^{-11}	$2.246634E-01$	$2.420382E+11$
	10^{-13}	$2.389069E-01$	$2.898674E+13$
0.01	10^{-11}	$1.106077E+00$	$2.012352E+10$
	10^{-13}	$1.155148E+00$	$2.491178E+12$
0.003	10^{-11}	$2.910582E+00$	$5.114160E+09$
	10^{-13}	$3.007602E+00$	$6.534474E+11$
0.001	10^{-11}	$7.488766E+00$	$1.274075E+09$
	10^{-13}	$7.683588E+00$	$1.721008E+11$

9.4 Conclusion

While the application of convergence acceleration to ODEs is not a new idea [BuSt66], [Gr63], its true application has never been fully appreciated. Most methods developers seem to consider convergence acceleration as a way to confirm the performance of a numerical method rather than to be the numerical method itself. In this presentation, it has clearly been shown that convergence acceleration is more than simple confirmation to find the order of a numerical method, but is a pathway to extreme accuracy. While extreme accuracy, 10^{-7}–10^{-9} relative error, may not always our the goal, is comforting to know that it is certainly possible if desired.

More importantly, if, as a numerical methods developer, we are aware of convergence acceleration, then becomes our responsibility and duty to apply convergence acceleration in order to achieve the most accurate results possible.

References

[AbHa02] Aboanber, A.E., Hamada, Y.M.: PWS: an efficient code system for solving space-independent nuclear reactor dynamics. Ann. Nucl. Energ. **29**, 2159–2172 (2002)

[BuSt66] Burlisch, R., Stoer, J.: Numerical treatment of ordinary differential equations by extrapolation methods. Numerische Math. **8**, 1–13 (1966)

[FeEtAl03] Ferguson, N., et al.: Rapid amyloid fiber formation from the fast-folding WW domain FBP28. Proc. Natl. Acad. Sci. **100**, 9814 (2003)

[FeLeSa09] Feynman, R.P., Leighton, R.B, Sands, M.: The Feynman Lectures on Physics, Vol 1: Mainly Mechanics and Heat. Addison-Wesley, Reading (2009)

[Ke65] Keepin, R.G.: Physics of Nuclear Kinetics. Addison-Wesley, Reading (1965)

[KnEtAl09] Knowles, T.P.J., et al.: An analytical solution to the kinetics of breakable filament assembly. Science **326**, 1533–1537 (2009)

[Le95] Lewins, J.D.: Reactivity oscillations and stability with delayed neutrons. Ann. Nucl. Energ. **22**, 411–414 (1995)

[Mi77] Mitchell, B.: Taylor series methods for the solution of the point reactor kinetic equations. Ann. Nucl. Energ. **4**, 169–176 (1977)

[Gr63] Gragg, W.B.: Repeated extrapolation to the limit in the numerical solution of ordinary differential equations. Thesis UCLA (1963)
[ShGlTh03] Shampine, L.F., Gladwell, I., Thompson, S.: Solving ODEs with MATLAB. Cambridge University Press, London (2003)
[Si03] Sidi, A.: Practical Extrapolation Methods. Cambridge University Press, Cambridge (2003)
[Vi67] Vigil, J.C.: Solution of the reactor kinetics, equations by analytic continuation. Nucl. Sci. Eng. **29**, 392–401 (1967)
[Vi71] Vigil, J.C.: ANCON User's Manual, LA-4616, UC-32. Mathematics and Computers, TID-4500 (1971)

Chapter 10
On the Fractal Pattern Phenomenology of Geological Fracture Signatures from a Scaling Law

I. Gioveli, A.J. Strieder, B.E.J. Bodmann, M.T. Vilhena, and A.S. Athayde

10.1 Introduction

Geologic fractures geometry and their connectivity are considered the main features when it comes to prospection of hydrocarbon reservoirs and also for the search of aquifers, since it is becoming increasingly important in the question of freshwater supply [OdEtAl99], [Be00], [PuEtAl01], [DaEtAl06]. Although there is a common consensus that fractures are commonly caused by stress–strain exceeding the admissible rock strength, it is save to say that there does not exist yet a quantitative and concise theory that relates fractures patterns to their origin by compression, or tension during genesis [Ha85], [RaHu87]. There exist mathematical model approaches such as the Mohr–Coulomb model, which is far from being a useful formulation in order to simulate fracture patterns similar to those found in rocky areas.

The complexity on the fracture pattern genesis as well as the dynamics of geological fracture pattern formation gives us the motivation to focus our attention on determining scaling laws, as a first step into a direction that shall reveal in the future the dynamics that leads to the observed fracture signature. In fact in this work we show that the fractal dimension of the geological fractures at different scales is a manifestation of a clean scaling law for the fracture directions.

I. Gioveli
Federal University of Fronteira Sul, Santo Ângelo, RS, Brazil
e-mail: izagio@gmail.com

A.J. Strieder (✉) • A.S. Athayde
Federal University of Pelotas, Pelotas, RS, Brazil
e-mail: adelirstrieder@uol.com.br; alexandre.athayde.ufpel@gmail.com

B.E.J. Bodmann • M.T. Vilhena
Federal University of Rio Grande do Sul, Porte Alegre, RS, Brazil
e-mail: bardo.bodmann@ufrgs.br; vilhena@mat.ufrgs.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_10,

It is notable that the directional step length range fits with considerable accuracy and affine feature in the log-log plane of step length by number of fractures with grid line intersections (to be introduced in the next section), suggesting a self-affine mechanism for fracture genesis. Hence, the present discussion is an attempt to translate the affine property of fracture direction occurrence into a fractal-discrete scheme, which represents the interactions of the shear or stress field with the considerable complex set of boundary conditions, established by the geological profile at each observable scale.

Fractal geometry has been a useful guide for understanding many natural patterns since it seems to be a common optimization solution used by Nature. The scale invariant fracture pattern is one of the many examples found in the geological scenarios where a fractal geometry is verified. In fact, the fracture hierarchy is composed by successive generations of fractures in different directions resulting from the multiple ramification of their antecedent (similar to a Cantor set construction), which reflects somehow the influence of the stress field modification by the presence of fractures of a given scale. From generation to generation, lengths diminish suggesting an underlying fractal geometry, which we confirm by our fractal analysis presented in Sects. 10.3–10.4.

In its consecutive generations ($n = 0, 1, 2 \ldots$), the hierarchy begins with the largest fracture scale, which gives rise to successive fracture scale subdivisions. The total of generations is finite by the fact that the analysis underlying images are limited by their resolution (i.e., the granular structure of outcrops, aerial photographs, remote sensing images among others) [Tu97]. Motivated by the fractal architecture of the fracture hierarchy scheme, which exhibits geometrically approximate self-similarity [Hi89], [Tu97], the present discussion is dedicated to the question whether affine characteristics and self-similar structure imposed by observational findings permit some sort of "reverse engineering" which may in a future work lead to an fracture pattern description implemented in a fractal discrete scheme [At99], [Ba04], [Ba84], [CaEtAl94], [CaCh95], [CaChCo03], [MoBoVa02], [PrVM03]. Such a procedure could replace the usual continuous formalisms based on mathematical spectral analysis which is in general too complicated when complex boundary conditions are involved.

The use of the fractal dimension method for geological fracture patterns is not new in the literature, for instance, in [VoKr04] it was used to analyze the apparent fracture intensity with their spatial orientation. General aspects of fracture systems in geological media with scaling laws were considered in [BoEtAl01], where the principal concern focused on the spatial distribution of fractures, the fracture intensity, and their self-similar appearance in different scales (see also [LaEtAl02], [NSEtAl05], [OrMaLa06]). To the best of our knowledge the perspectives that arise from our present discussion are new, since they may be considered a first step towards an approach that in a long term may open pathways for a dynamical fracture pattern genesis simulation beyond the phenomenological implementation, presented in this work.

Our article presents a study of the fractal dimension for an anisotropic fracture system which is typical for a selected geological site (Sect. 10.2), more specifically, homogeneous structural areas in Central Brazil (see Fig. 10.1). The fractal

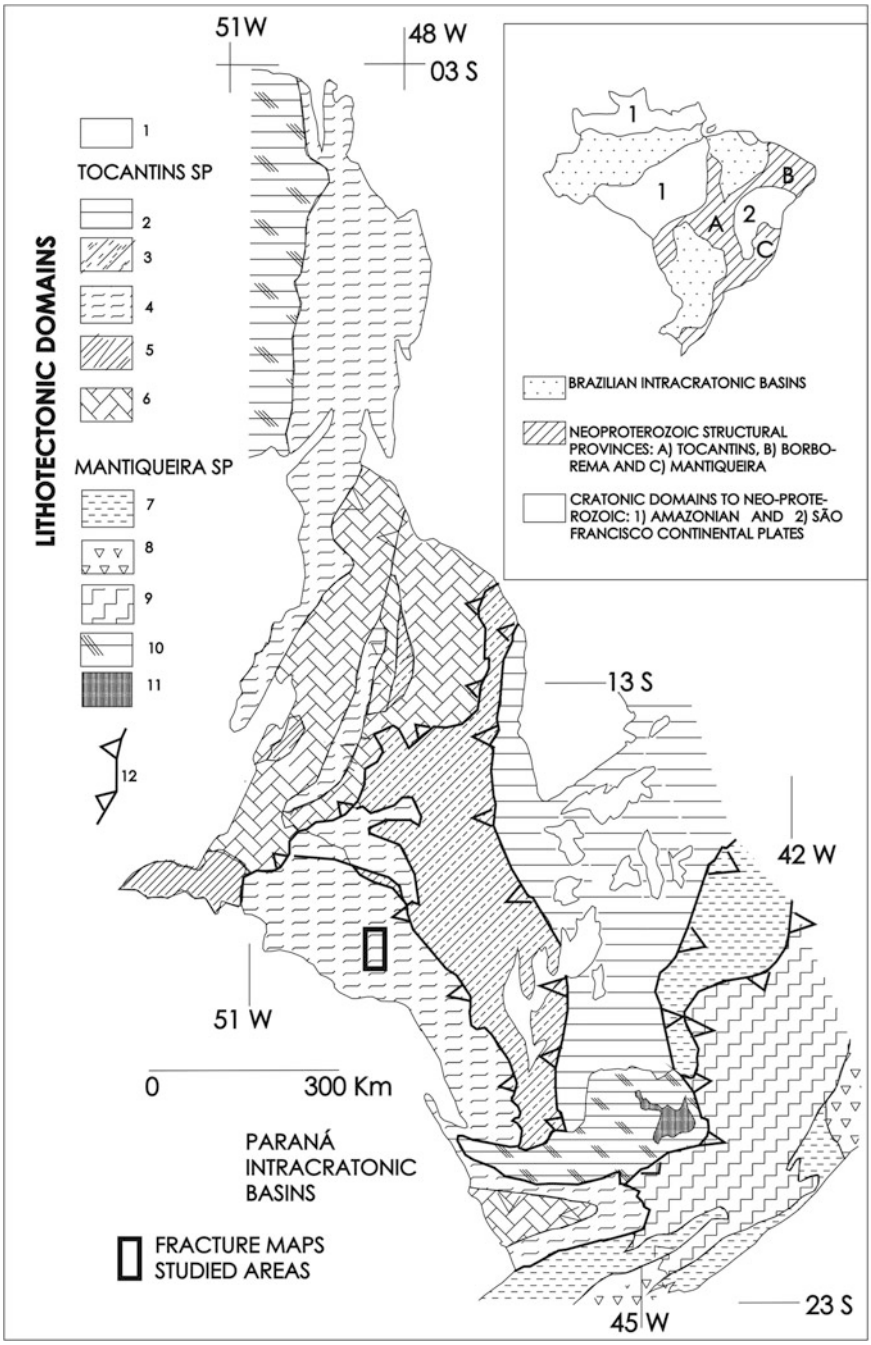

Fig. 10.1 (continued)

dimension was obtained by the Cantor Dust method [VeEtAl90], [VeEtAl91] for the fracture orientations which permits to establish the spatial fracture frequency distribution an essential input for the determination of the rock quality index as shown in [HuPr79].

10.2 Geological Setting of the Studied Areas

The studied areas are located in Central Brazil, in the Tocantins Structural Province (TSP, [AlEtAl81]). The TSP is a Neoproterozoic continental collision chain show predominantly north–south alignment. Of principal interest is the Abadiânia Nappe Thrust Sheet [StSu99], whose pseudo-stratigraphic $D1$ units are, from bottom to top: (a) Abadiânia Supersuite (Araxá metasedimentary suite and their Tectonic Block fragments suite), that correspond to an ophiolitic melange of the Neoproterozoic collision; and (b) a series of gneissifed units including Padre Souza Gneiss Suite, Maratá Lithodeme, Brumado Gneiss Suite, and others that are still being distinguished. The regional extent of the Abadiânia Nappe ($D2$ structure) from west to east developed a number of ESE tectonic inflexions, branching thrust faults and folds ($D3$ structures) in order to accommodate $D3$ deformation (see [StSu99] for details). The final deformation stage (retrogressive $D4$) corresponds to localized trans-current faults, controlled by a local stress field related to tectonic inflexions (see Fig. 10.2). One of the criteria for the choice of the area was good visibility of fractures over a considerable range of scales in order to analyze the fracture patterns for possible self-similarity [St93]. The Central Brazil (TSP) was not subject to younger deformational episodes, which is also the case for the areas located north, central, and south from the Serra do Fundão Inflexion. The fracture maps (shown in Fig. 10.2) for the Serra do Fundão in the scale 1 : 97000 were taken from [St93]. They originate from aerial photographs (1:110 000 scale) and printed LANDSAT TM5 images (221-072-X, band 5, scale 1:100 000; taken at the 16th of September of 1990; [St93]). Each fracture map represents the exact extension of the drainage linear segments, as observed in the aerial photographs and images.

Fig. 10.1 (continued) Geological setting of studied area in Central Brazil. (1) Phanerozoic sedimentary covers. Tocantins Structural Province; (2) undeformed low-grade meta-sediments; (3) Deformed low-grade meta-sediments; (4) Medium-grade meta-sediments (ophiolitic melange) and orthogneisses; (5) Neoproterozoic gneisses, granites, and volcano-sedimentary sequences; (6) Archean and Mesoproterozoic gneisses, granites, and volcano-sedimentary sequences. Mantiqueira Structural Province. (7) Deformed low-grade meta-sediments; (8) gneisses and granites of the Coastal Complex; (9) Granulytic gneisses of the Juiz de Fora Complex; (10) Cratonic granite-gneissic units of the (A) Amazonian and (B) São Francisco cratons. (11) Quadrilátero Ferrífero Greenstone Belt. (12) Main thrust faults

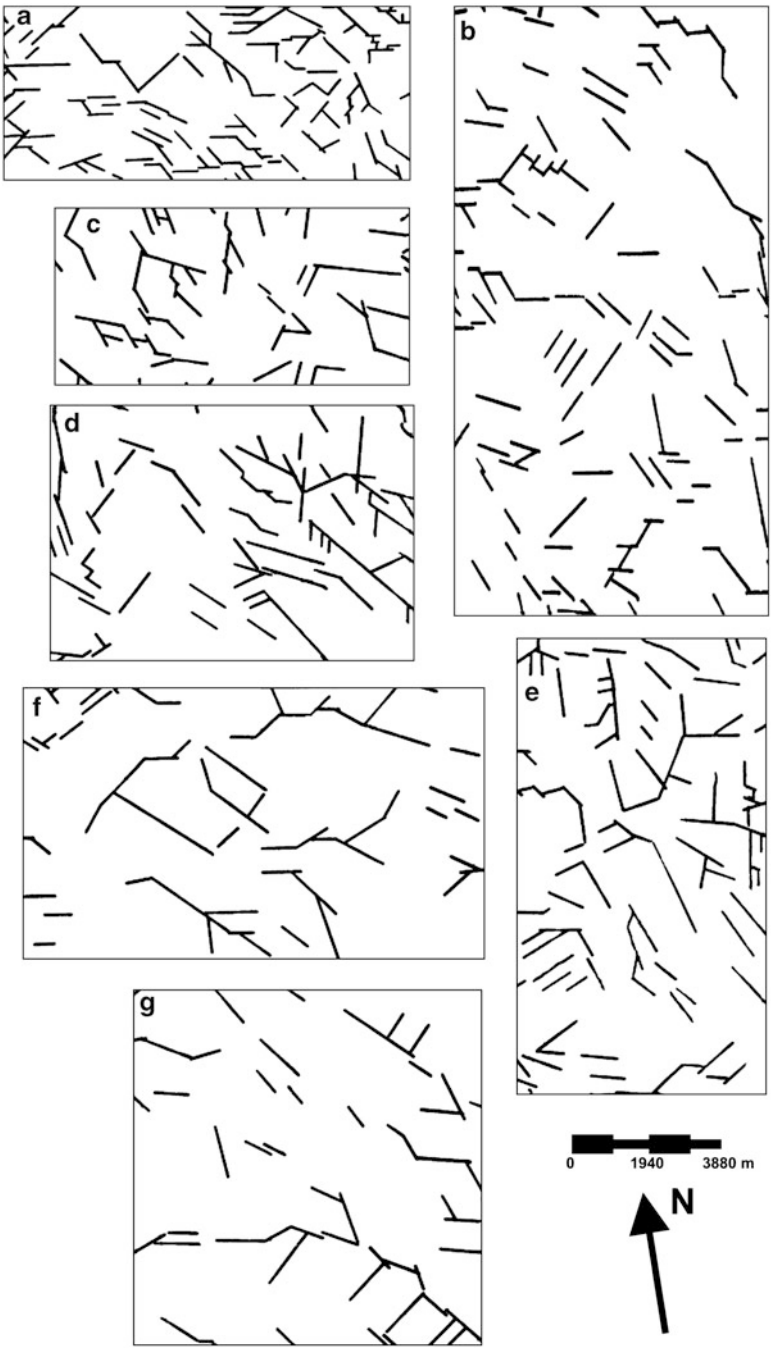

Fig. 10.2 Fracture maps in the Serra do Fundão region (GO, Brazil; scale 1:97 000), compiled from aerial photographs of reference Strider (1993). The maps are subdomains of the larger domain indicated by the box in the Fig. 10.1 and are located as arranged in the figure

10.3 The Fractal Dimension and Self-similarity Analysis

In order to perform the fractal dimension analysis we introduce in this section the adopted procedure, which if proven significant indicates the afore announced self-similarity. From the formal fractal construction point of view, a self-similar object with fractal dimension D may be divided into N smaller copies of itself scaled down by a factor r, where the number N is given in terms of the fractal dimension and the scaling factor by $N = \frac{1}{r^D}$. From the geometrical construction of the fractal the number of copies N and the ratio r are known, but the fractal dimension, that determines the scaling law, may be determined from the relation

$$D = -\frac{\log N}{\log r} .$$

Note that the construction of the fractal pattern with its self-similarity is not known a priori but shall be modeled in a progressive investigation, where the present fractal dimension analysis constitutes a first step. Already in [LiBZ05] it was pointed out that for natural objects the evidence for self-similarity may not be obtained by comparison of scales and simple counting of copies N per scale, in other words N and r, that are known in a theoretical construction of a fractal, are not directly accessible by observation, so that one has to resort to other techniques. For natural objects the dimensional Hausdorff–Besicovitch conception may be used, which may be implemented by a scaling law regression [GoMuMa98] from the box counting [Hi89], [Ba95] or the Cantor Dust method [VeEtAl90], [VeEtAl91], that yield the associated fractal dimension. It is noteworthy that each of these methods gives rise to a different fractal dimension for the same object [GiEtAl93], due to the fact that each method projects on a specific property that obeys a scaling law and in this sense they are complementary rather than contradictory. Since form observation the fracture orientation appears as a significant property in this paper the Cantor Dust method is used for the determination of the fractal dimension of the fracture maps. This dimension is a measure for the fracture density anisotropy concerning the trigonometric fracture orientation. For completeness we present in the following both methods and their resulting fractal dimensions.

The Cantor Dust method is applied on two-dimensional surfaces, i.e. the digitalized fracture maps of the Serra do Fundão region), following the reasoning of [VeEtAl90], [VeEtAl91]. To this end a C++ language program code was developed to automatically calculate $log(R)$ (length of steps) and $log(p)$ (ratio between total number of fractures per grid line intersections and total step number). The program code creates the orthogonal grid lines with varying step length (R) and trigonometric orientation to determine the fractal dimensions in different orientations. The Cantor Dust fractal dimension is then determined by the expression:

$$D = 1 - \frac{\log p}{\log R} . \tag{10.1}$$

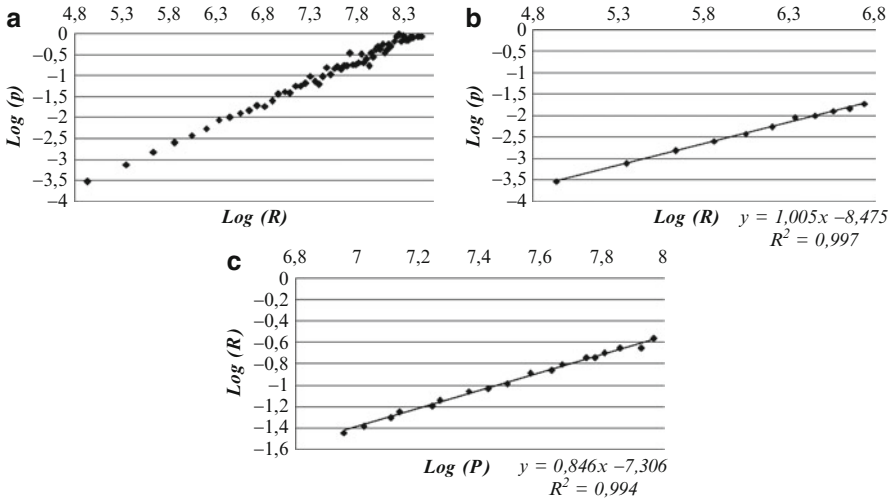

Fig. 10.3 The $log(p)$–$log(R)$ diagram and related fractal dimension by the Cantor Dust method. (**a**) Diagram for the whole range of p in a given area. (**b**) Diagram for the lowermost p values showing inclination close to unity. (**c**) Final diagram after eliminating lower and uppermost p values and showing the inclination of the linear correlation

The fractal dimension is determined from the linear correlation between $log(p)$–$log(R)$ (10.1) as shown in Fig. 10.3. In order to avoid errors in the fractal dimension due to physical limitations of the image (as granularity, for instance), lowermost p values are eliminated, when they show inclination close to unity (Fig. 10.3b); further p values larger than 0.8 are also eliminated when they show inclination close to 0 (for a detailed discussion of the cut application, see [VeEtAl91]). After application of the cuts the fractal dimension is obtained with a correlation coefficients larger than 0.97, which indicates a clear signature for a scaling law (see Fig. 10.3c).

As already announced before the Cantor Dust method is used in order to measure the anisotropy of the pattern, therefore the fracture maps are rotated sequentially by 10° counter-clockwise and for each orientation from 0 to 180° the fractal dimension is determined. A comparison between the respective orientations may be obtained using a polar plot, where for each angle the positive length $1 - D$ (i.e., the inclination of the regressions) is shown in Fig. 10.4. The anisotropy may be cast into two parameter form using the best fit with an ellipse for each fracture map by a procedure introduced in [HaFl98] and implemented with the *Matlab* program library. In the polar diagrams, the greatest and the smallest axis of the best fit ellipse indicates the directions of the higher and the lower inclinations of the $log(p)$–$log(R)$ diagrams. The measure for anisotropy is then defined by the axis ratio of the best fit ellipse. Thus, the larger the axial ratio, the larger the anisotropy of the fracture maps. Figure 10.5 shows the best fit ellipses for the set of fracture maps. Table 10.1 summarizes the found anisotropy parameters for each fracture map which may be related to the Serra do Fundão region in Fig. 10.2. According to our findings, the axis ratios range from 1.049 to 1.157.

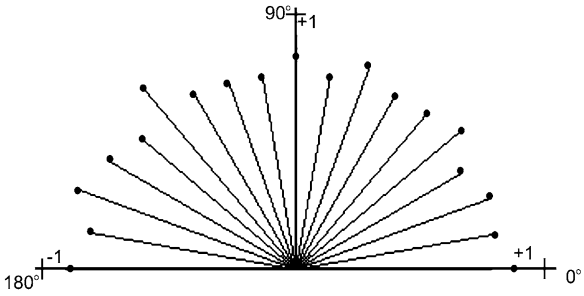

Fig. 10.4 Polar diagram of $1 - D$ for different grid orientations

The inclinations found in the $log(p)$–$log(R)$ plot from the Cantor Dust method show values in the range of 0.5 and 0.9 which corresponds to fractal dimensions between 0.5 and 0.1 for each trigonometric orientation and for each selected area. Figure 10.5 shows that the number of fractures intersecting with the reference grid is indeed anisotropic, where the ellipses of anisotropy show higher inclination if there is a larger number of fractures in other directions than the grid orientation; conversely it shows lower inclinations for a lower number of intersections. These directions are different for each fracture map area (see Fig. 10.2), where by inspection one observes differences in the preferred orientation for fractures in each area.

The second method mentioned above—the box counting method—is quite often used in order to analyze self-similarity. In the context of fracture patterns there exist applications in the literature, as, for instance, [Hi89], [AnEtAl98], [Ba95], [Ba01]. Note that by virtue box-counting captures the fracture length distribution but is not sensitive to the orientation of the pattern, which we believe to be the principal signature in the process of formation. A formal definition of the box-counting method may be found in [Fa97]. The basic procedure for the box-counting method makes use of a cover of the object by two- or three-dimensional boxes of edge length δ, where a total of N_δ boxes enclose the object completely. Upon rescaling the length to a fraction of the preceding one establishes a relation between δ and N_δ. In case of an apparent self-similarity the $log(\delta) - log(N_\delta)$ plot yields a scaling law, i.e., a linear correlation, a manifestation of a fractal dimension.

This definition is closely related to the question of the significance of the dimension determined by box-counting. The number of boxes (in the present case squares) of length δ that intersect the pattern is related to the dispersion or geometrical irregularity of the arrangement at the scale defined by δ. The numerical value of the fractal dimensions reflects the rapidity with which these irregularities evolve in the limit $\delta \to 0$. For practical applications there are limitations that have to be taken into account with respect to the maximum and minimum size of the boxes as exposed in detail in [GoMuMa98]. As an illustration we show this influence using the Koch curve as a test object with known fractal dimension $D = 1.2618$. The box-counting procedure was realized using the public domain

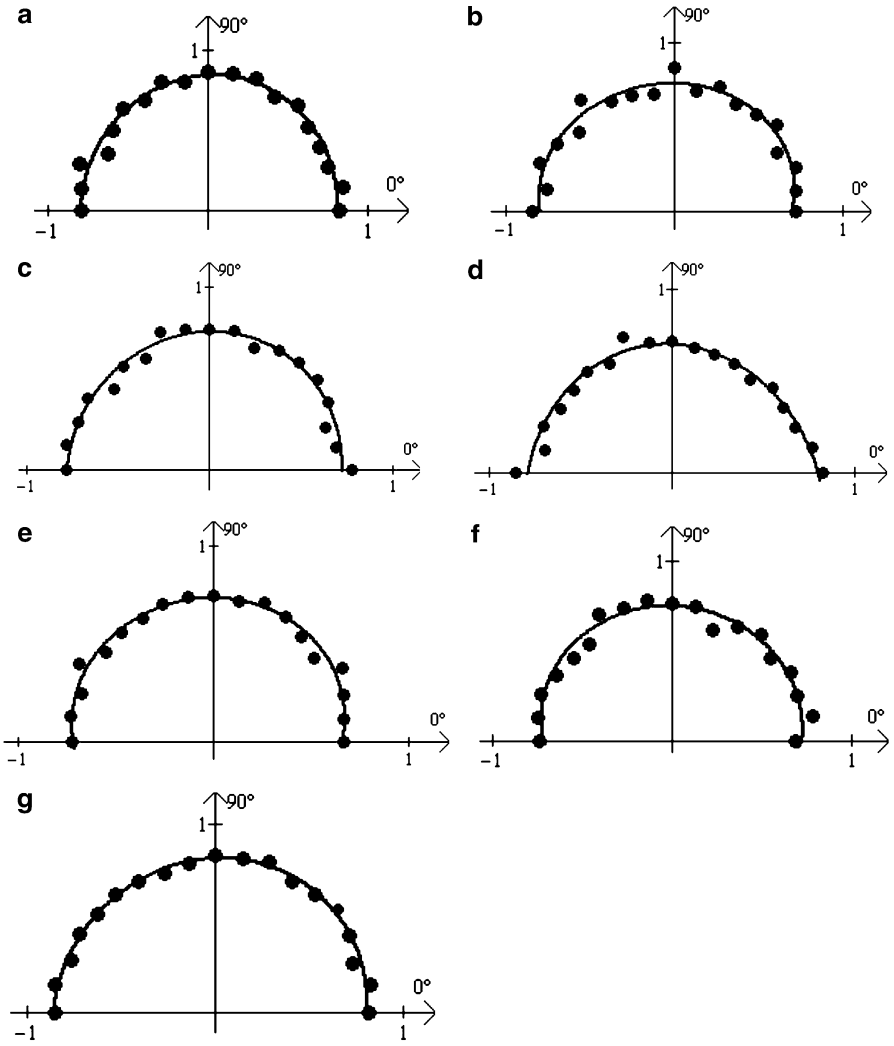

Fig. 10.5 Best fit ellipse polar plots of the fracture maps (scale 1:97 000) for the Serra do Fundão region (GO, Brazil). The values on the ellipses represent the inclination in the $log(p)$–$log(R)$ plot (**a**)–(**g**) for the respective fracture maps of the areas (**a**)–(**g**) in Fig. 10.2

software *FracAnalysis*. The fractal dimension and the box-counting method are related by the inclination

$$D = -\frac{\log N_\delta}{\log \delta} ,$$

which is illustrated in Fig. 10.6 for different maximum and minimum sizes, as well as different scale changes.

Table 10.1 Best fit ellipse parameters for fracture maps (scale 1:97 000) for the Serra do Fundão region (GO, Brazil)

Fracture map	Large axis direction	Small axis direction	Axis ratio	Ellipse origin
A	74°	164°	1.049	(−0.008; 0.012)
B	7°	97°	1.171	(−0.067; 0.108)
C	58°	148°	1.05	(−0.002; −0.029)
D	102°	12°	1.157	(0.045; −0.236)
E	7°	97°	1.117	(−0.043; 0.1064)
F	170°	80°	1.12	(0.026; 0.098)
G	43°	133°	1.06	(0.01; 0.023)

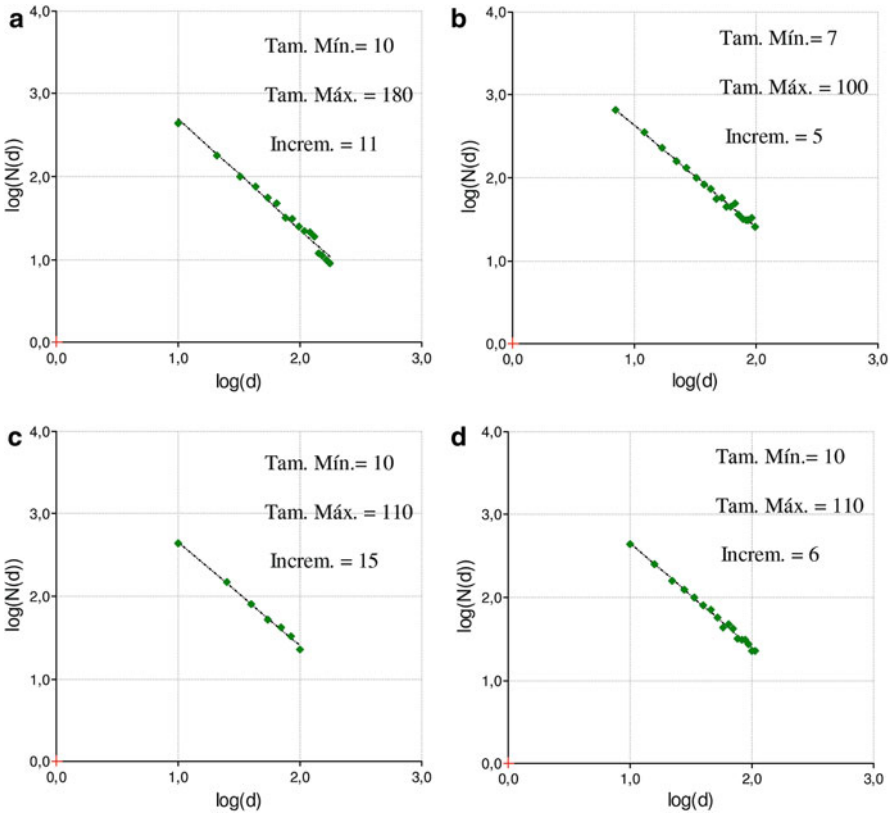

Fig. 10.6 Fractal dimension analysis of Koch's curve, for different box sizes and box rescaling. The minimum and maximum box sizes and increments are (**a**) min = 10, max =180, incr = 11; (**b**) min = 7, max = 100, incr = 5; (**c**) min = 10, max = 110, incr =15; (**d**) min = 10, max = 110, incr = 6

Table 10.2 Fractal dimension of the fracture pattern maps (scale 1:97 000) of the Serra do Fundão, (Goias, Brazil)

Map	A	B	C	D	E	F	G
Fractal dimension	1.3978	1.2608	1.2510	1.3443	1.3246	1.2450	1.1511
No. of fractures	134	128	65	77	108	59	53

The first two plots Fig. 10.6a, b are evaluated using 15 and 20 points, the non-adequateness of the maximum and minimum box sizes imposes a visible error. The third plot shows already a reasonable result for the fractal dimension, which is confirmed by the last plot where the number of box sizes was increased. The obtained result is in agreement with the theoretical expected value and also validates the program as a useful tool, once the minimum box size is not smaller than the smallest visible characteristic length of the curve, which analogously applies for the maximum box size in comparison with the largest observable characteristic length (see also [KuUmPa97]).

Application of the box-counting method to the thrusts in Japan yielded fractal dimensions between 1.05 and 1.60 [Hi89], the same method applied to fracture patterns observed in a copper mine in Arizona (USA) showed fractal dimensions between 1.34 and 1.92 [GhDa93]. Another work [Ba95] considered fracture patterns in the Yucca Mountain region (USA) and obtained numerical values between 1.12 and 1.16. The differences in the findings for the fractal dimensions are likely to be related to the different histories that caused the fracture patterns.

Further analyses are based on the digitalized fracture maps of the Serra do Fundão region (Goias, Brazil) and made use of the program *FracAanalysis*. The following procedure was adopted: The smallest admissible box size was chosen such as to be slightly larger than the smallest fracture size, whereas the largest box size was fixed in order to contain the whole pattern following the prescription in [Ba95]. The box size scaling was chosen in step/sizes such that the *log–log* plot contained between 15 and 20 points. The results for the box-counting method are shown in Table 10.2.

From the fact that the double logarithmic plots (Fig. 10.6) show a clear scaling law permits to interpret the scaling in Table 10.2 in terms of fractal dimensions that range between 1.1511 and 1.3978. Here the larger fractal dimension corresponds to the fracture map with a more dense and more complex structure, so that one may conclude that box-counting captures mainly fracture length density.

From our explanations concerning the two methods, Cantor Dust and box-counting, it becomes apparent that both methods are rather complementary and not contradictory, since they measure different properties of a specific map. As expected Cantor Dust is a method that indicates anisotropy which for a surface map has only an angular degree of freedom so that the fractal dimension shall be between 0 and 1, whereas the fractal dimensions by box-counting range from 1 to 2. Note that the same reasoning as in Cantor's Dust method was proposed by Buffon in 1777 to determine the numerical value of π but making use of non-isotropy of the method.

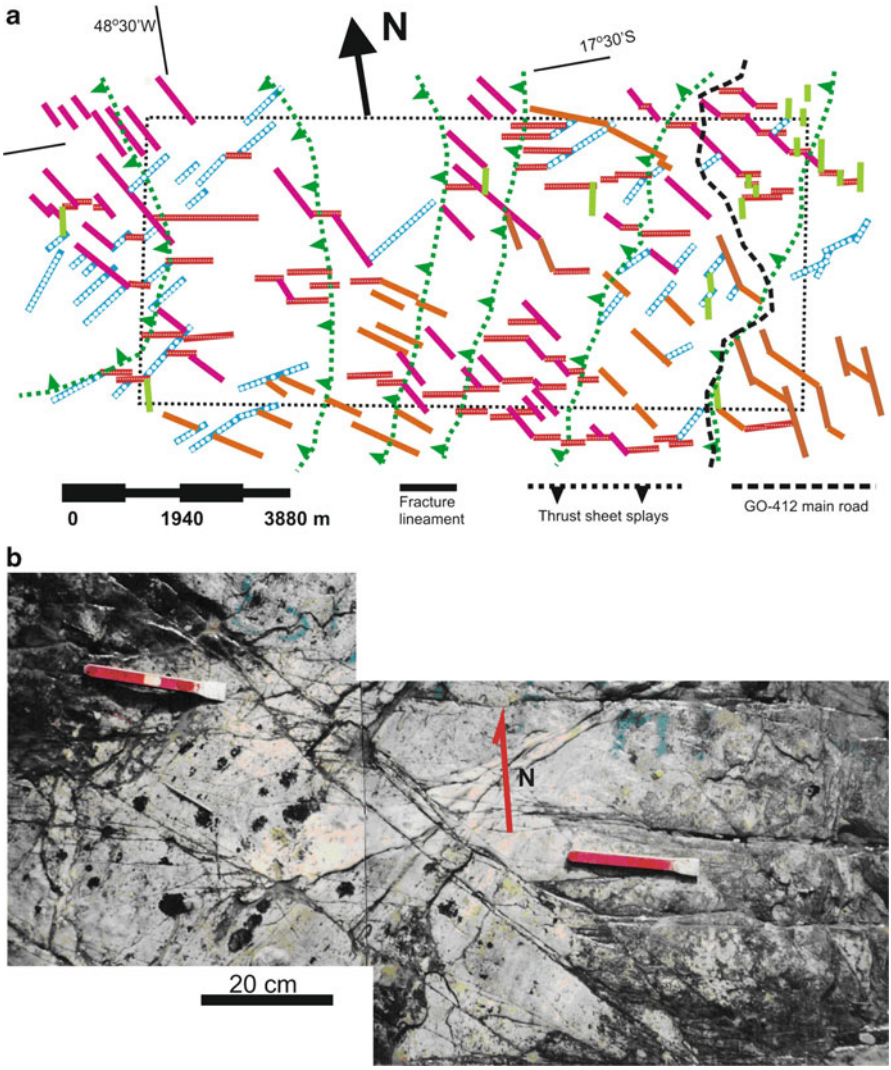

Fig. 10.7 Fracture maps for area A in the Serra do Fundão region (GO). (**a**) Fracture lineament map interpreted from satellite images and aerial photographs. (**b**) Outcrop scale fracture pattern in the selected area

Figure 10.7 shows a significant correlation of fracture alignments for certain areas, which indicates that for the further study the Cantor Dust analysis is the adequate tool for the fractal dimension analysis. Anisotropy is verified determining the inclination in the $\log(p)$–$\log(R)$ plots for a sequence of orientations as shown in the following sections.

10.4 Structural Fracture Analysis

In the last section scaling laws for fracture length density as well as anisotropy was verified using quantities that are related to the technique that allows to determine the fracture pattern associated fractal dimension. Moreover it is desirable to translate the procedure variables into those that refer directly to an observational quantity, the fracture frequency (f) and the respective length scale R (i.e., the step length). The fracture frequency (f) is a measure of the fracture quantity in a given rock mass, which depending on the consideration may be expressed in either of three ways: (a) the number of fractures per unit volume, (b) the number of fractures per unit area, and (c) the number of fractures per unit length in a given direction [TeEtAl05], where the latter is of interest for the remaining discussion.

Recalling that the linear fracture frequency can be defined by the ratio between the number of fractures that intersect a unit sample step of a given direction and the length R of this unit step, one may directly relate f, R, and p, where the latter is known to count the total number of fractures per grid line intersections and total step number (see (10.1)), so that $f = p/R$. The fractal dimension from (10.1) may be cast in to the form

$$D = -\frac{\log f}{\log R}.$$

For the purpose of simulating or constructing fracture patterns the scaling law that relates the fracture frequency and the length scale is then given by the inverse proportionality $f = R^{-D}$. An alternative interpretation of the linear fracture frequency is taking its inverse which is the fracture spacing in a given direction. The direct proportionality of fracture spacing and scaling is a manifestation of the aforementioned reasoning, that oriented fractures of one generation are the boundaries for the fractures of the subsequent scale a symmetry feature that a genuine dynamical equation for fracture pattern formation shall obey.

However, Fig. 10.5 and Table 10.1 show that fracture systems vary from one site to another in the Serra do Fundão region (GO). This anisotropic fracture distribution is due to the different fracture patterns present in each area. In fracture map for area A one identifies a strike slip duplex pattern, as can be seen in the map (Fig. 10.7a), and in the outcrop (Fig. 10.7b). The fracture lineaments map (Fig. 10.7a) can be seen as sets of straight lines, where each set may be parametrized by a linear equation $y = ax + b$, if one aligns the y-axis from south to north and the x-axis from west to east, which yields the fracture lineament in vector form $\lambda = (x, y)^T$ (see also [Pi56]). From the angular distribution, one recognizes six fracture families as shown in the rose diagram (Fig. 10.8). Each family may be characterized by an average fracture direction and a mean fracture length. The findings are given in Table 10.3.

From a comparison of Tables 10.1 and 10.4 one observes that the small axis direction for the best fit ellipse in fracture map A is close to the mean direction for fracture family 2 (Table 10.3). Fracture family 2 shows the highest fracture frequency or equivalently resulting in the lower intersection number for this direction. The highest inclination (highest f) in Fig. 10.5a is between 70 and 90°

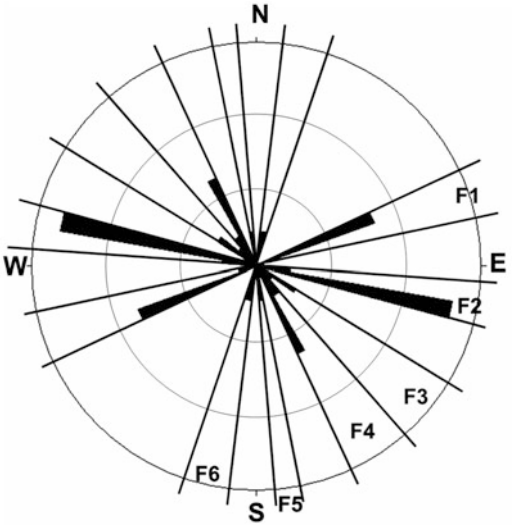

Fig. 10.8 Rose diagram for fracture lineaments in area A in the Serra do Fundão (GO) region. The outer circle represents 30% of the total (135) fracture lineaments in the map area

Table 10.3 Results for vector analysis of fracture lineaments in map A in the Serra do Fundão (GO) region. Fracture families are distinguished by the rose diagram. The geographic azimuth angles are transformed into trigonometric ones

	Fracture family range in °	Number of fractures	Mean direction°	Mean length *m*	Mean direction dispersion	Mean length variation
Total	000–180	135	120.42	390.83	3.46	36.85
Family 1	010–025	27	021.38	749.19	0.68	59.22
Family 2	165–175	42	169.05	594.99	0.23	42.47
Family 3	135–160	18	145.46	977.05	2.17	105.99
Family 4	120–135	33	122.35	682.14	0.96	52.30
Family 5	095–100	6	098.78	792.12	0.28	115.26
Family 6	075–085	9	079.97	330.83	0.35	37.75

indicating the highest number of intersections. The classification into families may now be used for simulating a fracture pattern, as shown in the next section.

10.5 Fracture Lineament Map Simulation

The fracture lineament map is then simulated according to the fractal dimension and taking into account the identified fracture lineament families, represented by their respective line equation (Table 10.4). The mean direction and length and also their deviation is then used to define an angular coefficient (a) for each fracture

Table 10.4 Equations defining each fracture lineament family present in map A in the Serra do Fundão (GO) region. Variation coefficients are presented and defined according to Table 10.3, and n is a real number

	Fracture family range in $^\circ$	Fracture lineament equation	Number of simulated fractures
Family 1	010–025	$y = (0.3914 \pm 0.011913)x \mp n7.0313$	24
Family 2	165–175	$y = \pm n1.9413$	38
Family 3	135–160	$y = (-0.6882 \pm 0.037977)x \mp n14.2278$	25
Family 4	120–135	$y = (-1.57848 \pm 0.016716)x \mp n28.9462$	20
Family 5	095–100	$y = (-6.477923 \pm 0.004911)x \mp n58.1902$	7
Family 6	075–085	$y = (5.652498 \pm 0.0061)x \mp n53.1228$	9

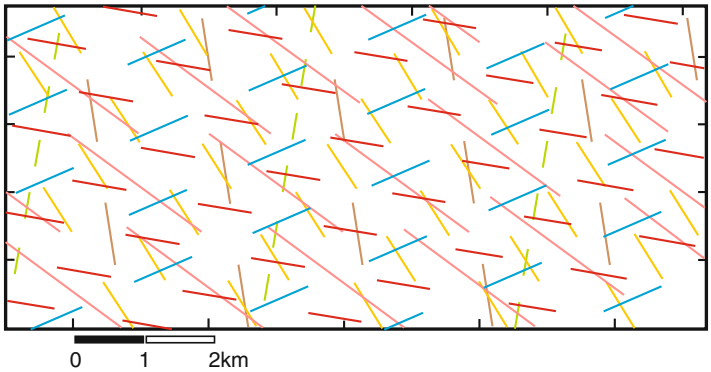

Fig. 10.9 Simulation of fracture map for area A in the Serra do Fundão region (GO, Brazil)

lineament family. Due to the finiteness of the lines it is sufficient to generate a specific position (x, y) for the whole set of lineaments of a specific family. Since such a simplification might cause that line segments overlap, such a coincidence may be reduced in the simulation making use of the constraint given by the fracture spacing $S = R^D$. Moreover, taking into account that lines of one family are parallel in the model simulation, b is related to fracture spacing by $b = R^D\sqrt{a^2 + 1}$. Thus, calculating b for each fracture family and replacing a and b in the linear equation yields the third column in Table 10.4.

The simulation was implemented in form of a *Matlab* code using the afore established rules. For each fracture family a representative was created to construct the lineament map as shown in Fig. 10.9.

10.6 Conclusion

In the present work we analyzed the fracture pattern in rocks for the Serra do Fundão region (GO, Brazil), which is the signature for the dynamics of rock evolution for that region. Our discussion showed that fractal analysis is beyond a mere

classification scheme but closely related to the dynamics by virtue of a clear self-similarity signature. We came to this conclusion by the structural geology analysis, where we distinguished different fracture sets that after parametrization was used to simulate a fracture lineaments map. These by inspection reproduced some characteristics of the original map, i.e. the fracture directional arrangement. The analytical fracture simulation took into account fracture direction and frequency, and the fractal dimension.

Those quantities are also relevant for rock qualification indexing that consider the fracture frequency in their computation. As was shown by the Cantor Dust approach, the fractal dimension can be related to the fracture frequency and fracture spacing commonly used in structural geology and rock mechanics. Cantor Dust is a one-dimensional embedding related to anisotropy (depending on the angle only) and thus the fractal dimension shall be $0 < D < 1$, whereas box-counting measures the length density distribution which implies in a fractal dimension $1 < D < 2$. From our explanation we showed that those different findings are a consequence of the procedure's complementary character, because each procedure captures a different aspect from the same phenomenon.

We are completely aware of the fact that our discussion is far from being complete. Nevertheless, the clear relation of the scaling law and the reasoning that fractures of larger scales play the role of boundaries in smaller scales and thus contemplates a more profound understanding of the fracture pattern signature. In this sense, a reproduction of our model theory could be implemented in laboratory, starting with a non-homogeneous and flat material sheet that is stretched beyond its admissible elastic limit and thus create a fracture pattern that should have a self-similar signature.

From the theoretical point of view we consider our contribution as a first step into a direction where the fracture dynamics may be understood in terms of some sort of inverse engineering, where the fractal scheme may substitute the otherwise necessary dynamical equation system with its constraints and boundary conditions in order to establish, which is known as a dynamical model. Such an approach may open pathways for a progressive modeling of fracture pattern pathology and in future hopefully for genesis of rock formation with its dependence on cooling and implications on shear stress fields. Although this may appear speculative, we at least believe to have shown with the present work that fracture analysis by self-similarity contains far more information than taxonomic ones, an impression that works known from the literature quite often transmit.

Of course it would be desirable to shed further light on some of the presented ideas that we postpone to a future work. An immediate challenge resulting from the presented discussion is the exploration of the self-similarity underlying scale invariance which we will cast into a simplified dynamical model for fracture pattern generation. In parallel, simulations will be improved implementing the combination of findings by the Cantor Dust and the box-counting method. Works in this direction may turn useful in the future once a more detailed comprehension of rock history will give further information for prospection site selections.

References

[AlEtAl81] Almeida, F.F.M., Hasui, Y., Neves, B.B.B., Fuck, R.A.: Brazilian structural provinces, an introduction. Earth Sci. Rev. **17**, 1–29 (1981)

[AnEtAl98] Angulo-Brown, F., Ramirez-Guzman, A.H., Yepez, E., Rudoif-Navarro, A., Paviamiller, C.G.: Fractal geometry and seismicity in the Mexican subduction zone. Geofísica Int. **37**, 29–33 (1998)

[At99] Atkins, A.G.: Scaling laws for elastoplastic fracture. Int. J. Fract. **95**, 51–65 (1999)

[Ba01] Babadagli, T.: Fractal analysis of 2-D fracture networks of geothermal reservoirs in south-western Turkey. J. Volcanol. Geoth. Res. **112**, 83–103 (2001)

[Ba95] Barton, C.C.: Fractal analysis of scaling and spatial clustering of fractures. In: Barton, C.C., La Pointe, P.R. (eds.) Fractals in the Earth Sciences, pp. 141–178. Plenum, New York (1995)

[Ba84] Bažant, Z.P.: Size effect in blunt fracture: concrete, rock, metal. J. Eng. Mech. **110**, 518–535 (1984)

[Ba04] Bažant, Z.P.: Scaling theory for quasi-brittle structural failure. PNAS **101**(37), 13400–13407 (2004)

[Be00] Berkowitz, B.: Scaling of fracture connectivity in geological formation. Geophys. Res. Lett. **27**, 2061–2064 (2000)

[BoEtAl01] Bonnet, E., Bour, O., Odling, N.E., Davy, P., Main, I., Cowie, P., Berkowitz, B.: Scaling of fracture systems in geological media. Rev. Geophys. **39**, 347–383 (2001)

[CaEtAl94] Cairns, D.S., Ilcewicz, L.B., Walker, T., Minguet, P.J.: Fracture scaling parameters of inhomogeneous microstructure in composite structures. J. Compos. Mater. **28**, 1598–1615 (1994)

[CaCh95] Carpinteri, A., Chiaia, B.: Multifractal nature of concrete fracture surfaces and size effects on nominal fracture energy. Mater. Struct. **28**, 435–443 (1995)

[CaChCo03] Carpinteri, A., Chiaia, B., Cornetti, P.: On the mechanics of quasi-brittle materials with a fractal microstructure. Eng. Fract. Mech. **70**, 2321–2349 (2003)

[DaEtAl06] Davy, P., Darcel, C., Bour, O., Munier, R., de Dreuzy, J.R.: Reconstructing the 3D fracture distribution model from core—10cm—to lineament—10km—scales. Geophys. Res. Abstr. **8**, 07751 (2006)

[Fa97] Falconer, K.J.: Techniques in Fractal Geometry. Wiley, New York (1997)

[GhDa93] Ghosh, A., Daemen, J.J.H.: Fractal characteristics of rock discontinuities. Eng. Geol. **34**, 1–9 (1993)

[GiEtAl93] Gillespie, P.A., Howard, C.B., Walsh, J.J., Watterson, J.: Measurement and characterisation of spatial distribution of fractures. Tectonophysics **226**, 113–141 (1993)

[GoMuMa98] Gonzato, G., Mulargia, F., Marzocchi, W.: Practical application of fractal analysis: problems and solutions. Geophys. J. Int. **132**, 275–282 (1998)

[HaFl98] Halíř, R., Flusser, J.: Numerically stable direct least squares fitting of ellipses, http://autotrace.sourceforge.net/WSCG98.pdf (1998)

[Ha85] Hancock, P.L.: Brittle microtectonics, principles and practice. J. Struct. Geol. **7**, 437–457 (1985)

[Hi89] Hirata, T.: Fractal dimension of fault systems in Japan: fractal structure in rock fracture geometry at various scales. Pure Appl. Geophys. **131**, 157–170 (1989)

[HuPr79] Hudson, J.A., Priest, S.D.: Discontinuities and rock mass geometry. Int. J. Rock Mech. Min. Sci. Geomech. Abstr. **16**, 339–362 (1979)

[KuUmPa97] Kulatilake, P.H.S.W., Um, J., Pan, G.: Requirements for accurate estimation of fractal parameters for self-affine roughness scaling method. Rock Mech. Rock Eng. **30**, 181–206 (1997)

[LaEtAl02] Laubach, S.E., Reed, R.M., Gale, J.F.W., Ortega, O.J., Doherty, E.H.: Fracture characterization based on microfracture surrogates, Pottsville Sandstone, Black Warrior Basin, Alabama. Gulf Coast Assoc. Geol. Soc. Trans. **52**, 585–596 (2002)

[LiBZ05] Libicki, E., Ben-Zion, Y.: Stochastic branching models of fault surfaces and estimated fractal dimensions. Pure Appl. Geophys. **162**, 1077–1111 (2005)
[MoBoVa02] Morel, S., Bouchaud, E., Valentin, G.: Size effect in fracture: roughening of crack surfaces and asymptotic analysis. Phys. Rev. B **65**, 104101 (2002)
[NSEtAl05] Nieto-Samaniego, A.F., Alaniz-Alvarez, S.A., Tolson, G., Oleschko, K., Korvin, G., Xu, S.S., Pérez-Venzor, A.: Spatial distribution, scaling and self-similar behavior of fracture arrays in the Los Planes Fault, Baja California Sur, Mexico. Pure Appl. Geophys. **162**, 805–826 (2005)
[OdEtAl99] Odling, N.E., Gillespie, P., Bourgine, B., Castaing, C., Chilés, J.P., Christensen, N.P., Fillion, E., Genter, A., Olsen, C., Thrane, L., Trice, R., Aarseth, E., Walsh, A.A., Watterson, J.: Variations in fracture system geometry and their implications for fluid flow in fractured hydrocarbon reservoir. Petrol. Geosci. **5**, 373–384 (1999)
[OrMaLa06] Ortega, O.J., Marrett, R.A., Laubach, S.E.: A scale-independent approach to fracture intensity and average spacing measurement. AAPG Bull. **90**, 193–208 (2006)
[Pi56] Pincus, H.J.: Some vector and arithmetic operations on two-dimensional orientation variates, with applications to geological data. J. Geol. **64**, 533–557 (1956)
[PrVM03] Prado, E.P., Van Mier, J.G.M.: Effect of particle structure on mode I fracture process in concrete. Eng. Fract. Mech. **70**, 1793–1807 (2003)
[PuEtAl01] Putot, C., Chastanet, J., Cacas, M.C., Daniel, J.M.: Fractography in sedimentary rocks: tension joint sets and fracture swarms. Rev. Institut Français du Pétrole **56**, 431–449 (2001)
[RaHu87] Ramsay, J.G., Huber, M.I.: The Techniques of Modern Structural Geology. Academic, Oxford (1987)
[St93] Strieder, A.J.: Deformação e Metamorfismo na Região de Santa Cruz de Goiás. Correlação Tectono-Estratigráfica e Evolução Tectónica Regional. Doctoral Dissertation, Institute of Geosciences-UNB, Brasília (DF), Brazil (1993)
[StSu99] Strieder, A.J., Suita, M.T.F.: Neoproterozoic geotectonic evolution of Tocantins Structural Province, Central Brazil. J. Geodyn. **28**, 267–289 (1999)
[TeEtAl05] Telles, I.A., Vargas, E.A. Jr., Lira, W.W.M., Martha, L.F.: Uma Ferramenta Computacional para a Geração de Sistemas de Fraturas em Meios Rochosos, http://www.tecgraf.pucrio.br/publications/artigo_2005_ferramenta_computacional_geracao_sistemas_fraturas.pdf (2005)
[Tu97] Turcotte, D.L.: Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge (1997)
[VeEtAl90] Velde, B., Dubois, J., Touchard, G., Badri, A.: Fractal analysis of fractures in rocks: the Cantor's Dust method. Tectonophysics **179**, 345–352 (1990)
[VeEtAl91] Velde, B., Dubois, J., Moore, D., Touchard, G.: Fractal patterns of fractures in granites. Earth Planet. Sci. Lett. **104**, 25–35 (1991)
[VoKr04] Vollanda, S., Kruhl, J.H.: Anisotropy quantification: the application of fractal geometry methods on tectonic fracture patterns of a Hercynian fault zone in NW Sardinia. J. Struct. Geol. **26**, 1499–1510 (2004)

Chapter 11
Spectral Boundary Homogenization Problems in Perforated Domains with Robin Boundary Conditions and Large Parameters

D. Gómez, M.E. Pérez, and T.A. Shaposhnikova

11.1 Introduction and Formulation of the Problem

Let Ω be a bounded domain in $\mathbb{R}^3$, with a smooth boundary $\partial\Omega$. We assume that $\gamma = \Omega \cap \{x_1 = 0\} \neq \emptyset$ is a domain on the plane $\{x_1 = 0\}$. We denote by G_0 the ball of radius 1 centered at the origin of coordinates. For a domain B, and for $\delta > 0$, we denote by $\delta B = \{x \,|\, \delta^{-1}x \in B\}$. We set

$$\tilde{G}_\varepsilon = \bigcup_{z\in\mathbb{Z}'} (a_\varepsilon G_0 + \varepsilon z) = \bigcup_{j\in\mathbb{Z}'} G_\varepsilon^j,$$

where $\mathbb{Z}'$ is the set of points of the form $z = (0, z_2, z_3)$ with integer components z_2, z_3; $a_\varepsilon = C_0\varepsilon^\alpha$, C_0 is a positive number, ε is a small positive parameter that we shall make converge towards zero, and α is a parameter, $\alpha \geq 1$. We define

$$G_\varepsilon = \bigcup_{j\in\Upsilon_\varepsilon} G_\varepsilon^j, \quad \text{where} \quad \Upsilon_\varepsilon = \{j \in \mathbb{Z}' : G_\varepsilon^j \subset \tilde{G}_\varepsilon, \overline{G}_\varepsilon^j \subset \Omega, \rho(\partial\Omega, \overline{G}_\varepsilon^j) \geq 2\varepsilon\}.$$

The number of G_ε^j with index $j \in \Upsilon_\varepsilon$ is $|\Upsilon_\varepsilon| = O(\varepsilon^{-2})$.

In what follows, we set

$$\Omega_\varepsilon = \Omega \setminus \overline{G}_\varepsilon, S_\varepsilon = \partial G_\varepsilon, \partial\Omega_\varepsilon = \partial\Omega \cup S_\varepsilon.$$

D. Gómez
Dpto. Matemáticas, Estadísitica y Computación, Universidad de Cantabria, Santander, 39005 Spain
e-mail: gomezdel@unican.es

M.E. Pérez (✉)
Dpto. Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander, 39005 Spain
e-mail: meperez@unican.es

T.A. Shaposhnikova
Department of Differential Equations, Moscow State University, Moscow, 119992 Russia
e-mail: shaposh.tan@mail.ru

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_11,

We consider the space $H^1(\Omega_\varepsilon, \partial\Omega)$ to be the completion with respect to the norm $H^1(\Omega_\varepsilon)$ of the set of functions $u \in \mathscr{C}^\infty(\overline{\Omega}_\varepsilon)$, u vanishing in a neighborhood of $\partial\Omega$.

Let us consider the eigenvalue problem

$$\begin{cases} -\Delta u^\varepsilon = \lambda^\varepsilon u^\varepsilon & \text{in } \Omega_\varepsilon, \\ u^\varepsilon = 0 & \text{on } \partial\Omega, \\ \dfrac{\partial u^\varepsilon}{\partial \nu} + \varepsilon^{-\kappa} a u^\varepsilon = 0 & \text{on } S_\varepsilon, \end{cases} \tag{11.1}$$

where ν denotes the unit outward normal vector ν to $\partial\Omega_\varepsilon$ on S_ε, $a \equiv a(x)$ is a strictly positive continuously differentiable function in $\overline{\Omega}$ and κ is any real parameter.

The variational formulation of (11.1) is: find λ^ε, $u^\varepsilon \in H^1(\Omega_\varepsilon, \partial\Omega)$, $u^\varepsilon \neq 0$, such that

$$\int_{\Omega_\varepsilon} \nabla u^\varepsilon \nabla v \, dx + \varepsilon^{-\kappa} \int_{S_\varepsilon} a u^\varepsilon v \, ds = \lambda^\varepsilon \int_{\Omega_\varepsilon} u^\varepsilon v \, dx, \qquad \forall v \in H^1(\Omega_\varepsilon, \partial\Omega). \tag{11.2}$$

For each fixed $\varepsilon > 0$, problem (11.2) is a standard spectral problem in the couple of spaces $H^1(\Omega_\varepsilon, \partial\Omega) \subset L^2(\Omega_\varepsilon)$, with a discrete spectrum. Let us consider

$$\lambda_1^\varepsilon \leq \lambda_2^\varepsilon \leq \cdots \leq \lambda_k^\varepsilon \leq \cdots \xrightarrow{k\to\infty} \infty \tag{11.3}$$

the sequence of its eigenvalues repeated according to their multiplicities. Let us consider $\{u_k^\varepsilon\}_{k=1}^\infty$ the set of associated eigenfunctions which form an orthonormal basis in $L^2(\Omega_\varepsilon)$.

The convergence of the spectrum of (11.2) towards that of the homogenized problem has been proved in [GoPeSh12]. The homogenized problem depends on the different values/relations of the parameters κ and α. For the sake of completeness, we gather in Theorem 1 below the results obtained in Sect. 9 of [GoPeSh12] along with the corresponding homogenized problems, namely, problems (11.5)–(11.9).

As is well known, problem (11.5) ((11.6), (11.7), (11.8) and (11.9), respectively), has a discrete spectrum; let us consider

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots \xrightarrow{k\to\infty} \infty \tag{11.4}$$

the sequence of its eigenvalues repeated according to their multiplicities and we denote by $\{u_k\}_{k=1}^\infty$ the set of associated eigenfunctions which form an orthonormal basis in $L^2(\Omega)$.

Theorem 1. *For each fixed k, $k = 1, 2, 3 \ldots$, λ_k^ε in (11.3) converge towards λ_k in (11.4) when $\varepsilon \to 0$, where $\{\lambda_k\}_{k=1}^\infty$ are the eigenvalues of*

- *the Dirichlet problem*

$$-\Delta u = \lambda u \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{11.5}$$

when $\alpha \geq 1$ and $\kappa < 2(\alpha - 1)$ or $\alpha > 2$ and $\kappa \in \mathbb{R}$;

- *the Dirichlet problem in* $\Omega^- \cup \Omega^+$

$$-\Delta u = \lambda u \text{ in } \Omega^- \cup \Omega^+, \quad u = 0 \text{ on } \partial\Omega \cup \gamma, \tag{11.6}$$

when $\alpha \in [1,2)$ *and* $\kappa > 2(\alpha - 1)$;
- *the problem*

$$\begin{cases} -\Delta u = \lambda u & \text{in } \Omega^- \cup \Omega^+, \\ u = 0 & \text{on } \partial\Omega, \\ [u] = 0, \quad \left[\dfrac{\partial u}{\partial x_1}\right] = 4\pi C_0 u & \text{on } \gamma, \end{cases} \tag{11.7}$$

when $\alpha = 2$ *and* $\kappa > 2$, *where the brackets mean the jump across* γ;
- *the problem*

$$\begin{cases} -\Delta u = \lambda u & \text{in } \Omega^- \cup \Omega^+, \\ u = 0 & \text{on } \partial\Omega, \\ [u] = 0, \quad \left[\dfrac{\partial u}{\partial x_1}\right] = 4\pi C_0^2 a u & \text{on } \gamma, \end{cases} \tag{11.8}$$

when $\alpha \in [1,2)$ *and* $\kappa = 2(\alpha - 1)$;
- *the problem*

$$\begin{cases} -\Delta u = \lambda u & \text{in } \Omega^- \cup \Omega^+, \\ u = 0 & \text{on } \partial\Omega, \\ [u] = 0, \quad \left[\dfrac{\partial u}{\partial x_1}\right] = 4\pi C_0 h u & \text{on } \gamma, \end{cases} \tag{11.9}$$

when $\alpha = \kappa = 2$, *where* $h \equiv h(x)$ *is the strictly positive continuously differentiable function defined by*

$$h(x) = \frac{a(x)C_0}{a(x)C_0 + 1}, \quad x \in \overline{\Omega}. \tag{11.10}$$

This result does not provide bounds for convergence rates of eigenvalues and the associated eigenfunctions, since it is obtained from general convergence results for nonlinear stationary problems, and convergence rates for the solutions of these stationary problems rely on the usual assumption of smoothness of the solution of the limiting problem. Since we are dealing with eigenvalue problems such an assumption makes no sense (see Remark 3 in [GoPeSh12] in this connection).

The aim of this paper is to obtain precise bounds for discrepancies of the eigenvalues and the associated eigenfunctions in terms of the eigenvalue number and the parameter ε. We emphasize that obtaining these bounds proves to be

essential in determining, e.g., the time in terms of ε in which certain solutions of the associated evolution problems can be approached through time-dependent functions constructed from the homogenized problem (see [Pe08] and [Pe11]). Associated evolution problems arise, e.g., in Ecology: see [GoPeSh12] for further references on the model and related works in the literature.

For the proofs, we use a result from the spectral perturbation theory (Lemma 4) for ε-dependent Hilbert spaces and operators, which provides convergence for the spectrum when convergence of the associated stationary problems is known. Since we are dealing with a linear problem, in this paper we obtain a certain smoothness for the solution of the stationary problem (11.18) (cf. Lemma 5). Consequently, avoiding the assumptions on smoothness of solutions in [GoPeSh12] we obtain lower order powers or ε in the bounds for the discrepancies, but in these bounds we can control the dependence on the data f in the norm of $L^2(\Omega)$ (cf. (11.23), (11.24) and (11.29)), which is a usual topology for the spectral problems here considered. To prove the above-mentioned smoothness, we use a variant of results on interior estimates of Sobolev norms for solutions of second order elliptic equations with Dirichlet boundary conditions (cf. [Mo66] and [Sh08]), and Sobolev embedding theorems which also imply some restriction on the dimension of the space under consideration.

Section 11.2 contains some preliminary results on the solutions of the stationary problems (11.16) and (11.18) used to prove the convergence in the rest of the paper. Section 11.3 contains the convergence results for the stationary problems (cf. Theorem 2) and the spectral convergence (cf. Theorem 3) for the case $\alpha = 2$ and $\kappa > 2$. Section 11.4 contains the corresponding convergence results for the case $\alpha \in [1,2)$ and $\kappa = 2(\alpha - 1)$. Both cases provide a critical relation between the parameters α and κ. These critical relations amount to a critical size of the cavities for the different values of the parameter κ (in which the dimension of the space is also involved), implying a nontrivial average on the transmission condition for the flux throughout γ in the way outlined in Fig. 11.1. The critical case where $\alpha = \kappa = 2$ is considered in [GoEtAl12]: the average on γ contains a nonlinear dependence on $a(x)$ (cf. problem (11.9)). Nevertheless, we outline that certain proofs of results in [GoEtAl12] are further developed in this paper (cf., e.g., Lemmas 3 and 5). In addition, the proof of the convergence in Theorem 1 in [GoPeSh12] has been performed only for the critical case where $\alpha = \kappa = 2$; here, we perform in detail the proof for the other critical cases (cf. problems (11.7) and (11.8)) and, at the same time, we provide bounds for convergence rates of the eigenelements. For brevity, we avoid proofs for the rest of the cases (cf. problems (11.5) and (11.6)) but we also provide the above-mentioned convergence rates.

We emphasize that the spectral problems here considered differ from others in the literature. As a matter of fact, asymptotics for the eigenvalues of (11.1) for the case of one single hole has been considered in, e.g., [Ro93]; for a spatial distribution of the holes periodically distributed let us mention [OlSh95b]; similarly, for a two-dimensional domain and $\kappa \in (0,1)$ we refer to [Oz96]. For the geometrical

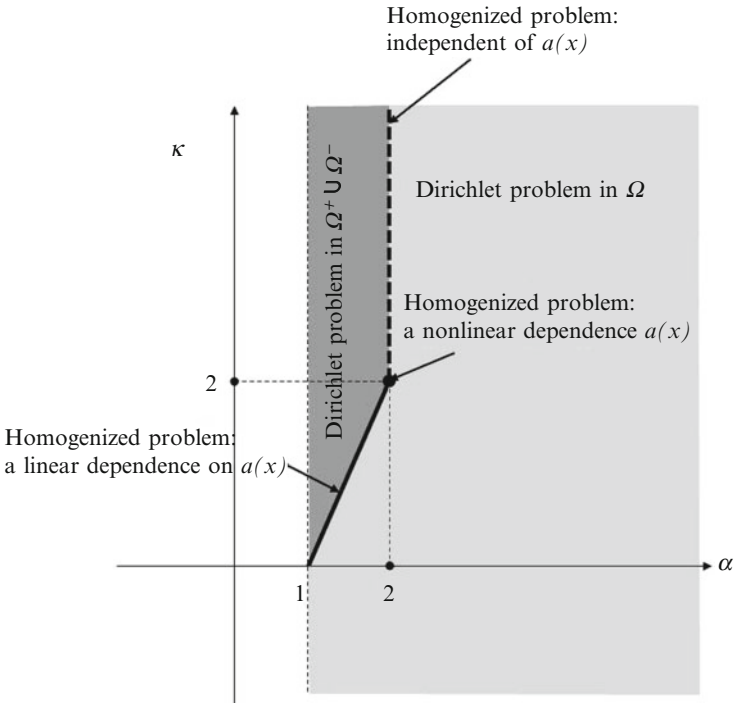

Fig. 11.1 Sketch of spectral homogenized problems depending on α and κ

configuration here considered and $\kappa = 0$ we mention [LoEtAl97]. Let us refer to the above-mentioned papers for further references on spectral problems in perforated domains.

11.2 Preliminary Results

Let us introduce some notation and results which will prove to be useful for the rest of the paper. In the lemmas below, and in what follows, C denotes a constant independent of ε. Also, in these lemmas, the constant C does not depend on the functions w appearing in their statements. See Lemma 1 in [OlSh95a] and Lemma 1 in [LoEtAl97] for the proof of Lemma 1 and 2, respectively.

Lemma 1. *There exists an operator $\mathscr{P}_\varepsilon$ from $H^1(\Omega_\varepsilon, \partial\Omega)$ into $H_0^1(\Omega)$, such that for $w \in H^1(\Omega_\varepsilon, \partial\Omega)$ we set $\mathscr{P}_\varepsilon w = \tilde{w}$ the function which satisfies: $\tilde{w}(x) = w(x)$ for $x \in \Omega_\varepsilon$, and*

$$\|\tilde{w}\|_{H^1(\Omega)} \leq C\|w\|_{H^1(\Omega_\varepsilon)} \quad \textit{and} \quad \|\nabla\tilde{w}\|_{L^2(\Omega)} \leq C\|\nabla w\|_{L^2(\Omega_\varepsilon)}. \tag{11.11}$$

Lemma 2. *Let P_ε^j be the center of the ball G_ε^j and let T_ε^j denote the ball of radius $b_0\varepsilon$ with center P_ε^j $(j \in \Upsilon_\varepsilon)$; $0 < b_0 < 1$. Then,*

$$\Big| \sum_{j\in\Upsilon_\varepsilon} \int_{\partial T_\varepsilon^j} w\,ds - 4\pi b_0^2 \int_\gamma w\,d\hat{x} \Big| \le C\varepsilon^{1/2} \|w\|_{H^1(\Omega)}, \quad w \in H_0^1(\Omega).$$

Here, and in what follows, $\hat{x}$ denotes $\hat{x} = (x_2, x_3)$.

Lemma 3. *Let Π_ε be $\Pi_\varepsilon = \Omega \cap \{-\varepsilon/2 < x_1 < \varepsilon/2\}$. Then, for all $w \in H_0^1(\Omega)$,*

$$\|w\|_{L^2(\Pi_\varepsilon)} \le C\varepsilon^{1/2} \|\nabla w\|_{L^2(\Omega)}, \tag{11.12}$$

$$\Big| \frac{1}{\varepsilon} \int_{\Pi_\varepsilon} w\,dx - \int_\gamma w\,d\hat{x} \Big| \le C\varepsilon^{1/2} \|\nabla w\|_{L^2(\Omega)} \tag{11.13}$$

and

$$\|w\|_{L^4(\Pi_\varepsilon)} \le C\varepsilon^{1/8} \|\nabla w\|_{L^2(\Omega)}. \tag{11.14}$$

Proof. We refer to Lemma 2.4 in Sect. II.3 of [MaKh74] for the proof of (11.12) and (11.13). Besides, we observe that (11.14) can be obtained as a consequence of the Hölder inequality, namely, from:

$$\|w\|_{L^q(\Pi_\varepsilon)} \le \|w\|^{\lambda}_{L^p(\Pi_\varepsilon)} \|w\|^{1-\lambda}_{L^r(\Pi_\varepsilon)} \qquad \text{for } w \in L^r(\Pi_\varepsilon)$$

where $p \le q \le r$ and $1/q = \lambda/p + (1-\lambda)/r$, $0 < \lambda < 1$ (cf. Sect. 7.1 in [GiTr01], for example). Now, taking $p = 2$, $q = 4$, $r = 6$ and $\lambda = 1/4$, we deduce that $\|w\|_{L^4(\Pi_\varepsilon)} \le \|w\|^{1/4}_{L^2(\Pi_\varepsilon)} \|w\|^{3/4}_{L^6(\Pi_\varepsilon)}$, and, by (11.12) and the embedding of $H_0^1(\Omega)$ into $L^6(\Omega)$, we obtain (11.14).

For the proofs of Theorems 3, 5, 7, 9, 11, and 13, we use the technique in Sect. III.4 of [OlShYo92] (cf. also [LoEtAl97] and [OlSh95a] for further references). For the sake of completeness, we introduce a general strong result from the spectral perturbation theory: see Theorems 1.4 and 1.7 in Chap. III of [OlShYo92] for its proof.

Lemma 4. *Let H_ε and H_0 be two separable Hilbert spaces with the scalar products $(\cdot,\cdot)_\varepsilon$ and $(\cdot,\cdot)_0$, respectively. Let $A^\varepsilon \in \mathscr{L}(H_\varepsilon)$ and $A^0 \in \mathscr{L}(H_0)$. Let $\mathscr{W}$ be a subspace of H_0 such that $Im A^0 = \{v \,|\, v = A^0 u \,:\, u \in H_0\} \subset \mathscr{W}$. We assume that the following properties are satisfied:*

(C1) *There exists an operator $\mathscr{R}^\varepsilon \in \mathscr{L}(H_0, H_\varepsilon)$ and a constant $\mathrm{a} > 0$ such that, for any $f \in \mathscr{W}$, $\|\mathscr{R}^\varepsilon f\|_\varepsilon$ converge towards $\mathrm{a}\|f\|_0$ as $\varepsilon \to 0$.*

(*C*2) *A^ε and A^0 are positive, compact, and self-adjoint operators on H_ε and H_0, respectively. Besides, the norms $\|A^\varepsilon\|_{\mathscr{L}(H_\varepsilon)}$ are bounded by a constant independent of ε.*

(*C*3) *For any $f \in \mathscr{W}$, $\|A^\varepsilon \mathscr{R}^\varepsilon f - \mathscr{R}^\varepsilon A^0 f\|_\varepsilon \to 0$ as $\varepsilon \to 0$.*

(*C*4) *The family of operators A^ε is uniformly compact, i.e., for any sequence f^ε in H_ε such that $\sup_\varepsilon \|f^\varepsilon\|_\varepsilon$ is bounded by a constant independent of ε, we can extract a subsequence $f^{\varepsilon'}$ verifying $\|A^{\varepsilon'} f^{\varepsilon'} - \mathscr{R}^{\varepsilon'} w^0\|_{\varepsilon'} \to 0$, as $\varepsilon' \to 0$, for certain $w^0 \in \mathscr{W}$.*

Let $\{\mu_i^\varepsilon\}_{i=1}^\infty$ ($\{\mu_i^0\}_{i=1}^\infty$, respectively) be the sequence of the eigenvalues of A^ε (A^0, respectively) with the usual convention of repeated eigenvalues. Let $\{w_i^\varepsilon\}_{i=1}^\infty$ and ($\{w_i^0\}_{i=1}^\infty$, respectively) be the corresponding eigenfunctions which are assumed to be an orthonormal basis in H_ε (H_0, respectively).

Then, for each fixed k there exists a constant C_k independent of ε and there is $\varepsilon_k > 0$ such that for $\varepsilon \le \varepsilon_k$,

$$|\mu_k^\varepsilon - \mu_k^0| \le C_k \sup \|A^\varepsilon \mathscr{R}^\varepsilon u - \mathscr{R}^\varepsilon A^0 u\|_\varepsilon$$

where the sup *is taken over all the functions u in the eigenspace associated with μ_k^0, u such that $\|u\|_0 = 1$. In addition, for any eigenvalue μ_k^0 of A^0 with multiplicity s ($\mu_k^0 = \mu_{k+1}^0 = \cdots = \mu_{k+s-1}^0$), and for any w eigenfunction corresponding to μ_k^0, with $\|w\|_0 = 1$, there exists w^ε, w^ε being a linear combination of eigenfunctions $\{w_j^\varepsilon\}_{j=k}^{j=k+s-1}$ of A^ε corresponding to $\{\mu_j^\varepsilon\}_{j=k}^{j=k+s-1}$, such that*

$$\|w^\varepsilon - \mathscr{R}^\varepsilon w\|_\varepsilon \le C_k \|A^\varepsilon \mathscr{R}^\varepsilon w - \mathscr{R}^\varepsilon A^0 w\|_\varepsilon, \tag{11.15}$$

for a certain constant C_k independent of ε.

For $f \in L^2(\Omega)$, we consider $u_\varepsilon \in H^1(\Omega_\varepsilon, \partial\Omega)$ the solution of

$$\int_{\Omega_\varepsilon} \nabla u_\varepsilon \nabla v \, dx + \varepsilon^{-\kappa} \int_{S_\varepsilon} a u_\varepsilon v \, ds = \int_{\Omega_\varepsilon} f v \, dx, \qquad \forall v \in H^1(\Omega_\varepsilon, \partial\Omega). \tag{11.16}$$

It satisfies the following estimates (see Lemma 2.7 and Theorem 2.1 in [GoPeSh12]):

$$\begin{aligned} \|\nabla u_\varepsilon\|_{L^2(\Omega_\varepsilon)} + \varepsilon^{-\kappa/2} \|u_\varepsilon\|_{L^2(S_\varepsilon)} &\le C \|f\|_{L^2(\Omega_\varepsilon)}, \\ \|\tilde{u}_\varepsilon\|_{H^1(\Omega)} &\le C \|f\|_{L^2(\Omega_\varepsilon)}. \end{aligned} \tag{11.17}$$

Let $b \equiv b(x)$ be a strictly positive continuously differentiable function in $\overline{\Omega}$. For $f \in L^2(\Omega)$, we consider $u \in H_0^1(\Omega)$ the solution of

$$\int_\Omega \nabla u \nabla v \, dx + \int_\gamma b u v \, d\hat{x} = \int_\Omega f v dx, \quad \forall v \in H_0^1(\Omega). \tag{11.18}$$

Lemma 5. *Let u be the solution of (11.18) with $f \in L^2(\Omega)$. Then,*

$$\|u\|_{H^1(\Omega)} \leq C\|f\|_{L^2(\Omega)}, \tag{11.19}$$

$$\|u\|_{W^{1,4}(\Omega)} \leq C\|f\|_{L^2(\Omega)} \tag{11.20}$$

and

$$\|u\|_{L^\infty(\Omega)} \leq C\|f\|_{L^2(\Omega)}. \tag{11.21}$$

Proof. Taking $v = u$ in (11.18) and using Poincaré inequality we get (11.19).

In order to show (11.20), we consider the function $\psi(x) = u(x)\exp(g(x))$ where g is defined by

$$g(x) = \begin{cases} -b(0,\hat{x})x_1 & \text{if } x_1 > 0 \\ 0 & \text{if } x_1 \leq 0. \end{cases}$$

Thus, ψ is a solution of

$$\begin{cases} -\partial_{x_i}(a^{ij}\partial_{x_j}\psi + b^i\psi) = f \text{ in } \Omega \\ \psi = 0 \qquad\qquad\qquad\quad \text{on } \partial\Omega, \end{cases}$$

with

$$a^{ij}(x) = \exp(-g(x)), \quad b^i(x) = -\exp(-g(x))\partial_{x_i}g(x), \quad i,j = 1,2,3.$$

By the smoothness of b, we can show that $a^{ij} = a^{ji}$ are bounded from below and from above by strictly positive constants, $a^{ij} \in C^{0,1}(\Omega)$ and $b^i \in L^\infty(\Omega)$. Then, on account of the Dirichlet condition for ψ on the boundary, we use an adaptation of the proof in Theorem 1 in [Sh08] for $n = 3$, $q = 4$, and $p = 12/7$. This can be summarized as follows: by means of local maps, locally, the problem for ψ can be transformed into another problem posed in a domain of $\mathbb{R}^3$ with a Dirichlet condition on a plane. Then, a suitable extension of the transformed solution gives the solution of a problem within the framework of Theorem 1 in [Sh08], and we can apply both the interior $W^{1,4}$ smoothness of the solution in Theorem 5.5.4' of Sect. V.5 in [Mo66] and the interior Sobolev estimates in Theorem 1 of [Sh08]. Consequently, $\psi \in W_0^{1,4}(\Omega)$ and we can write

$$\|\psi\|_{W^{1,4}(\Omega)} \leq C(\|\psi\|_{L^1(\Omega)} + \|f\|_{L^{12/7}(\Omega)}).$$

Now, by definition of ψ, the smoothness of b, the embedding of $L^r(\Omega)$ into $L^s(\Omega)$ for $1 \leq s \leq r \leq \infty$ and (11.19), we obtain

$$\|u\|_{W^{1,4}(\Omega)} \le C\|\psi\|_{W^{1,4}(\Omega)} \le C(\|u\|_{L^1(\Omega)} + \|f\|_{L^{12/7}(\Omega)}) \le C\|f\|_{L^2(\Omega)},$$

and (11.20) holds.

Finally, (11.21) can be obtained directly from (11.20) and the embedding of $W^{1,4}(\Omega)$ into $L^\infty(\Omega)$.

In order to prove the convergence, we introduce the test function W_ε. Let P_ε^j be the center of the ball G_ε^j and we denote by T_ε^j the ball of radius $\varepsilon/4$ with center P_ε^j. Let us consider the functions w_ε^j $(j \in \Upsilon_\varepsilon)$ as the solutions of the following problems

$$\begin{cases} \Delta w_\varepsilon^j = 0 \text{ in } T_\varepsilon^j \setminus \overline{G_\varepsilon^j}, \\ w_\varepsilon^j = 1 \quad \text{on } \partial G_\varepsilon^j, \\ w_\varepsilon^j = 0 \quad \text{on } \partial T_\varepsilon^j. \end{cases} \tag{11.22}$$

We define the function $W_\varepsilon \in H^1(\mathbb{R}^3)$ by extending by 1 for $x \in \overline{G_\varepsilon}$ and by 0 for $x \in \mathbb{R}^3 \setminus \bigcup_{j\in\Upsilon_\varepsilon} T_\varepsilon^j$. As a matter of fact, for $\alpha = 2$ w_ε^j, used in Sect. 11.3, reads

$$w_\varepsilon^j(x) = \frac{|x - P_\varepsilon^j|^{-1} - (\varepsilon/4)^{-1}}{(C_0\varepsilon^2)^{-1} - (\varepsilon/4)^{-1}},$$

$0 \le W_\varepsilon \le 1$, and the weak convergence $W_\varepsilon \rightharpoonup 0$ in $H_0^1(\Omega)$, as $\varepsilon \to 0$, holds.

11.3 Convergence Results for $\alpha = 2$ and $\kappa > 2$

Theorem 2. *Let $\alpha = 2$, $\kappa > 2$, and $f \in L^2(\Omega)$. Let W_ε be the function defined by (11.22). Let u_ε be the solution of (11.16) and u the solution of (11.18) with $b(x) = 4\pi C_0$. Then, we have*

$$\|u_\varepsilon - u + W_\varepsilon u\|_{H^1(\Omega_\varepsilon)}^2 + \varepsilon^{-\kappa}\|u_\varepsilon\|_{L^2(S_\varepsilon)}^2 \le C\varepsilon^{\min(1/8,(\kappa-2)/2)}\|f\|_{L^2(\Omega)}^2, \tag{11.23}$$

$$\|u_\varepsilon - u\|_{L^2(\Omega_\varepsilon)}^2 \le C\varepsilon^{\min(1/8,(\kappa-2)/2)}\|f\|_{L^2(\Omega)}^2. \tag{11.24}$$

Proof. Let us consider (11.16) and (11.18) with $v = u_\varepsilon - u + W_\varepsilon u \in H^1(\Omega_\varepsilon, \partial\Omega)$ and $v = \tilde{u}_\varepsilon - u + W_\varepsilon u \in H_0^1(\Omega)$ as test functions, respectively. Subtracting both equalities and taking into account that $W_\varepsilon = 1$ in $\overline{G_\varepsilon}$, we obtain

$$\|\nabla(u_\varepsilon - u + W_\varepsilon u)\|_{L^2(\Omega_\varepsilon)}^2 + \varepsilon^{-\kappa}\int_{S_\varepsilon} a\,u_\varepsilon^2\,ds = I_1 + I_2 + I_3$$

where

$$I_1 = \int_{G_\varepsilon} \nabla u \nabla \tilde{u}_\varepsilon dx,$$

$$I_2 = -\int_{G_\varepsilon} f \tilde{u}_\varepsilon dx,$$

$$I_3 = \int_{\Omega_\varepsilon} \nabla(W_\varepsilon u)\nabla(u_\varepsilon - u + W_\varepsilon u)dx + 4\pi C_0 \int_\gamma u(\tilde{u}_\varepsilon - u + W_\varepsilon u)d\hat{x}.$$

Now, considering the volume of G_ε, (11.20), (11.17), and (11.12), we obtain

$$|I_1| \le \|\nabla u\|_{L^4(G_\varepsilon)} |G_\varepsilon|^{1/4} \|\nabla \tilde{u}_\varepsilon\|_{L^2(\Omega)} \le C\varepsilon \|f\|^2_{L^2(\Omega)},$$

$$|I_2| \le \|f\|_{L^2(G_\varepsilon)} \|\tilde{u}_\varepsilon\|_{L^2(\Pi_\varepsilon)} \le C\|f\|_{L^2(G_\varepsilon)} \varepsilon^{1/2} \|\nabla \tilde{u}_\varepsilon\|_{L^2(\Omega)} \le C\varepsilon^{1/2} \|f\|^2_{L^2(\Omega)}.$$

Let us estimate I_3. Using

$$\int_{\Omega_\varepsilon} \nabla(W_\varepsilon u)\nabla w dx = \int_{\Omega_\varepsilon} \nabla W_\varepsilon \nabla(uw)dx - \int_{\Omega_\varepsilon} \nabla W_\varepsilon \nabla u w dx + \int_{\Omega_\varepsilon} W_\varepsilon \nabla u \nabla w dx$$

for $w = u_\varepsilon - u + W_\varepsilon u$, the Green formula for the first integral on the right-hand side above, and the definition of W_ε we have that $I_3 = I_{3a} + I_{3b} + I_{3c}$, where

$$I_{3a} = \sum_{j\in\Upsilon_\varepsilon} \int_{\partial T_\varepsilon^j} \partial_\nu w_\varepsilon^j u(u_\varepsilon - u + W_\varepsilon u)ds + 4\pi C_0 \int_\gamma u(\tilde{u}_\varepsilon - u + W_\varepsilon u)d\hat{x},$$

$$I_{3b} = \sum_{j\in\Upsilon_\varepsilon} \int_{\partial G_\varepsilon^j} \partial_\nu w_\varepsilon^j u(u_\varepsilon - u + W_\varepsilon u)ds \quad \text{and}$$

$$I_{3c} = -\int_{\Omega_\varepsilon} \nabla W_\varepsilon \nabla u\,(u_\varepsilon - u + W_\varepsilon u)dx + \int_{\Omega_\varepsilon} W_\varepsilon \nabla u \nabla(u_\varepsilon - u + W_\varepsilon u)dx.$$

Taking into account the explicit computation of the normal derivatives of w_ε^j, Lemma 2 and the trace theorem in $H^1(\Omega)$, we have

$$|I_{3a}| \le \left|\frac{C_0 2^4}{1 - 4C_0\varepsilon}\right| \left| \sum_{j\in\Upsilon_\varepsilon} \int_{\partial T_\varepsilon^j} u(u_\varepsilon - u + W_\varepsilon u)\,ds - \frac{\pi}{4}\int_\gamma u(\tilde{u}_\varepsilon - u + W_\varepsilon u)\,d\hat{x}\right|$$

$$+\Big|\frac{4\pi C_0}{1-4C_0\varepsilon}-4\pi C_0\Big|\Big|\int_\gamma u(\tilde{u}_\varepsilon-u+W_\varepsilon u)\,d\hat{x}\Big|$$
$$\leq\ C\varepsilon^{1/2}\|\nabla(u(\tilde{u}_\varepsilon-u+W_\varepsilon u))\|_{L^2(\Omega)}.$$

Then, from the embedding of $H_0^1(\Omega)$ into $L^6(\Omega)$, the boundedness of W_ε in $H^1(\Omega)$, (11.17), (11.19), (11.20), and (11.21), it follows that

$$\begin{aligned}|I_{3a}|\ &\leq C\varepsilon^{1/2}(\|\nabla u\|_{L^4(\Omega)}\|\tilde{u}_\varepsilon-u+W_\varepsilon u\|_{L^4(\Omega)}+\|u\|_{L^\infty(\Omega)}\|\nabla(\tilde{u}_\varepsilon-u+W_\varepsilon u)\|_{L^2(\Omega)})\\ &\leq C\varepsilon^{1/2}\|f\|^2_{L^2(\Omega)}.\end{aligned}$$

Besides, by the explicit form of the normal derivatives of w_ε^j and the definition of W_ε, we can rewrite I_{3b} as

$$I_{3b}=\frac{1}{\varepsilon^2C_0(1-4C_0\varepsilon)}\int_{S_\varepsilon}uu_\varepsilon ds.$$

Thus, computing the area of S_ε and using (11.21) and (11.17) we get

$$|I_{3b}|\leq C\varepsilon^{-2}\|u\|_{L^\infty(\Omega)}|S_\varepsilon|^{1/2}\|u_\varepsilon\|_{L^2(S_\varepsilon)}\leq C\varepsilon^{(\kappa-2)/2}\|f\|^2_{L^2(\Omega)}.$$

In a similar way,

$$\begin{aligned}|I_{3c}|\leq\ &\|\nabla W_\varepsilon\|_{L^2(\Omega)}\|\nabla u\|_{L^4(\Pi_\varepsilon)}\|\tilde{u}_\varepsilon-u+W_\varepsilon u\|_{L^4(\Pi_\varepsilon)}\\ &+|\Pi_\varepsilon|^{1/4}\|\nabla u\|_{L^4(\Pi_\varepsilon)}\|\nabla(\tilde{u}_\varepsilon-u+W_\varepsilon u)\|_{L^2(\Pi_\varepsilon)},\end{aligned}$$

and by the boundedness of W_ε in $H^1(\Omega)$, (11.20), (11.19), (11.17), and (11.14) we get $|I_{3c}|\leq C\varepsilon^{1/8}\|f\|^2_{L^2(\Omega)}$.

Now, gathering all the above estimates, we conclude that

$$\|\nabla(u_\varepsilon-u+W_\varepsilon u)\|^2_{L^2(\Omega_\varepsilon)}+\varepsilon^{-\kappa}\|u_\varepsilon\|^2_{L^2(S_\varepsilon)}\leq C\varepsilon^{\min(1/8,(\kappa-2)/2)}\|f\|^2_{L^2(\Omega)}. \quad (11.25)$$

To obtain (11.23) from (11.25), we consider the Poincaré inequality for the H^1-extension of $u_\varepsilon-u+W_\varepsilon u$ to Ω given by Lemma 1, $\mathscr{P}_\varepsilon(u_\varepsilon-u+W_\varepsilon u)\in H_0^1(\Omega)$, which satisfies (11.11) for $w=u_\varepsilon-u+W_\varepsilon u$.

Finally, from (11.23), the definition of W_ε, (11.12), and (11.19), we can write

$$\begin{aligned}\|u_\varepsilon-u\|^2_{L^2(\Omega_\varepsilon)}\ &\leq\|u_\varepsilon-u+W_\varepsilon u\|^2_{L^2(\Omega_\varepsilon)}+\|W_\varepsilon u\|^2_{L^2(\Omega_\varepsilon)}\\ &\leq C(\varepsilon^{\min(1/8,(\kappa-2)/2)}\|f\|^2_{L^2(\Omega)}+\|u\|^2_{L^2(\Pi_\varepsilon)})\end{aligned}$$

$$\leq C(\varepsilon^{\min(1/8,(\kappa-2)/2)}\|f\|^2_{L^2(\Omega)} + \varepsilon\|\nabla u\|^2_{L^2(\Omega)})$$
$$\leq C\varepsilon^{\min(1/8,(\kappa-2)/2)}\|f\|^2_{L^2(\Omega)}.$$

Consequently, (11.24) holds and the theorem is proved.

Theorem 3. *Let $\alpha = 2$ and $\kappa > 2$. Let $\{\lambda_k^\varepsilon\}_{k=1}^\infty$ and $\{\lambda_k\}_{k=1}^\infty$ be the eigenvalues of problem (11.1) and (11.7), respectively. Then, for each fixed k there exists a constant C_k independent of ε such that, for sufficiently small $\varepsilon > 0$,*

$$|\lambda_k^\varepsilon - \lambda_k|^2 \leq C_k\varepsilon^{\min(1/8,(\kappa-2)/2)}. \tag{11.26}$$

Moreover, for any eigenvalue λ_k of (11.7) with multiplicity s ($\lambda_k = \lambda_{k+1} = \cdots = \lambda_{k+s-1}$), and for any u eigenfunction corresponding to λ_k, with $\|u\|_{L^2(\Omega)} = 1$, there exists $\tilde{u}^\varepsilon$, $\tilde{u}^\varepsilon$ being a linear combination of eigenfunctions $\{u_k^\varepsilon\}_{r=k}^{r=k+s-1}$ of (11.1) corresponding to $\{\lambda_k^\varepsilon\}_{r=k}^{r=k+s-1}$, such that

$$\|\tilde{u}^\varepsilon - u\|^2_{L^2(\Omega_\varepsilon)} \leq C_k\varepsilon^{\min(1/8,(\kappa-2)/2)}. \tag{11.27}$$

Proof. Let us define $\mathscr{H}^\varepsilon = L^2(\Omega_\varepsilon)$, $\mathscr{H}^0 = L^2(\Omega)$ with the usual scalar products. Let us introduce the operators $\mathscr{A}^\varepsilon : \mathscr{H}^\varepsilon \to \mathscr{H}^\varepsilon$, $\mathscr{A}^0 : \mathscr{H}^0 \to \mathscr{H}^0$. For $f^\varepsilon \in \mathscr{H}^\varepsilon$, we set $\mathscr{A}^\varepsilon f^\varepsilon = u_\varepsilon$ where $u_\varepsilon \in H^1(\Omega_\varepsilon, \partial\Omega)$ is the unique solution of (11.16). Consequently, the eigenelements of $\mathscr{A}^\varepsilon$ are $\{((\lambda_k^\varepsilon)^{-1}, u_k^\varepsilon)\}_{k=1}^\infty$ with $\{(\lambda_k^\varepsilon, u_k^\varepsilon)\}_{k=1}^\infty$ the eigenelements of (11.2). In the same way, for $f \in \mathscr{H}^0$, we set $\mathscr{A}^0 f = u$ where $u \in H_0^1(\Omega)$ is the unique solution of (11.18) for $b(x) = 4\pi C_0$. Consequently, the eigenelements of $\mathscr{A}^0$ are $\{((\lambda_k)^{-1}, u_k)\}_{k=1}^\infty$ with $\{(\lambda_k, u_k)\}_{k=1}^\infty$ the eigenelements of (11.7).

We define $\mathscr{R}^\varepsilon : L^2(\Omega) \to L^2(\Omega_\varepsilon)$, as the restriction operator, namely, $(\mathscr{R}^\varepsilon f)(x) = f(x)$ if $x \in \Omega_\varepsilon$. We also define $\mathscr{W} = H_0^1(\Omega)$ which contains the space $Im(\mathscr{A}^0)$.

On account of (11.17) and (11.24), it is self-evident that the properties $(C1)$–$(C3)$ of Lemma 4 are satisfied. Let us prove property $(C4)$ in Lemma 4. In order to do this, for the $f^\varepsilon \in L^2(\Omega_\varepsilon)$, as stated in property $(C4)$, we consider $\hat{f}^\varepsilon \in L^2(\Omega)$ the extension of f^ε by zero in G_ε. We have that $\|\hat{f}^\varepsilon\|_{L^2(\Omega)}$ is bounded by a constant independent of ε and consequently, there is a subsequence $\varepsilon' \to 0$ and a certain $f^0 \in L^2(\Omega)$ such that $\hat{f}^{\varepsilon'} \rightharpoonup f^0$ in $L^2(\Omega)$. Considering $u_{\varepsilon'} = \mathscr{A}^{\varepsilon'}\mathscr{R}^{\varepsilon'}\hat{f}^{\varepsilon'}$ and $w^0 \in H_0^1(\Omega)$ solution of (11.18) for $b(x) = 4\pi C_0$ and $f = f^0$, we rewrite the proof in Theorem 2 with minor modifications, and we obtain that $\|u_{\varepsilon'} - w^0\|_{L^2(\Omega_\varepsilon)} \to 0$, as $\varepsilon' \to 0$. Consequently, property $(C4)$ also holds.

Now, applying Lemma 4, we have that for each fixed k,

$$\left|(\lambda_k^\varepsilon)^{-1} - (\lambda_k)^{-1}\right| \leq C_k \sup\|u_{\varepsilon,k} - u_{0,k}\|_{L^2(\Omega_\varepsilon)} \tag{11.28}$$

where the supremum is taken over all the functions f_k in the eigenspace associated with $(\lambda_k)^{-1}$, f_k are such that $\|f_k\|_{L^2(\Omega)} = 1$, $u_{\varepsilon,k} = \mathscr{A}^\varepsilon \mathscr{R}^\varepsilon f_k$, and $u_{0,k} = \mathscr{R}^\varepsilon \mathscr{A}^0 f_k$. But (11.24) allows us to assert that

$$\left\|u_{\varepsilon,k} - u_{0,k}\right\|^2_{L^2(\Omega_\varepsilon)} \leq C_k \varepsilon^{\min(1/8,(\kappa-2)/2)} \|f_k\|^2_{L^2(\Omega)} \leq C_k \varepsilon^{\min(1/8,(\kappa-2)/2)}$$

for a certain constant C_k independent of ε. From this last inequality, (11.28) reads $|(\lambda_k^\varepsilon)^{-1} - (\lambda_k)^{-1}|^2 \leq C_k \varepsilon^{\min(1/8,(\kappa-2)/2)}$ which ensures the boundedness of $(\lambda_k^\varepsilon)^{-1}$ by a constant independent of ε and consequently the estimate for the eigenvalues (11.26) holds.

Finally, let us note that the estimate for the eigenfunctions (11.27) also holds applying (11.15) and (11.24).

11.4 Convergence Results for $\alpha \in [1,2)$ and $\kappa = 2(\alpha - 1)$

Theorem 4. *Let $\alpha \in [1,2)$, $\kappa = 2(\alpha-1)$ and $f \in L^2(\Omega)$. Let u_ε be the solution of (11.16) and u the solution of (11.18) with $b(x) = 4\pi C_0^2 a(x)$. Then, we have*

$$\|u_\varepsilon - u\|^2_{H^1(\Omega_\varepsilon)} + \varepsilon^{-\kappa} \|u_\varepsilon - u\|^2_{L^2(S_\varepsilon)} \leq C\varepsilon^q \|f\|^2_{L^2(\Omega)} \tag{11.29}$$

where $q = 1/4$ if $\alpha = 1$ and $q = \min\{(3\alpha-2)/4, 3(\alpha-1)/2, (2-\alpha)/2\}$ if $\alpha \in (1,2)$.

Proof. Let us consider formulas (11.16) and (11.18) with $v = u_\varepsilon - u \in H^1(\Omega_\varepsilon, \partial\Omega)$ and $v = \tilde{u}_\varepsilon - u \in H_0^1(\Omega)$ as test functions, respectively. Subtracting both equalities, we obtain

$$\|\nabla(u_\varepsilon - u)\|^2_{L^2(\Omega_\varepsilon)} + \varepsilon^{-\kappa} \int_{S_\varepsilon} a\,(u_\varepsilon - u)^2\, ds = R_1 + R_2 + R_3$$

where

$$R_1 = \int_{G_\varepsilon} \nabla u \nabla(\tilde{u}_\varepsilon - u) dx, \qquad R_2 = -\int_{G_\varepsilon} f(\tilde{u}_\varepsilon - u) dx \qquad \text{and}$$

$$R_3 = 4\pi C_0^2 \int_\gamma a\, u(\tilde{u}_\varepsilon - u) d\hat{x} - \varepsilon^{-\kappa} \int_{S_\varepsilon} au(u_\varepsilon - u)\, ds.$$

Now, considering the volume of G_ε, (11.19), (11.20), (11.17), and (11.12), we obtain

$$|R_1| \leq \|\nabla u\|_{L^4(G_\varepsilon)} |G_\varepsilon|^{1/4} \|\nabla(\tilde{u}_\varepsilon - u)\|_{L^2(\Omega)} \leq C\varepsilon^{(3\alpha-2)/4} \|f\|^2_{L^2(\Omega)}, \tag{11.30}$$

$$|R_2| \le \|f\|_{L^2(G_\varepsilon)} \varepsilon^{1/2} \|\nabla(\tilde{u}_\varepsilon - u)\|_{L^2(\Omega)} \le C\varepsilon^{1/2} \|f\|^2_{L^2(\Omega)}. \tag{11.31}$$

Let us estimate R_3. First, let us note that if $\alpha = 1$, then $\kappa = 0$ and by Lemma 2 $|R_3| \le C\varepsilon^{1/2} \|\nabla(a\, u(\tilde{u}_\varepsilon - u))\|_{L^2(\Omega)}$. But, from the smoothness of a, we can write

$$\begin{aligned} \|\nabla(a\, u(\tilde{u}_\varepsilon - u))\|_{L^2(\Omega)} &\le C\|\nabla(u(\tilde{u}_\varepsilon - u))\|_{L^2(\Omega)} \\ &\le C[\|u\|_{L^\infty(\Omega)} \|\nabla(\tilde{u}_\varepsilon - u)\|_{L^2(\Omega)} \\ &\quad + \|\nabla u\|_{L^4(\Omega)} \|\tilde{u}_\varepsilon - u\|_{L^4(\Omega)}]. \end{aligned}$$

Thus, using (11.21), (11.17), (11.19), (11.20), and the embeddings of $L^r(\Omega)$ into $L^s(\Omega)$ for $1 \le s \le r \le \infty$ and $H^1_0(\Omega)$ into $L^6(\Omega)$, we conclude

$$\|\nabla(a\, u(\tilde{u}_\varepsilon - u))\|_{L^2(\Omega)} \le C\|f\|^2_{L^2(\Omega)} \tag{11.32}$$

and

$$|R_3| \le C\varepsilon^{1/2} \|f\|^2_{L^2(\Omega)} \quad \text{for } \alpha = 1. \tag{11.33}$$

We assume that $\alpha \in (1,2)$ and write $Y_\varepsilon = \sum_{j \in \Upsilon_\varepsilon} (\varepsilon Y + \varepsilon j) \setminus \overline{G}^j_\varepsilon = \sum_{j \in \Upsilon_\varepsilon} Y^j_\varepsilon \setminus \overline{G}^j_\varepsilon$, where $Y = (-1/2, 1/2)^3$. We introduce the function $\theta_\varepsilon(x)$ as a solution of the problem

$$\begin{cases} \Delta \theta_\varepsilon = \mu_\varepsilon & x \in \varepsilon Y \setminus a_\varepsilon \overline{G_0}, \\ \partial_\nu \theta_\varepsilon = -1 & x \in \partial(a_\varepsilon G_0), \\ \partial_\nu \theta_\varepsilon = 0 & x \in \partial(\varepsilon Y) \setminus \partial(a_\varepsilon G_0), \end{cases}$$

where $\mu_\varepsilon = -\dfrac{4\pi C_0^2 \varepsilon^{2(\alpha-1)-1}}{1 - (a_\varepsilon \varepsilon^{-1})^3 |G_0|}$. We assume that $\int\limits_{\varepsilon Y \setminus a_\varepsilon G_0} \theta_\varepsilon dx = 0$. Then, by rewriting the computation in Sect. 1.4 of [OlSh96] with minor modifications, (cf. also estimate (36) and Lemmas 1 and 2 in [OlSh96]), we deduce that

$$\|\nabla \theta_\varepsilon\|^2_{L^2(Y_\varepsilon)} \le C\varepsilon^{3\alpha-2}. \tag{11.34}$$

We denote by $\theta^j_\varepsilon(x)$ the solution of the problem posed in $Y^j_\varepsilon \setminus G^j_\varepsilon$.

By means of θ_ε, the integral on S_ε can be transformed into a volume integral. Thus, we can write

$$\int\limits_{S_\varepsilon} a\, u(\tilde{u}_\varepsilon - u) ds = -\sum_{j \in \Upsilon_\varepsilon} \int\limits_{Y^j_\varepsilon \setminus G^j_\varepsilon} \nabla \theta^j_\varepsilon \nabla(a\, u(\tilde{u}_\varepsilon - u)) dx - \mu_\varepsilon \sum_{j \in \Upsilon_\varepsilon} \int\limits_{Y^j_\varepsilon \setminus G^j_\varepsilon} a\, u(\tilde{u}_\varepsilon - u) dx$$

and $R_3 = R_{3a} + R_{3b} + R_{3c}$, where

$$R_{3a} = (4\pi C_0^2 + \varepsilon^{1-\kappa}\mu_\varepsilon) \int_\gamma a\, u(\tilde{u}_\varepsilon - u)\, d\hat{x},$$

$$R_{3b} = \varepsilon^{-\kappa} \sum_{j\in\Upsilon_\varepsilon} \int_{Y_\varepsilon^j \setminus G_\varepsilon^j} \nabla\theta_\varepsilon^j \nabla(a\, u(\tilde{u}_\varepsilon - u))\, dx$$

and

$$R_{3c} = \varepsilon^{1-\kappa}\mu_\varepsilon \Big[\frac{1}{\varepsilon} \sum_{j\in\Upsilon_\varepsilon} \int_{Y_\varepsilon^j \setminus G_\varepsilon^j} a\, u(\tilde{u}_\varepsilon - u)\, dx - \int_\gamma a\, u(\tilde{u}_\varepsilon - u)\, d\hat{x}\Big].$$

By definition of μ_ε, the trace theorem in $H^1(\Omega)$, (11.32), and (11.34), it follows that

$$|R_{3a}| \le C(a_\varepsilon \varepsilon^{-1})^3 \|\nabla(au(\tilde{u}^\varepsilon - u))\|_{L^2(\Omega)} \le C\varepsilon^{3(\alpha-1)} \|f\|^2_{L^2(\Omega)},$$
$$|R_{3b}| \le C\varepsilon^{(3\alpha-2)/2-\kappa} \|\nabla(au(\tilde{u}^\varepsilon - u))\|_{L^2(\Omega)} \le C\varepsilon^{(2-\alpha)/2} \|f\|^2_{L^2(\Omega)}.$$

In order to estimate R_{3c}, let us define $\tilde{\Pi}_\varepsilon$ as $\tilde{\Pi}_\varepsilon = \Pi_\varepsilon \setminus \bigcup_{j\in\Upsilon_\varepsilon} Y_\varepsilon^j$. Then

$$\begin{aligned}
&\frac{1}{\varepsilon} \sum_{j\in\Upsilon_\varepsilon} \int_{Y_\varepsilon^j \setminus G_\varepsilon^j} a\, u(\tilde{u}_\varepsilon - u) dx - \int_\gamma a\, u(\tilde{u}_\varepsilon - u) d\hat{x} \\
&\qquad = \frac{1}{\varepsilon} \int_{\Pi_\varepsilon} a\, u(\tilde{u}_\varepsilon - u) dx - \int_\gamma a\, u(\tilde{u}_\varepsilon - u) d\hat{x} \\
&\qquad\qquad - \frac{1}{\varepsilon} \int_{G_\varepsilon} a\, u(\tilde{u}_\varepsilon - u) dx - \frac{1}{\varepsilon} \int_{\tilde{\Pi}_\varepsilon} a\, u(\tilde{u}_\varepsilon - u) dx. \qquad (11.35)
\end{aligned}$$

The two first terms on the right-hand side of (11.35) can be estimated directly by (11.13) and (11.32):

$$\begin{aligned}
\Big|\frac{1}{\varepsilon} \int_{\Pi_\varepsilon} a\, u(\tilde{u}_\varepsilon - u) dx - \int_\gamma a\, u(\tilde{u}_\varepsilon - u) d\hat{x}\Big| &\le C\varepsilon^{1/2} \|\nabla(au(\tilde{u}_\varepsilon - u))\|_{L^2(\Omega)} \\
&\le C\varepsilon^{1/2} \|f\|^2_{L^2(\Omega)}.
\end{aligned}$$

Besides, using (11.12), (11.32), $|G_\varepsilon| \le C\varepsilon^{3\alpha-2}$ and $|\tilde{\Pi}_\varepsilon| \le C\varepsilon^2$, we obtain

$$\left|\frac{1}{\varepsilon}\int_{G_\varepsilon} a\,u(\tilde{u}_\varepsilon - u)dx\right| \le C\varepsilon^{-1}|G_\varepsilon|^{1/2}\|a\,u(\tilde{u}_\varepsilon - u)\|_{L^2(\Pi_\varepsilon)} \le C\varepsilon^{3(\alpha-1)/2}\|f\|^2_{L^2(\Omega)},$$

$$\left|\frac{1}{\varepsilon}\int_{\tilde{\Pi}_\varepsilon} a\,u(\tilde{u}_\varepsilon - u)dx\right| \le C\varepsilon^{-1}|\tilde{\Pi}_\varepsilon|^{1/2}\|a\,u(\tilde{u}_\varepsilon - u)\|_{L^2(\Pi_\varepsilon)} \le C\varepsilon^{1/2}\|f\|^2_{L^2(\Omega)}.$$

Thus, by definition of R_{3c} and μ_ε, $|R_{3c}| \le C\varepsilon^{\min(1/2,3(\alpha-1)/2)}\|f\|^2_{L^2(\Omega)}$.

Gathering the above estimates, we get

$$|R_3| \le C\varepsilon^{\min((2-\alpha)/2,3(\alpha-1)/2)}\|f\|^2_{L^2(\Omega)} \quad \text{for } \alpha \in (1,2). \tag{11.36}$$

Now, by (11.30), (11.31), (11.33), and (11.36), we deduce that

$$\|\nabla(u_\varepsilon - u)\|^2_{L^2(\Omega_\varepsilon)} + \varepsilon^{-\kappa}\|u_\varepsilon\|^2_{L^2(S_\varepsilon)} \le C\varepsilon^q\|f\|^2_{L^2(\Omega)}, \tag{11.37}$$

where $q = 1/4$ if $\alpha = 1$ and $q = \min((3\alpha-2)/4, 3(\alpha-1)/2, (2-\alpha)/2)$ if $\alpha \in (1,2)$.

To obtain (11.29) from (11.37), we consider the Poincaré inequality for the H^1-extension of $u_\varepsilon - u$ to Ω given by Lemma 1, $\mathscr{P}_\varepsilon(u_\varepsilon - u) \in H^1_0(\Omega)$, which satisfies (11.11) for $w = u_\varepsilon - u$ and the theorem is proved.

Theorem 5. *Let $\alpha \in [1,2)$ and $\kappa = 2(\alpha-1)$. Let $\{\lambda^\varepsilon_k\}_{k=1}^\infty$ and $\{\lambda_k\}_{k=1}^\infty$ be the eigenvalues of problem (11.1) and (11.8), respectively. Then, for each fixed k there exists a constant C_k independent of ε such that, for sufficiently small $\varepsilon > 0$,*

$$|\lambda^\varepsilon_k - \lambda_k|^2 \le C_k\varepsilon^q,$$

where $q = 1/4$ if $\alpha = 1$ and $q = \min((3\alpha-2)/4, 3(\alpha-1)/2, (2-\alpha)/2)$ if $\alpha \in (1,2)$. Moreover, for any eigenvalue λ_k of (11.8) with multiplicity s ($\lambda_k = \lambda_{k+1} = \cdots = \lambda_{k+s-1}$), and for any u eigenfunction corresponding to λ_k, with $\|u\|_{L^2(\Omega)} = 1$, there exists $\tilde{u}^\varepsilon$, $\tilde{u}^\varepsilon$ being a linear combination of eigenfunctions $\{u^\varepsilon_k\}_{r=k}^{r=k+s-1}$ of (11.1) corresponding to $\{\lambda^\varepsilon_k\}_{r=k}^{r=k+s-1}$, such that

$$\|\tilde{u}^\varepsilon - u\|^2_{L^2(\Omega_\varepsilon)} \le C_k\varepsilon^q.$$

Proof. By rewriting the reasoning in proof of Theorem 3 with minor modifications, Theorem 5 is proved. Now, in order to apply Lemma 4, for $f \in \mathscr{H}^0 = L^2(\Omega)$, we set $\mathscr{A}^0 f = u$ where $u \in H^1_0(\Omega)$ is the unique solution of (11.18) for $b(x) = 4\pi C_0^2 a(x)$. Consequently, the eigenelements of $\mathscr{A}^0$ are $\{((\lambda_k)^{-1}, u_k)\}_{k=1}^\infty$ with $\{(\lambda_k, u_k)\}_{k=1}^\infty$ the eigenelements of (11.8). Besides, on account of (11.17) and (11.29), we can check

that the properties $(C1)$–$(C4)$ of Lemma 4 are satisfied. Thus, similar arguments to the proof of Theorem 3 allow us to prove Theorem 5.

11.5 Bounds for Other Values of α and κ

Let us note that the technique here introduced can be extended to the rest of the values of α and κ for the dimension 3 of the space. For completeness, we state the precise bounds obtained for the discrepancies of the stationary problems (cf. Theorems 6, 8, 10, and 12) and the corresponding spectral problems (Theorems 7, 9, 11, and 13). By rewriting the reasoning in proof of Theorems 2 and 4 (3 and 5) with minor modifications, we prove Theorems 6, 8, 10, and 12 (7, 9, 11, and 13, respectively); cf. [GoPeSh12] for some details about the stationary problems. Theorems 12 and 13 are proved in [GoEtAl12].

Theorem 6. *Let $\alpha \geq 1$, $\kappa < 2(\alpha - 1)$ and $f \in L^2(\Omega)$. Let u_ε be the solution of (11.16) and u the weak solution of the Dirichlet problem*

$$-\Delta u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \tag{11.38}$$

Then, we have

$$\|u_\varepsilon - u\|^2_{H^1(\Omega_\varepsilon)} + \varepsilon^{-\kappa}\|u_\varepsilon - u\|^2_{L^2(S_\varepsilon)} \leq C\varepsilon^{\min(1/2,(3\alpha-2)/4,2(\alpha-1)-\kappa)}\|f\|^2_{L^2(\Omega)}.$$

Theorem 7. *Let $\alpha \geq 1$ and $\kappa < 2(\alpha - 1)$. Let $\{\lambda_k^\varepsilon\}_{k=1}^\infty$ and $\{\lambda_k\}_{k=1}^\infty$ be the eigenvalues of problem (11.1) and (11.5), respectively. Then, for each fixed k there exists a constant C_k independent of ε such that, for sufficiently small $\varepsilon > 0$,*

$$|\lambda_k^\varepsilon - \lambda_k|^2 \leq C_k\varepsilon^{\min(1/2,(3\alpha-2)/4,2(\alpha-1)-\kappa)}.$$

Moreover, for any eigenvalue λ_k of (11.5) with multiplicity s ($\lambda_k = \lambda_{k+1} = \cdots = \lambda_{k+s-1}$), and for any u eigenfunction corresponding to λ_k, with $\|u\|_{L^2(\Omega)} = 1$, there exists $\tilde{u}^\varepsilon$, $\tilde{u}^\varepsilon$ being a linear combination of eigenfunctions $\{u_k^\varepsilon\}_{r=k}^{r=k+s-1}$ of (11.1) corresponding to $\{\lambda_k^\varepsilon\}_{r=k}^{r=k+s-1}$, such that

$$\|\tilde{u}^\varepsilon - u\|^2_{L^2(\Omega_\varepsilon)} \leq C_k\varepsilon^{\min(1/2,(3\alpha-2)/4,2(\alpha-1)-\kappa)}.$$

Theorem 8. *Let $\alpha > 2$, $\kappa \in \mathbb{R}$ and $f \in L^2(\Omega)$. Let u_ε be the solution of (11.16) and u the weak solution of the Dirichlet problem (11.38). Then, we have*

$$\|u_\varepsilon - u\|^2_{H^1(\Omega_\varepsilon)} + \varepsilon^{-\kappa}\|u_\varepsilon\|^2_{L^2(S_\varepsilon)} \leq C\varepsilon^{\min(1,(\alpha-2)/2)}\|f\|^2_{L^2(\Omega)}.$$

Theorem 9. *Let $\alpha > 2$ and $\kappa \in \mathbb{R}$. Let $\{\lambda_k^\varepsilon\}_{k=1}^\infty$ and $\{\lambda_k\}_{k=1}^\infty$ be the eigenvalues of problem (11.1) and (11.5), respectively. Then, for each fixed k there exists a constant C_k independent of ε such that, for sufficiently small $\varepsilon > 0$,*

$$|\lambda_k^\varepsilon - \lambda_k|^2 \leq C_k \varepsilon^{\min(1,(\alpha-2)/2)}.$$

Moreover, for any eigenvalue λ_k of (11.5) with multiplicity s ($\lambda_k = \lambda_{k+1} = \cdots = \lambda_{k+s-1}$), and for any u eigenfunction corresponding to λ_k, with $\|u\|_{L^2(\Omega)} = 1$, there exists $\tilde{u}^\varepsilon$, $\tilde{u}^\varepsilon$ being a linear combination of eigenfunctions $\{u_k^\varepsilon\}_{r=k}^{r=k+s-1}$ of (11.1) corresponding to $\{\lambda_k^\varepsilon\}_{r=k}^{r=k+s-1}$, such that

$$\|\tilde{u}^\varepsilon - u\|_{L^2(\Omega_\varepsilon)}^2 \leq C_k \varepsilon^{\min(1,(\alpha-2)/2)}.$$

Theorem 10. *Let $\alpha \in [1,2)$, $\kappa > 2(\alpha-1)$ and $f \in L^2(\Omega)$. Assume that, in a neighborhood of $\{x_1 = 0\}$, Ω coincides with the domain $(-T,T) \times \Theta$ of $\mathbb{R}^3$, where T is a fixed positive constant and Θ is the domain $\Omega \cap \{x_1 = 0\} \subset \mathbb{R}^2$. Let u_ε be the solution of (11.16) and u the weak solution of the Dirichlet problem in $\Omega^- \cup \Omega^+$:*

$$-\Delta u = f \text{ in } \Omega^- \cup \Omega^+, \quad u = 0 \text{ on } \partial\Omega \cup \gamma.$$

Then, we have

$$\|u_\varepsilon - u\|_{L^2(\Omega)}^2 \leq C\varepsilon^{\min(\kappa-2(\alpha-1),2-\alpha)} \|f\|_{L^2(\Omega)}^2.$$

Theorem 11. *Let $\alpha \in [1,2)$ and $\kappa > 2(\alpha-1)$. Let us assume that the domain Ω satisfies the assumption in Theorem 10. Let $\{\lambda_k^\varepsilon\}_{k=1}^\infty$ and $\{\lambda_k\}_{k=1}^\infty$ be the eigenvalues of problem (11.1) and (11.6), respectively. Then, for each fixed k there exists a constant C_k independent of ε such that, for sufficiently small $\varepsilon > 0$,*

$$|\lambda_k^\varepsilon - \lambda_k|^2 \leq C_k \varepsilon^{\min(\kappa-2(\alpha-1),2-\alpha)}.$$

Moreover, for any eigenvalue λ_k of (11.6) with multiplicity s ($\lambda_k = \lambda_{k+1} = \cdots = \lambda_{k+s-1}$), and for any u eigenfunction corresponding to λ_k, with $\|u\|_{L^2(\Omega)} = 1$, there exists $\tilde{u}^\varepsilon$, $\tilde{u}^\varepsilon$ being a linear combination of eigenfunctions $\{u_k^\varepsilon\}_{r=k}^{r=k+s-1}$ of (11.1) corresponding to $\{\lambda_k^\varepsilon\}_{r=k}^{r=k+s-1}$, such that

$$\|\tilde{u}^\varepsilon - u\|_{L^2(\Omega_\varepsilon)}^2 \leq C_k \varepsilon^{\min(\kappa-2(\alpha-1),2-\alpha)}.$$

Theorem 12. *Let $\alpha = \kappa = 2$ and $f \in L^2(\Omega)$. Let W_ε and h be the functions defined by (11.22) and (11.10), respectively. Let u_ε be the solution of (11.16) and u the solution of (11.18) for $b(x) = 4\pi C_0 h(x)$. Then, we have*

$$\|u_\varepsilon - u + W_\varepsilon h u\|^2_{H^1(\Omega_\varepsilon)} + \varepsilon^{-2}\|u_\varepsilon - u + hu\|^2_{L^2(S_\varepsilon)} \le C\varepsilon^{1/8}\|f\|^2_{L^2(\Omega)}$$

and

$$\|u_\varepsilon - u\|^2_{L^2(\Omega_\varepsilon)} \le C\varepsilon^{1/8}\|f\|^2_{L^2(\Omega)}.$$

Theorem 13. *Let $\alpha = \kappa = 2$. Let $\{\lambda_k^\varepsilon\}_{k=1}^\infty$ and $\{\lambda_k\}_{k=1}^\infty$ be the eigenvalues of problem (11.1) and (11.9), respectively. Then, for each fixed k there exists a constant C_k independent of ε such that, for sufficiently small $\varepsilon > 0$,*

$$|\lambda_k^\varepsilon - \lambda_k|^2 \le C_k\varepsilon^{1/8}.$$

Moreover, for any eigenvalue λ_k of (11.9) with multiplicity s ($\lambda_k = \lambda_{k+1} = \cdots = \lambda_{k+s-1}$), and for any u eigenfunction corresponding to λ_k, with $\|u\|_{L^2(\Omega)} = 1$, there exists $\tilde{u}^\varepsilon$, $\tilde{u}^\varepsilon$ being a linear combination of eigenfunctions $\{u_k^\varepsilon\}_{r=k}^{r=k+s-1}$ of (11.1) corresponding to $\{\lambda_k^\varepsilon\}_{r=k}^{r=k+s-1}$, such that

$$\|\tilde{u}^\varepsilon - u\|^2_{L^2(\Omega_\varepsilon)} \le C_k\varepsilon^{1/8}.$$

Acknowledgments This work has been partially supported by the Spanish project MTM2009-12628. The authors are grateful to professor S.V. Shaposhnikov for fruitful discussions.

References

[GiTr01] Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Springer, Berlin (2001)

[GoPeSh12] Gómez, D., Pérez, M.E., Shaposhnikova, T.A.: On homogenization of nonlinear Robin type boundary conditions for cavities along manifolds and associated spectral problems. Asymptotic Anal. **80**, 289–322 (2012)

[GoEtAl12] Gómez, D., Lobo, M., Pérez, E., Shaposhnikova, T.A.: On correctors for spectral problems in the homogenization of Robin boundary conditions with very large parameters. International Journal of Applied Mathematics, **26**(3), (2013)

[LoEtAl97] Lobo, M., Oleinik, O.A., Pérez, M.E., Shaposhnikova, T.A.: On homogenization of solutions of boundary value problem in domains perforated along manifolds. Ann. Scuola Norm. Sup. Pisa Cl. Sci. Ser. 4 **25**, 611–629 (1997)

[MaKh74] Marchenko, V.A., Khruslov, E.Ya.: Boundary Value Problems in Domains with a Fine-Grained Boundary. Naukova Dumka, Kiev (1974) (Russian)

[Mo66] Morrey, C.B.: Multiple Integrals in the Calculus of Variations. Springer, New York (1966)

[OlSh95a] Oleinik, O.A., Shaposhnikova, T.A.: On homogenization problem for the Laplace operator in partially perforated domains with Neumann's condition on the boundary of cavities. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei Mat. Appl. **6**, 133–142 (1995)

[OlSh95b] Oleinik, O.A., Shaposhnikova, T.A.: On the averaging for a partially perforated domain with a mixed boundary condition with a small parameter on the cavity boundary. Differ. Equat. **31**, 1086–1098 (1995)

[OlSh96] Oleinik, O.A., Shaposhnikova, T.A.: On homogenization of the Poisson equation in partially perforated domains with arbitrary density of cavities and mixed type conditions on their boundary. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei Mat. Appl. Ser. 9 **7**, 129–146 (1996)

[OlShYo92] Oleinik, O.A., Shamaev, A.S., Yosifian, G.A.: Mathematical Problems in Elasticity and Homogenization. North-Holland, Amsterdam (1992)

[Oz96] Ozawa, S.: Random media with many small Robin holes. Proc. Jpn. Acad. Ser. A Math. Sci. **72**, 4–5 (1996)

[Pe08] Pérez, M.E.: Long time approximations for solutions of wave equations via standing waves from quasimodes. J. Math. Pures Appl. **90**, 387–411 (2008)

[Pe11] Pérez, M.E.: Long time approximations for solutions of evolution equations from quasimodes: perturbation problems. Math. Balkanica **25**, 95–130 (2011)

[Ro93] Roppongi, S.: Asymptotics of eigenvalues of the Laplacian with small spherical Robin boundary. Osaka J. Math. **30**, 783–811 (1993)

[Sh08] Shaposhnikov, S.V.: On interior estimates of the Sobolev norms of solutions of elliptic equations. Math. Notes **83**, 285–289 (2011)

Chapter 12
A Finite Element Formulation of the Total Variation Method for Denoising a Set of Data

P.J. Harris and K. Chen

12.1 Introduction

The total variation method for removing the noise in a set of data, typically digital image data, formulates the underlying optimization problem in terms of a nonlinear differential equation. Generally, this equation has to be solved numerically, and the finite difference method has been widely used; unfortunately the subsequent system cannot be solved by a Newton-type method directly. However, it is possible to obtain an alternative numerical formulation using the finite element method. Indeed, as shall be shown here, the properties of the Galerkin finite element method make it an ideal choice of method for solving the differential equation due to its ability of implicitly reducing the nonlinearity.

12.2 Formulation of the Nonlinear Differential Equation

Suppose that we have a function $z(x,y)$ which, at the integer values of x and y defines the intensity levels of a data-set (typically a digital image). However, we assume that this datum contains an amount of random noise so that

$$z(x,y) = u(x,y) + \eta(x,y)$$

P.J. Harris (✉)
University of Brighton, Brighton, UK
e-mail: p.j.harris@brighton.ac.uk

K. Chen
University of Liverpool, Liverpool, UK
e-mail: K.Chen@liverpool.ac.uk

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_12,

where $u(x,y)$ represents the unknown underlying exact data and $\eta(x,y)$ represents the (also unknown) random error of Gaussian white noise in the data. The objective of this work is to try and get the best estimate possible of $u(x,y)$. Here it has been assumed that u is scaled such that $0 \leq u(x,y) \leq 1$.

The total variation model chooses u to minimize

$$\int\int_{\Omega} |\nabla u| \, dx \, dy$$

subject to the constraint

$$\|u - z\|^2 = \sigma^2$$

where σ is the magnitude of the noise in the data [AuVe97], [YaChYu12], [DoVo97], [RuOsFa92]. Applying the method of Lagrange multipliers to this problem leads to the required solution u satisfying the nonlinear differential equation

$$-\alpha \nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta}} \right) + (u - z) = 0 \tag{12.1}$$

subject to the boundary condition

$$\frac{\partial u}{\partial n} = 0$$

on the whole of the boundary $\partial\Omega$ where $\mathbf{n}$ is the unit normal vector to $\partial\Omega$. Here β is small parameter which has been introduced to avoid computational problems which would arise if $\|\nabla u\| = 0$.

12.3 Finite Element Method

In this section we introduce the finite element method for solving the differential equation (12.1). A more complete description of the finite element method can be found in one of the many textbooks on the subject, such as [ZiTa91].

For the Galerkin FEM, we now approximate u by

$$\tilde{u} = \sum_{i=1}^{N} u_i \phi_i(x,y) \tag{12.2}$$

where $\{\phi_1(x,y), \phi_2(x,y), \ldots, \phi_N(x,y)\}$ and $\{u_1, u_2, \ldots, u_N\}$ are, respectively, a set of known basis functions and a set of constants to be determined. Here N is the total number of data-points and in the case of an image will be the total number of pixels.

Since (12.2) is not, in general, the exact solution of (12.1) it will not satisfy (12.1) but will instead satisfy

$$-\alpha\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)+(u-z)=r(x,y) \tag{12.3}$$

where $r(x,y)$ is a residual function. The constants $u_1\ldots u_N$ are now chosen to make the residual term small in some sense. The Galerkin method used here requires that we choose the set of constants $\{u_1,u_2,\ldots,u_N\}$ such that this residual is orthogonal to all of the basis functions. This, in turn, means that the residual function cannot be written as a linear combination of the basis functions, and if the basis function span a large enough subspace of the solution, space then this will make the residual "small."

Applying the Galerkin method to (12.3), and recalling that the inner product of the residual with the basis functions is zero, leads to the system of equations

$$\int\int_\Omega\left[-\alpha\nabla\cdot\left(\frac{\nabla\tilde{u}}{\sqrt{|\nabla\tilde{u}|^2+\beta}}\right)+(\tilde{u}-z)\right]\phi_i\,dx\,dy=0\quad i=1,2,\ldots,N \tag{12.4}$$

Applying Green's theorem to the first term on the left-hand side of (12.4) yields

$$\begin{aligned}\int\int-\alpha\nabla\cdot\left(\frac{\nabla\tilde{u}}{\sqrt{|\nabla\tilde{u}|^2+\beta}}\right)\phi_i\,dx\,dy=&\int\int_\Omega\left(\frac{\alpha\nabla\tilde{u}}{\sqrt{|\nabla\tilde{u}|^2+\beta}}\right)\cdot\nabla\phi_i\,dx\,dy\\&+\int_{\partial\Omega}\left(\frac{\alpha\nabla\tilde{u}\cdot\mathbf{n}}{\sqrt{|\nabla\tilde{u}|^2+\beta}}\right)\phi_i\,dC.\end{aligned} \tag{12.5}$$

As the approximate solution $\tilde{u}$ is assumed to satisfy the boundary condition $\partial\tilde{u}/\partial n=\partial u/\partial n=0$, the line integral around the boundary in (12.5) is zero. Substituting (12.5) into (12.4) gives

$$\int\int_\Omega\alpha\left(\frac{\nabla\tilde{u}}{\sqrt{|\nabla\tilde{u}|^2+\beta}}\right)\cdot\nabla\phi_i+(\tilde{u}-z)\phi_i\,dx\,dy=0. \tag{12.6}$$

At this point it is worth noting that the boundary conditions are automatically incorporated into the finite element formulation, so unlike the finite difference method, there is no need to give any special treatment to the equations for data points which lie on the boundary of the domain.

For the finite element method, the domain Ω is now divided into sub-regions or elements connected together at vertices or nodes. Here the domain is first divided into small squares where the location of each vertex of the square corresponds to a data-point (or pixel in the image). In theory, it is possible to use quadrilateral finite element with bilinear basis functions in each element. However, if such elements

are used the integrals appearing in (12.6) cannot be evaluated analytically and need to be found numerically. As the solution process for this problem is an iterative one (due to the nonlinearity of the governing equations) it is not very desirable to have to evaluate all of the integrals numerically at each iteration due to the computational cost involved. An alternative is to divide each square, along one of the diagonals, into two triangles and use linear basis functions in each triangle. In this case the integrals in (12.6) can be evaluated analytically and this will speed up the iterative solution process.

Substituting the basis functions into (12.6) and integrating exactly yields a system of nonlinear algebraic equations that need to be solved for the coefficient u_i appearing in (12.2). Although the nonlinear equations can be formed by considering each basis function in turn, in practice, the equations are assembled in an element-by-element fashion, as in the standard linear finite element method. Further details of the computational process can be found, for example, in [ZiTa91].

Having formed the system of nonlinear equations, we now have to consider methods for solving them. Much of the previous work using finite difference methods has used the fixed-point iteration method to solve the system. In this work we are going to employ Newton's method to solve the system. In practice, the accuracy of the final numerical solution depends on the two parameters α and β appearing in (12.1). Whilst both have an effect on the overall accuracy of the method, β has a more significant effect on the convergence (or otherwise) of Newton's method. If β is chosen too large, then Newton's method for solving the nonlinear equations converges quickly for almost any initial guess of the solution, but to a very inaccurate solution. However, if β is too small, then unless the initial guess is close to the final solution Newton's method does not converge (it tends to oscillate between two inaccurate solutions rather than diverge to infinity).

An alternative algorithm is to start with a large value of β, say β_1 and to use the data $z(x,y)$ as the initial guess. This will yield an approximate solution $\tilde{u}_1(x,y)$ which will be closer to the desired solution $u(x,y)$ than $z(x,y)$. The value of β is now reduced to β_2 and using $\tilde{u}_1(x,y)$ as the initial solution we obtain $\tilde{u}_2(x,y)$ which is a further improvement in the solution. This process is repeated until either a minimum value of β is reached, or there is no significant improvement in the accuracy of the solution when β is reduced further.

Varying the parameter α in (12.1) affects the overall accuracy of the method, and we investigate how α affects the accuracy in numerical results section below.

12.4 Numerical Results

The results presented here are for an $n \times n$ grid of data where the ideal values are such that the central points are 1 and the surrounding data edges are zero. We then introduce random errors into the data so that we can investigate the accuracy and efficiency of our finite element for recovering the original data. A typical example is shown in Fig. 12.1.

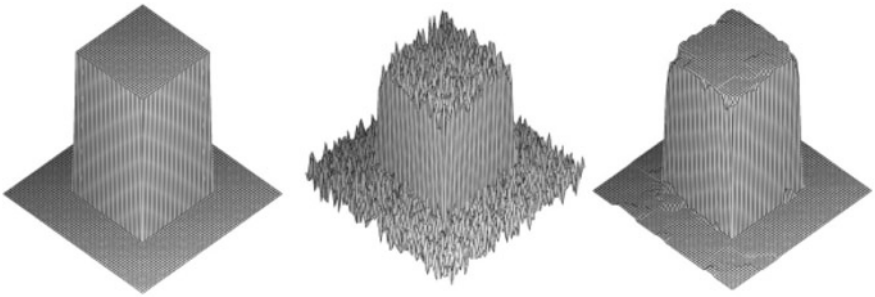

Fig. 12.1 A typical ideal data-set (*left*), a data-set with random errors or noise (*middle*) and the data-set obtained using the finite element method to denoise the data (*right*)

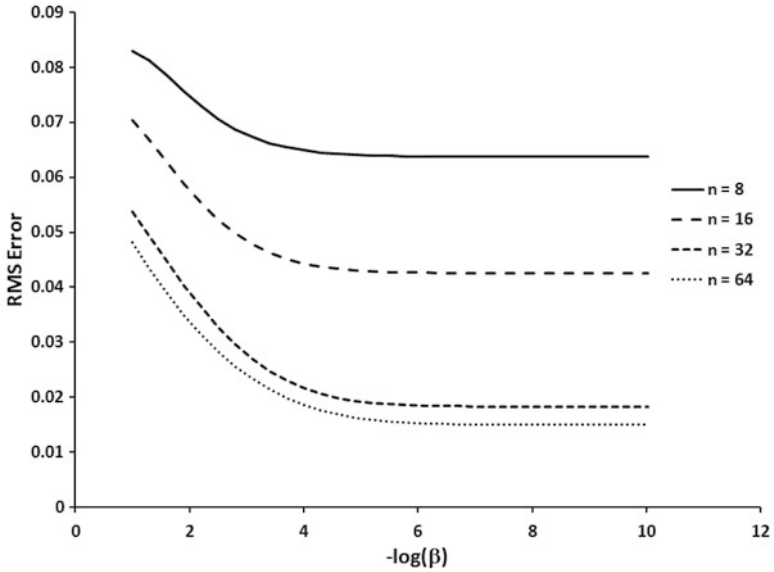

Fig. 12.2 How the RMS error in the finite element solution varies with β for different sized data-sets

The plot on the left of Fig. 12.1 shows the original, ideal data. The plot in the middle shows that data after some random errors (of maximum magnitude 0.1) have been introduced and the image on the right shows the plot obtained after applying our finite element method. Here the stopping criteria was when β was smaller than 10^{-10}.

Figure 12.2 shows how the method converges as the parameter β is reduced for difference sized data-sets.

This graph shows that the overall error is smaller for larger images, although this may simply be a artifact of the way in which the RMS error is computed. However,

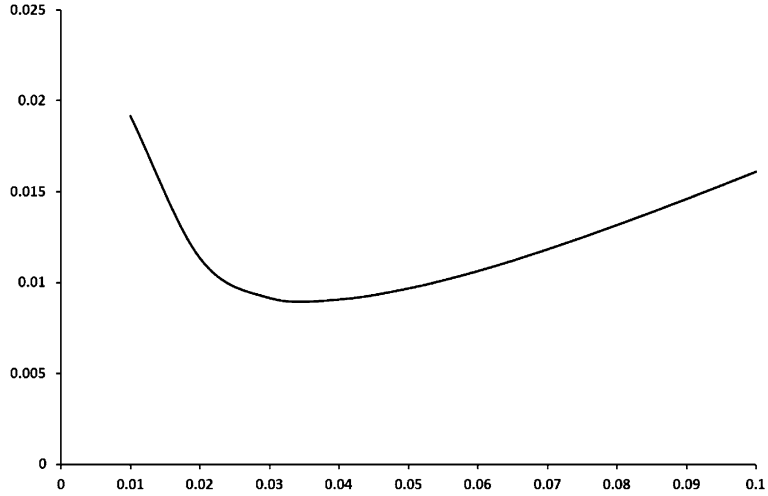

Fig. 12.3 How the RMS error in the finite element solution for a 32×32 data-set varies with α

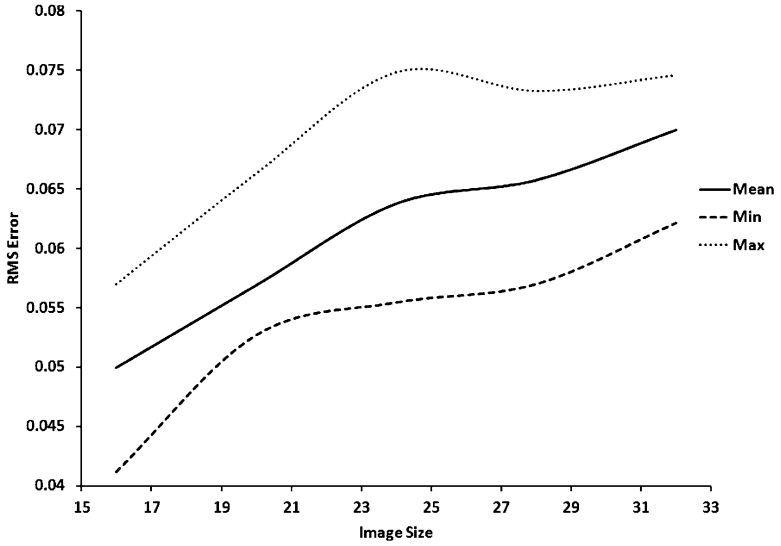

Fig. 12.4 How the min, max, and mean of the optimal value of α varies with image size

in all cases it shows that once β is smaller than around 10^{-6} there is no significant reduction in the size of the RMS error, and so there is no computational advantage to using such values.

Figure 12.3 shows how the RMS error in the final 32×32 data-set varies with the parameter α.

This shows that there is an optimal value of the parameter α. However, the exact value of this parameter is problem dependent and a data-set with different random errors would not necessarily have the same optimal value of α. A suggested value of α can be obtained by analyzing a large number of different cases and taking the average value of α. Figure 12.4 shows the results of doing this with ten different data-sets of different sizes.

The results presented in this graph seem to indicate that the optimal value of α is dependent on the size of the data-sets. However, as this needs to be investigated further, perhaps using a large number of data-sets to get a better estimate of the mean optimal value of α.

12.5 Conclusions

The results presented in this paper show that the finite element method can be used to obtain an accurate solution to the nonlinear differential equation which arises in total-variation denoising problems. The nature of the differential equation, which can be expressed as the divergence of a nonlinear vector-valued term, is such that it is relatively simple to apply the standard Galerkin method to this equation. Further, as part of this process, the homogenous boundary condition for the differential equation is automatically incorporated into the formulation meaning that unlike the finite difference method, we do not have to give any special treatment to the grid-points or nodes which lie on the boundary of the domain of the equation.

The results in this paper show that the system of nonlinear algebraic equations which result from using the finite element method to solve (12.1) can be solved using a algorithm based on Newton's method. The results also show that once the parameter β appearing in (12.1) is reduced below approximately 10^{-6} then there is no significant improvement in the accuracy of the solution obtained. Further, the results show that the value of the parameter α appearing in (12.1) has a major effect on the accuracy of the method and that in each case there is definite optimal value. However, this value appears to be different for every example, and so an average value has to be used. More work is needed to establish what the optimal value is, how it is related to the size of the data-set and how FEM may be advantageously applied to solving other types of variational models [ChSh85].

References

[AuVe97] Aubert, G., Vese, L.: A variational method in image recovery. SIAM J. Numer. Anal. **34**, 1948–1979 (1997)

[ChSh85] Chan, T.F., Shen, J.: Image Processing and Analysis—Variational, PDE, Wavelet, and Stochastic Methods. SIAM Publications (1985)

[DoVo97] Dobson, D.C., Vogel, C.R.: Convergence of an iterative method for total variation denoising. SIAM J. Numer. Anal. **34**, 1779–1791 (1997)

[RuOsFa92] Rudin, I.L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)

[YaChYu12] Yang, F., Chen, K., Yu, B.: Homotopy curve tracking for total variation image restoration. J. Comp. Math. **30**, 177–196 (2012)

[ZiTa91] Zienkiewicz, O.C., Taylor, R.L.: The Finite Element Method, vols. 1, 2. McGraw-Hill, London (1988, 1991)

Chapter 13
On the Convergence of the Multi-group Isotropic Neutron LTS_N Nodal Solution in Cartesian Geometry

E.B. Hauser, R.P. Pazos, and M.T. Vilhena

13.1 Introduction

The Discrete Ordinate Nodal approach (nodal S_N approximation) is a well-known technique to work out multidimensional neutron transport problems. There exists a vast amount of literature concerning the numerical nodal methods [BaLa90, BaLa92, ShBe09, Wi71], but the analytical ones are scarce, restricted, for instance, to the LTS_N nodal approach. Briefly speaking, the basic idea of this methodology encompasses the transverse integration of the multi-group neutron transport equation in a multidimensional Cartesian geometry domain, resulting in a coupled system of one-dimensional S_N equations for the average angular fluxes, which are then analytically solved by the Laplace Transform technique (LTS_N method), as in [BaVi91]. This methodology was applied to these sort of problems, without losing generality, for isotropic scattering and multigroup energy models. However, to the best of our knowledge, a convergence analysis of the nodal methods is not found in the literature, except for the isotropic scattering and a monoenergetic LTS_N nodal approach. Therefore, we extend the convergence analysis for this nodal technique, [HaPaVi05], now assuming the multigroup model, specializing the study for the two-dimensional problem. The basic idea relies on the definition of an error with a proper norm both for the LTS_N nodal solution and for the Gaussian quadrature approximation of the integral scattering term appearing in the neutron transport

E.B. Hauser
Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: eliete@pucrs.br

R.P. Pazos
University of Santa Cruz do Sul, Santa Cruz do Sul, RS, Brazil
e-mail: rpazos@unisc.br

M.T. Vilhena (✉)
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: vilhena@mat.ufrgs.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_13,

equation. From these definitions, we can determine an error bound estimate for the two-dimensional LTS_N neutron nodal solution, which guarantees the convergence of the discussed solution to the exact one, when N goes to infinity [FrNa90].

13.2 The Two-Group Discrete Ordinate (S_N) Approximation to the Transport Equation in X, Y Geometry

We consider the steady-state multigroup Boltzmann transport equation in two-dimensional Cartesian geometry

$$\mu \frac{\partial \psi_g}{\partial x}(x,\, y, \nu) + \eta \frac{\partial \psi_g}{\partial y}(x,\, y, \nu) + h(\vec{r}, \nu)\ \psi_g(x,\, y, \nu) = q_g(x,\, y, \nu) + \int_V \psi_g(x,\, y, \nu')\, k(\nu, \nu')\, d\nu', \quad (13.1)$$

where g is the energy group, (x, y) represents the particle position in the domain $X = [0, a] \times [0, b]$, $\nu = (\mu, \eta)$ is a point referred to angular coordinates in

$$V = \{\, \nu \mid \mu^2 + \eta^2 \leq 1 \,\},$$

$\psi_g(x, y, \nu)$ is the density flux function, $h(x, y, \nu)$ is the collision frequency, $k(x, y, \nu, \nu')$ is the scattering kernel, and $q_g(x, y, \nu)$ is the source function.

The discrete ordinate method (S_N) is a technique used for obtaining numerical solutions to the integro-differential equation (13.1). In S_N equations the flux scalar is approximated by quadrature formulas, as has been shown in [LeMi84]. Thus, we consider the discrete ordinates S_N approximation to (13.1), with linearly isotropic scattering and two energy groups; that is,

$$\mu_m \frac{\partial}{\partial x}\Psi_{m,g}(x,y) + \eta_m \frac{\partial}{\partial y}\Psi_{m,g}(x,y) + \sigma_{t,g}\Psi_{m,g}(x,y) = \frac{1}{4}\left[\sigma_{s,1,g}\sum_{n=1}^{N} w_n \Psi_{n,1}(x,y) + \sigma_{s,2,g}\sum_{n=1}^{N} w_n \Psi_{n,2}(x,y)\right] + Q_g(x,y), \quad (13.2)$$

where $\Psi_{m,g}(x,y) = \Psi_g(x,y,\mu_m,\eta_m)$ is the angular flux per group in the discrete direction (μ_m, η_m), $m = 1 : M$ with $M = N(N+2)/2$, for any even N quadrature set index, ω_m are the angular quadrature weights, $\sigma_{t,g}, \sigma_{s,1,g}, \sigma_{s,2,g}$ are the total t and scattering s macroscopic cross section per group, and $Q_{m,g}(x,y)$ is the isotropic interior source term defined in the discrete direction. We assume that the solutions of (13.1) and (13.2) satisfy the same boundary conditions in the discrete directions.

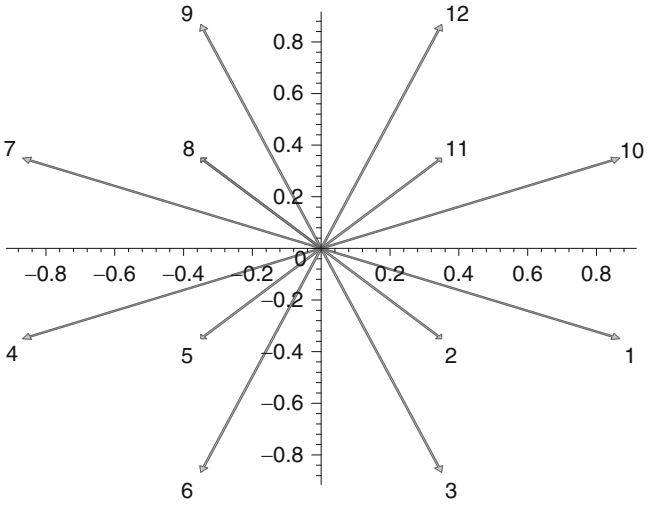

Fig. 13.1 S_4 discrete directions with level symmetry

In the sequel, we prove that, under suitable restrictions, the solution of (13.2) converges to the solution of (13.1) as $N \to \infty$ for N an even number. We choose the set with level symmetry for the discrete directions $\Omega_m = (\mu_m, \eta_m)$, [LeMi84], as illustrated in Fig. 13.1 for $N = 4$.

13.3 The Multigroup Nodal LTS_N Formulation in a Rectangle

We integrate (13.2) with respect to y in the interval $[0, b]$ and obtain

$$\mu_m \frac{d\Psi_{mx,g}}{dx}(x) + \sigma_{t,g}\Psi_{mx,g}(x) - \frac{1}{4}\left[\sigma_{s,1,g}\sum_{n=1}^{N} w_n \Psi_{nx,1}(x) + \sigma_{s,2,g}\sum_{n=1}^{N} w_n \Psi_{nx,2}(x)\right] = S_{mx,g}(x), \quad (13.3)$$

where the y-edge average angular flux in the discrete direction $\Omega_m = (\mu_m, \eta_m)$ is defined as

$$\Psi_{mx,g}(x) = \frac{1}{b}\int_0^b \Psi_{m,g}(x,y)dy. \quad (13.4)$$

The source term $S_{mx,g}(x)$, which includes the interior source and the transverse leakage terms, is

$$S_{mx,g}(x) = Q_{x,g}(x) - \frac{\eta_m}{b}\left[\Psi_{m,g}(x,b) - \Psi_{m,g}(x,0)\right],$$

where

$$Q_{x,g}(x) = \frac{1}{b}\int_0^b Q_g(x,y)dy.$$

The transverse integrated S_N equations for the spatial y-direction are obtained in a similar fashion.

Equation (13.3) form a system of $4M$ linear ordinary differential equations in the $8M$ unknown functions: $\Psi_{mx,g}(x)$, $\Psi_{my,g}(y)$, $\Psi_{m,g}(x)$, $\Psi_{m,g}(y)$, $g = 1,2$. Therefore, for the x-direction we write

$$\begin{aligned}
&\frac{d}{dx}\Psi_{mx,1}(x) + \frac{\sigma_{t,1}}{\mu_m}\Psi_{mx,1}(x) \\
&\qquad - \frac{1}{4\mu_m}\left[\sigma_{s,1,1}\sum_{n=1}^{N} w_n\Psi_{nx,1}(x) + \sigma_{s,2,1}\sum_{n=1}^{N} w_n\Psi_{nx,2}(x)\right] = \frac{S_{mx,1}(x)}{\mu_m}, \\
&\frac{d}{dx}\Psi_{mx,2}(x) + \frac{\sigma_{t,2}}{\mu_m}\Psi_{mx,2}(x) \\
&\qquad - \frac{1}{4\mu_m}\left[\sigma_{s,1,2}\sum_{n=1}^{N} w_n\Psi_{nx,1}(x) + \sigma_{s,2,2}\sum_{n=1}^{N} w_n\Psi_{nx,2}(x)\right] = \frac{S_{mx,2}(x)}{\mu_m}.
\end{aligned} \tag{13.5}$$

We apply the Laplace transformation with respect to x in (13.5). For $g = 1,2$ we use

$$\mathscr{L}\left\{S_{mx,g}(x)\right\} = \overline{S}_{mx,g}(s), \quad \mathscr{L}\left\{\Psi_{mx,g}(x)\right\} = \overline{\Psi}_{mx,g}(s),$$
$$\mathscr{L}\left\{\frac{d\Psi_{mx,g}}{dx}(x)\right\} = s\overline{\Psi}_{mx,g}(s) - \Psi_{mx,g}(0).$$

For $m = 1 : M$, we obtain the algebraic system of $2M$ linear equations

$$\begin{aligned}
&s\overline{\Psi}_{mx,1}(s) + \frac{\sigma_{t,1}}{\mu_m}\overline{\Psi}_{mx,1}(s) - \frac{\sigma_{s,1,1}}{4\mu_m}\sum_{n=1}^{N} w_n\overline{\Psi}_{nx,1}(s) \\
&\qquad - \frac{\sigma_{s,2,1}}{4\mu_m}\sum_{n=1}^{N} w_n\overline{\Psi}_{nx,2}(s) = \Psi_{mx,1}(0) + \frac{\overline{S}_{mx,1}(s)}{\mu_m}, \\
&s\overline{\Psi}_{mx,2}(s) + \frac{\sigma_{t,2}}{\mu_m}\overline{\Psi}_{mx,2}(s) - \frac{\sigma_{s,2,2}}{4\mu_m}\sum_{n=1}^{N} w_n\overline{\Psi}_{nx,2}(s) \\
&\qquad - \frac{\sigma_{s,1,2}}{4\mu_m}\sum_{n=1}^{N} w_n\overline{\Psi}_{nx,1}(s) = \Psi_{m,2}(0) + \frac{\overline{S}_{mx,2}(s)}{\mu_m}.
\end{aligned} \tag{13.6}$$

One may cast (13.6) in the matrix form

$$[s\mathbf{I}-\mathbf{A}_{\mathbf{x}}]\begin{bmatrix}\overline{\Psi}_{mx,1}(s)\\ \overline{\Psi}_{mx,2}(s)\end{bmatrix}=\begin{bmatrix}\Psi_{mx,1}(0)\\ \Psi_{mx,2}(0)\end{bmatrix}+\frac{1}{\mu_m}\begin{bmatrix}\overline{S}_{mx,1}(s)\\ \overline{S}_{mx,2}(s)\end{bmatrix}, \tag{13.7}$$

where $\mathbf{I}$ is the identity matrix. In (13.7), for each $g = 1.2$, we defined the $2M$-dimensional vector functions

$$\begin{aligned}\overline{\Psi}_{mx,g}(s)&=\left[\ \Psi_{1x,g}(s)\ \ \overline{\Psi}_{2x,g}(s)\cdots\overline{\Psi}_{Mx,g}(s)\ \right]^T,\\ \Psi_{mx,g}(0)&=\left[\ \Psi_{1x,g}(0)\ \ \Psi_{2x,g}(0)\cdots\Psi_{Mx,g}(0)\ \right]^T,\\ \overline{S}_{mx,g}(s)&=\left[\ \overline{S}_{1x,g}(s)\ \ \overline{S}_{2x,g}(s)\cdots\overline{S}_{Mx,g}(s)\ \right]^T.\end{aligned}$$

In (13.7), we create a novel way to define the $2M\times 2M$ matrix $\mathbf{A}_{\mathbf{x}}$, namely

$$\mathbf{A}_{\mathbf{x}}=\begin{bmatrix}\mathbf{A}_{\mathbf{x},\mathbf{11}}&\mathbf{A}_{\mathbf{x},\mathbf{12}}\\ \mathbf{A}_{\mathbf{x},\mathbf{21}}&\mathbf{A}_{\mathbf{x},\mathbf{22}}\end{bmatrix}.$$

$\mathbf{A}_{\mathbf{x}}$ is composed of $M\times M$ sub-matrices $\mathbf{A}_{\mathbf{x},\mathbf{g}',\mathbf{g}}$, $g',g=1,2$, which are

$$\mathbf{A}_{\mathbf{x},\mathbf{11}}=\begin{bmatrix}-\dfrac{4\sigma_{t,1}-\sigma_{s,1,1}\omega_1}{4\mu_1}&\dfrac{\sigma_{s,1,1}\omega_2}{4\mu_1}&\cdots&\dfrac{\sigma_{s,1,1}\omega_M}{4\mu_1}\\ \dfrac{\sigma_{s,1,1}\omega_1}{4\mu_2}&-\dfrac{4\sigma_{t,1}-\sigma_{s,1,1}\omega_2}{4\mu_2}&\cdots&\dfrac{\sigma_{s,1,1}\omega_M}{4\mu_2}\\ \vdots&\vdots&\ddots&\vdots\\ \dfrac{\sigma_{s,1,1}\omega_1}{4\mu_M}&\dfrac{\sigma_{s,1,1}\omega_2}{4\mu_M}&\cdots&-\dfrac{4\sigma_{t,1}-\sigma_{s,1,1}\omega_M}{4\mu_M}\end{bmatrix},$$

$$\mathbf{A}_{\mathbf{x},\mathbf{22}}=\begin{bmatrix}-\dfrac{4\sigma_{t,2}-\sigma_{s,2,2}\omega_1}{4\mu_1}&\dfrac{\sigma_{s,2,2}\omega_2}{4\mu_1}&\cdots&\dfrac{\sigma_{s,2,2}\omega_M}{4\mu_1}\\ \dfrac{\sigma_{s,2,2}\omega_1}{4\mu_2}&-\dfrac{4\sigma_{t,2}-\sigma_{s,2,2}\omega_2}{4\mu_2}&\cdots&\dfrac{\sigma_{s,2,2}\omega_M}{4\mu_2}\\ \vdots&\vdots&\ddots&\vdots\\ \dfrac{\sigma_{s,2,2}\omega_1}{4\mu_M}&\dfrac{\sigma_{s,2,2}\omega_2}{4\mu_M}&\cdots&-\dfrac{4\sigma_{t,2}-\sigma_{s,2,2}\omega_M}{4\mu_M}\end{bmatrix},$$

$$\mathbf{A}_{\mathbf{x},\mathbf{21}}=\begin{bmatrix}\dfrac{\sigma_{s,1,2}\omega_1}{4\mu_1}&\dfrac{\sigma_{s,1,2}\omega_2}{4\mu_1}&\cdots&\dfrac{\sigma_{s,1,2}\omega_M}{4\mu_1}\\ \dfrac{\sigma_{s,1,2}\omega_1}{4\mu_2}&\dfrac{\sigma_{s,1,2}\omega_2}{4\mu_2}&\cdots&\dfrac{\sigma_{s,1,2}\omega_M}{4\mu_2}\\ \vdots&\vdots&\ddots&\vdots\\ \dfrac{\sigma_{s,1,2}\omega_1}{4\mu_M}&\dfrac{\sigma_{s,1,2}\omega_2}{4\mu_M}&\cdots&\dfrac{\sigma_{s,1,2}\omega_M}{4\mu_M}\end{bmatrix},$$

$$\mathbf{A_{x,12}} = \begin{bmatrix} \frac{\sigma_{s,2,1}\omega_1}{4\mu_1} & \frac{\sigma_{s,2,1}\omega_2}{4\mu_1} & \cdots & \frac{\sigma_{s,2,1}\omega_M}{4\mu_1} \\ \frac{\sigma_{s,2,1}\omega_1}{4\mu_2} & \frac{\sigma_{s,2,1}\omega_2}{4\mu_2} & \cdots & \frac{\sigma_{s,2,1}\omega_M}{4\mu_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{s,2,1}\omega_1}{4\mu_M} & \frac{\sigma_{s,2,1}\omega_2}{4\mu_M} & \cdots & \frac{\sigma_{s,2,1}\omega_M}{4\mu_M} \end{bmatrix}.$$

The solution of (13.7) is

$$\begin{bmatrix} \overline{\Psi}_{mx,1}(s) \\ \overline{\Psi}_{mx,2}(s) \end{bmatrix} = [s\mathbf{I} - \mathbf{A_x}]^{-1} \left(\begin{bmatrix} \Psi_{mx,1}(0) \\ \Psi_{mx,2}(0) \end{bmatrix} + \frac{1}{\mu_m} \begin{bmatrix} \overline{S}_{mx,1}(s) \\ \overline{S}_{mx,2}(s) \end{bmatrix} \right). \tag{13.8}$$

In order to determine the angular flux, we apply the inverse Laplace transformation to (13.8). The result is

$$\begin{bmatrix} \Psi_{mx,1}(x) \\ \Psi_{mx,2}(x) \end{bmatrix} = \mathscr{L}^{-1} \left\{ [s\mathbf{I} - \mathbf{A_x}]^{-1} \left(\begin{bmatrix} \Psi_{mx,1}(0) \\ \Psi_{mx,2}(0) \end{bmatrix} + \frac{1}{\mu_m} \begin{bmatrix} \overline{S}_{mx,1}(s) \\ \overline{S}_{mx2}(s) \end{bmatrix} \right) \right\}.$$

Therefore, we obtain

$$\begin{bmatrix} \Psi_{mx,1}(x) \\ \Psi_{mx,2}(x) \end{bmatrix} = \mathscr{L}^{-1} \left\{ [s\mathbf{I} - \mathbf{A_x}]^{-1} \right\} \begin{bmatrix} \Psi_{mx,1}(0) \\ \Psi_{mx,2}(0) \end{bmatrix} + \frac{1}{\mu_m} \mathscr{L}^{-1} \left\{ [s\mathbf{I} - \mathbf{A_x}]^{-1} \right\} * \begin{bmatrix} \overline{S}_{mx,1}(x) \\ \overline{S}_{mx,2}(x) \end{bmatrix},$$

where $*$ denotes the convolution operation.

Furthermore, in order to determine $\mathscr{L}^{-1} \left\{ [s\mathbf{I} - \mathbf{A_x}]^{-1} \right\}$, we assume that the matrix $\mathbf{A_x} = \mathbf{V_x D_x V_x}^{-1}$ can be diagonalized and write

$$\begin{aligned} \mathscr{L}^{-1} \left\{ [s\mathbf{I} - \mathbf{A_x}]^{-1} \right\} &= \mathscr{L}^{-1} \left\{ \left[s\,\mathbf{V_x V_x}^{-1} - \mathbf{V_x D_x V_x}^{-1} \right]^{-1} \right\} \\ &= \mathscr{L}^{-1} \left\{ \left[\mathbf{V_x} (s\mathbf{I} - \mathbf{D_x}) \mathbf{V_x}^{-1} \right]^{-1} \right\} = \mathbf{V_x} \mathscr{L}^{-1} \{ [s\mathbf{I} - \mathbf{D_x}]^{-1} \} \mathbf{V_x}^{-1}. \end{aligned} \tag{13.9}$$

$\mathbf{D_x}$ is an $2M-$ order diagonal matrix of the eigenvalues of $\mathbf{A_x}$ and $\mathbf{V_x}$ is the matrix whose columns are the $2M$ eigenvectors of $\mathbf{A_x}$.

Applying the inverse Laplace transformation, we find that

$$\mathscr{L}^{-1} \left\{ (s\mathbf{I} - \mathbf{D_x})^{-1} \right\} = e^{\mathbf{D_x} x},$$

which, substituted in (13.9), leads to

$$\mathscr{L}^{-1}\left\{(sI-A_x)^{-1}\right\} = \mathbf{V}_x e^{\mathbf{D}_x x}\mathbf{V}_x^{-1}.$$

As a result, the analytical solution for the two-group S_N equations with isotropic scattering (13.2) is

$$\begin{bmatrix}\Psi_{mx,1}(x)\\ \Psi_{mx,2}(x)\end{bmatrix} = \left[\mathbf{V}_\mathbf{x}\, e^{\mathbf{D}_\mathbf{x} x}\mathbf{V}_\mathbf{x}{}^{-1}\right]\begin{bmatrix}\Psi_{mx,1}(0)\\ \Psi_{mx,2}(0)\end{bmatrix} + \frac{1}{\mu_m}\left[\mathbf{V}\, e^{\mathbf{D}_\mathbf{x} x}\mathbf{V}_\mathbf{x}{}^{-1}\right] * \begin{bmatrix}\overline{S}_{mx,1}(x)\\ \overline{S}_{mx,2}(x)\end{bmatrix}.$$

In a similar fashion, we obtain

$$\begin{bmatrix}\Psi_{my,1}(y)\\ \Psi_{my,2}(y)\end{bmatrix} = \left[\mathbf{V}_\mathbf{y}\, e^{\mathbf{D}_\mathbf{y} y}\mathbf{V}_\mathbf{y}{}^{-1}\right]\begin{bmatrix}\Psi_{my,1}(0)\\ \Psi_{my,2}(0)\end{bmatrix} + \frac{1}{\mu_m}\left[\mathbf{V}\, e^{\mathbf{D}_\mathbf{y} y}\mathbf{V}_\mathbf{y}{}^{-1}\right] * \begin{bmatrix}\overline{S}_{my,1}(y)\\ \overline{S}_{my,2}(y)\end{bmatrix}.$$

At this point, based on the physics of shielding problems, we assume that the neutron flux attenuates exponentially with increasing distance from the source, a hypothesis also assumed in [BaVi91], [HaViBa09], [HaEtAl08], [HaPaVi05], where the only approximation involved is in the transverse leakage terms.

For instance, in (13.4) we use

$$\Psi_{mx,1}(x,0) = \sum_{m=1}^{M} C_m\, e^{-(\operatorname{sign}\mu_m)\sigma_a x},$$

where $\sigma_a = \sigma_t - \sigma_s$, $\operatorname{sign}\mu_m = 1$ if $\mu_m > 0$ and $\operatorname{sign}\mu_m = -1$ if $\mu_m < 0$. The solution is completely determined when we apply the boundary conditions.

13.4 Error Bounds for the Discrete Ordinates Nodal Method and Two Energy Groups

In this section we extend the mathematical analysis of the error bound estimate and convergence to the mono-energetic nodal-LTS_N solution in a rectangle proposed in [HaEtAl08]. We discuss the conditions for the convergence of the discrete ordinates nodal method, LTS_N, mentioned in [Ze90], [KaLeHe82], [Kh97], [PaVi99], [KhSb05a], [KhSb05b], [HaEtAl08]. We define the so-called errors of the approximated flux and the error in the quadrature formula and then establish a relationship between both errors in order to give a global estimate of the approximated flux. We denote the base spaces by $\mathbf{E_g} = L^1(\mathbf{X_g} \times \mathbf{V})$ and the approximating spaces by $\mathbf{E}_{M,\mathbf{g}} = \prod_{m=1}^{M} L^1(\mathbf{X_g} \times \Omega_m)$. The solutions are studied in the Banach subspaces defined by

$$W_\mathbf{g} = \left\{ \psi \in \mathbf{E_g} \;\middle|\; \mu\frac{\partial\psi}{\partial x} + \eta\frac{\partial\psi}{\partial y} \in \mathbf{E_g} \right\}$$

and, for all $m = 1 : M$,

$$W_{m,\mathbf{g}} = \left\{ \{\psi_m\} \in \mathbf{E}_{M,\mathbf{g}} \;\middle|\; \mu_m \tfrac{\partial \psi_m}{\partial x} + \eta_m \tfrac{\partial \psi_m}{\partial y} \in L^1(\mathbf{X_g}) \right\}.$$

We define the error functions in the approximate flux, for the energy level $\mathbf{g}$ as

$$\varepsilon_{m,\mathbf{g}}(\vec{r}) = \psi(\vec{r}, \Omega_m) - \Psi_{m,\mathbf{g}}(\vec{r}),$$

and error function in the quadrature formula for the energy level $\mathbf{g}$ as

$$\tau_{m,\mathbf{g}}(\vec{r}) = \int_V \sigma_s(\Omega_m, \nu')\psi(\vec{r}, \nu')\, d\nu' - \sum_{m=1}^{M} \omega_m k_{mn} \Psi_{m,\mathbf{g}}(\vec{r}, \Omega_m).$$

After subtracting (13.2) from (13.1) and doing some algebraic manipulations, we obtain the relationships

$$\mu_m \frac{\partial \varepsilon_{m,\mathbf{g}}}{\partial x}(\vec{r}) + \eta_m \frac{\partial \varepsilon_{m,\mathbf{g}}}{\partial y}(\vec{r}) + h_m(\vec{r})\, \varepsilon_{m,\mathbf{g}}(\vec{r}) = \sum_{n=1}^{M} w_n k_{mn}\, \varepsilon_{n,\mathbf{g}}(\vec{r}) + \tau_{m,\mathbf{g}}(\vec{r}).$$

Now, multiplying both sides of the last equation by $\varepsilon_{m,\mathbf{g}}(\vec{r})$ and integrating in the domain $\mathbf{X}$, we arrive at

$$\frac{\mu_m}{2} \int\int_{\mathbf{X}} \frac{\partial\, \varepsilon^2_{m,\mathbf{g}}}{\partial\, x}(\vec{r})\, d\vec{r} + \frac{\eta_m}{2} \int\int_{\mathbf{X}} \frac{\partial\, \varepsilon^2_{m,\mathbf{g}}}{\partial\, y}(\vec{r})\, d\vec{r} + \int\int_{\mathbf{X}} h_m(\vec{r})\, \varepsilon^2_{m,\mathbf{g}}(\vec{r})\, d\vec{r}$$
$$= \sum_{n=1}^{M} \omega_m k_{nm} \int\int_{\mathbf{X}} \varepsilon_{n,\mathbf{g}}(\vec{r})\, \varepsilon_{m,\mathbf{g}}(\vec{r})\, d\vec{r} + \int\int_{\mathbf{X}} \varepsilon_{m,\mathbf{g}}(\vec{r})\, \tau_{m,\mathbf{g}}(\vec{r})\, d\vec{r};$$

hence,

$$\int\int_{\mathbf{X}} h_m(\vec{r})\, \varepsilon^2_{m,\mathbf{g}}(\vec{r})\, d\vec{r} = \sum_{n=1}^{M} \omega_m k_{nm} \int\int_{\mathbf{X}} \varepsilon_{n,\mathbf{g}}(\vec{r})\, \varepsilon_{m,\mathbf{g}}(\vec{r})\, d\vec{r}$$
$$+ \int\int_{\mathbf{X}} \varepsilon_{m,\mathbf{g}}(\vec{r})\, \tau_{m,\mathbf{g}}(\vec{r})\, d\vec{r} + \frac{\mu_m}{2} \int_0^b \left[\varepsilon^2_{m,\mathbf{g}}(0,y) - \varepsilon^2_{m,\mathbf{g}}(a,y)\right] dy$$
$$+ \frac{\eta_m}{2} \int_0^a \left[\varepsilon^2_{m,\mathbf{g}}(x,0) - \varepsilon^2_{m,\mathbf{g}}(x,b)\right] dx. \tag{13.10}$$

At this point, we introduce new definitions for the scalar product and norm:

$$\alpha_m = (\alpha)_{m=1:M}, \quad \langle \alpha \,|\, \beta \rangle = \int\int_{\mathbf{X}} w_m\, \alpha_m(\vec{r})\, \beta_m(\vec{r})\, d\vec{r}, \quad \|\alpha\|^2 = \langle \alpha \,|\, \alpha \rangle.$$

We assume (see [KaLeHe82] and [PaVi99]) that there is c_0 such that $h(\vec{r},\nu) > \lambda' - c_0\,|\,(\nu)\,|, \forall\,(\vec{r},\nu) \in \mathbf{X} \times \mathbf{V}$, where $\lambda' = \inf\left\{ \lim_{|\nu|\longrightarrow 0} h(\vec{r},\nu)\,,(\vec{r})\,\varepsilon\,\mathbf{X}\right\}$. Then, by (13.10),

$$\int\int_{\mathbf{X}} \left(\lambda' - c_0\,|\,(\Omega_m)\,|\right)\,\varepsilon^2_{m,\mathbf{g}}(\vec{r})\,d\vec{r}$$
$$\leq \sum_{n=1}^{M} \omega_m\,k_{nm} \int\int_{\mathbf{X}} \varepsilon_{n,\mathbf{g}}(\vec{r})\,\varepsilon_{m,\mathbf{g}}(\vec{r})\,d\vec{r} \int\int_{\mathbf{X}} \varepsilon_{m,\mathbf{g}}(\vec{r})\,\tau_{m,\mathbf{g}}(\vec{r})\,d\vec{r}$$
$$+ \frac{\mu_m}{2}\int_0^b \left[\varepsilon^2_{m,\mathbf{g}}(0,y) - \varepsilon^2_{m,\mathbf{g}}(a,y)\right]\,dy + \frac{\eta_m}{2}\int_0^a \left[\varepsilon^2_{m,\mathbf{g}}(x,0) - \varepsilon^2_{m,\mathbf{g}}(x,b)\right]\,dx.$$

Multiplying each term in the above inequality by w_m and summing up with respect to m, we obtain

$$\sum_{m=1}^{M} \omega_m\,\left(\lambda' - c_0\,|\,(\mu_m,\eta_m,\xi_m)\,|\right)\int\int_{\mathbf{X}} \varepsilon^2_{m,\mathbf{g}}(\vec{r})\,d\vec{r}$$
$$\leq \sum_{m=1}^{M}\sum_{n=1}^{M} \omega_m\,\omega_n\,k_{nm} \int\int_{\mathbf{X}} \varepsilon_{n,\mathbf{g}}(\vec{r})\,\varepsilon_{m,\mathbf{g}}(\vec{r})\,d\vec{r}$$
$$+ \sum_{m=1}^{M} \omega_m \int\int_{\mathbf{X}} \varepsilon_{m,\mathbf{g}}(\vec{r})\,\tau_{m,\mathbf{g}}(\vec{r})\,d\vec{r}$$
$$+ \sum_{m=1}^{M} \omega_m\,\frac{|\,\mu_m\,|}{2}\int_0^b \left|\,\varepsilon^2_{m,\mathbf{g}}(0,y) - \varepsilon^2_{m,\mathbf{g}}(a,y)\,\right|\,dy$$
$$+ \sum_{m=1}^{M} \omega_m\,\frac{|\,\eta_m\,|}{2}\int_0^a \left|\,\varepsilon^2_{m,\mathbf{g}}(x,0) - \varepsilon^2_{m,\mathbf{g}}(x,b)\,\right|\,dx.$$

We now define

$$F_{1,m,\mathbf{g}} = \sqrt{\frac{|\,\mu_m\,|}{2}\int_0^b \left|\,\varepsilon^2_{m,\mathbf{g}}(0,y) - \varepsilon^2_{m,\mathbf{g}}(a,y)\,\right|\,dy}\,,$$
$$F_{2,m,\mathbf{g}} = \sqrt{\frac{|\,\eta_m\,|}{2}\int_0^a \left|\,\varepsilon^2_{m,\mathbf{g}}(0,y) - \varepsilon^2_{m,\mathbf{g}}(a,y)\,\right|\,dx}\,.$$

Applying the Cauchy–Schwarz inequality using a parameter K_0 associated with $k(\vec{r}\,\nu,\nu')$ and the quadrature weights, we have

$$\left(\lambda' - c_0 \max_{m=1:M} |\,\Omega_m\,|\right)\,\|\,\varepsilon^2_{\mathbf{g}}\,\| \leq K_0\|\varepsilon_{\mathbf{g}}\|^2 + \|\,\varepsilon_{\mathbf{g}}\,\|\,\|\,\tau_{\mathbf{g}}\,\| + \|\,F_{1,\mathbf{g}}\,\| + \|\,F_{2,\mathbf{g}}\,\|.$$

Choosing δ so that

$$\| \varepsilon_{\mathbf{g}} \| \cdot \| \tau_{\mathbf{g}} \| \leq \frac{1}{2} \left(\delta \| \varepsilon_{\mathbf{g}} \|^2 + \frac{\| \tau_{\mathbf{g}} \|^2}{\delta} \right),$$

we get

$$\left(\lambda' - c_0 \max_{m=1:M} |\Omega_m| \right) \| \varepsilon_{\mathbf{g}}^2 \| \leq K_0 \| \varepsilon_{\mathbf{g}} \|^2 + \frac{1}{2} \left(\delta \| \varepsilon_{\mathbf{g}} \|^2 + \frac{\| \tau_{\mathbf{g}} \|^2}{\delta} \right) + \| F_{1,\mathbf{g}} \| + \| F_{2,\mathbf{g}} \|.$$

So, for each energy level $\mathbf{g}$, we obtain the expression

$$\| \varepsilon_{\mathbf{g}} \|^2 \leq \frac{\frac{\| \tau_{\mathbf{g}} \|^2}{2\delta} + \| F_{1,\mathbf{g}} \| + \| F_{2,\mathbf{g}} \|}{\left(\lambda' - c_0 \max_{m=1:M} |\Omega_m| \right) - K_0 - \frac{\delta}{2}},$$

Finally, defining

$$\| \varepsilon \| = \sqrt{\| \varepsilon_{\mathbf{1}} \|^2 + \| \varepsilon_{\mathbf{2}} \|^2},$$
$$\| \tau \| = \sqrt{\| \tau_{\mathbf{1}} \|^2 + \| \tau_{\mathbf{2}} \|^2},$$

we obtain the relationship between the global error ε in the approximate flux and the error τ in the quadrature formula, with the latter depending on the boundary conditions, in the form

$$\| \varepsilon \|^2 \leq \frac{\frac{\| \tau \|^2}{2\delta} + \| F_1 \| + \| F_2 \|}{\left(\lambda' - c_0 \max_{m=1:M} |\Omega_m| \right) - K_0 - \frac{\delta}{2}}.$$

The last inequality guarantees the convergence of the two-group nodal LTS_N solution when N increases significantly.

13.5 Conclusions

We have constructed the analytical nodal method, LTS_N, to solve numerically the two-group energy S_N equations with linearly isotropic scattering in a homogeneous x, y geometry. Moreover, we determined an error bound estimate for the two-group energy LTS_N neutron nodal solution, which guarantees the convergence of the discussed solution to the exact one, when N goes to infinity. Further, we defined

a proper norm for both the LTS_N nodal solution and for the Gaussian quadrature approximation of the integral scattering term appearing in the neutron transport equation. The analysis of the convergence of the method gives error bounds of the approximated angular flux in terms of truncation error in the quadrature formula and of the boundary conditions.

We have restricted ourselves to S_N with two energy groups. This is the reason why the resulting spectrum contains $2M$ eigenvalues. In general, though, the S_N problem is allowed to have an arbitrary number G of energy groups. The resulting spectrum will then contain $G \times M$ eigenvalues that have an associated basis for the $G \times M-$ dimensional representation of the solution.

References

[BaVi91] Barichello, L.B., Vilhena, M.T.B.: A new analytical approach to solve the neutron transport equation. Kerntechnik **56**, 334–336 (1991)

[BaLa90] Barros, R.C., Larsen, E.W.: A numerical method for one-group slab-geometry discrete ordinates problems. Nucl. Sci. Eng. **104**, 199–208 (1990)

[BaLa92] Barros, R.C., Larsen, E.W.: A spectral nodal method for one-group X,Y-geometry discrete ordinates problems. Nucl. Sci. Eng. **111**, 34–45 (1992)

[FrNa90] Frigyes, R., Nagy, B.: Functional Analysis. Dover, New York (1990)

[HaPaVi05] Hauser, E.B., Pazos, R.P., Vilhena, M.T.: An error bound estimate of the LTS_N nodal solution in Cartesian geometry. Ann. Nucl. Energ. **32**, 1146–1156 (2005)

[HaEtAl08] Hauser, E.B., Pazos, R.P, Vilhena, M.T., Barros, R.C.: The error bounds for the three-dimentional nodal LTS_N method. In: Proceedings of 16th International Conference on Nuclear Engineering, vol. 1, pp. 1–12, Orlando, FL (2008)

[HaViBa09] Hauser, E.B., Vilhena, M.T., Barros, R.C.: A Laplace transform exponential method for monoenergetic three-dimensional fixed source discrete ordinates problems in Cartesian geometry. IJNEST **5**, 80–89 (2009)

[KaLeHe82] Kaper, H.G., Lekkerkerker, C.G., Hejtmanek, J.: Spectral Methods in Linear Transport Theory. Birkhäuser, Basel (1982)

[Kh97] Kharroubi, M.H.: Mathematical Topics in Neutron Transport Theory. New Aspects. World Scientific, Singapore (1997)

[KhSb05a] Kharroubi, M.H., Sbihi, M.: Critical spectrum and spectral mapping theorems in transport theory. Semigroup Forum **70**, 406–435 (2005)

[KhSb05b] Kharroubi, M.H., Sbihi, M.: Spectral mapping theorems for neutron transport. L^1-theory. Semigroup Forum **72**, 249–282 (2005)

[LeMi84] Lewis, E., Miller, W.: Computational Methods of Neutron Transport. Wiley, New York (1984)

[PaVi99] Panta, R.P., Vilhena, M.T.B.: Convergence in transport theory. Appl. Numer. Math. **30**, 79–92 (1999)

[ShBe09] Sharipov, F., Bertoldo, G.: Numerical solution of the linearized Boltzmann equation for an arbitrary intermolecular potential. J. Comput. Phys. **228**, 3345–3357 (2009)

[Wi71] Williams, M.M.R.: Mathematical Methods in Particle Transport Theory. Butterworth, London (1971)

[Ze90] Zeidler, E.: Nonlinear Functional Analysis and Applications, vols. 1–2. Springer, Berlin (1990)

Chapter 14
Numerical Integration with Singularity by Taylor Series

H. Hirayama

14.1 Introduction

We consider the integration of the product of a smooth function $f(x)$ and a function $K(x;c)$ with a singularity in the finite integration interval $[a,b]$; that is,

$$I(a,b,c) = \int_a^b K(x;c) f(x)\, dx. \tag{14.1}$$

This type of integral is difficult to evaluate by the usual numerical methods when $K(x;c)$ is a singular function such as $|x-c|^{\alpha}(\log|x-c|)^n$, with $\alpha > -1$ a real number and $n > 0$ an integer, or $(x-c)^{-1}$ (the Cauchy principal-value case) or $(x-c)^{-n}$, with $n > 1$ an integer (the Hadamard finite-part case).

For functions with singularities at the end-points of the integration interval, the integral can be computed numerically by means of a transformation of variable—for example, the double exponential formula method [TaMo74].

The same can be done when the singularity in $|x-c|^{\alpha}(\log|x-c|)^n$ lies within the integration interval, after dividing the interval into two subintervals at the singular point.

Many numerical integration methods, such as the Chebyshev integration technique or Gauss-type numerical integration, need to reconstruct their formulas in accordance with the kind of singularity that K has. Therefore, the integration program needs to be adjusted for each type of singularity, which is rather inconvenient.

In this paper, we use Taylor series to split the integral (14.1) into a singular part, which is computed analytically, and a part without singularity (or with a weak singularity), which can be computed by standard numerical methods. As mentioned

H. Hirayama (✉)
Department of Vehicle System Engineering, Faculty of Creative Engineering,
Kanagawa Institute of Technology, Japan
e-mail: hirayama@sd.kanagawa-it.ac.jp

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_14,

earlier, we consider integrals with algebraic and logarithmic singularity, Cauchy principal-value integrals, and Hadamard finite-part integrals; specifically,

$$I_1 = \int_a^b |x-c|^{\alpha} (\log|x-c|)^n f(x)\,dx, \tag{14.2}$$

$$I_2 = p.v. \int_a^b \frac{f(x)}{x-c}\,dx, \tag{14.3}$$

$$I_3 = f.p. \int_a^b \frac{f(x)}{(x-c)^n}\,dx, \tag{14.4}$$

where $n > 1$ is a positive integer and $\alpha > -1$ is a real number.

14.2 Taylor Series

We begin by explaining the basic ideas behind the expansion of functions in Taylor series (see [Ra81], [He74], and [HiEtAl07] for details).

Without loss of generality, we consider Taylor series around the origin. Any other case can be reduced to this one through a translation of the variable. For convenience and later use, we list three such expansions:

$$f(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + f_4 x^4 \cdots, \tag{14.5}$$

$$g(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3 + g_4 x^4 \cdots, \tag{14.6}$$

$$h(x) = h_0 + h_1 x + h_2 x^2 + h_3 x^3 + h_4 x^4 \cdots. \tag{14.7}$$

14.2.1 The Arithmetic of Taylor Series

Arithmetic operations with Taylor series are defined naturally and without difficulty.

1. *Addition and subtraction.* If $h(x) = f(x) \pm g(x)$, the coefficients of f, g, and h (see (14.5)–(14.7)) satisfy

$$h_i = f_i \pm g_i.$$

2. *Multiplication.* If $h(x) = f(x)g(x)$, then

$$h_n = \sum_{k=0}^{n} f_i g_{n-i}.$$

3. *Division.* If $h(x) = \dfrac{f(x)}{g(x)}$, then

$$h_0 = \frac{f_0}{g_0}, \quad h_n = \frac{1}{g_0}\left(f_n - \sum_{k=0}^{n-1} h_k g_{n-k}\right) \quad (n \geq 1).$$

14.2.2 Basic Functions of Taylor Series

Many basic functions satisfy simple differential equations. Using these equations, we can easily compute such functions of Taylor series.

1. *Exponential function.* If $h(x) = e^{f(x)}$, then

$$\frac{dh(x)}{dx} = h(x)\frac{df(x)}{dx}.$$

Substituting (14.5) and (14.7) in this differential equation and comparing the coefficients on both sides, we get

$$h_0 = e^{f_0}, \quad h_n = \frac{1}{n}\sum_{k=1}^{n} k h_{n-k} f_k \quad (n \geq 1).$$

2. *Logarithmic function.* If $h(x) = \ln f(x)$, then

$$f(x)\frac{dh(x)}{dx} = \frac{df(x)}{dx}.$$

Substituting (14.5) and (14.7) in this differential equation and comparing the coefficients, we arrive at

$$h_0 = \log f_0, \quad h_n = \frac{1}{nf_0}\left(nf_n - \sum_{j=1}^{n-1} j h_j f_{n-j}\right).$$

Similar differential equations and coefficient relationships may be obtained without difficulty between the coefficients of the Taylor series for other elementary transcendental functions.

3. *Integration and differentiation.* These two operations can be performed on Taylor series in the expected way; thus,

$$h(x) = \frac{df(x)}{dx}, \quad h(x) = \int_0^x f(t)dt$$

yield, respectively,

$$h_0 = 0, \quad h_i = (i+1)f_{i+1} \quad (i = 1, \ldots, n-1),$$

$$h_0 = 0, \quad h_i = \frac{1}{i} f_{i-1} \quad (i = 1, \ldots, n).$$

14.2.3 Numerical Example

As an illustration, we compute the Taylor series at $x = 2$ for the function

$$f(x) = \frac{\sin x}{1+x^2}. \tag{14.8}$$

This is performed by the C++ program

```
1 : #include "taylor_template.h"//define Taylor series
2 : typedef taylor_template<double> taylor ;
3 : void main()
4 : {
5 :     taylor x, y ;
6 :     x = taylor( 2.0, 2.0, 1.0); // x=2.0+1.0(x-2)
7 :     y = sin(x)/(1+x*x) ;
8 :     cout « y « endl ;
9 : }
```

where taylor(a,b,c) generates $b + c(x-a)$.

The output giving the Taylor polynomial of degree 14 for $f(x)$ in (14.8) is

$$\begin{aligned} &0.181859 - 0.228717(x-2) + 0.0556719(x-2)^2 + 0.0150774(x-2)^3 \\ &- 0.0156188(x-2)^4 + 0.00878601(x-2)^5 - 0.00415762(x-2)^6 \\ &+ 0.00158541(x-2)^7 - 0.000432293(x-2)^8 + 2.85231 \times 10^{-5}(x-2)^9 \\ &+ 6.359 \times 10^{-5}(x-2)^{10} - 5.65745 \times 10^{-5}(x-2)^{11} + 3.2542 \times 10^{-5}(x-2)^{12} \\ &- 1.47187 \times 10^{-5}(x-2)^{13} + 5.26657 \times 10^{-6}(x-2)^{14}. \end{aligned}$$

14.3 Integration of Singular Functions

The Taylor series of the function $f(x)$ given by (14.1) is written around a generic point c as

$$f(x) = f_0 + f_1(x-c) + f_2(x-c)^2 + \cdots + f_m(x-c)^m + \cdots. \tag{14.9}$$

Using (14.9), we split each of the integrals (14.2)–(14.4) into a part, computed analytically, where the integrand contains the original singularity, and a part (computed numerically), where the integrand is a smooth function.

14.3.1 Integrals with Algebraic and Logarithmic Singularity

Integral (14.2) with algebraic and logarithmic singularity becomes

$$I_1 = \int_a^b |x-c|^\alpha (\log|x-c|)^n (f_0 + \cdots + f_m(x-c)^m) dx + \int_a^b |x-c|^\alpha (\log|x-c|)^n (f(x) - \{f_0 + \cdots + f_m(x-c)^m\}) dx. \quad (14.10)$$

The logarithmic factor in the first integral on the right-hand side in (14.10) can be removed by repeated application of integration by parts with the help of the formulas

$$\int_a^b |x-c|^\alpha (x-c)^m dx = \left[\frac{|x-c|^\alpha (x-c)^{m+1}}{\alpha+m+1} \right]_a^b,$$

$$\int_a^b |x-c|^\alpha (x-c)^m (\log|x-c|)^n dx = \left[\frac{|x-c|^\alpha (x-c)^{m+1}}{\alpha+m+1} (\log|x-c|)^n \right]_a^b - \frac{n}{\alpha+m+1} \int_a^b |x-c|^\alpha (x-c)^m (\log|x-c|)^{n-1} dx.$$

The second integrand on the right-hand side in (14.10) is

$$|x-c|^\alpha (\log|x-c|)^n (f(x) - \{f_0 + f_1(x-c) + \cdots + f_m(x-c)^m\}) = |x-c|^\alpha (\log|x-c|)^n O((x-c)^{m+1}),$$

which is an m-times differentiable function. If we take m large enough, then the integral of this function can be computed by any one of a number of numerical procedures.

Since loss of significant digits occurs near $x = c$, it is difficult to compute the second integrand in (14.10) with enough accuracy. This can be avoided by taking a sufficiently large number of terms in the series

$$f(x) - \{f_0 + f_1(x-c) + \cdots + f_m(x-c)^m = f_{m+1}(x-c)^{m+1} + f_{m+2}(x-c)^{m+2} + \cdots + f_{m+k}(x-c)^{m+k} + \cdots.$$

14.3.2 Cauchy Principal Value Integral

The Cauchy principal value integral (14.3) can be written as

$$\begin{aligned} p.v.\int_a^b \frac{f(x)}{x-c}\,dx &= \lim_{\varepsilon\to 0+}\left(\int_a^{c-\varepsilon}+\int_{c+\varepsilon}^b\right)\frac{f(x)}{x-c}\,dx \\ &= p.v.\int_a^b \frac{f_0}{x-c}\,dx+\int_a^b \frac{f(x)-f_0}{x-c}\,dx \\ &= f_0\ln\left|\frac{b-c}{a-c}\right|+\int_a^b \frac{f(x)-f_0}{x-c}\,dx. \end{aligned} \tag{14.11}$$

As this formula shows, the singularity is completely removed from the computation of the integral, which was not possible in the case of an integrand with an algebraic and logarithmic singular point. Here, the Taylor series expansion leads to

$$\frac{f(x)-f_0}{x-c} = f_1+f_2(x-c)+f_3(x-c)^2+\cdots,$$

which permits us to attain a sufficiently high computational accuracy by means of numerical integration. In the example given in Sect. 14.4, we compute $\dfrac{f(x)-f_0}{x-c}$ for $|x-c|>1/10$, and $f_1+f_2(x-c)+f_3(x-c)^2+\cdots$ otherwise.

14.3.3 Hadamard Finite–Part Integral

The Hadamard finite part integral (14.4) is written as

$$\begin{aligned} f.p.\int_a^b \frac{f(x)}{(x-c)^n}dx &= f.p.\int_a^b \sum_{k=0}^{n-2}\frac{f_k}{(x-c)^{n-k}}dx+p.v.\int_a^b \frac{f_{n-1}}{(x-c)}dx \\ &\quad+\int_a^b \frac{1}{x-c}\left(f(x)-\sum_{k=0}^{n-1}f_k(x-c)^k\right)dx \\ &= \sum_{k=0}^{n-2}\frac{f_k}{n-k+1}\left(\frac{1}{(a-c)^{n-k+1}}-\frac{1}{(b-c)^{n-k+1}}\right) \\ &\quad+f_{n-1}\log\left|\frac{b-c}{a-c}\right|+\int_a^b \frac{1}{x-c}\left(f(x)-\sum_{k=0}^{n-1}f_k(x-c)^k\right)dx. \end{aligned}$$

As seen from this formula, the singularity has been removed completely from the integral, just as it was in the case of the Cauchy principal value, and numerical

integration methods can be used to complete the evaluation of the integral. The loss of significant digits near $x = c$ for the integrand in the last term can be avoided by taking a sufficiently large number of terms in the Taylor series.

14.4 Numerical Examples

14.4.1 Integration with Algebraic and Logarithmic Singularity

There are papers in the literature dealing with the numerical integration of functions with algebraic singularities (see [HaTo91]) and with logarithmic ones (see [HaTo87b]), but the author is not aware of any that treat functions exhibiting both at the same time. We give a simple example of such a case.

Consider

$$\int_{-1}^{1} \frac{\log|x|}{\sqrt{|x|}} e^x dx = -8.16418166413206192974141914914955390\ldots. \quad (14.12)$$

Taking the Taylor polynomial of degree 9 for e^x around $x = 0$, we have

$$\begin{aligned} e^x \approx 1 + x + 0.5x^2 + 0.166667x^3 + 0.0416667x^4 \\ + 0.00833333x^5 + 0.00138889x^6 + 0.000198413x^7 \\ + 2.48016 \times 10^{-5}x^8 + 2.75573 \times 10^{-6}x^9. \end{aligned}$$

Using this expansion and performing the integration by means of the double exponential numerical method with 31 sample points, we obtain the result $2.68e-11$. This becomes -8.164181664132062366 when the analytic part of the calculation is added to it. The final number is in agreement to 15 decimal places with (14.12), which was computed by taking the Taylor polynomial of degree 20.

14.4.2 Cauchy Principal Value Integral

The Cauchy principal value integral has been studied extensively (see [Bi90a], [El79], and [OgSuMo93]). As a numerical illustration, we choose Hasegawa's example [HaTo87a]

$$p.v. \int_{-1}^{1} \frac{e^{4(x-1)}}{x - \frac{1}{2}} dx = 0.67053144165072524849322194979263006 44\ldots. \quad (14.13)$$

The Taylor polynomial of degree 8 for the function $e^{4(x-1)}$ at $x = 0.5$ is

$$\begin{aligned} e^{4(x-1)} \approx\ & 0.135335 + 0.541341(x-0.5) + 1.08268(x-0.5)^2 \\ & + 1.44358(x-0.5)^3 + 1.44358(x-0.5)^4 + 1.15486(x-0.5)^5 \\ & + 0.769907(x-0.5)^6 + 0.439947(x-0.5)^7 + 0.219974(x-0.5)^8. \end{aligned}$$

The numerical integration part is carried out by means of the double exponential method with 132 sample points and gives the value 0.819212. Adding the analytic part, we arrive at 0.670531441650725646, a result that agrees to 15 decimal places with (14.13). The latter was computed with the Taylor polynomial of degree 20.

14.4.3 Hadamard Finite Part Integral

For this type, we choose Bialecki's example (see [Bi90b] and [Pa81])

$$\begin{aligned} f.p. \int_{-1}^{1} \frac{(1-x)^{1/4}(1+x)^{-1/4}}{(x-\frac{1}{10})} dx &= -\frac{50\pi}{3^{3/3}11^{5/4}} \\ &= -1.5090274451745640506248\ldots. \end{aligned} \tag{14.14}$$

The Taylor polynomial of degree 6 for the function $(x-1)^{1/4}(x+1)^{-1/4}$ at $x = 0.1$ is

$$\begin{aligned} & 0.95107 - 0.480338(x-0.1) + 0.0727785(x-0.1)^2 - 0.164181(x-0.1)^3 \\ & \quad + 0.0326109(x-0.1)^4 - 0.0975269(x-0.1)^5 + 0.0137508(x-0.1)^6; \end{aligned}$$

therefore,

$$\begin{aligned} & f.p. \int_{-1}^{1} \frac{(x-1)^{1/4}(x+1)^{-1/4}}{(x-0.1)^2} dx \\ & \quad \approx f.p. \int_{-1}^{1} \frac{0.95107}{(x-0.1)^2} dx - p.v. \int_{-1}^{1} \frac{0.480338}{x-0.1} dx \\ & \qquad + \int_{-1}^{1} \left\{ \frac{(x-1)^{1/4}(x+1)^{-1/4}}{(x-0.1)^2} - 0.95107 - 0.480338(x-0.1) \right\} dx. \end{aligned}$$

The first two terms on the right-hand side are computed analytically, and the rest are evaluated numerically by means of the double exponential method with 37 sample points. The numerical part produces the result 0.315936140492676 that changes

to -1.509027445174564 when the analytic part is added to it. The final number coincides to 16 decimal places with (14.14), which is computed with the Taylor polynomial of degree 20.

14.5 Conclusion

The singularities in the integrands of the Cauchy principal-value integral and Hadamard finite-part integral are easily removed when Taylor series are used. We have shown that these types of integrals can be evaluated without difficulty by many numerical integration methods. In the case of an algebraic and logarithmic singular point, we showed that the singularity can be weakened, making such integrals computable by standard numerical integration methods. This efficient manner of computation is facilitated by the use of Taylor series.

We point out that other problems can be treated equally successfully by the method proposed in this chapter, which are otherwise unsolvable; for example, the Cauchy principal value integral [DaRa75]

$$T(f) = p.v.\frac{1}{2\pi}\int_{-\pi}^{\pi} \cot\left(\tfrac{1}{2}(\theta-\phi)\right) f(\phi)d\phi.$$

Errors can be estimated by means of error analysis techniques for numerical integration methods.

References

[Bi90a] Bialecki, B.: A Sinc-Hunter quadrature rule for Cauchy principal value integrals. Math. Comput. **55**, 665–581 (1990)

[Bi90b] Bialecki, B.: A Sinc quadrature rule for Hadamard finite-part integral. Numer. Math. **57**, 263–269 (1990)

[DaRa75] Davis, P.J., Rabinowitz, P.: Methods of Numerical Integration. Academic, New York (1975)

[El79] Elliott, D.: Gauss type quadrature rule for Cauchy principal value integrals. Math. Comput. **33**, 301–309 (1979)

[HaTo87a] Hasegawa, T., Torii, T.: An automatic quadrature for Cauchy principal value integrals. Inform. Process. Soc. Jpn. **25**, 857–913 (1984) (Japanese)

[HaTo87b] Hasegawa, T., Torii, T.: An automatic scheme for indefinite integration of function with a logarithmic singularity. Inform. Process. Soc. Jpn. **28**, 907–914 (1987) (Japanese)

[HaTo91] Hasegawa, T., Torii, T.: An automatic quadrature for indefinite Integral of algebraic singular integrand. JSIAM **1**, 1–11 (1991) (Japanese)

[He74] Henrici, P.: Applied Computational Complex Analysis, vol. 1. Wiley, New York (1974)

[HiEtAl07] Hirayama, H., Tateno, H., Asano, N., Kawaguchi, T.: How to use Taylor series library. Tohoku Univ. Information Synergy Center SENAC **40**, 29–68 (2007) (Japanese)

[OgSuMo93] Ogata, H., Sugihara, M., Mori, M.: A DE-type quadrature rule for Cauchy principal-value integrals and Hadamard finite-part integrals. JSIAM **3**, 309–322 (1993) (Japanese)

[Pa81] Paget, D.F.: Numerical evaluation of Hadamard finite-part integrals. Numer. Math. **36**, 447–453 (1981)

[Ra81] Rall, L.B.: Automatic Differentiation-Technique and Applications. Lecture Notes in Computer Science, vol. 120. Springer, Berlin (1981)

[TaMo74] Takahasi, H., Mori, M.: Double exponential formula for numerical integration. Publ. RIMS, Kyoto Univ. **9**, 121–141 (1974)

Chapter 15
Numerical Solutions of the 1D Convection–Diffusion–Reaction and the Burgers Equation Using Implicit Multi-stage and Finite Element Methods

C.A. Ladeia and N.M.L. Romeiro

15.1 Introduction

In the last decades, developments in computational mechanics motivated extensive research on numerical solutions that had an important impact on society [OdEtAl03]. In particular, we are interested in procedures that can be adapted to problems involving convective, diffusive, and reactive processes. These problems have a vast applicability (see [GoCoCa00], [TaShDe07], [KuEsDa04]), such as the simulation of pollution effects in rivers; modeling of the evolution of oil and natural gas reserves in the underground; modeling of heat transfer problems, dispersion of pollutants; modeling of cosmological scenarios, analysis in seismology; phenomenology of turbulence; the theory of shock waves; and in many other applications.

Usually, the studies employ implicit multi-stage methods combined with the finite element method to increase the convergence region of the obtained results (see [DoRoHu00], [Ve04], [RoSa07], [TiYu11]). In this discussion, we consider the implicit multi-stage method of second-order R_{11} and fourth-order R_{22}, for the discretization of the temporal domain and we use three formulations of the finite element method type for the discretization of the spatial domain, i.e., least squares (LSFEM), Galerkin (GFEM), and *streamline-upwind* Petrov–Galerkin (SUPG) to solve the 1D convection–diffusion–reaction and the Burgers equation.

C.A. Ladeia (✉) • N.M.L. Romeiro
State University of Londrina, Rodovia Celso Garcia Cid-PR 445 Km 380-Campus Universitario CEP 86.057-970, Londrina, Paraná, Brazil
e-mail: cibele_mat_uel@yahoo.com.br; nromeiro@uel.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_15,

15.2 Statement of the Problems

15.2.1 1D Convection–Diffusion–Reaction Equation

We consider the 1D convection–diffusion–reaction problem, consisting in finding $u(x,t) : \Omega \to \mathbb{R}$ such that

$$u_t(x,t) + vu_x(x,t) - Du_{xx}(x,t) + \sigma u(x,t) = f(x,t), \quad \text{in } \Omega, \tag{15.1}$$

$$u(0,t) = u(l,t) = 0 \quad \text{on } \Gamma, \tag{15.2}$$

$$u(x,0) = u_0(x) \quad \forall x \in \Omega, \tag{15.3}$$

where $\Omega \subset \mathbb{R}$ is an open bounded domain with boundary $\Gamma = \partial\Omega$. The coefficients of (15.1) are $v : \Omega \to \mathbb{R}$, the velocity field; $D \geq 0$, the diffusion coefficient; $\sigma : \Omega \to \mathbb{R}$, the linear reaction coefficient; $f : \Omega \to \mathbb{R}$, the source term and (15.2) a Dirichlet boundary, and (15.3) the initial condition. We can rewrite (15.1) as $u_t + \mathscr{L}(u) = f$, where the spatial differential operator is defined as

$$\mathscr{L}(u) = vu_x - Du_{xx} + \sigma u \tag{15.4}$$

and $\mathscr{L} = \mathscr{L}_{conv} + \mathscr{L}_{dif} + \mathscr{L}_{reac}$ represents the sum of the linear convective, diffusive, and reactive operators, respectively.

15.2.2 Burgers Equation

Here, we consider the Burgers equation problem

$$u_t(x,t) + u(x,t)u_x(x,t) - \varepsilon u_{xx}(x,t) = f(x,t) \quad \text{in } \Omega, \tag{15.5}$$

$$u(0,t) = u(l,t) = 0 \quad \text{on } \Gamma, \tag{15.6}$$

$$u(x,0) = u_0(x) \quad \forall x \in \Omega\,. \tag{15.7}$$

The coefficients of (15.5) are given by $\varepsilon = 1/Re$, the coefficient of viscosity of the fluid, Re the Reynolds number. Further, $u(x,t)$ is the x-component of the fluid velocity field, $f : \Omega \to \mathbb{R}$, the source term and (15.6) a Dirichlet boundary condition, and (15.7) the initial condition, where u_0 is a known function. We can rewrite (15.5) as

$$u_t + \mathscr{L}(u) = f,$$

where the spatial operator is defined as

$$\mathscr{L}(u) = uu_x - \varepsilon u_{xx}\,, \tag{15.8}$$

and $\mathscr{L} = \mathscr{L}_{conv} + \mathscr{L}_{dif}$ represents the sum of the nonlinear and linear convective and diffusive operators.

15.3 Numerical Methods

15.3.1 Time Discretization

We consider the time parts of (15.1) and (15.5). The time variable is discretized using the implicit multi-stage methods of second order R_{11} and fourth order R_{22} [HuRoDo02]. The implicit multi-stage method is given in incremental form by

$$\frac{\Delta u}{\Delta t} - \mathbf{W}\triangle u_t = \mathbf{w}u_t^n, \tag{15.9}$$

where the unknown $\Delta u \in \mathbb{R}^n$ is a vector with dimension n. The vector Δu_t is the partial derivative of Δu with respect to time. The time derivatives in (15.9) are replaced by spatial derivatives using the differential equations (15.4). The coefficients in $\mathscr{L}$ are assumed smooth for the accuracy analysis.

$$\frac{\Delta u}{\Delta t} + \mathbf{W}\mathscr{L}(\Delta u) = \mathbf{w}[f^n - \mathscr{L}(u^n)] + \mathbf{W}\Delta f.$$

Here, Δu is defined in (15.9), where $\mathbf{W}$, Δf and $\mathbf{w}$ depends on each particular method. We will linearize the convective term of (15.8), which will become a pointwise linear operator. For illustration we show the compact form for the methods R_{11} and R_{22}.

R_{11} (Crank–Nicolson):

$$\Delta u = u^{n+1} - u^n; \quad \Delta f = f^{n+1} - f^n;$$
$$\mathbf{W} = 1/2; \quad \mathbf{w} = 1.$$

R_{22}:

$$\Delta u = \left\{ \begin{array}{c} u^{n+1/2} - u^n \\ u^{n+1} - u^{n+1/2} \end{array} \right\};$$

$$\Delta f = \left\{ \begin{array}{c} f^{n+1/2} - f^n \\ f^{n+1} - f^{n+1/2} \end{array} \right\};$$

$$\mathbf{W} = \frac{1}{24}\left[\begin{array}{cc} 7 & -1 \\ 13 & 5 \end{array} \right]; \quad \mathbf{w} = \frac{1}{2}\left\{ \begin{array}{c} 1 \\ 1 \end{array} \right\}.$$

15.3.2 Spatial Discretization

We shall now construct a finite-dimensional subspace V_h of $V = H_0^1(0,l)$ formed by piecewise linear functions of the set of m elements of V denoted by $V_h = [\varphi_0, \ldots, \varphi_m]$. The basis functions φ_j are from the finite element method considering a partition $x_0 < x_1 < x_2 \ldots < x_{m-1} < x_m$.

15.3.3 Finite Element Method via Least Squares

Using the implicit multi-stage method defined above for the time discretization of (15.1), the least squares method is applied at each t_{n+1} in (15.9) $n = 0, 1, 2, \ldots, N$, u^n, which are assumed to be known. Let the set of test solutions $V = H_0^1(0,L)$ and the functional

$$\mathscr{F} : V \to \mathbb{R}, \quad u^{n+1} \to \mathscr{F}(u^{n+1}).$$

To minimize the functional $\mathscr{F}$ with respect to u^{n+1} for $n = 0, 1, 2, \ldots, N$, we use the Gâteaux derivative [BeNa08]. Thus, we can solve the variational problem where $u^{n+1} \in V$ is to be found such that

$$a_M(u^{n+1}, w) = F_M(w) \quad \forall w \in V.$$

The problem (15.1)–(15.3) is the solved using LSFEM and considering the subspace $V_h \subset V$, for $n = 0, 1, 2, \ldots, N$. The problem consists then in finding an approximate solution $u_h^{n+1} \in V_h$ such that

$$a_M(u_h^{n+1}, w_h) = F_M(w_h) \quad \forall w_h \in V.$$

15.3.4 Finite Element Method via Galerkin Procedure

Using the implicit multi-stage method defined above for the time discretization of (15.1), the Galerkin method is applied at each t_{n+1} in (15.9) $n = 0, 1, 2, \ldots, N$, u^n and are assumed to be known. Let the set $V = H_0^1(0,L)$, then the weak formulation of the problem is to find $u^{n+1} \in V$ such that $a_G(u^{n+1}, w) = F_G(w)$, $\forall w \in V$. To solve the problem (15.1)–(15.3) using GFEM, we consider the subspace $V_h \subset V$, for $n = 0, 1, 2, \ldots, N$. Thus, the problem consists in finding an approximate solution $u_h^{n+1} \in V_h$ such that

$$a_G(u_h^{n+1}, w_h) = F_G(w_h)\,, \quad \forall w_h \in V_h\,. \tag{15.10}$$

15.3.5 Finite Element Method via Streamline-Upwind Petrov–Galerkin Procedure

The SUPG stabilization for (15.1) is attained by finding $u_h \in V_h$ such that

$$a_G(u_h, w_h) + E_{\text{SUPG}}(u_h, w_h) = F_G(w_h) \quad \forall w_h \in V_h,$$

where $E_{\text{SUPG}}(u_h, w_h)$ indicates the terms of perturbation that are added to the standard variational formulation (15.10). These terms assure that consistency and numerical stability is given by the expression

$$E_{\text{SUPG}}(u_h, w_h) = \sum_{e_j} (\mathscr{P}(w_h), \tau \mathscr{R}(u_h))_{\Omega^j},$$

where $\mathscr{P}(w)$ is a certain operator applied to the test function, τ is the stabilization parameter, and $\mathscr{R}$ is the residual of the differential equation defined by [DoRpHu03]

$$\mathscr{P}(w) = v\frac{\partial w_h}{\partial x},$$

$$\mathscr{R} = v\frac{\partial u_h}{\partial x} - D\frac{\partial^2 u_h}{\partial x^2} + \sigma u_h - f,$$

$$\tau = \left(\frac{2v}{h} + \frac{4D}{h^2} + \sigma\right)^{-1} = \frac{h}{2v}\left(1 + \frac{1}{Pe} + \frac{h\sigma}{2v}\right)^{-1}.$$

Here, h is the size of the grid, Pe is the Péclet number and v, D and σ are the coefficients defined in equation (15.1). To solve the problem (15.1)–(15.3) using SUPG, one considers the subspace $V_h \subset V$ for $n = 0, 1, 2, \ldots, N$ and determines an approximate solution $u_h^{n+1} \in V_h$ such that

$$a_G(u_h^{n+1}, w_h) + E_{\text{SUPG}}(u_h^{n+1}, w_h) = F_G(w_h) \quad \forall w_h \in V_h.$$

Next, we linearize the convective term in (15.5), which changes the size of the element in each stage using the information from the previous step [KuEsDa04] that casts the Burgers equation into a linear local problem.

15.3.6 Linearization of the Convective Term

Upon multiplying both sides of (15.5) by a test function $w \in V$ and integrating out the x-degree of freedom yields

$$\int_0^l (u_t + uu_x - \varepsilon u_{xx} - f) w \, dx = 0. \tag{15.11}$$

A numerical solution to problem (15.5)–(15.7) is constructed in the region $0 \leq x \leq l$ with boundary conditions specified at $x = 0$ and $x = l$. To this end, we consider the finite dimensional subspace V_h, where the basis functions φ_j are from the finite element method considering a partition $x_0 < x_1 < x_2 \ldots < x_{m-1} < x_m$ of size

$$h_j = x_j - x_{j-1}.$$

We now construct a test function u_h, and the parameters that are to describe the function u_h are the values $u_0, u_1, u_2 \ldots, u_m$ at the nodes x_j. Therefore, we can write the approximate equation (15.11)

$$\sum_{j=0}^{m} \int_0^l \left(\frac{\partial u_j}{\partial t} \varphi_i(x) + \eta \frac{\partial \varphi_j(x)}{\partial x} \varphi_i(x) u_j \right.$$
$$\left. - \varepsilon \frac{\partial^2 \varphi_j(x)}{\partial x^2} \varphi_i(x) u_j - f \varphi_i u_j \right) dx = 0 \quad \forall \varphi_i, \varphi_j \in V_h,$$

where $\eta = u_0 \Delta t / h_j$ and Δt is the time step, and $w_h = \varphi_i(x)$, $i = 0, 1, 2, \ldots, m$. Thus, the Burgers equation becomes a 1D linear local problem.

Now, we consider the development for the 1D convection–diffusion–reaction equation in this case $\sigma = 0$, $D = \varepsilon$, and $v = \eta$. For the Burgers equation, the value of the stabilization parameter τ, which is used by SUPG [DoRpHu03], is

$$\tau = \left((2u/h)^2 + 9 \left(4\varepsilon / (h^2) \right)^2 \right)^{-1/2},$$

where h is the size of the grid and $\varepsilon = 1/Re$, with Re and u defined in (15.5).

15.4 Numerical Results

15.4.1 1D Convection–Diffusion–Reaction Equation

Consider the 1D convection–diffusion–reaction problem (15.1)–(15.3) with the function $f(x,t) = 0$ and the initial condition given by a Gaussian distribution

$$u(x,0) = \exp\left\{ - \left(\frac{x - x_0}{\ell} \right)^2 \right\}.$$

For a linear decay term, $-\sigma u$, the analytical solution on $-\infty < x < \infty$ is [DoRpHu03]

$$u(x,t) = \frac{\exp(-\sigma t)}{\gamma(t)} \exp\left\{ - \left(\frac{x - x_0 - vt}{\ell \gamma(t)} \right)^2 \right\}, \tag{15.12}$$

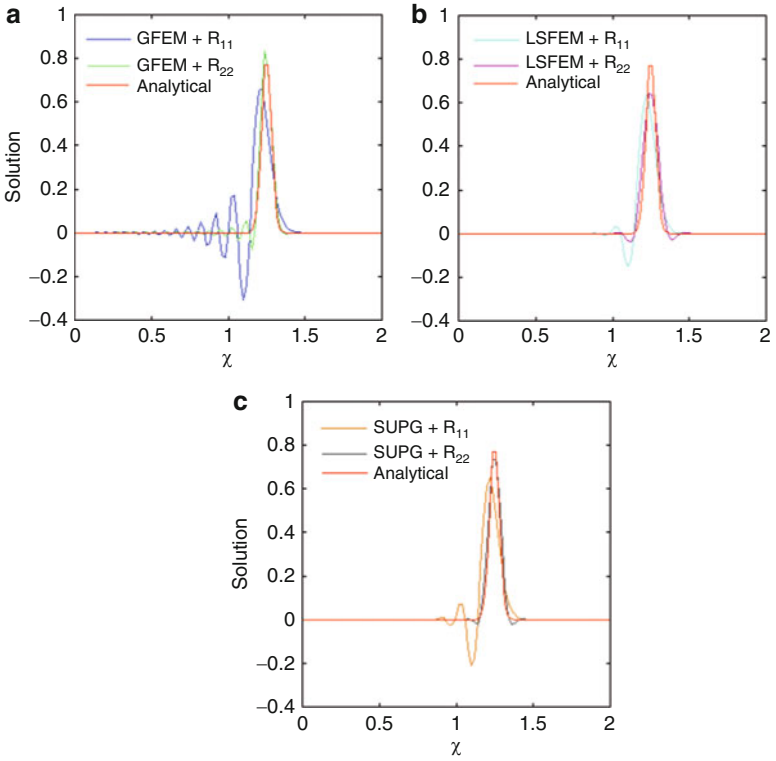

Fig. 15.1 Comparisons of the Padé approximants R_{11} and R_{22} together with the formulations (**a**) GFEM, (**b**) LSFEM, and (**c**) SUPG

where

$$\gamma(t) = \sqrt{1 + \frac{4Dt}{\ell^2}}.$$

For this example we consider $0 \leq x \leq l$, $l = 2$ the domain of the 1D problem. For illustration, we present some results with 100 linear elements and

$$x_0 = 1/4, \quad \ell = 1/25, \quad v = 1, \quad \sigma = 0.1, \quad C = 1, \quad Pe = 100,$$

where v and σ are the coefficients of (15.1) and C and Pe are the Courant and Péclet numbers, respectively.

In Fig. 15.1 we present comparisons between the Padé approximants of R_{11} and R_{22} modified by the formulations GFEM, LSFEM, and SUPG, with $\Delta t = \Delta x = 0.02$. The analysis of stability and convergence are shown for the time limit $t = 1$ and compared with the analytical solution (15.12). One observes in Fig. 15.1 that the implicit multi-stage method of fourth-order R_{22} modified by the formulations

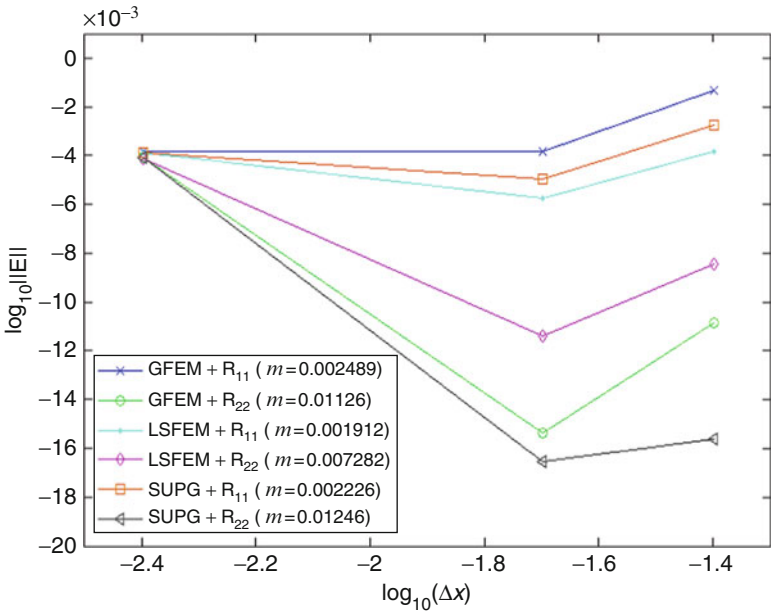

Fig. 15.2 Convergence of the numerical results with the grid refinement for the example 1D convection-diffusion-reaction problem

GFEM, LSFEM, and SUPG smoothed out the numerical oscillations. We present the errors between the methods evaluated for function grid refinement ($h = 1/50$, $h = 1/680$ and $h = 1/1000$) in Fig. 15.2 and for function time step refinement ($\Delta t = 0.5$, $\Delta t = 0.05$ and $\Delta t = 0.01$) in Fig. 15.3 using the L^2-norm.

15.4.2 The Burgers Equation

We consider an analytical solution for the Burgers equation (15.5) and (15.6) given by [KuEsDa04]

$$u(x,t) = \frac{2\varepsilon\pi \exp(\pi^2 \varepsilon t)\sin(\pi x)}{a + \exp(-\pi^2 \varepsilon t)\cos(\pi x)}, \quad a > 1,$$

with initial condition

$$u(x,0) = \frac{2\varepsilon\pi \sin(\pi x)}{a + \cos(\pi x)}, \quad a > 1,$$

where $\varepsilon = 1/Re$ is the coefficient of viscosity of the fluid and Re represents the Reynolds number. Let $0 \leq x \leq 1$ be the domain with the boundary conditions

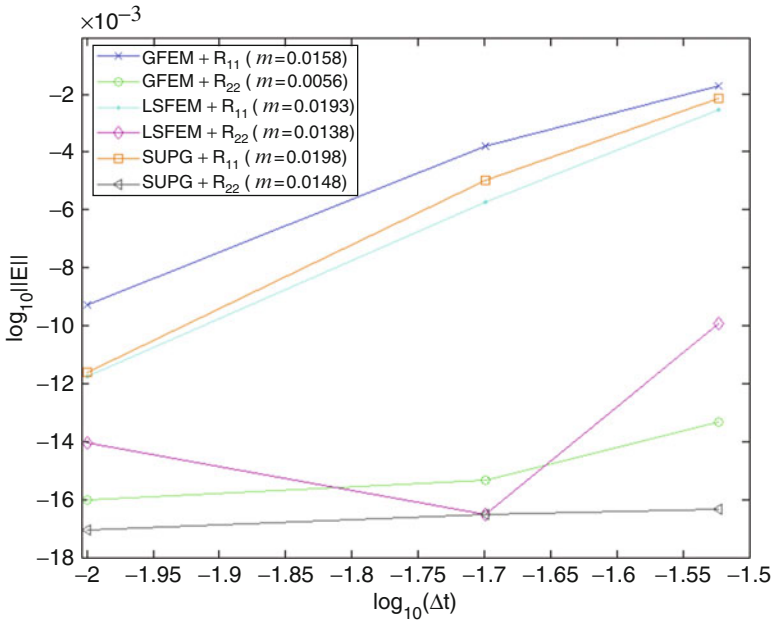

Fig. 15.3 Convergence of the numerical results with the time step refinement for the example 1D convection-diffusion-reaction problem

$u(0,t) = u(1,t) = 0$. To illustrate some results we used 50 linear elements and $Re = 10000$.

Figure 15.4 presents comparisons between the Padé approximants of R_{11} and R_{22} modified by the formulations GFEM, LSFEM, and SUPG, respectively, with $\Delta t = \Delta x = 0.02$ and as an analysis of stability and convergence we present the results of the formulations for the upper time limit $t = 1$ and compare these findings with the analytical solution (15.12). One observes in Fig. 15.5, that the implicit multi-stage method of fourth-order R_{22}, modified by the formulations GFEM, LSFEM and SUPG smoothed out numerical oscillations. We present the errors between the methods, evaluated for function grid refinement ($h = 2/50$, $h = 2/100$, and $h = 2/500$), in Fig. 15.5 and for function time step refinement ($\Delta t = 0.03$, $\Delta t = 0.02$ and $\Delta t = 0.01$) in Fig. 15.6 using the L^2 norm.

15.5 Conclusions

We conclude that the implicit multi-stage method of fourth-order R_{22}, when complemented by the finite element methods studied here, proved efficient since the Padé approximant R_{22} increased the convergence region of the numerical solutions.

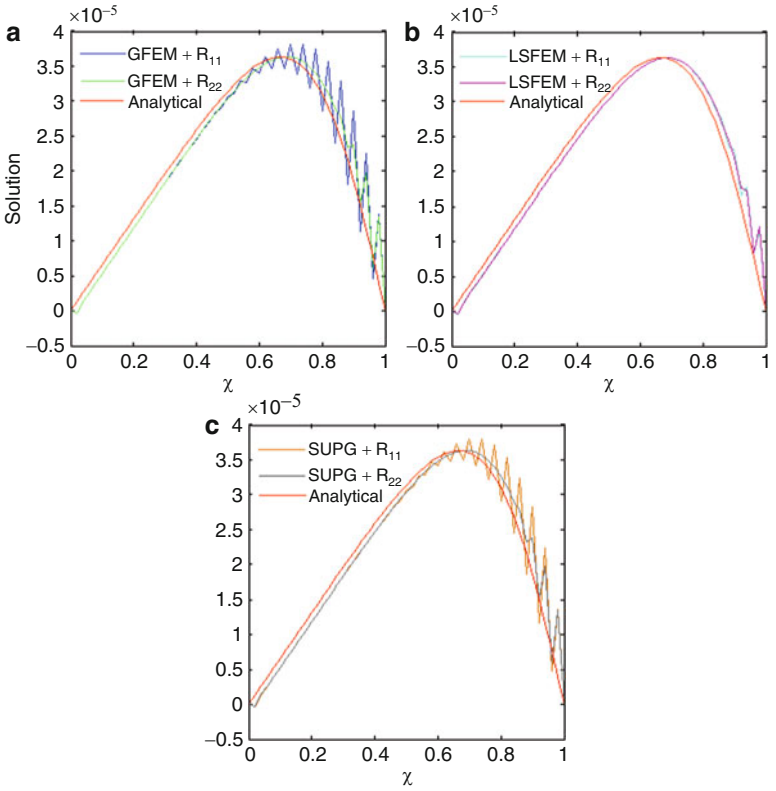

Fig. 15.4 Comparisons between the Padé approximants R_{11} and R_{22} modified by the formulations (**a**) GFEM, (**b**) LSFEM and (**c**) SUPG

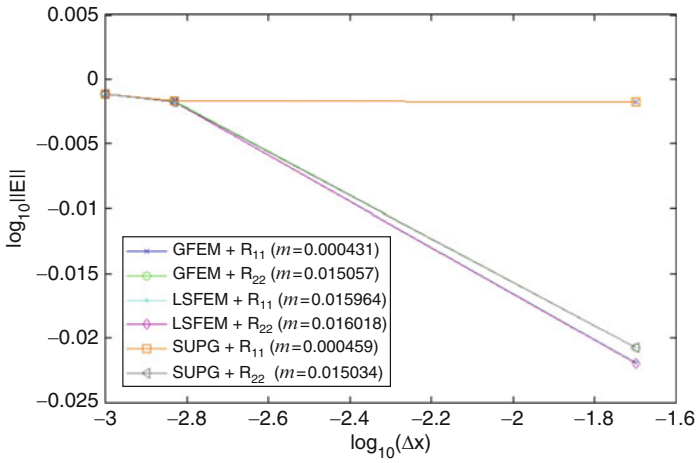

Fig. 15.5 Convergence of the numerical results with grid refinement for the example the Burgers equation

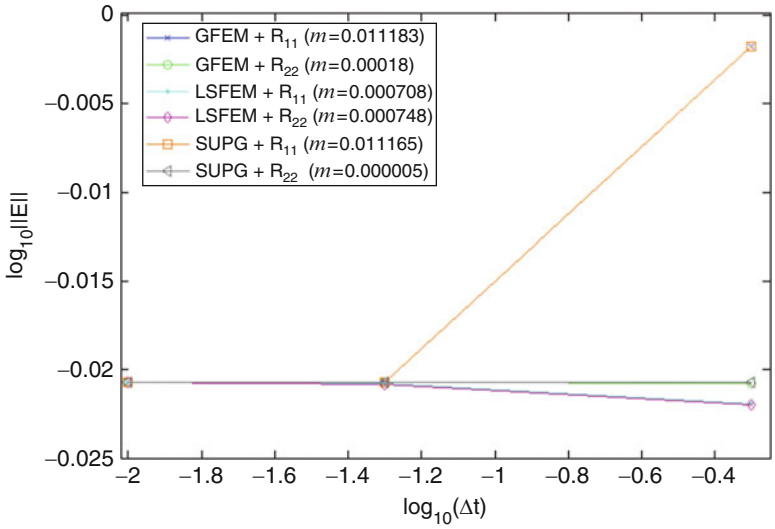

Fig. 15.6 Convergence of the numerical results with time step refinement for the example the Burgers equation

We also note that the LSFEM eliminated the oscillations of numerical solutions more efficiently than the methods GFEM and SUPG.

References

[BeNa08] Behmardi, D., Nayeri, D.E.: Introduction of Fréchet and Gâteaux derivative. Appl. Math. Sci. **2**, 975–980 (2008)

[DoRoHu00] Donea, J., Roig, B., Huerta, A.: Higher-order accurate time-stepping schemes for convection-difusion problems. Comput. Meth. Appl. Mech. Eng. **182**, 249–275 (2000)

[DoRpHu03] Donea, J., Roig, B., Huerta, A.: Finite Element Methods for Flow Problems. Wiley, Chichester (2003)

[GoCoCa00] Gomes, H., Colominas, I., Casteleiro, E.M.: Finite element model and applications. Int. J. Numer. Meth. Eng. **00**, 1–6 (2000)

[HuRoDo02] Huerta, A., Roig, B., Donea, J.: Time-accurate solution of stabilized convection–diffusion–reaction equations. II: accuracy analysis and examples. Comm. Numer. Meth. Eng. **18**, 575–584 (2002)

[KuEsDa04] Kutluay, S., Esen A., Dag, I.: Numerical solutions of the Burgers' equation by the least-squares quadratic B-spline finite element method. J. Comput. Appl. Math. **167**, 21–33 (2004)

[OdEtAl03] Oden, J.T., Belytschko, T., Babuska, I., Hughes, J.R.: Research directions in computational mechanics. Comput. Meth. Appl. Mech. Eng. **192**, 913–922 (2003)

[RoSa07] Rodríguez–Ferran, A., Sandoval, M.L.: Numerical performance of incomplete factorizations for 3D transient convection-diffusion problems. Adv. Eng. Software **38**, 439–450 (2007)

[TaShDe07] Tabatabaei, A.E.A.H.H., Shakour, E., Dehghan, M.: Some implicit methods for the numerical solution of Burgers' equation. Appl. Math. Comput. **191**, 560–570 (2007)
[TiYu11] Tian, Z.F., Yu, P.X.: A High-order exponential scheme for solving 1D unsteady convection–difusion equations. J. Comput. Appl. Math. **235**, 2477–2491 (2011)
[Ve04] Venutelli, M.: Time-stepping Padé–Petrov–Galerkin models for hydraulic jump simulation. Math. Comput. Simulat. **66**, 585–604 (2004)

Chapter 16
Analytical Reconstruction of Monoenergetic Neutron Angular Flux in Non-multiplying Slabs Using Diffusion Synthetic Approximation

R.S. Mansur and R.C. Barros

16.1 Introduction

In this chapter we describe two analytical reconstruction schemes for the neutron angular flux for fixed-source one-speed transport problems in slab geometry [DuHa76]. To be more specific, the spatial reconstruction scheme expresses the diffusion solution in each spatial cell as a linear combination of two basis functions. The latter are determined by spectral analysis with boundary conditions assigned at the cell edges that are given by the spectral nodal results. Thus one obtains a system of two linear algebraic equations in two unknown expansion coefficients, which then yield the exact diffusion solution at each point in the cell. Note that the spectral nodal method for diffusion (SND) is free from all spatial truncation errors. Moreover, the angular reconstruction scheme yields an approximate angular flux at any angular direction $-1 \leq \mu \leq 1$, $\mu \neq 0$, where $\mu = \cos\theta$, with $0 \leq \theta \leq \pi$ being the polar angle, $\theta \neq \pi/2$ [LeMi93]. To achieve this goal, we substitute the local solution within each discretized cell of the spatial grid as determined by the spatial reconstruction scheme, into the integral source terms of the analytical first-order form of the neutron transport equation in slab geometry with linearly anisotropic scattering [LeMi93] and solve the resulting approximate differential equation analytically. This method is referred to as a synthetic method since it uses a lower-order model, which is diffusion, to solve a higher-order equation, which is the neutron transport equation.

Now we outline the content of the remainder of this chapter: in the next section we describe the spatial and the angular reconstruction schemes of the SND coarse-mesh solution. Further we present numerical results to a multilayer fixed source

R.S. Mansur • R.C. Barros (✉)
University of the State of Rio de Janeiro, Programa de Pós-graduação em Ciências Computacionais, Rua São Francisco Xavier 524, 20550-013, Rio de Janeiro, RJ, Brazil
e-mail: ralph@ime.uerj.br; rcbarros@pq.cnpq.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_16,

problem in slab geometry with linearly anisotropic scattering, and then we give a number of concluding remarks.

16.2 The Spatial and the Angular Reconstruction Schemes of the SND Coarse-Mesh Numerical Solution

16.2.1 The Spatial Reconstruction Scheme

Let us consider Fig. 16.1 which represents a spatial discretization grid wherein each cell is also termed node Ω_i.

We begin by describing the spectral analysis we perform to the diffusion equation inside node Ω_i in order to obtain the present SND method. Therefore, it is convenient to write the one-speed slab-geometry diffusion equation in Ω_i in the form

$$\frac{d}{dx}J(x) + \Sigma_{a_i}\phi(x) = Q_i, \tag{16.1}$$

$$J(x) = -D_i\frac{d}{dx}\phi(x),\ 0 \le x \le \Omega_i, \tag{16.2}$$

where (16.1) is the neutron continuity equation and (16.2) is the classical Fick's law, the essence of diffusion theory. Moreover we have defined

$\phi(x)$: Neutron scalar flux;
$J(x)$: Total current;
Σ_{a_i}: Macroscopic absorption cross section;
D_i: Diffusion coefficient;
Q_i: Uniform and isotropic interior source.

At this point we apply the operator $\frac{1}{h_i}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} dx$ in (16.1) and (16.2) to obtain the discretized balance equations

$$\frac{J_{i+\frac{1}{2}} - J_{i-\frac{1}{2}}}{h_i} + \Sigma_{ai}\,\bar{\phi}_i = Q_i \tag{16.3}$$

$$\bar{J}_i = -\frac{D_i}{h_i}\left(\phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}\right). \tag{16.4}$$

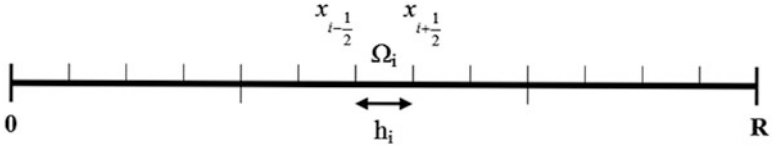

Fig. 16.1 Spatial node Ω_i of width h_i

Here we have defined the average values within the spatial cell Ω_i as

$$\bar{g}_i = \frac{1}{h_i} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} g(x)\, dx, \quad g = J \text{ or } g = \phi. \tag{16.5}$$

Now we write two expressions for solutions of (16.1) and (16.2) in Ω_i as

$$\phi(x) = a\, e^{-\Sigma_{a_i} x/\nu} + \phi^p, \tag{16.6}$$

$$J(x) = b\, e^{-\Sigma_{a_i} x/\nu} + J^p, \tag{16.7}$$

where the superscript p denotes particular solutions and for simplicity we have dropped the subscript i. Then we substitute (16.6) and (16.7) into (16.1) and (16.2) to obtain the system of two algebraic equations

$$\Sigma_{a_i} \left(a\, e^{-\frac{\Sigma_{a_i} x}{\nu}} \right) - \frac{\Sigma_{a_i}}{\nu} \left(b\, e^{-\frac{\Sigma_{a_i} x}{\nu}} \right) = Q_i - \Sigma_{a_i} \phi^p, \tag{16.8}$$

$$-\frac{D_i \Sigma_{a_i}}{\nu} \left(a\, e^{-\frac{\Sigma_{a_i} x}{\nu}} \right) + b\, e^{-\frac{\Sigma_{a_i} x}{\nu}} = -J^p. \tag{16.9}$$

We determine the particular solutions by setting the right-hand sides of (16.8) and (16.9) equal to zero. Therefore, we obtain

$$\phi^p = \frac{Q_i}{\Sigma_{a_i}},$$

$$J^p = 0.$$

By solving the resulting homogeneous system for nontrivial solution, we obtain

$$\nu = \pm \sqrt{D_i\, \Sigma_{a_i}}. \tag{16.10}$$

As it stands, (16.6) and (16.7) appear as

$$\phi(x) = a\, e^{-\sqrt{\frac{\Sigma_{a_i}}{D_i}} x} + \frac{Q_i}{\Sigma_{a_i}}, \tag{16.11}$$

$$J(x) = b\, e^{-\sqrt{\frac{\Sigma_{a_i}}{D_i}} x}. \tag{16.12}$$

Furthermore, the undetermined homogeneous system appears as

$$\Sigma_{a_i}\, a - \sqrt{\frac{\Sigma_{a_i}}{D_i}}\, b = 0,$$

$$-\sqrt{D_i\, \Sigma_{a_i}}\, a + b = 0.$$

By choosing $a = 1$, we obtain $b = \sqrt{D_i \Sigma_{a_i}}$, and (16.11) and (16.12) can be written as

$$\phi(x) = e^{-\sqrt{\frac{\Sigma_{a_i}}{D_i}}x} + \frac{Q_i}{\Sigma_{a_i}}, \tag{16.13}$$

$$J(x) = \sqrt{D_i \Sigma_{a_i}}\, e^{-\sqrt{\frac{\Sigma_{a_i}}{D_i}}x}.$$

Therefore, the expressions for the local analytical general solutions of (16.1) and (16.2), considering (16.10), are given by

$$\phi(x) = C_1 e^{-\sqrt{\frac{\Sigma_{a_i}}{D_i}}x} + C_2 e^{\sqrt{\frac{\Sigma_{a_i}}{D_i}}x} + \frac{Q_i}{\Sigma_{a_i}}, \tag{16.14}$$

$$J(x) = C_1 \sqrt{D_i \Sigma_{a_i}}\, e^{-\sqrt{\frac{\Sigma_{a_i}}{D_i}}x} + C_2 \sqrt{D_i \Sigma_{a_i}}\, e^{\sqrt{\frac{\Sigma_{a_i}}{D_i}}x},\ x \in \Omega_i. \tag{16.15}$$

Furthermore, to derive the discretized equations of the SND method, we write two auxiliary equations in the form

$$\bar{\phi}_i = \frac{\gamma_i}{2}\left(\phi_{i+\frac{1}{2}} + \phi_{i-\frac{1}{2}}\right) + G(Q_i), \tag{16.16}$$

$$\bar{J}_i = \frac{\beta_i}{2}\left(J_{i+\frac{1}{2}} + J_{i-\frac{1}{2}}\right). \tag{16.17}$$

In order to determine expressions for γ_i and β_i , we first substitute (16.13) into (16.16). By using definition (16.5) and defining the diffusion length $L_i = \sqrt{\frac{D_i}{\Sigma_{ai}}}$, we obtain

$$\bar{\phi}_i = \frac{L_i}{h_i}\left(e^{-\frac{x_{i-\frac{1}{2}}}{L_i}} - e^{-\frac{x_{i+\frac{1}{2}}}{L_i}}\right) + \frac{Q_i}{\Sigma_{ai}},$$

which can be substituted into the auxiliary equation (16.16) to obtain

$$G(Q_i) = \frac{(1-\gamma_i)\, Q_i}{\Sigma_{ai}},$$

and then

$$\gamma_i = 2\,\frac{L_i}{h_i}\tanh\left(\frac{h_i}{2L_i}\right). \tag{16.18}$$

Using a similar procedure, we obtain an expression for β_i which is identical to (16.18). Therefore, the auxiliary equations (16.16) and (16.17) appear as

$$\bar{\phi}_i = \frac{\gamma_i}{2}\left(\phi_{i+\frac{1}{2}} + \phi_{i-\frac{1}{2}}\right) + \frac{(1-\gamma_i)\,Q_i}{\Sigma_{ai}}, \tag{16.19}$$

$$\bar{J}_i = \frac{\gamma_i}{2}\left(J_{i+\frac{1}{2}} + J_{i-\frac{1}{2}}\right). \tag{16.20}$$

Now, we substitute (16.19) and (16.20) into (16.3) and (16.4) to obtain

$$J_{i+\frac{1}{2}} - J_{i-\frac{1}{2}} = h_i\,\gamma_i\,Q_i - \frac{h_i\,\gamma_i\,\Sigma_{ai}}{2}\left(\phi_{i+\frac{1}{2}} + \phi_{i-\frac{1}{2}}\right), \tag{16.21}$$

$$J_{i+\frac{1}{2}} + J_{i-\frac{1}{2}} = -\frac{2\,D_i}{h_i\,\gamma_i}\left(\phi_{i+\frac{1}{2}} + \phi_{i-\frac{1}{2}}\right). \tag{16.22}$$

By summing (16.21) and (16.22) and then subtracting (16.21) from (16.22), we obtain, respectively,

$$J_{i+\frac{1}{2}} = -\frac{D_i}{h_i\,\gamma_i}\left(\phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}\right) - \frac{h_i\,\gamma_i\,\Sigma_{ai}}{4}\left(\phi_{i+\frac{1}{2}} + \phi_{i-\frac{1}{2}}\right) + \frac{h_i\,\gamma_i}{2}\,Q_i, \tag{16.23}$$

$$J_{i-\frac{1}{2}} = -\frac{D_i}{h_i\,\gamma_i}\left(\phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}\right) + \frac{h_i\,\gamma_i\,\Sigma_{ai}}{4}\left(\phi_{i+\frac{1}{2}} + \phi_{i-\frac{1}{2}}\right) - \frac{h_i\,\gamma_i}{2}\,Q_i. \tag{16.24}$$

To proceed, we substitute the subscript i for the subscript $(i+1)$ in (16.24) to write

$$J_{i+\frac{1}{2}} = -\frac{D_{i+1}}{h_{i+1}\,\gamma_{i+1}}\left(\phi_{i+\frac{3}{2}} - \phi_{i+\frac{1}{2}}\right) + \frac{h_{i+1}\,\gamma_{i+1}\,\Sigma_{a_{i+1}}}{4}\left(\phi_{i+\frac{3}{2}} + \phi_{i+\frac{1}{2}}\right) - \frac{h_{i+1}\,\gamma_{i+1}}{2}\,Q_{i+1}. \tag{16.25}$$

By comparing (16.25) and (16.23), we obtain an equation involving three consecutive node–edge scalar fluxes on the left-hand side, and the sources Q_i and Q_{i+1} on the right-hand side. This equation is used for the interior nodes Ω_i, $i = 2 : I-1$.

For the first spatial cell, we set $i = 1$ in (16.24) to write

$$J_{\frac{1}{2}} = -\frac{D_1}{h_1\,\gamma_1}\left(\phi_{\frac{3}{2}} - \phi_{\frac{1}{2}}\right) + \frac{h_1\,\gamma_1\,\Sigma_{a1}}{4}\left(\phi_{\frac{3}{2}} + \phi_{\frac{1}{2}}\right) - \frac{h_1\,\gamma_1}{2}\,Q_1. \tag{16.26}$$

Furthermore, we introduce the generalized left boundary condition as

$$J_{\frac{1}{2}} = I_0 - \alpha_0\phi_{\frac{1}{2}}, \tag{16.27}$$

where I_0 is the isotropic incident flux at $x = 0$ and α_0 is an appropriate constant, which is defined in Table 16.1. Using (16.26) into (16.27), we obtain an equation

Table 16.1 Boundary conditions

	Prescribed	Reflexive	Vacumm	Zero scalar-flux
α_m	0.5	0	0.5	∞
I_m	I^{a}	0	0	0

[a]Input data

involving $\phi_{\frac{1}{2}}$ and $\phi_{\frac{3}{2}}$ on the left-hand side, and the source Q_1 and I_0 on the right-hand side. This equation is used for the first spatial cell Ω_1.

For the last spatial cell, we set $i = I$ in (16.23), and we obtain

$$J_{I+\frac{1}{2}} = -\frac{D_I}{h_I\,\gamma_I}\left(\phi_{I+\frac{1}{2}} - \phi_{I-\frac{1}{2}}\right) - \frac{h_I\,\gamma_I\,\Sigma_{aI}}{4}\left(\phi_{I+\frac{1}{2}} + \phi_{I-\frac{1}{2}}\right) + \frac{h_I\,\gamma_I}{2}\,Q_I. \tag{16.28}$$

Similarly, we substitute the generalized boundary condition at $x = R$ (see Fig. 16.1), that is,

$$J_{I+\frac{1}{2}} = \alpha_R \phi_{I+\frac{1}{2}} - I_R,$$

into (16.28), where I_R is the isotropic incident flux at $x = R$ and α_R is an appropriate constant, which is also defined in Table 16.1. The result involves $\phi_{I+\frac{1}{2}}$ and $\phi_{I-\frac{1}{2}}$ on the left-hand side, and the source Q_I and I_R on the right-hand side. This equation is used for the last spatial cell Ω_I.

At this point we remark that the parameters α_m and I_m, $m \in \{0, R\}$, depend on the boundary conditions assigned to the diffusion equation and are given in Table 16.1.

The expressions for the interior interfaces and boundaries result in an algebraic linear system with $I+1$ equations in $I+1$ unknowns ϕ_i, $i = 1/2 : I + 1/2$. The coefficient matrix of this linear system is symmetric, tridiagonal and has the characteristic of being diagonal dominant, which discards the use of pivoting strategies to solve the system for the scalar flux at the node edges [BuFa85].

As mentioned earlier in this chapter, the numerical solution generated by the SND method has no spatial truncation errors. Therefore, it generates accurate numerical results, even for coarse discretization grids set up on the domain.

To proceed further with the spatial reconstruction scheme, we first set $x_{i-\frac{1}{2}} = 0$ and $x_{i+\frac{1}{2}} = h_i$, $i = 1 : I$, in (16.14), which is the analytical solution inside Ω_i. As the values of $\phi(0)$ and $\phi(h_i)$ are determined by the SND method, we solve the system

$$\phi(0) = C_1 + C_2 + \frac{Q_i}{\Sigma_{a_1}}, \tag{16.29}$$

$$\phi(h_i) = C_1\, e^{\frac{h_i}{L_i}} + C_2\, e^{\frac{h_i}{L_i}} + \frac{Q_i}{\Sigma_{a_1}}, \tag{16.30}$$

for the two unknowns C_1 and C_2.

To conclude the spatial reconstruction scheme, we substitute the constants C_1 and C_2 back into the local general solution (16.14), so we can evaluate the scalar flux at any point $x \in \Omega_i$. We note that all spatial cells of the same material zone that have the same width may have different pairs of constants C_1 and C_2 , since the right-hand side of system (16.14) is generally different in each Ω_i, as it depends on the coarse-mesh results generated by the SND method in accordance with (16.29) and (16.30).

16.2.2 The Angular Reconstruction Scheme

Once we have described a technique to perform spatial reconstruction within each spatial node Ω_i, $i = 1 : I$, of the coarse-mesh discretization grid, we proceed to describing a synthetic method to perform angular reconstructions for any possible value (except $\mu = 0$) of the direction-of-motion variable μ.

The monoenergetic neutron transport equation in slab geometry with linearly anisotropic scattering appears as

$$\mu \frac{\partial}{\partial x}\psi(x,\mu) + \Sigma_T \psi(x,\mu) = \frac{\Sigma_{s_0}}{2}\int_{-1}^{1} \psi(x,\mu')d\mu' + \frac{3}{2}\mu\Sigma_{s_1}\int_{-1}^{1} \mu'\psi(x,\mu')d\mu' + \frac{Q}{2}, \quad (16.31)$$

where

$\psi(x,\mu)$: Neutron angular flux in direction μ;
Σ_T : Total macroscopic cross section;
Σ_{s_0} : Zero'th-order term of the differential scattering macroscopic cross section;
Σ_{s_1}: First-order term of the differential scattering macroscopic cross section.

The zero'th-order term of the differential scattering macroscopic cross section and the total macroscopic cross section are defined, respectively, by $\Sigma_{s_0} = \Sigma_T - \Sigma_a$ and $\Sigma_T = \frac{1}{3D} + \Sigma_{s_1}$ [LeMi93].
Therefore, let us use the standard definition of neutron scalar flux and total current

$$\phi(x) = \int_{-1}^{1} \psi(x,\mu)d\mu,$$

$$J(x) = \int_{-1}^{1} \mu\,\psi(x,\mu)d\mu.$$

Moreover, using the expression for the local general solutions (16.14) and (16.15) in (16.31), we solve the resulting equation analytically for the neutron angular flux $\psi(x,\mu)$, $x \in \Omega_i$. For $0 < \mu \leq 1$, the result is

$$\psi(x,\mu) = \psi(0,\mu)\, e^{-\Sigma_T \frac{x}{\mu}} + \frac{1}{\mu} \sum_{\ell=1}^{2} \frac{C_\ell\, \xi_\ell \left(e^{-\Sigma_T \frac{x}{\mu}} - e^{-\Sigma_a \frac{x}{\nu_\ell}} \right)}{\frac{\Sigma_a}{\nu_\ell} - \frac{\Sigma_T}{\mu}} + \frac{Q}{2\Sigma_a} \left(1 - e^{-\Sigma_T \frac{x}{\mu}} \right). \quad (16.32)$$

where we have defined $\xi_\ell = \frac{\Sigma_{s_0}}{2} + \frac{3}{2}\Sigma_{s_1}\, \mu\, \nu_\ell$.

Furthermore, for $-1 \leq \mu < 0$, we obtain the result

$$\psi(x,\mu) = \psi(h,\mu)\, e^{-\Sigma_T \frac{h-x}{|\mu|}} + \frac{1}{|\mu|} \sum_{\ell=1}^{2} \frac{C_\ell\, \xi_\ell \left(e^{-\Sigma_a \frac{x}{\nu_\ell}} - e^{-\Sigma_T \frac{h-x}{|\mu|}} \cdot e^{-\Sigma_a \frac{h}{\nu_\ell}} \right)}{\frac{\Sigma_a}{\nu_\ell} - \frac{\Sigma_T}{|\mu|}} + \frac{Q}{2\Sigma_a} \left(1 - e^{-\Sigma_T \frac{h-x}{|\mu|}} \right).$$

Here, $x = 0$ is the left-hand edge of node Ω_i and $x = h_i$ is the right-hand edge of node Ω_i , i.e., $0 \leq x \leq h_i$.

Equations (16.32) and (16.33) are the analytical solutions of the approximate neutron transport equation for all direction-of-motion variables $-1 \leq \mu \leq 1$, $\mu \neq 0$, wherein we have considered the local general solution for the scalar flux and total current given by (16.14) and (16.15). We remark that the values of C_1 and C_2 are determined by solving the linear system of (16.29) and (16.30), as described for the spatial reconstruction scheme. Moreover, we see from (16.32) and (16.33) that the present angular reconstruction scheme is not valid for the particular case of $\mu = 0$.

16.3 Numerical Results

Let us consider a heterogeneous model problem that consists of a three-layer slab of thickness $R = 50$ cm and two material zones, represented in Fig. 16.2. Vacuum boundary conditions apply at $x = 0$ and $x = 50$. The central layer has a constant unit source and the cross sections for the two material zones are listed in Table 16.2.

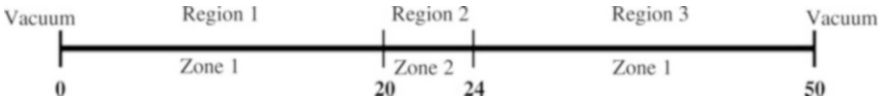

Fig. 16.2 Model problem

Table 16.2 Parameters for the two material zones

Zone	D	Σ_a	Σ_T	Σ_{s_1}
1	0.33333	0.1	1.0	0.8
2	0.37037	0.2	0.9	0.6

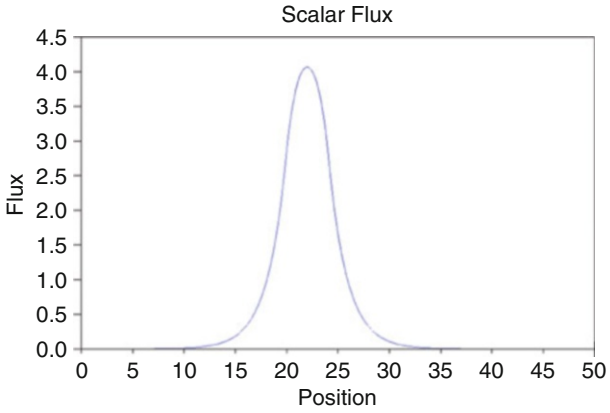

Fig. 16.3 Scalar flux

Table 16.3 Numerical results generated by the synthetic angular reconstruction (S_{16})

Positon (cm)	Direction S_{16} Gauss–Legendre	Synthetic angular reconstruction	$DD - S_{16}^{a}$	Percent relative deviation
2	0.0950125	0.0101181	0.0105915	4.47
8	−0.4580168	0.089064	0.0884181	0.73
15	−0.8656312	0.7213077	0.6725081	7.26
20	−0.0950125	1.5299474	1.5604457	1.95
22	0.0950125	1.7115275	1.7738038	3.51
24	−0.7554044	0.1495803	0.1463598	2.20
49	−0.4580168	0.000703	0.0007441	5.52
Execution time:		18.266 s	818.064 s	–

[a]Diamond difference—Gauss–Legendre S_{16} [Ly11]

We ran the SND code on a spatial grid composed of one node per layer and the results for the scalar flux profile as generated with the present spatial reconstruction are plotted in Fig. 16.3, where we used a step of 0.01 cm to build the graph. The numerical results generated by the angular synthetic reconstruction are listed in Table 16.3. In the first column of the table, we present various positions of the domain where the angular flux were calculated in several directions, positive and negative, of the conventional Gauss–Legendre S_{16} angular quadrature [LeMi93], that are listed in the second column of Table 16.3. The third column lists the numerical values for the angular flux generated by the offered synthetic diffusion angular reconstruction method and in the fourth column we have the numerical values generated by the Diamond Difference (DD) method using the Gauss–Legendre S_{16} angular quadrature [LeMi93]. As we see in the fifth column, the percent relative deviations were generally small, and for the numerical results listed in Table 16.3, the maximum value of 7.26% occurred at position 15 cm in direction $\mu = -0.8656312$.

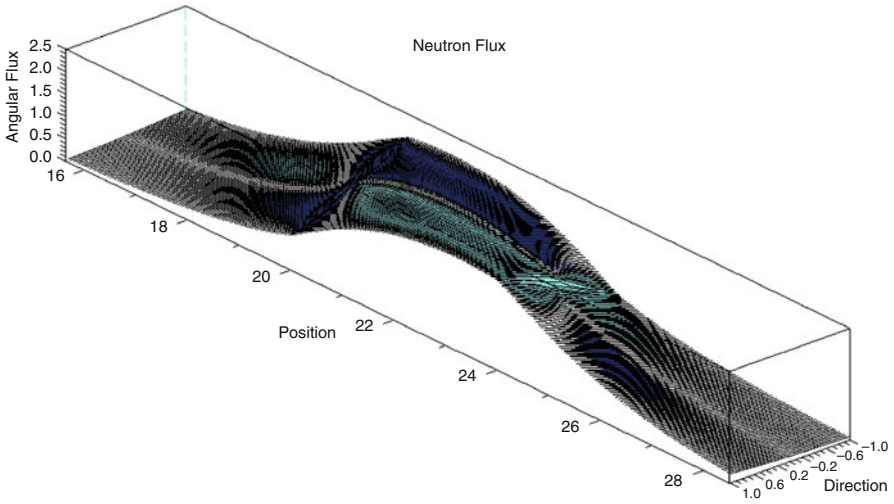

Fig. 16.4 Angular flux for $16 \leq x \leq 28$

The results for the neutron angular flux profile as generated with the present angular reconstruction are plotted in Fig. 16.4, for $16 \leq x \leq 28$, region where a unit neutron source is located. The graph indicates that, from the center of the region to the left side of the domain, the angular flux is higher for the negative directions; whereas from the center to right, the flux is higher for the positive directions.

16.4 Conclusions

The SND method generates numerical solution with no spatial truncation errors for one-speed slab-geometry diffusion problems, regardless of the spatial grid setup on the domain, but apart from computational finite arithmetic considerations. Therefore, one may be able to solve slab-geometry one-speed diffusion problems with many fewer spatial cells than standard numerical methods, e.g., the classical finite-difference method. On the other hand, as a drawback of coarse-mesh numerical methods, we note that they do not generate localized quantities that frequently are needed, as the grid points may be considerably away from each other. Therefore, we have described in this chapter two numerical algorithms to reconstruct the coarse-mesh solution within each discretization spatial node, i.e., the spatial reconstruction scheme and the synthetic angular reconstruction scheme. According to the numerical results generated for the model problem considered in the previous section by the present reconstruction schemes, we conclude that they are reasonably accurate with respect to the direct calculations. In addition, as we see in the last row of Table 16.3, the offered synthetic angular reconstruction scheme generates accurate results in much less computational running time than the

$DD - S_{16}$ method. Before closing this chapter, we remark that for highly absorbing optically thin multilayer slabs, the present synthetic angular reconstruction scheme may not generate accurate results since diffusion theory is not a good model for such problems.

References

[BuFa85] Burden, R.L., Faires, J.D.: Numerical Analysis. Prindle, Weber & Schmidt, Boston (1985)

[DuHa76] Duderstadt, J.J., Hamilton, L.J.: Nuclear Reactor Analysis. Wiley, New York (1976)

[LeMi93] Lewis, E.E., Miller, W.F.: Computational Methods of Neutron Transport. American Nuclear Society, La Grange Park (1993)

[Ly11] Lydia, E.J.: Um método de Matriz Resposta com Esquema Iterativo de Inversão Nodal Parcial para Cálculos Unidimensionais de Transporte de Nêutrons Monoenergéticos na Formulação de Ordenadas Discretas com Espalhamento Linearmente Anisotrópico, Doctoral dissertation, IME-UERJ (2011)

Chapter 17
On the Fractional Neutron Point Kinetics Equations

M. Schramm, C.Z. Petersen, M.T. Vilhena, B.E.J. Bodmann, and A.C.M. Alvim

17.1 Introduction

In neutron diffusion theory, equations that govern the dynamics of space–time and the neutron population are called kinetics equations. The kinetics equations are divided into point kinetics equations and space kinetics equations. In this work we will emphasize the point kinetics model, more specifically, variations in the neutron density for small time scales, or equivalently changes in criticality due to changes of nuclear parameter in small time intervals. The point kinetics equations describe only the behavior of the neutron density with time, assuming total separability of time from spatial degrees of freedom but with an a priori known spatial shape of the density. The point kinetics model, although derived already decades ago, plays still a significant role in reactor physics and is used to estimate the power response of the reactor, allowing for control and intervention in the power plant operation, that may also be helpful to avoid the occurrence of incidents or accidents.

M. Schramm • M.T. Vilhena • B.E.J. Bodmann (✉)
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: marceloschramm@hotmail.com; vilhena.mat@ufrgs.br; bardo.bodmann@ufrgs.br

C.Z. Petersen
Federal University of Pelotas, Pelotas, RS, Brazil
e-mail: claudio.petersen@ufpel.edu.br

A.C.M. Alvim
Federal University of Rio de Janeiro, RJ, Brazil
e-mail: alvim@con.ufrj.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_17,

The present discussion is an attempt to provide a new method that solves the point kinetics equations. The new aspect is a fractional kinetics model, which reproduces the classical model and thus allows to capture effects that differ from the usually employed hypothesis of Fick. The fractional point kinetics model presented here is derived thoroughly and solved analytically, which hopefully will mark the beginning of an extensive theoretical research for future validation and applications of this kind of approach in nuclear reactor theory.

The subject of fractional calculus has gained considerable importance during the last two decades, primarily due to its applications in various fields of science and engineering [EdFoSi02], [OlSp74], [MiRo93], [SaKiMa93], [Po99], [Hi00], [KiSrTr06], [Ma06]. However, its use in the area of reactor physics is new, and as a first step into a new direction, we determine an analytical solution for the fractional neutron point kinetics equation (FNPK). Classical diffusion theory provides a description that is strictly valid for the neutron flux when the following assumptions are satisfied [La65]:

1. The absorption probability is considerably smaller than the scattering.
2. The variation of the spatial distribution of neutrons is linear.
3. The scattering is isotropic.

The incompatibility of one of these hypotheses is clear evidence that Fick's law needs to be modified. In this sense, the justification of this work is to analyze and validate a new model of non-Fickian diffusion solutions. Note that hitherto classical diffusion solutions are still a standard in reactor physics although nuclear interaction parameter hide properties of local quantum degrees of freedom in opposition to the nonlocal character of Fick's law.

The main idea of the forthcoming discussion is to correct some problems in nonclassical diffusion phenomena arising from the highly heterogeneous configuration in nuclear reactors, by solving a fractional diffusion model as the constituent equation for the neutron density. The fractional diffusion model presented here can be applied to a large range of neutron cross sections, which is not always the case in approaches by classical models that make use of the neutron diffusion equation. Limitations of these approaches occur typically in the presence of neutron draining close to the border between nuclear fuel and the control rods, or strong neutron absorption in a boron loaded refrigerant.

The modified laws proposed next allow to broaden the scope and improve the diffusion theory with respect to the classical framework that describes neutronics, for example, the transient behavior in a highly heterogeneous reactor core assembly. In order to improve the classical diffusion theory, we reason that the fractional point kinetics equation can improve the predictions and, presumably in some cases, be similar to results from more complicated transport theory approaches. Here, we solve in closed form the kinetics model based on diffusion theory but driven by fractional derivatives. To this end we start from the fractional point kinetics equations model, considering one group of delayed neutron precursors and constant reactivity. Further, we resort to the decomposition method recently

applied successfully to a similar type of problem [PeEtAl11a]. Other works along this line may be found in [BoEtAl10], [Ce10], [PeEtAl11b], [Pe11], [AzEtAl11], [BoEtAl12], [SeViGo12], [VaSeVi12].

17.2 Derivation of the Fractional Neutron Point Kinetics Equations

Our starting point is the time-dependent neutron diffusion equation without external sources; that is,

$$\begin{aligned} -\nabla \cdot J(r,E,t) - \Sigma_t(r,E)\phi(r,E,t) \\ + \int \sum_j f_p^j(E)(1-\hat{\beta}_j)\nu_j \Sigma_{fj}(r,E')\phi(r,E',t)\, dE' \\ + \int \Sigma_s(r,E' \to E)\phi(r,E',t)\, dE' \\ + \sum_i f_i(E)\lambda_i \hat{C}_i(r,t) = \frac{1}{v(E)} \frac{\partial \phi(r,E,t)}{\partial t}, \end{aligned} \tag{17.1}$$

where $J(r,E,t)$ is the current density, $\Sigma_t(r,E)$ is total cross section, $\phi(r,E,t)$ is the neutron flux, $\Sigma_s(r,E' \to E)$ is the energy-dependent scattering cross section, λ_i is the decay constant of the neutron precursor group i, $\hat{C}_i(r,t)$ is the concentration of delayed neutrons precursors of group i, $\hat{\beta}_j$ is the delayed neutron fraction for the isotope j and $v(E)$ is the neutron speed. Here, $f_p^j(E)$ is the probability that a prompt neutron will appear as result of fission and $f_i(E)$ is the probability that a delayed neutron will appear due to the decay of an isotope of the precursor group i. Note that $\int f_p^j(E)\, dE = \int f_i(E)\, dE = 1$, $i = 1,2,\ldots I$ and $j = 1,2,\ldots J$. The expression for concentration of delayed neutrons precursors is given by

$$\int \sum_j \hat{\beta}_{ji} \nu_j \Sigma_{fj}(r,E',t)\phi(r,E',t)\, dE' - \lambda_i \hat{C}_i(r,t) = \frac{\partial \hat{C}_i(r,t)}{\partial t},$$

where $\hat{\beta}_{ji}$ is the delayed neutron fraction for isotope j that will result from an isotope of the precursor group i. Notice that $\sum_i \hat{\beta}_{ji} = \hat{\beta}_j$.

In many physical problems it has been observed that a diffusion processes does not follow the Fick law. Such a phenomenon is referred as non-classical diffusion. Specifically, in the case of a nuclear reactor, some anomalous diffusion phenomena occur due to the highly heterogeneous configuration. The result is an anomalous diffusion process that cannot be accurately described by a Fickian diffusion process. Recently, Nec and Nepomnyashchy [NeNe07] have proposed fractional modifications to Cattaneo's constituent equation [ChEtAl08] as a phenomenological model to describe anomalous diffusion processes. According to these ideas, a version for the fractional derivative equation of the current density is given by

$$\tau^k \frac{\partial^k J(r,E,t)}{\partial t^k} + J(r,E,t) = -D\nabla\phi(r,E,t)\,, \tag{17.2}$$

where τ is the relaxation time and k is the anomalous diffusion order. For processes of sub-diffusion, $0 < k < 1$, while for processes of super-diffusion, $1 < k < 2$. The fractional derivative operator $\frac{\partial^k}{\partial t^k}$ is defined by the Riemann–Liouville prescription [OlSp74]. In the limit $\tau^k \to 0$, Fick's law is recovered and for $k \to 1$ the Cattaneo equation is obtained, as shown explicitly further down.

We extend the usual derivative by applying the operator $\tau^k \frac{\partial^k}{\partial t^k}$ to (17.1) and adding the ensuing equation to (17.1) itself. Then, omitting, for convenience, the explicit space, time, and energy dependences in the current density and scalar neutron flux, we arrive at

$$\begin{aligned}
&-\nabla\cdot\left(\tau^k\frac{\partial^k J}{\partial t^k} + J\right) - \Sigma_t\left(\tau^k\frac{\partial^k\phi}{\partial t^k} + \phi\right) \\
&\quad + \int\sum_j f_j^p\left(1-\hat{\beta}_j\right)\nu_j\Sigma_{fj}\left(\tau^k\frac{\partial^k\phi}{\partial t^k} + \phi\right)dE' \\
&\quad + \int\Sigma_s\left(\tau^k\frac{\partial^k\phi}{\partial t^k} + \phi\right)dE' \\
&\qquad + \sum_i f_i\lambda_i\left(\tau^k\frac{\partial^k\hat{C}_i}{\partial t^k} + \hat{C}_i\right) = \frac{1}{v}\left(\tau^k\frac{\partial^{k+1}\phi}{\partial t^{k+1}} + \frac{\partial\phi}{\partial t}\right).
\end{aligned} \tag{17.3}$$

Replacing expression (17.2) in (17.3), integrating out volume and energy and further recalling that

$$\int f_j^p(E)\,dE = \int f_i(E)\,dE = 1,$$

$$\int\Sigma_s(r,E'\to E)\,dE = \Sigma_t(r,E') - \Sigma_a(r,E'),$$

we rewrite (17.3) as

$$\begin{aligned}
&\int\int\left[\nabla\cdot D\nabla\phi + \left(\sum_j(1-\hat{\beta}_j)\nu_j\Sigma_{fj} - \Sigma_a\right)\phi\right]dE\,dV \\
&\qquad + \sum_i\int\lambda_i\hat{C}_i\,dV + \tau^k\frac{d^k}{dt^k}\sum_i\int\lambda_i\hat{C}_i\,dV \\
&= \frac{d}{dt}\int\int\frac{1}{v}\phi\,dE\,dV + \tau^k\frac{d^{k+1}}{dt^{k+1}}\int\int\frac{1}{v}\phi\,dE\,dV \\
&\qquad + \tau^k\frac{d^k}{dt^k}\int\int\left(\Sigma_a - \sum_j(1-\beta_{ij})\nu_j\Sigma_{fj}\right)\phi\,dE\,dV\,.
\end{aligned} \tag{17.4}$$

We assume that the neutron flux and the delayed neutron precursors can be factorized as a time-dependent amplitude function times a shape function; that is,

$$\phi(r,E,t) = n(t)G(r,E,t), \quad \hat{C}_i(r,t) = C_i(t)F_i(r,t). \tag{17.5}$$

Here, $n(t)$ is the neutron density, $C_i(t)$ is the delayed neutron precursor concentration for group i, and $G(r,E,t)$ and $F_i(r,t)$ are the functions satisfying

$$\int\int \frac{G(r,E,t)}{v(E)}\, dE\, dV = \int F_i(r,t)\, dV = 1,$$

$$\int\int \frac{\phi(r,E,t)}{v(E)}\, dE\, dV = n(t), \quad \int \hat{C}_i(r,t)\, dV = C_i(t).$$

Rewriting (17.4) in terms of the new definitions (17.5) yields

$$\begin{aligned} n(t)\int\int \left[\nabla\cdot D\nabla G + \left(\sum_j \nu_j \Sigma f j\right) G\right] dE\, dV \\ - n(t)\sum_j \int\int \hat{\beta}_j \nu_j \Sigma_{fj} G\, dE\, dV \\ + \sum_i \lambda_i C_i + \tau^k \sum_i \lambda_i \frac{d^k C_i(t)}{dt^k} = \frac{dn(t)}{dt} + \tau^k \frac{d^{k+1} n(t)}{dt^k} \\ + \tau^k \frac{d^k n(t)}{dt^k} \int\int \left(\Sigma_a - \sum_j (1-\hat{\beta}_j)\nu_j \Sigma_{fj}\right) G\, dE\, dV, \end{aligned} \tag{17.6}$$

which, on defining reactor-specific parameters, reduces to the FNPK equation for $n(t)$ and for $C_i(t)$:

$$\begin{aligned} n(t)\frac{\rho(t)-\beta(t)}{\Lambda(t)} + \sum_i \lambda_i C_i(t) + \tau^k \sum_i \lambda_i \frac{d^k C_i(t)}{dt^k} \\ = \frac{dn(t)}{dt} + \tau^k \frac{d^{k+1} n(t)}{dt^{k+1}} + \tau^k \frac{d^k n(t)}{dt^k}\left(\frac{1}{l(t)} - \frac{1-\beta(t)}{\Lambda(t)}\right), \\ n(t)\frac{\beta_i(t)}{\Lambda(t)} - \lambda_i C_i(t) = \frac{dC_i(t)}{dt}. \end{aligned}$$

The effective kinetics parameters may be identified by comparison with (17.6) and represent the reactivity function ρ, the fraction β_i of delayed neutrons for precursor group i, the total fraction β of delayed neutrons, the neutron generation time Λ, and the mean neutron lifetime l; specifically,

$$\rho(t) = \frac{\int\int \left[\nabla \cdot D\nabla G + \left(\sum_j \nu_j \Sigma_{fj} - \Sigma_a\right) G\right] \, dE\, dV}{\int\int \sum_j \nu_j \Sigma_{fj} G \, dE\, dV} \,,$$

$$\beta_i(t) = \frac{\int\int \sum_j \hat{\beta}_{ji} \nu_j \Sigma_{fj} G \, dE\, dV}{\int\int \sum_j \nu_j \Sigma_{fj} G \, dE\, dV} \,,$$

$$\beta(t) = \frac{\sum_j \int\int \hat{\beta}_j \nu_j \Sigma_{fj} G \, dE\, dV}{\int\int \sum_j \nu_j \Sigma_{fj} G \, dE\, dV} = \sum_i \beta_i(t) \,,$$

$$\Lambda(t) = \frac{1}{\int\int \sum_j \nu_j \Sigma_{fj} G \, dE\, dV} \,,$$

$$l(t) = \frac{1}{\int\int \Sigma_a G \, dE\, dV} \,.$$

This non-Fickian model includes three additional terms with respect to the classical point kinetics equations containing the fractional derivatives $\frac{d^{k+1}n(t)}{dt^{k+1}}$, $\frac{d^k n(t)}{dt^k}$, and $\frac{d^k C_i(t)}{dt^k}$. The physical meaning of these terms suggests that for sub-diffusion processes the first term is an important contribution to the fast changes in the neutron density, while the second term represents an important contribution when the changes in the neutron density are relatively slow, for example, during the start-up of a nuclear plant involving operational re-conditioning of the reactor. The third term becomes more important, for example, when the reactor is on shut down. It may further be useful for the processes in accelerated driven systems (ADS) that are characterized by a low fraction of delayed neutrons.

Note that for $\tau^k \to 0$ we recover the classical neutron point kinetics (NPK) model,

$$n(t)\frac{\rho(t) - \beta(t)}{\Lambda(t)} + \sum_i \lambda_i C_i(t) = \frac{dn(t)}{dt} \,,$$

and for $k \to 1$ we get Cattaneo's classical model

$$n(t)\frac{\rho(t) - \beta(t)}{\Lambda(t)} + \sum_i \lambda_i C_i(t) + \tau \sum_i \lambda_i \frac{dC_i(t)}{dt} = \frac{dn(t)}{dt}\left(1 + \tau\frac{1}{l(t)} - \tau\frac{1-\beta(t)}{\Lambda(t)}\right) + \tau\frac{d^2 n(t)}{dt^2}.$$

It is noteworthy that the formulation for the neutron precursors concentrations $C_i(t)$ is identical to the classic model, even with a non-Fickian closure. Usually, the parameters $\beta_i(t)$, $\beta(t)$, $\Lambda(t)$, and $l(t)$ are not considered time-dependant in point kinetics models. In general the reactivity term shall have time dependence, or even $n(t)$ dependence, when one considers a reactor dynamics model, unlike other kinetics parameters.

17.3 The Solution of the FNPK Equations

In what follows, we solve the FNPK equations for constant kinetics parameters $\rho(t)$, $\beta_i(t)$, $\beta(t)$, $\Lambda(t)$ and $l(t)$. The fractional model is considered for I delayed neutron precursor groups and given in [PaLaMa11]; that is,

$$\begin{aligned}\frac{dn(t)}{dt} &= -\tau^k \frac{d^{k+1}n(t)}{dt^{k+1}} - \tau^k \left[\frac{1}{l} + \left(\frac{1-\beta}{\Lambda}\right)\right] \frac{d^k n(t)}{dt^k} + \frac{\rho-\beta}{\Lambda} n(t) \\ &\quad + \sum_{i=1}^{I} \lambda_i C_i(t) + \tau^k \sum_{i=1}^{I} \lambda_i \frac{d^k C_i(t)}{dt^k}, \qquad (17.7) \\ \frac{dC_i(t)}{dt} &= \frac{\beta_i}{\Lambda} n(t) - \lambda_i C_i.\end{aligned}$$

The initial conditions are $n(0) = n_0$ and $C_i(0) = \frac{\beta_i}{\lambda_i \Lambda} n_0$, respectively.

To this end, we make use of the decomposition method [Ad94]. In some cases the Adomian decomposition is known to have slow convergence for large time periods, so that one may avoid an extensive number of terms in the recursion scheme by resorting to an analytic continuation method. The analytic continuation consists in solving for short time steps, so that only a few terms of the decomposition series are needed and represent already a reasonable result. Moreover, the solution for one time step is evaluated at the upper time step limit and defines the initial condition for the next step.

The Adomian decomposition method consists in expanding the neutron density and concentration of delayed neutrons precursors in a truncated series. The number of terms (R) considered in the decomposition series is determined by some convergence criterion:

$$n(t) = \sum_{r=0}^{R} n_r(t), \quad C_i(t) = \sum_{r=0}^{R} C_{ir}(t).$$

These expansions are inserted into the FNPK equations and a set of first-order recursive differential equations are constructed according to the following prescription.

$$\begin{aligned}\frac{d(n_0 + \ldots + n_R)}{dt} &= -\tau^k \frac{d^{k+1}(n_0 + \ldots + n_R)}{dt^{k+1}} \\ &\quad -\tau^k \left[\frac{1}{l} - \left(\frac{1-\beta}{\Lambda}\right)\right] \frac{d^k(n_0 + \ldots + n_R)}{dt^k} \\ &\quad + \frac{\rho-\beta}{\Lambda}(n_0 + \ldots + n_R) + \sum_{i=1}^{I} \lambda_i \frac{(C_{i0} + \ldots + C_{iR})}{dt^k}\end{aligned}$$

$$+\tau^k \sum_{i=1}^{I} \lambda_i \frac{d^k\left(C_{i0}+\ldots+C_{iR}\right)}{dt^k},$$

$$\frac{d\left(C_{i0}+\ldots+C_{iR}\right)}{dt} = \frac{\beta_i}{\Lambda}\left(n_0+\ldots+n_R\right) - \lambda_i\left(C_{i0}+\ldots+C_{iR}\right).$$

The recursive equation system is given by

$$\begin{aligned}
&\frac{dn_r(t)}{dt} - \frac{(\rho-\beta)}{\Lambda} n_r(t) - \sum_{i=1}^{I} \lambda_i C_{ir}(t) = s_r(t),\\
&\frac{dC_{ir}(t)}{dt} - \frac{\beta_i}{\Lambda} n_r + \lambda_i C_{ir} = 0,\\
&s_r(t) = -\tau^k \frac{d^{k+1} n_{r-1}(t)}{dt^{k+1}} - \tau^k \left[\frac{1}{l} + \left(\frac{1-\beta}{\Lambda}\right)\right] \frac{d^k n_{r-1}(t)}{dt^k}\\
&\qquad\qquad + \tau^k \sum_{i=1}^{I} \lambda_i \frac{d^k C_{i(r-1)}(t)}{dt^k}.
\end{aligned} \tag{17.8}$$

Casting (17.8) in matrix form yields

$$\frac{d\mathbf{Y}_r(t)}{dt} - \mathbf{A}\mathbf{Y}_r(t) = \mathbf{S}_r(t)\,, \tag{17.9}$$

where

$$\mathbf{Y}_r(t) = \begin{pmatrix} n_r(t) \\ C_{1r}(t) \\ \vdots \\ C_{Ir}(t) \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \frac{\rho-\beta}{\Lambda} & \lambda_1 & \ldots & \lambda_I \\ \frac{\beta_1}{\Lambda} & -\lambda_1 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ \frac{\beta_I}{\Lambda} & 0 & \ldots & -\lambda_I \end{pmatrix}, \quad \mathbf{S}_r(t) = \begin{pmatrix} s_r(t) \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

for $r = 0 : R$, considering $s_0(t) = 0$. The first solution $\mathbf{Y}_0(t)$ is determined by the homogeneous problem with the initial conditions of the original problem.

$$\frac{d\mathbf{Y}_0(t)}{dt} - \mathbf{A}\mathbf{Y}_0(t) = \mathbf{0}, \quad \mathbf{Y}_0(0) = \left(n(0)\; C_1(0)\; \ldots\; C_I(0)\right)^T, \tag{17.10}$$

where T denotes transposition.

The solution of (17.10) is obtained using known solutions for first-order differential equation systems.

$$\mathbf{Y}_0(t) = \exp(\mathbf{A}t)\mathbf{Y}_0(0).$$

In the present case the eigenvalues of $\mathbf{A}$ are in general distinct, so that the exponential matrix may be expressed by

$$\exp(\mathbf{A}t) = \mathbf{X}\exp(\mathbf{D}t)\mathbf{X}^{-1},$$

where $\mathbf{X}$ is the matrix containing the eigenvectors of $\mathbf{A}$, $\mathbf{X}^{-1}$ is its inverse and $\mathbf{D}$ is the diagonal matrix with the eigenvalues of $\mathbf{A}$. In order to find the recursive system solutions, the initial conditions from the second recursion step on are all zero. In order to solve the recursive system (17.9) we propose to rewrite the system, so that the fractional derivatives are implemented using the integral Riemann–Liouville definition of fractional derivatives [OlSp74], given by

$$\frac{d^q f(t)}{dt^q} = \frac{1}{\Gamma(1-q)}\frac{d}{dt}\int_0^t \frac{f(x)}{(t-x)^q}dx, \quad \mathrm{Re}(q) > 0, \tag{17.11}$$

where $\Gamma(1-q)$ is the Gamma function and q the fractional derivative.

The source term $s_r(t)$ given by (17.8) was evaluated according to (17.11) and is treated as a step function, namely

$$\mathbf{S}_r(t) = \left(\hat{s}_r\ 0\ \ldots\ 0\right)^T = \mathbf{S}_r,$$

so that the solution of (17.8) may be obtained by the Laplace transformation. Then the transformed recursive system, written in matrix form (with zero initial conditions), is

$$p\overline{\mathbf{Y}}_r(p) - \mathbf{A}\overline{\mathbf{Y}}_r(p) = \overline{\mathbf{S}}_r(p), \tag{17.12}$$

where

$$\overline{\mathbf{Y}}_r(p) = L[\mathbf{Y}_r(t), t \to p], \quad \overline{\mathbf{S}}_r(p) = L[\mathbf{S}_r, t \to p] = \frac{1}{p}\mathbf{S}_r.$$

The solution for (17.12) is found by Cramer's rule. In order to find the general solution, we used the Heaviside inversion method to find the inverse Laplace transform of $\overline{\mathbf{Y}}_r(p)$. Once found the solution of the homogeneous system, the global solution may be found by repeating the non-homogeneous recursive calculation with a recursion depth compatible with a desired precision so that one obtains the final solution of the problem. It is noteworthy that the solution for the neutron density and for the precursor concentrations found by the outlined method are sums of exponentials with the same arguments, like in the classic point kinetic case.

17.4 Numerical Results

In order to analyze the effect of anomalous diffusion and the relaxation time on the behavior of the neutron density, the mathematical model (17.7) was solved in analytical form using the following nuclear parameters obtained from [KiAl04]; that is,

$$\beta = \sum_{i=1}^{I} \beta_i = 0.007, \quad \lambda = \frac{\beta}{\frac{\sum_{i=1}^{I} \beta_i}{\lambda_i}}, \quad \Lambda = 0.00002, \quad l = 0.00024.$$

The values of β_i and λ_i are given in Table 17.1. In the sequel, the numerical results of three cases are reported.

17.4.1 Case A

For the first test case, we calculate the neutron density and concentration of delayed neutron precursors varying two parameters, the relaxation time (τ) and the order of anomalous diffusion (k). The results are shown in Table 17.2 for the neutron density and concentration of delayed neutron precursors. The reactivity was calculated at two points, sub-critical $(\rho = -0.003)$ and super-critical $(\rho = 0.003)$ for different times. The findings indicate the behavior of the neutron density and concentration

Table 17.1 Nuclear data for the case studies

Delayed neutrons	
$\beta_i \times 10^{-3}$	$\lambda_i(s^{-1})$
0.266	0.0127
1.491	0.0317
1.316	0.1550
2.849	0.311
0.896	1.40
0.182	3.87
$k_{eff} = 1.000008$	

Table 17.2 Neutron density and concentration of delayed neutron precursors for the fractional model with $\tau = 10^{-4}$, $k = 0.96$, $\rho = -0.003$ and $\rho = 0.003$

Density(n)/Concentration(C)	$\rho = -0.003$	$\rho = 0.003$
$n(t = 0.1\,\text{s})$	0.899653	1.16243
$n(t = 1\,\text{s})$	0.695748	1.69812
$n(t = 10\,\text{s})$	0.556141	2.93924
$n(t = 100\,\text{s})$	0.0637804	549.61
$C(t = 0.1\,\text{s})$	43.1399	43.1881
$C(t = 1\,\text{s})$	42.3437	44.6792
$C(t = 10\,\text{s})$	34.1038	74.445
$C(t = 100\,\text{s})$	3.91323	13,938

Table 17.3 Neutron density and concentration of delayed neutron precursors for different relaxation times (τ) for $\rho = 0.003$ and $k = 0.96$

Density(n)/Concentration(C)	$\tau = 10^{-4}s$	$\tau = 10^{-5}s$	$\tau = 10^{-6}s$	$\tau = 0$
$n(t = 0.1\,\text{s})$	1.16243	1.13885	1.13633	1.13602
$n(t = 1\,\text{s})$	1.69812	1.6793	1.67726	1.67701
$n(t = 10\,\text{s})$	2.93924	2.92908	2.92797	2.92783
$n(t = 100\,\text{s})$	549.61	496.13	490.892	490.254
$C(t = 0.1\,\text{s})$	43.1881	43.1839	43.1834	43.1834
$C(t = 1\,\text{s})$	44.6792	44.6382	44.6338	44.6332
$C(t = 10\,\text{s})$	74.445	74.2811	74.2632	74.261
$C(t = 100\,\text{s})$	13,938	12583.5	12450.9	12434.7

of delayed neutron precursors with variation in reactivity. One observes a decreased density and precursors concentration for $(\rho = -0.003)$ with increasing time. Likewise, there is an expected growth in density and precursor concentration with time.

17.4.2 Case B

This case considers a varying relaxation time (τ) and reactivity (ρ), maintaining the anomalous diffusion order (k). The results are shown in Table 17.3 for the neutron density using a supercritical reactivity and an order of the anomalous diffusion $k = 0.96$.

Note that for relaxation times there is an increase in reactivity as well as concentration. With decreasing relaxation time one observes a decrease in the neutron density and concentration of delayed neutron precursors. Moreover, with decreasing relaxation time the solution approaches the classical model.

A graphical illustration of the results presented in Table 17.3 is shown in Fig. 17.1 for the neutron density.

17.4.3 Case C

In the third and final test case, the order of anomalous diffusion is varied, comparing the classical model $(k = 1)$ with the fractional model and using fixed values for the reactivity (ρ) and the relaxation time (τ), respectively. The results are shown in Table 17.4 for different times.

One notices an increase in neutron density and concentration of delayed neutron precursors as time increases. The effect of anomalous diffusion is to lower the neutron density as well as the concentration of delayed neutrons with an increase in anomalous diffusion. Moreover, one perceives an approximation of the solution

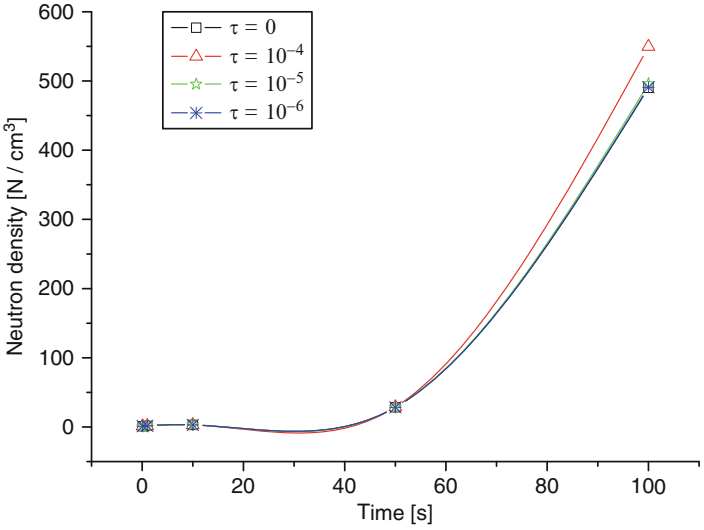

Fig. 17.1 Neutron density for $\rho = 0.003$ and $k = 0.96$

Table 17.4 Neutron density and concentration of delayed neutron precursors for different orders for anomalous diffusion, $\rho = 0.003$ and $\tau = 10^{-4}$

Density(n)/ Concentration(C)	$k = 0.96$	$k = 0.97$	$k = 0.98$	$k = 0.99$	$k = 1$
$n(t = 0.1\,\text{s})$	1.16243	1.15428	1.14725	1.1412	1.13602
$n(t = 1\,\text{s})$	1.69812	1.69131	1.68563	1.68091	1.67701
$n(t = 10\,\text{s})$	2.93924	2.93541	2.93231	2.92982	2.92783
$n(t = 100\,\text{s})$	549.61	527.26	510.967	499.015	490.254
$C(t = 0.1\,\text{s})$	43.1881	43.1866	43.1854	43.1843	43.1834
$C(t = 1\,\text{s})$	44.6792	44.6644	44.652	44.6417	43.1834
$C(t = 10\,\text{s})$	74.445	74.3832	74.3331	74.2929	74.261
$C(t = 100\,\text{s})$	13,938	13,372	12959.3	12656.6	12434.7

of the fractional model to the classical model once the order of anomalous diffusion approaches the unit value $k = 1$. The results presented in Table 17.4 are illustrated in Fig. 17.2.

17.5 Concluding Remarks

From our case studies one finds that the solutions are located between those of classical kinetics and results from transport approaches. One may reason in a reverse fashion that the justification for using a fractional derivative may be based on the following argument. Phenomena in multiplicative media, like neutron chain

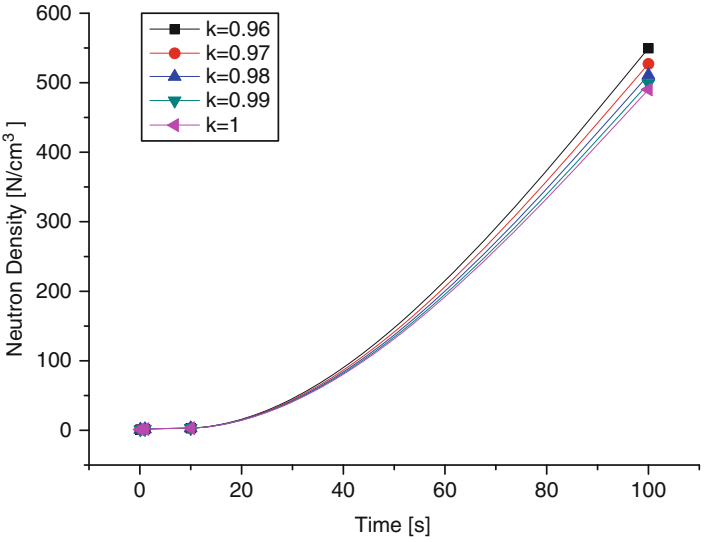

Fig. 17.2 Neutron density for $\rho = 0.003$ and $\tau = 10^{-4}$

reactions, have a close analogy with the instructions for a Cantor set construction, which in turn is related to a fractal dimension and thus to some scaling law. From the definition of the fractional derivative it becomes evident that the order of the derivative is related to a scaling of the differential with the "effective volume" which does not necessarily coincide with an integer.

We are aware of the fact that we have neglected so far the important question, what optimization criterion should be used in order to uniquely determine the order fractional derivative. In other words to find the answer as to what derivative supplies with the most adequate solution if compared to classical diffusion or transport theory. In classical diffusion where Fick's hypothesis is used, there is put severe restriction on the constant value of D, where the ratio of nth derivative of the current density by the $n+1$th derivative of the scalar flux remains constant independent of n. It is not at all plausible why a diffusion process in multiplicative media shall obey such a restriction. Investigations in this direction will be focused on in future work.

Another question that we have not answered in our discussion is the existence and the specific form of the adjoint flux. At the present stage of the work it remains open whether this flux may be represented in a factorized form, in this particular case the adjoint flux would coincide with the classical expression. However, from the differences found between the classical solution and the fractional ones one may reason that the classical adjoint flux represents at least a first approximation for the fractional one, especially because the time dependence of G and F is weak. Theoretical developments in this direction are already in progress.

References

[Ad94] Adomian, G.: Solving Frontier Problems of Physics: The Decomposition Method. Kluwer, Boston (1994)

[AzEtAl11] Azevedo, F.S., Sauter, E., Thompson, M., Vilhena, M.T.: Existence theory and simulations for one-dimensional radiative flows. Ann. Nucl. Energ. **38**, 1115–1124 (2011)

[BoEtAl10] Bodmann, B., Vilhena, M.T., Ferreira, L.S., Bardaji, J.B.: An analytical solver for the multi-group two-dimensional neutron–diffusion equation by integral transform techniques. Nuovo Cimento C **1**, 1–10 (2010)

[BoEtAl12] Borges, V., Fernandes, J.C.L., Bodmann, B., Vilhena, M.T., Rodriguez, B.A.: A closed-form formulation for the build-up factor and absorbed energy for photons and electrons in the compton energy range in Cartesian geometry. World J. Nucl. Sci. Tech. **2**, 23–28 (2012)

[Ce10] Ceolin, C.: Solução Analítica da Equação Cinética de Difusão Multigrupo de Nêutrons em Geometria Cartesiana Unidimensional pela Técnica da Transformada Integral, M.Sc. dissertation, PROMEC/UFRGS, Brazil (2010) (Portuguese)

[ChEtAl08] Chen, W.-B., Wang, J., Qiu, W.-Y., Ren, F.-Y.: Solutions for time-fractional diffusion equation with absorption: influence of different diffusion coefficients and external forces. J. Phys. A **41**, 045003–045012 (2008)

[EdFoSi02] Edwards, J.T., Ford, N.J., Simpson, A.C.: The numerical solution of linear multi-term fractional differential equations. J. Comput. Appl. Math. **148**, 401–418 (2002)

[Hi00] Hilfer, R.: Applications of Fractional Calculus in Physics. World Scientific, Singapore (2000)

[KiSrTr06] Kilbas, A.A., Srivastava, H.M., Trujillo, J.J.: Theory and Applications of Fractional Differential Equations, Amsterdam, Netherlands, Elsevier, Begell House, Redding (2006)

[KiAl04] Kinard, M., Allen, K.E.J.: Efficient numerical solution of the point kinetics equations in nuclear reactor dynamics. Ann. Nucl. Energ. **31**, 1039–1051 (2004)

[La65] Lamarsh, J.R.: Introduction to Nuclear Reactor Theory. Addison-Wesley, Longman Higher Education, Reading (1965)

[Ma06] Magin, R.L.: Fractional Calculus in Bioengineering. Begell House Publishers, Redding (2006)

[MiRo93] Miller, K.S., Ross, B.: An Introduction to the Fractional Integrals and Derivatives: Theory and Applications. Wiley, New York (1993)

[NeNe07] Nec, Y., Nepomnyashchy, A.A.: Turing instability in sub-diffusive reaction–diffusion system. J. Phys. A **40**, 14687–14702 (2007)

[OlSp74] Oldham, K.B., Spanier, J.: The Fractional Calculus. Academic, New York (1974)

[PaLaMa11] Paredes, E.F., Labarrios, M.P., Martinez, E.E.: Fractional neutron point kinetics equations for nuclear reactor dynamics. Ann. Nucl. Energ. **38**, 307–330 (2011)

[Pe11] Petersen, C.Z.: Solução Analítica das equações da Cinética Pontual e Espacial da Teoria de Difusão de Nêutrons pelas técnicas da GITT e Decomposição, PhD dissertation, PROMEC/UFRGS, Brazil (2011) (Portuguese)

[PeEtAl11a] Petersen, C.Z., Dulla, S., Vilhena, M.T., Ravetto, P.: An analytical solution of the point kinetics equations with time-variable reactivity by the decomposition method. Progr. Nucl. Energ. 1–4 (2011). doi:10.1016/j.pnucene.2011.01.001

[PeEtAl11b] Petersen, C.Z., Vilhena, M.T., Bodmann, B.E.J., Dulla, S., Ravetto, P.: On the exact solution for the multi-group kinetic neutron diffusion equation in a rectangle. In: International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering, Rio de Janeiro, RJ (2011)

[Po99] Podlubny, I.: Fractional Differential Equations. Academic, New York (1999)

[SaKiMa93] Samko, S.G., Kilbas, A.A., Marichev, O.I.: Fractional Integrals and Derivatives: Theory and Applications. Gordon and Breach, Linghorne (1993)

[SeViGo12] Segatto, C.F., Vilhena, M.T., Gonçalez, T.T.: On the analytical solution of the neutron Sn equation in a rectangle assuming an exponential exiting angular flux at boundary. Int. J. Nucl. Energ. **7**, 45–56 (2012)

[VaSeVi12] Vargas, R.M.F., Segatto, C.F., Vilhena, M.T.: On the analytical solution of the SN radiative transport equation in a slab for a space-dependent albedo coefficient. J. Phys. **369**, 1–10 (2012)

Chapter 18
On a Closed Form Solution of the Point Kinetics Equations with a Modified Temperature Feedback

J.J.A. Silva, B.E.J. Bodmann, M.T. Vilhena, and A.C.M. Alvim

18.1 Introduction

The point kinetics equations with temperature feedback corresponds to a stiff system of nonlinear differential equations for the neutron density and delayed precursor concentrations. These variables determine the time-dependent behavior of the power level of a nuclear reactor and are influenced, for example, by the position of the control rods. Computing solutions of the equations of point kinetics provide information on the dynamics of nuclear reactor operation and are useful, for example, in understanding the power fluctuations experienced during start-up or shut-down, when the control rods are adjusted. Recently, a large number of kinetics studies have been reported [PeEtAl11], [NaZa10], which modeled the time-dependent behavior of a nuclear reactor using point-kinetic equations. As pointed out by many authors, this system of point kinetics equations is still an important set of equations. Although its range of applicability has been severely restricted by the increasing importance of optimal power reactor cores with loose coupling, they remain very useful in terms of preliminary studies, especially when control aspects are considered. The presence of temperature feedback is useful to provide an estimate of the transient behavior of reactor power and other system variables of the reactor core that are very tightly coupled. In this paper, the point kinetics equations in the presence of Newtonian temperature feedback are reduced to a second order nonlinear differential equation in a simple form convenient for application of Adomian's method [Ad94], [Ad89]. The basic idea consists

J.J.A. Silva • A.C.M. Alvim
Federal University of Rio de Janeiro, RJ, Brazil
e-mail: shaolin.jr@gmail.com; aalvim@gmail.com

B.E.J. Bodmann (✉) • M.T. Vilhena
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: bardo.bodmann@ufrgs.br; mtmbvilhena@gmail.com

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_18,

in expanding the solution in series of functions and the nonlinear term is then defined by Adomian's polynomials. Substituting these expansions into the original equation, a recursive linear system is built, which is then solved analytically. This technique has been applied to a broad class of problems in physics, mathematics, and engineering. Here we show a solution in analytical form for the point kinetics equations with temperature feedback in the presence of one group delayed neutrons concentration. The reactor is assumed to be initially critical at some steady power level and a Newtonian feedback model is being assumed for the fuel temperature equation. The equation that is commonly used to model the temperature variation is added of a term depending on the square of the neutron population (Zn^2). Practical use of the method is tested with different types of step reactivity input, different time steps, and different values of Z, and compared with results of the literature, both numerical and analytical.

18.2 The Kinetic Model with Modified Temperature Feedback

The model used in this study starts from the point kinetics equations and one group of precursors as reported in [NaZa10], where $n(t)$ is the neutron population, $C(t)$ is the concentration of delayed neutron precursors, $T(t)$ is the temperature of the core, $\rho(T)$ is the reactivity (which depends on the temperature T), β is the delayed neutron fraction, L is the prompt neutrons generation time, and λ is the average decay constant of the precursors.

$$\frac{dn}{dt} = n'(t) = \left(\frac{\rho(T)-\beta}{L}\right) n(t) + \lambda C(t),$$
$$\frac{dC}{dt} = C'(t) = \frac{\beta}{L} n(t) - \lambda C(t).$$

This equation system is extended by a perturbation in form of a temperature feedback, where the perturbation may be understood as a change in the nuclear system configuration, as a consequence of a heat flow that induces a change in the temperature. Since the source of heat production are the nuclear processes, we assume that the thermal change rate may be related to the neutron density, with a second-order term $Zn(t)^2$:

$$\frac{dT}{dt} = T'(t) = Hn(t) + Z\left(n(t)\right)^2. \tag{18.1}$$

In this sense, the equation for the temperature change rate is a perturbation where the proportionality constant H is a parameter for the influence of the change of heat flow on the rate of temperature change and Z is another parameter. Thus, the

linear relation between temperature change rate and neutron density supplies with a feedback mechanism. As the equation system is formed by first order differential equations, it is necessary to know the initial conditions of each of the variables to determine a unique solution of the problem. Thus, we consider the reactor initially at equilibrium ($n'(0) = 0$), with known initial power and temperature ($n(0)$, $T(0)$). The equilibrium condition allows us to calculate the initial concentration of delayed neutrons precursors as:

$$C(0) = \frac{1}{\lambda}\left(\frac{\beta - \rho(0)}{L}\right) n(0).$$

In the model discussed in this work, the variation of reactivity with temperature is given by the equation

$$\rho(T) = \rho(0) - \alpha\left[T(t) - T(0)\right],$$

where $\rho(0)$ is the initial reactivity and α is the fuel temperature reactivity coefficient. After some algebraical effort, we arrived at almost the same final set of equations, the only difference being a nonlinear term depending on n^3 on the right-hand side of the equation; that is,

$$n'' + bn' + cn = \mathbb{S},$$
$$T' = Hn + Zn^2,$$

with $b = \lambda - [\rho(0) + \alpha T(0) - \beta]/L$, $c = -\lambda[\rho - \alpha T(0)]/L$, and

$$\mathbb{S} = -\frac{\alpha}{L}\left[T\left(n' - \lambda n\right) + Hn^2 + Zn^3\right] = -\frac{\alpha}{L}\left[P + A + B\right],$$

where $P = T(n' - \lambda n)$, $A = Hn^2$, and $B = Zn^3$.

The strategy adopted to find the solution consists of using the decomposition method first, finding the solution to the homogeneous equation, and then solving the nonlinear terms of the decomposition using Adomian polynomials. The initial step is to write n and T as sums:

$$n = \sum_{j=0}^{R-1} n_j, \quad T = \sum_{j=0}^{R-1} T_j, \quad \mathbb{S}_j = -\frac{\alpha}{L}\left(P_{j-1} + A_{j-1} - B_{j-1}\right).$$

In this way, we can write the homogeneous equation ($j = 0$) as

$$n_0'' + bn_0' + cn_0 = 0,$$

for which the solution is well known, namely

$$n_0(t) = k_1 e^{r_1 t} + k_2 e^{r_2 t},$$
$$r_1 = \frac{-b + \sqrt{b^2 + 4c}}{2}, \quad r_2 = \frac{-b - \sqrt{b^2 + 4c}}{2},$$
$$k_1 = \frac{n'(0) - n(0) r_2}{(r_1 - r_2)}, \quad k_2 = \frac{n(0) r_1 - n'(0)}{(r_1 - r_2)}.$$

From n_0, we can find T_0 by integrating (18.1):

$$\int_0^t T_0'(\tau) d\tau = \int H n_0(\tau) + Z n_0^2(\tau) d\tau,$$
$$T_0(t) = T(0) + H \int_0^t n_0(\tau) d\tau + Z \int_0^t n_0^2(\tau) d\tau.$$

Evaluating each of the integrals, we have

$$\int_0^t n_0(\tau) d\tau = \int_0^t \left[k_1 e^{r_1 \tau} + k_2 e^{r_2 \tau}\right] d\tau = \frac{k_1}{r_1}\left(e^{r_1 t} - 1\right) + \frac{k_2}{r_2}\left(e^{r_2 t} - 1\right)$$

and

$$\int_0^t n_0^2(\tau) d\tau = \int_0^t \left[k_1 e^{r_1 \tau} + k_2 e^{r_2 \tau}\right]\left[k_1 e^{r_1 \tau} + k_2 e^{r_2 \tau}\right] d\tau$$
$$= \frac{k_1^2}{2r_1}\left(e^{2r_1 t} - 1\right) + \frac{k_2^2}{2r_2}\left(e^{2r_2 t} - 1\right) + \frac{k_1 k_2}{r_1 + r_2}\left(e^{(r_1 + r_2)t} - 1\right).$$

Then the analytic expression of $T_0(t)$ is

$$T_0(t) = T(0) + H\left\{\frac{k_1}{r_1}\left(e^{r_1 t} - 1\right) + \frac{k_2}{r_2}\left(e^{r_2 t} - 1\right)\right\}$$
$$+ Z\left\{\frac{k_1^2}{2r_1}\left(e^{2r_1 t} - 1\right) + \frac{k_2^2}{2r_2}\left(e^{2r_2 t} - 1\right) + + \frac{k_1 k_2}{r_1 + r_2}\left(e^{(r_1 + r_2)t} - 1\right)\right\}.$$

The terms with index $j > 0$ are evaluated in a slightly different way. The expressions for n_0 and T_0 are completely analytical, but the equations for n_j and T_j, $j > 0$ are complex enough to justify the use of a numerical approach; hence,

$$n_1'' + b n_1' + c n_1 = -\frac{\alpha}{L}\left[P_0 + A_0 + B_0\right].$$

As the right-hand side of the previous equation depends on n_0 and T_0, we have set a fixed time step, ω, and considered $P_j(\omega)$, $A_j(\omega)$ and $B_j(\omega)$ constants, so

$$n_1'' + bn_1' + cn_1 = -\frac{\alpha}{L}\left[P_0(\omega) + A_0(\omega) - B_0(\omega)\right],$$
$$n_j'' + bn_j' + cn_j = Q_{j-1},$$

with $Q_j = -(\alpha/L)\left[P_j(\omega) + A_j(\omega) - B_j(\omega)\right]$. The solution of the equation above is even simpler if we set $n_1' = n_1'' = 0$ and $n_1(t) = n_1(\omega)$, that is, the solution obtained for n_1 is valid only for" $t = \omega$; therefore,

$$n_1(\omega) = \frac{Q_0}{c}, \quad n_1'(\omega) = 0, \quad n_1''(\omega) = 0,$$

and for any $j > 0$,

$$n_j(\omega) = \frac{Q_{j-1}}{c}, \quad n_j'(\omega) = 0.$$

18.2.1 Expansions of P_j

The terms P_j carry the products of the variables, $n(t)T(t)$ and $n'(t)T(t)$. In their construction terms are grouped together in a way such that each of the P_j depends only on n_p and T_p for which $p \leq j$. Thus, we can construct

$$\begin{aligned}
P_0 &= T_0(\lambda n_0 + n_0'),\\
P_1 &= T_1(\lambda n_0 + n_0') + T_0(\lambda n_1 + n_1') + T_1(\lambda n_1 + n_1'),\\
P_2 &= T_2(\lambda n_0 + n_0') + T_2(\lambda n_1 + n_1') + T_0(\lambda n_2 + n_2')\\
&\quad + T_1(\lambda n_2 + n_2') + T_2(\lambda n_2 + n_2'),
\end{aligned}$$

which can be expressed as the recursive relation

$$P_0 = T_0\left[\lambda n_0 + n_0'\right],$$
$$P_j = \sum_{p=0}^{j} T_j\left[\lambda n_p + n_p'\right] + \sum_{p=0}^{j-1} T_p\left[\lambda n_j + n_j'\right].$$

18.2.2 Expansion of A_j and B_j in Terms of Adomian Polynomials

George Adomian wrote in his books [Ad89], [Ad94] that although the expansion of the nonlinear terms is unique, there are numerous ways to group together the

terms of such an expansion. The general terms of the "fast conversion" expansion (accelerated polynomials) for n^2 e n^3 are (see pages 36 and 37 in [Ad89]).

$$A_0 = H\left[n_0^2\right], \quad B_0 = Z\left[n_0^3\right], \quad A_j = H\left[n_j^2 + 2n_j \sum_{p=0}^{j-1} n_p\right],$$

$$B_j = Z\left[n_j^3 + 3n_j\left(\sum_{p=0}^{j-1} n_p^2\right) + 3n_j^2\left(\sum_{p=0}^{j-1} n_p\right) + 6n_j\left(\sum_{q=0}^{j-2}\sum_{p=q+1}^{j-1} n_p n_q\right)\right].$$

18.2.3 Solution Algorithm

In the following we list the sequence of the algorithm that solves the problem in closed form.

- We start at some initial time (usually $t = 0$) and start the recursion with $j = 0$.
- Entries that are chosen or known beforehand are β, λ, L, H, Z, α, $n(t)$, $T(t)$, $C(t)$, ω, t_{max}, ρ, ρ', ε (precision, defined only for n).
- From those, we can find the parameters

$$b = \lambda + \frac{\beta - (u + \alpha T(t))}{L}, \quad c = \frac{u' + \lambda\,(u + \alpha T(t))}{L},$$

and

$$r_1 = \frac{-b + \sqrt{b^2 + 4c}}{2}, \quad r_1 = \frac{-b - \sqrt{b^2 + 4c}}{2},$$

$$k_1 = \frac{n'(t) - n(t) r_2}{(r_1 - r_2)}, \quad k_2 = \frac{n(t) r_1 - n'(t)}{(r_1 - r_2)}.$$

- Evaluate $n_0(t+\omega)$, $n_0'(t+\omega)$, and $T_0(t+\omega)$:

$$n_0(t+\omega) = k_1 e^{r_1(t+\omega)} + k_2 e^{r_2(t+\omega)},$$

$$n_0'(t+\omega) = k_1 r_1 e^{r_1(t+\omega)} + k_2 r_2 e^{r_2(t+\omega)},$$

$$T_0(t+\omega) = T(0) + H\left\{\frac{k_1}{r_1}\left(e^{r_1(t+\omega)} - e^{r_1 t}\right) + \frac{k_2}{r_2}\left(e^{r_2(t+\omega)} - e^{r_2 t}\right)\right\}$$

$$- Z\left\{\frac{k_1^2}{2r_1}\left(e^{2r_1(t+\omega)} - e^{2r_1 t}\right) + \frac{k_2^2}{2r_2}\left(e^{2r_2(t+\omega)} - e^{2r_2 t}\right) +\right.$$

$$\left. + \frac{k_1 k_2}{r_1 + r_2}\left(e^{(r_1+r_2)(t+\omega)} - e^{(r_1+r_2)t}\right)\right\}.$$

- Evaluate $P_0(t+\omega)$, $A_0(t+\omega)$, and $B_0(t+\omega)$:

$$P_0(t+\omega) = T_0(t+\omega)\left[\lambda n_0(t+\omega) + n_0'(t+\omega)\right],$$
$$A_0(t+\omega) = H\left[n_0^2(t+\omega)\right],$$
$$B_0(t+\omega) = Z\left[n_0^3(t+\omega)\right].$$

- Evaluate $n_1(t+\omega)$ and $T_1(t+\omega)$:

$$Q_0 = \frac{\alpha}{L}\left[P_0(t+\omega) + A_0(t+\omega) - B_0(t+\omega)\right],$$
$$n_1(t+\omega) = \frac{Q_0}{c},$$
$$T_1(t+\omega) = H\left(\frac{Q_0}{c}\right)(t+\omega) - Z\left(\frac{Q_0}{c}\right)^2(t+\omega).$$

- From $j=1$, the tolerance is tested. Thus, if $n_j(t+\omega) - n_{j-1}(t+\omega) \geqslant \varepsilon$, then $j = j+1$; otherwise, $j=0$ and $t = t+\omega$.

18.3 Results

The parameters used in the problem were $\beta = 0.0065$, $\lambda = 0.07741\,\mathrm{s}^{-1}$, $H = 0.05\,\mathrm{K/MWs}$, $\alpha = 5\times10^{-5}\,\mathrm{K}^{-1}$, $L = 0.0001\,\mathrm{s}$, and $\varepsilon = 10^{-8}$, with the initial conditions $n(0) = 10\,\mathrm{MW}$, $n'(0) = 0\,\mathrm{MW/s}$, $T(0) = 300\,\mathrm{K}$, $C(0) = (1/(\lambda L))[\beta - \rho(0)]n(0)$. Also, the parameter Z, the initial condition $\rho(0)$, and the length of the time step (for the numerical evaluation) were changed to observe their influence in the final solution. For each case, the reactor power evolution with time is shown in Figs. 18.1–18.11 for a selection of initial conditions $\rho(0)$.

For $\rho(0) = 0.2\beta$:
For a small positive reactivity, divergences were not observed regardless of the length of the time step. As expected, better approaches are obtained by using smaller Z and Δt. Note that for time steps smaller than $1s$, the results for $Z = \pm 0,001H$ are very close to the reference solution.

For $\rho(0) = 0.5\beta$:
As expected, smaller values of $|Z|$ give a better approach to the solution, for time steps $\Delta t \leq 1\,\mathrm{s}$.

For $\rho(0) = 0.8\beta$:
Something unexpected is observed for the case where $\rho(0) = 0.8\beta$. Regardless of the choice of Z, all the solutions diverge for $\Delta t = 10\,\mathrm{s}$. But for $\Delta t = 1\,\mathrm{s}$, the value of $Z = 0.001H$ provides an excellent approach. In smaller time steps, the error due to a bad assumption of Z becomes more evident.

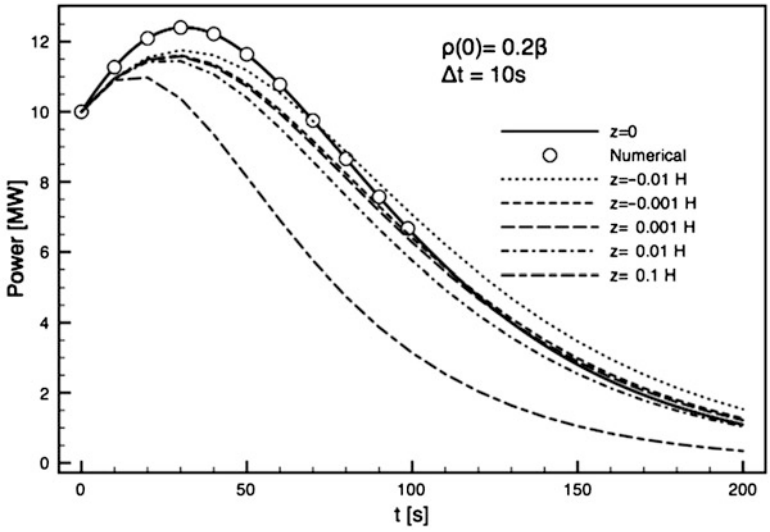

Fig. 18.1 Time evolution of power for $\rho(0) = 0.2\beta$ and 10 s time step

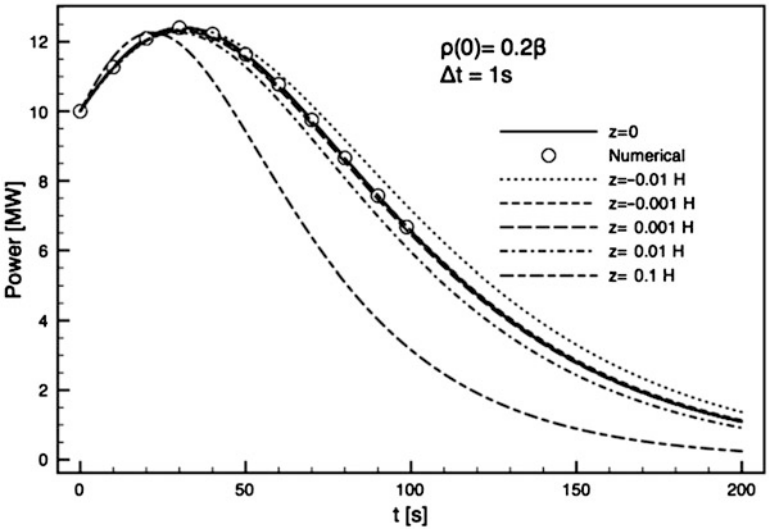

Fig. 18.2 Time evolution of power for $\rho(0) = 0.2\beta$ and 1 s time step

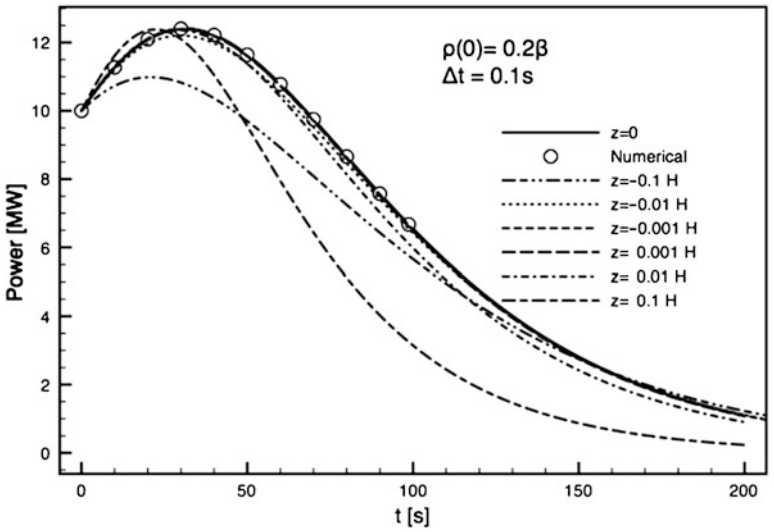

Fig. 18.3 Time evolution of power for $\rho(0) = 0.2\beta$ and 0.1 s time step

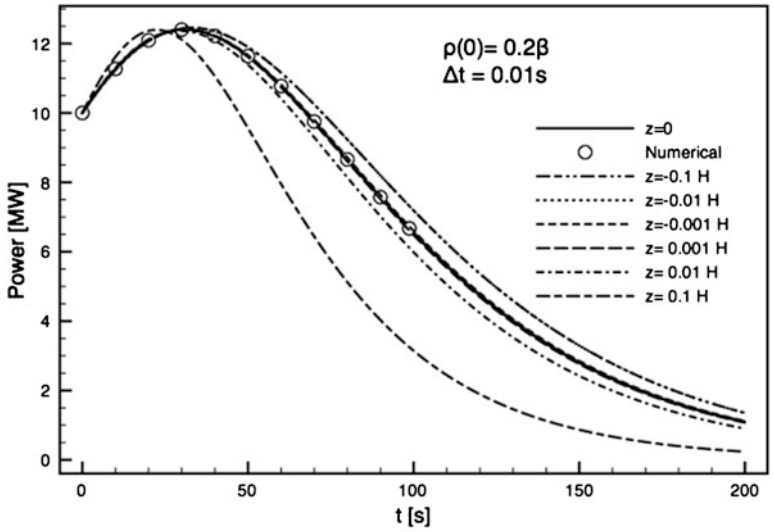

Fig. 18.4 Time evolution of power for $\rho(0) = 0.2\beta$ and 0.01 s time step

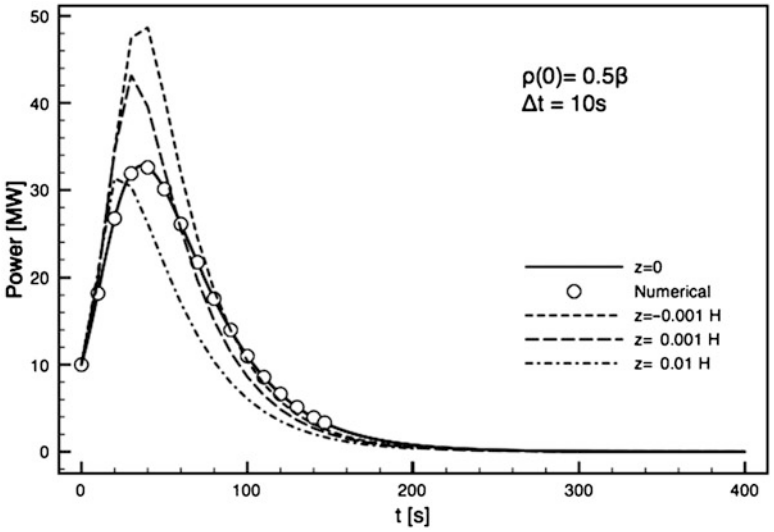

Fig. 18.5 Time evolution of power for $\rho(0) = 0.5\beta$ and 10 s time step

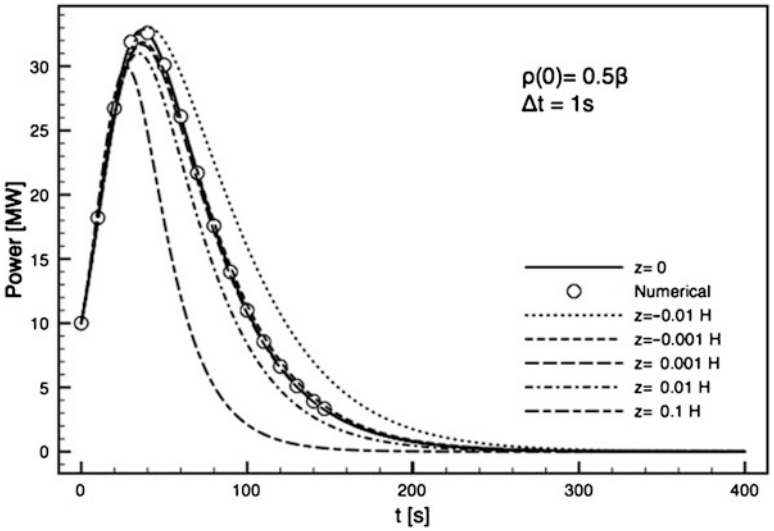

Fig. 18.6 Time evolution of power for $\rho(0) = 0.5\beta$ and 1 s time step

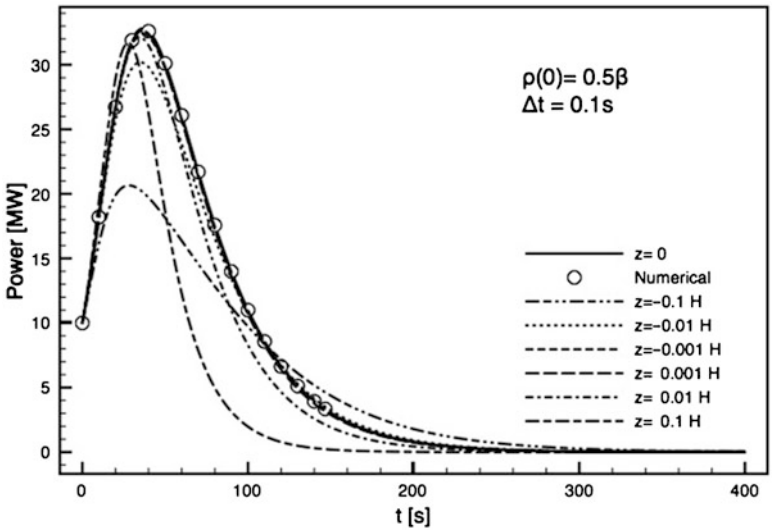

Fig. 18.7 Time evolution of power for $\rho(0) = 0.5\beta$ and 0.1 s time step

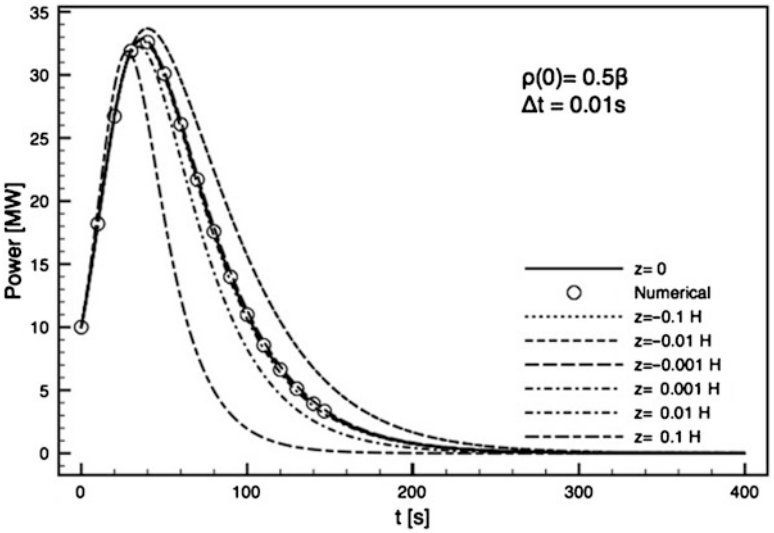

Fig. 18.8 Time evolution of power for $\rho(0) = 0.5\beta$ and 0.01 s time step

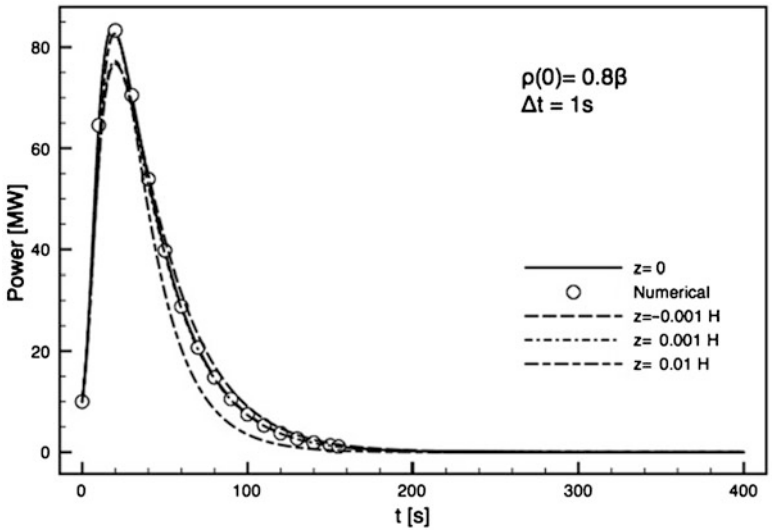

Fig. 18.9 Time evolution of power for $\rho(0) = 0.8\beta$ and 1 s time step

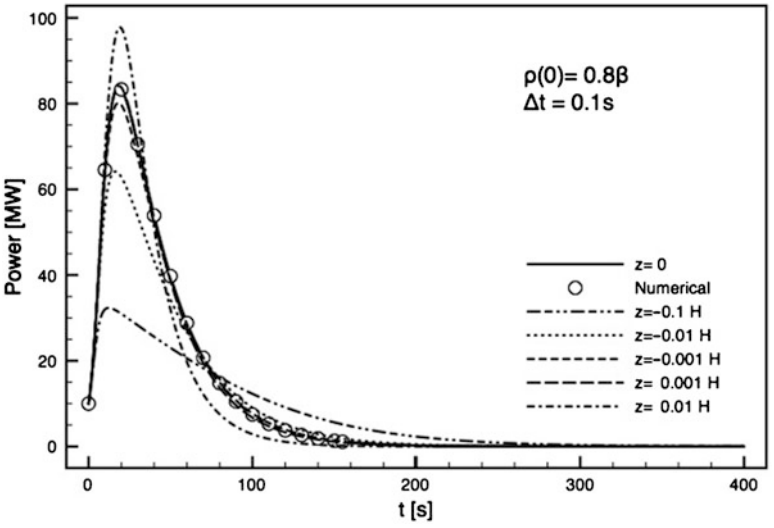

Fig. 18.10 Time evolution of power for $\rho(0) = 0.8\beta$ and 0.1 s time step

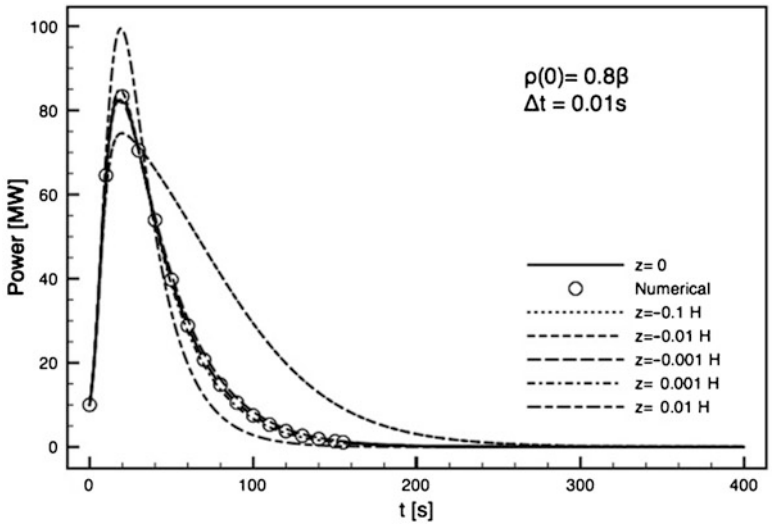

Fig. 18.11 Time evolution of power for $\rho(0) = 0.8\beta$ and 0.01 s time step

18.4 Conclusions

Comparing the model that takes into account the second order contribution of the neutron density with the one which uses only the first order term some conclusions can be drawn. The first is that, generally, the smaller the contribution of the n^2 term (which consists in a small $|Z|$), the more the solutions for first order and second order temperature models coincide.

Second, as the time step taken for the numerical evaluation decreases, the results improve (for small values of $|Z|$), but if the second order term is expressive, then the numerical procedure is likely to diverge, this is not the case for the Adomian-based approach. A mathematical proof of the convergence of the recursive scheme is a topic for future work.

References

[Ad89] Adomian, G.: Nonlinear Stochastic Systems Theory and Applications to Physics. Kluwer, Dordrecht (1989)

[Ad94] Adomian, G.: Solving Frontier Problems of Physics: The Decomposition Method. Kluwer, Dordrecht (1994)

[NaZa10] Nahla, A.A., Zayed, E.M.E.: Solution to the nonlinear point nuclear reactor kinetics equations. Progr. Nucl. Energ. **52**, 743–746 (2010)

[PeEtAl11] Petersen, C.Z., Dulla, S., Vilhena, M.T.M.B., Ravetto, P.: An analytical solution of the point kinetics equations with time-variable reactivity by the decomposition method. Progr. Nucl. Energ. **53**, 1091–1094 (2011)

Chapter 19
Eulerian Modeling of Radionuclides in Surficial Waters: The Case of Ilha Grande Bay (RJ, Brazil)

F.F. Lamego Simões Filho, A.S. de Aguiar, A.D. Soares, C.M.F. Lapa, and M.A.V. Wasserman

19.1 Introduction

The mathematical models that represent hydrodynamics and contaminant transport in water bodies are generally based on conceptual laws or principles expressed by differential equations. Numerical or numerical-analytical models translate mathematical equations to computational language (e.g., finite differences, finite elements, finite volumes, or probabilistic models) and have high predictive power and little loss of information. The uncertainty can be largely reduced with calibration process and model validation. For these reasons, the recommendation to move from box-model hydrological models (with high uncertainty level) to hydrodynamic process-oriented numerical modeling should be considered as an important issue for radionuclide transport.

The models are equation systems capable to quantify the flow and represent a practical way to forecast the behavior of water bodies. They are used to infer about known or hypothetical scenarios, allowing a better understanding of the system that is fundamental to decision makers, especially in accident situations. In case of accidental releases of liquid wastes from nuclear power plants, the previous knowledge about the advection and turbulent diffusion pathways in different scenarios is critical to providing the hydrodynamics basic information to simulate dispersion of radioactive pollutants. In this work we have used the Database System for

F.F. Lamego Simões Filho (✉) • C.M.F. Lapa • M.A.V. Wasserman
Institute of Nuclear Engineering, Rio de Janeiro, RJ, Brazil
e-mail: flamego@ien.gov.br; lapa@ien.gov.br; mwasserman@ien.gov.br

A.S. de Aguiar
Federal University of Rio de Janeiro, RJ, Brazil
e-mail: aguiargm@gmail.com

A.D. Soares
National Commission for Nuclear Energy, Rio de Janeiro, RJ, Brazil
e-mail: asoares@cnen.gov.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_19,

Environmental Hydrodynamics SisBAHIA that is a computational model applied to hydrodynamic circulation and advection–diffusion contaminant transport. It is suitable for natural or man-made water bodies under different meteorological, fluvial, lacustrine, or oceanographic scenarios and was developed by the Program on Coastal and Oceanographic Engineering of the Federal University of Rio de Janeiro since 1987.

19.2 Methodology and Modeling Approach

In all cases pertinent to modeling the transport of water constituents and determining their fate during a period of about a month, the focus will be in the far field; that is, in regions sufficiently far from the water outlets, away from the active turbulent mixing zones typical of the jets that form in the near field of the outlets. In these far regions, the plumes of constituents, including those of heated water, are passively transported by the prevailing currents. Thus, in a far field sense, the considered water constituents, including heat and particulate substances, can be treated as passive scalars. The passive scalar approach allows the decoupling of the transport modeling from the hydrodynamic circulation modeling. In this respect, the implicit hypothesis is that the hydrodynamic circulation in the far field is independent of the concentration distribution of a given constituent. The decoupling of the transport model from the hydrodynamic model allows us to neglect the baroclinic forcing in the latter. Therefore, in order to model the transport of constituents for a given scenario, the pertinent hydrodynamic circulation will be first modeled. This is because velocity fields and large-scale turbulence parameters, which are necessary input data for the transport models, are computed by the hydrodynamic models. The modeling approach is dependent on the features of the adopted modeling system that must comply with the physics of the problem. The models for the simulations of hydrodynamic circulation and transport of contaminants to be used in this project pertain to a system called SisBAHIA, as described below.

19.2.1 Hydrodynamical Modeling Approach

The hydrodynamics of most part of natural aquatic bodies is extremely complex due to the irregular geometric shape and also because of the diversity of features that produce the flow. The main forcing parameters are the winds, river discharges to the watersheds, tides, and water density. To get forcing data it is necessary to monitor in situ variations of water level, wind direction and speed, tide currents, temperature and salinity, because these parameters help to understand the hydrodynamic processes and establish the conceptual model. The main system attributes for hydrodynamics are:

- The FIST (filtered in space and time) hydrodynamic turbulence model is based on Large Eddy Simulation (LES) to simulate vortices.
- The model computes flow velocities either on three-dimensional (3D) or on two-dimensional on Horizontal or vertical averages (2DH).
- The spatial discretization is made through 4th order finite elements with two-quadratic squares or quadratic triangles or both.
- Sigma transformation is used to vertical discretization resulting in finite element mesh pile.
- Processing time is faster than 50 times the real time, i.e. one day of circulation is simulated in less than half hour.

The advection–diffusion contaminant transport modeling can be performed by two different modules of the system according to computational fluid dynamics (CFD) formulations. The Eulerian module works with fixed meshes as referential, while Lagrangian module uses adaptive meshes accompanying the movement of the particles of the pollutant. The general modeling approach was to include the whole bay in the modeling domain, and use finite element discretization techniques to model in proper detail the areas of interest around the Itaorna cove. Figure 19.1 illustrates these techniques, respectively, for the present situation, and for the situation foreseen the construction of Angra 3. The 3D spatial discretization is done via a vertical stack of sub-parametric finite element meshes using σ coordinate transformation along the vertical dimension. That is, if one looks from the top, one sees the horizontal plane of the domain discretized by a single mesh of finite elements (see Fig. 19.1). However, in fact, there will be a stack of meshes, one for every σ level. In this way, vertical discretization is done automatically once the user defines the number of desired σ levels (usually between 10 and 50). The 3D model is automatically activated if at least 5 σ levels are requested.

Elements in a mesh are sub-parametric, for that, the variables in each element are defined by quadratic Lagrangian polynomials whereas the element geometry is defined by linear Lagrangian polynomials. Elements in a mesh can be quadrilaterals and/or triangles. Quadrilaterals are preferred, because variables become bi-quadratic, and thus have a higher accuracy. This discretizing scheme is potentially of 4th order on the σ planes and of 2nd order on the σ dimension. In addition, the scheme allows very good representation of domains with complex geometries and bottom topography, as in the case of Bay of Ilha Grande. Temporal discretization is done through a second-order implicit factored scheme for nonlinear terms and a Crank–Nicholson scheme for linear terms. Phase errors are minimized because all terms in the numerical scheme are centered at the same instant, $t = (n + 1/2)\Delta t$. Phase errors are prone to occur in numerical schemes in which all terms are not centered in the same instant. Open boundaries elevations and current velocities can be prescribed in many different ways, including synthetic tides generated by given harmonic constants, and data measured or provided at discrete times. A different value, and/or phase shift, can be given for each node along any open boundary segment. Land boundaries can prescribe either normal or imposed directional fluxes or velocities. Fluxes or velocities can be constant or variable in time (a river

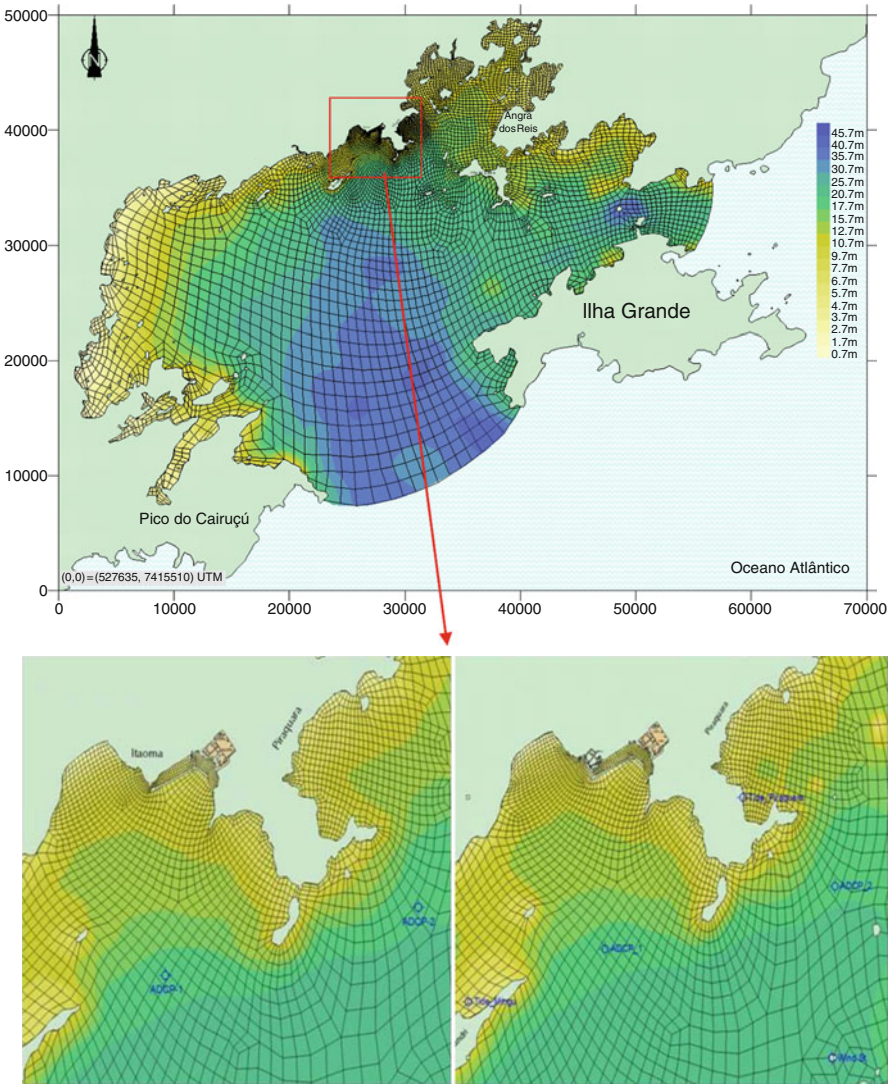

Fig. 19.1 Modeling domain considered in present (*left*) and future (*right*) situation, with the discretization mesh. The 3D domain is discretized by a stack of 21 finite element meshes. Each *crossing line* shown above represents a water column. See above the bathymetric map

discharge curve for instance). Leaky boundaries are allowed. Slip and no-slip boundaries are allowed, and the equivalent roughness along each boundary node can be prescribed. Surface and bottom boundary conditions for the 3D model, when zero velocity is the bottom boundary condition, and the wind stress is the free surface condition. The model accepts inputs of wind fields that can be variable in

space and time. The amplitude of the equivalent bottom roughness can be specified for each bottom node for computing the bottom stresses, reflecting the type of material (rock, sand, mud, vegetation, etc.). The computed friction coefficients of the bottom vary dynamically in time and space. A multi-scale model is employed to model turbulence with horizontal sub grid scale turbulent stresses based on filtering techniques, also known as Large Eddy Simulation (LES). Small scale horizontal and vertical turbulent stresses employ eddy viscosity approach. Eddy viscosity tensor is anisotropic and dynamically variable in space and time for each node.

19.2.2 Transport Modeling Approach

The main attributes of Eulerian transport modeling are:

- Eulerian advective–diffusive transport module with kinetic reactions is suitable for simulating the dispersion of dissolved substances.
- It is possible to apply this module for 2DH or to selected layers of 3D hydrodynamic output.
- Solve scale conflict with adaptive (changing) mesh only around the contaminants.
- Gain factor between modeling time and real time in processing are 5–8 times faster than FIST3D.

The Eulerian transport model in SisBAHIA solves the following conservation equation:

$$\frac{\partial C}{\partial t} + u_i \frac{\partial C}{\partial x_i} = \frac{1}{h}\frac{\partial}{\partial x_i}\left(hD_{ij}\frac{\partial C}{\partial x_i}\right) - (k_d - k_s)C + q_s(C_s - C), \quad i,j = 1,2 \quad (19.1)$$

where $C(x,y,t)$ is the concentration averaged over height of the water column or thickness of a surface layer $h(x,y,t)$, $u_i(x,y,t)$ is the velocity component in the x_i direction averaged over $h(x,y,t)$, $D_{ij}(x,y,t)$ is the turbulent diffusion and dispersion tensor averaged over $h(x,y,t)$, k_d is the time rate of mass consumption ($k_d > 0$) or production ($k_d < 0$), $k_s(x,y,t)$ is the time rate of removal of mass due to settling processes, $q_s(x,y,t)$ is the discharge per unit horizontal area at a source region, and $C_s(x,y,t)$ is the concentration at the source region. For the simulations of reference contaminants presented here the variable $h(x,y,t)$ is the whole water column. The time rate of removal of mass due to settling process is computed as

$$\begin{aligned} k_s &= \frac{\ln(0.1)}{h/V_S} \quad \text{if } \frac{\tau_0}{\tau_{0c}} \le 1 - a \text{ or } \frac{\tau_0}{\tau_{0c}} - 1 + a < 2a \times R[0,1], \\ k_s &= 0 \quad \text{if } \frac{\tau_0}{\tau_{0c}} \ge 1 + a \text{ or } \frac{\tau_0}{\tau_{0c}} - 1 + a > 2a \times R[0,1], \end{aligned} \quad (19.2)$$

where V_S is a constant characteristic settling velocity given by the user, $\tau_0(x,y,t)$ is the stress exerted by the flow at the bottom of the layer with thickness h, and τ_{0c} is

the critical bottom stress necessary to mobilize the particles settling with velocity V_S. The parameter a is a tolerance parameter between 0 and 0.5, and $R[0,1]$ is a random number with values between 0 and 1. If the user prescribes values for V_S, τ_{0c}, and a, the model computes k_s, which varies in time and space. When $\tau_0/\tau_{0c} < (1a)$ turbulence is weak and settling occurs ($k_s > 0$). When $\tau_0/\tau_{0c} > (1+a)$ turbulence is too strong and there is no settling, since $k_s = 0$. When $(1a) < \tau_0/\tau_{0c} < (1+a)$ the settling processes becomes probabilistic. Note that if $\tau_0/\tau_{0c} = 1$ there is a 50% chance of occurring settling. As $\tau_0/\tau_{0c} \to (1a)$ the chances of settling increase, and as $\tau_0/\tau_{0c} \to (1+a)$ the chances decrease. Since k_s varies in space and time it is not a rate constant as k_d, which is indeed a constant. k_s is a variable local rate of removal of suspended mass in the water column due to settling. Some models simply use $k_s = V_S/h$, which is the inverse of the maximum settling time (T_s) for a particle with a settling velocity V_S in a water column of height h. T_s can be considered a characteristic settling time. From a simple geometric reasoning, after a time T_s all particles should have settled. However, solving the equation for a still water situation one finds that after a time equal to T_s about 37% of the particles would remain in suspension. In addition, this simpler formulation allows settling even if, in reality, the flow is too turbulent for the occurrence of deposition in the bottom. The formulation in (19.2) is more realistic than the simplified formulation adopted in other models for two reasons:

- Mass is only removed from the water column, in a given position, when the flow is such that effective deposition in the bottom might occur. That is, when and so, the flow is quiescent enough for deposition to occur. The use of a tolerance value a is to account for the fact that usual criteria for defining critical bottom shear is not exact. The Shields curve, for instance, is just an adjusted curve in the middle of a cloud of experimental data.
- In a quiescent flow situation, 90% of the suspended particles will be deposited after a time equal to T_s. Theoretically 100% should have deposited, thus the model is still conservative, but not unrealistic.

The terrestrial boundary conditions imposed in present and future scenarios considered uptake and discharge in Itaorna cove, only discharge for Piraquara cove and included recirculation effects. At all other land boundary points the prescribed condition was of zero contaminant flux in the normal direction to the boundary. For open boundary points presenting inflow situations, the following conditions are used:

$$\begin{aligned} T &= T_0 + \frac{T - T_0}{2}\left(1 - \cos\left(\pi \frac{t - t_0}{\tau}\right)\right) \quad \text{when } t - t_0 \le \tau, \\ T &= T^* \quad \text{when } t - t_0 > \tau, \end{aligned} \tag{19.3}$$

where $T^*(t)$ are prescribed values, T_0 is the value of the concentration calculated at the boundary point in the instant t_0, which is the instant immediately before the outflow changed to inflow situation, and τ is a prescribed transition period, which depends on the modeler's experience or available data. Usual values for τ are in the

range of half an hour to two hours. This kind of condition is particularly useful in modeling estuarine boundary conditions. In outflow situations, the model simply computes the transport equation with no diffusive terms along the open boundary points.

19.3 Input Data and Boundary Conditions for Simulations

19.3.1 Bathymetry

The bathymetry of Ilha Grande Bay was defined through digitizing nautical charts edited by the bureau of Hydrography and Navigation (DHN chart numbers 1607, 1633, 1637, and 23100) added of the value of 0.68 cm to correct the reduction level used for navigation that in the present case correspond to the mean higher low water (MHLW). Thus, all depth values correspond to the mean level of the bay. These data were interpolated to generate a bathymetric map (Fig. 19.1) in which a depth value was assigned for each mesh node.

19.3.2 Astronomical Tide

The propagation of tide wave on the open borders was simulated from the measurements of water level inside the domain that allow to prescribe the boundary conditions for them. To simulate the Ilha Grande Bay model it was considered synthetic tides generated from the harmonic constants from Angra dos Reis Harbor. The specifications of tide height to the boundaries were calculated in each time step, using the harmonic constants shown in Table 19.1 from Angra dos Reis. It was simulated a time interval of 30 days that contained spring and neap tide cycles. Figure 19.2 shows the tide elevation curves from Angra dos Reis that were used as boundary condition for the performed simulations.

The positioning of open border 1 is almost perpendicular to tide front that propagates on the coast mostly form west to east, as well as is also perpendicular to open border 2, situated in a more sheltered zone. Thus, it is supposed to occur a phase lag between the two boundaries so the tide wave arrives first in the border one and sometime later in second border. This discrepancy was estimated to be around 600 s. During effluence conditions the boundary conditions are prescribed as the tide level oscillation. In order to do that it was used inverse modeling to estimate tide elevations on the borders, which are based in applying the harmonic constants from inside the domain (Table 19.1) and use the same overestimation percentage produced by model results to correct the border values. This was done because we do not have tide measurements from outside the domain to be used in modeling. On the other hand, during affluence time, the boundary condition adopted forced the flow to enter in normal direction (90°) to the border.

Table 19.1 Harmonic constants sorted by significance of amplitude (Angra dos Reis harbor station, Ilha Grande Bay)

Name	Period (s)	Amplitude (m)	Phase(degrees)
M2	44714.16439359	0.2869	1.3799
S2	43200.00000000	0.1649	1.4396
O1	92949.62999305	0.0967	1.4692
M4	22357.08219679	0.0332	0.5664
K1	86164.09076147	0.0535	2.4888
K2	43082.04523752	0.0516	1.2908
N2	45570.05368141	0.0356	2.1349
MS4	21972.02140437	0.0165	2.0408
MN4	22569.02607322	0.0144	6.0327
Q1	96726.08402232	0.0270	1.0818
L2	43889.83274041	0.0164	1.6310
P1	86637.20458000	0.0171	2.2640
2N2	46459.34813490	0.0098	2.2611
M3	29809.44292906	0.0121	3.4137
MU2	46332.00000000	0.0155	1.7054

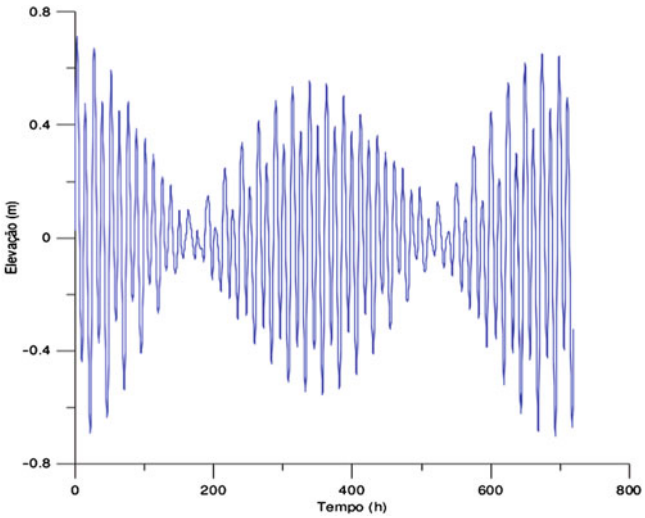

Fig. 19.2 Tide elevation curves for Angra dos Reis Harbor during one month, showing the forcing parameters used to model the Ilha Grande Bay, which are generated with the harmonic constants of Table 19.1

19.3.3 Wind Speed and Direction

The wind data for hydrodynamic circulation modeling could be supplied in several forms to the model. The data could be from time constant and uniform in space until

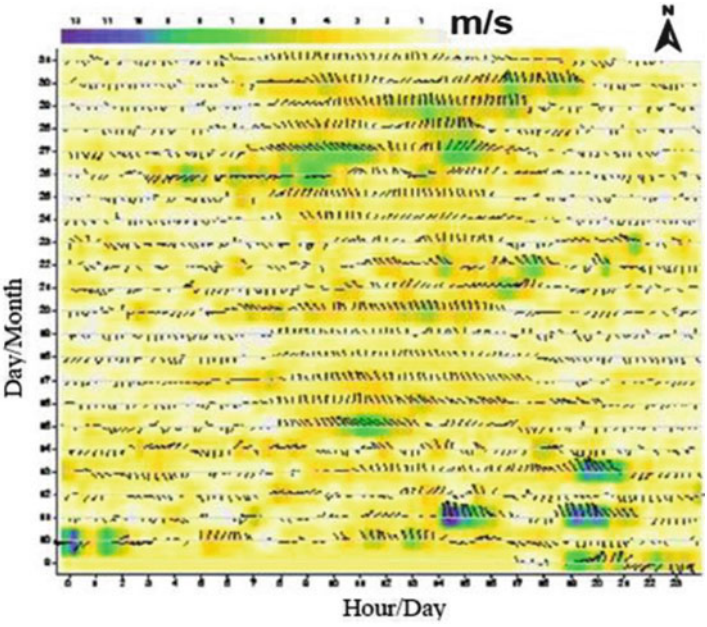

Fig. 19.3 Temporal series of usual winds measured at meteorological station supplied as input data to the model. The *arrows* have module proportional to wind speed showed by the color pattern and also point out the wind direction related to the geographical north

time variable and varied in space. The most common available wind input data are time variable but uniform in space. To simulate wind patterns typical of the area, it was selected the local wind regime, characterized by variations between night time and day time in wind speed and direction. The data used in this work were extracted from the meteorological station (B15) placed at the area of Nuclear Power Plants at Ponta Fina (Itaorna Beach) from a time series between 1995 and 1996. The place where the meteorological tower is situated is appropriated in terms of climate and geomorphologic aspects. The speed and direction of wind are measures each 15 min, but are manipulated to compose an hourly average value. The data analysis of this station, belonging to plant operator, allows distinguishing two different wind stages. A first one starts between 7 and 9 a.m. and finishes between 3 and $5p.m.$ with dominant winds showing N and ESSE directions and velocities from 4 to 8 m/s, which can reach 12 m/s. The second stage ranges from $5p.m.$ to $7a.m.$, when occurs the domination of smaller intensity winds (1–5 m/s) with WSW and SW directions. Figure 19.3 shows this well-marked behavior of wind regime of Ilha Grande Bay.

19.3.4 River Discharge

The watershed of the bay characterizes as a estuarine system where the mountains (Serra do Mar) are in direct contact with the sea and the coastal plains practically

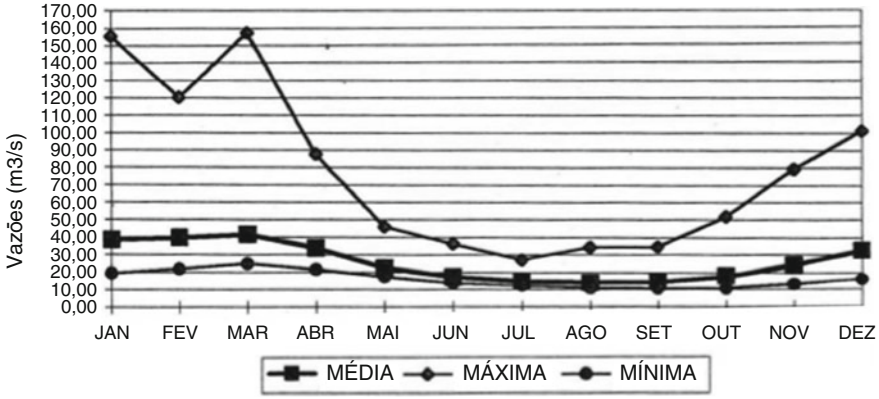

Fig. 19.4 Hydrograph of discharges from Mambucaba River

do not exist. The majority of the rivers is poorly drained and shows in general low average discharge. They present high steepness because their springs are in high altitude but with small extension, in order of 15 km, where the larger discharges occur during the summer. The most important one is Mambucaba River with drainage basin of 592 km^2 that corresponds to 78% of the watershed area. The average discharge is 27.5 m^3/s with the larger values between January and March, when maximum discharges reach 157 m^3/s and smaller values between June and October when the average and minimum discharges are respectively 14 and 10 m^3/s.

The Mambucaba River was used to prescribe the boundary conditions for terrestrial closed border. As mentioned before, it was prescribed zero for all the nodes, with exception of Mambucaba River Discharge (Annual Average Discharge, see Fig. 19.4) as well as the other two points (reception and discharge of seawater) accounting for a total flux of 120 m^3/s only for the scenario 2. The discharge input values are calculated taking also into account the cross-section area of the river, reception and discharge points. It was also considered the effect of lateral friction on closed borders that modifies the friction tension in the bottom, including a sliding index (between 0 and 1), prescribed in the present case to 0.7.

19.3.5 Hydrodynamic Model Remarks

The main aspects of hydrodynamic modeling are presented for present and future scenarios.

19.3.5.1 Future Scenario: Angra 1, 2, and 3 Operating with Discharges in Itaorna and Piraquara

Figure 19.5 presents typical current patterns, respectively, for flooding tides and ebbing tides. For this case, it is irrelevant to compare situations in spring and neap

tides because the visual aspect is practically the same. That is so, for the following reasons:

- The circulation patterns in Itaorna cove are dominated by the inflow discharges of Angra 1, 2, and 3 at the entrance of the breakwater, and the outflow discharge of Angra 3. Current patterns in Piraquara cove are mainly affected by the outflow discharge of Angra 1 and 2.
- Tidal components in the prevailing currents are very small, with magnitudes often smaller than 0.05 m/s. The changes in the magnitudes of flooding and ebbing tidal components within Itaorna cove and Piraquara cove, from spring to neap tides are subtle, in comparison with the prevailing circulation caused by the power plant discharges.

By examining Fig. 19.5, one sees that the recirculating cells formed by the effluent jet from Angra 3 are quasi steady, and quite insensitive to tidal conditions. The aspect of the recirculating cells remains practically the same during flood and ebb tides. It is interesting to note that during flooding tides the jet form Angra 3 opposes the natural flow in the channel to the North of Sandri Island, producing a stagnant zone in that region. Conversely, during ebbing tides, the jet from Angra 3 enhances the natural flow. A similar effect also occurs in Piraquara cove, when natural flooding currents are opposed by the effluent jet from Angra 1 and 2, while in ebbing tides the jet enhances the flow.

19.3.5.2 Present Scenario: Angra 1 and 2 Operating with Discharge in Piraquara Cove

Figure 19.6 present typical current patterns, respectively, for flooding tides and ebbing tides. As discussed in the previous scenario. Also for this case, it is irrelevant to compare situations in spring and neap tides because the visual aspect is practically the same, due to the same reasons. Contrary to what is observed in Piraquara cove for previous scenario (see Fig. 19.5), as one can see in Fig. 19.6, flooding currents do not create a stagnant zone in front of the stronger effluent jet. Flooding currents in Itaorna are enhanced by the water intake at the entrance of the breakwater. Conversely, ebbing currents are opposed by water intake, and a stagnant zone appears in front of breakwater.

19.4 Transport Model Remarks

This section presents the results of the transport modeling of radioactive elements present in the released liquid wastes according with the both hydrodynamic scenarios. In the first, the nuclear plants shut down without any further pumping of sea water, while the other one still keep pumping and discharging operations at the same rates. For the sake of conciseness, this paper will present only the dispersion

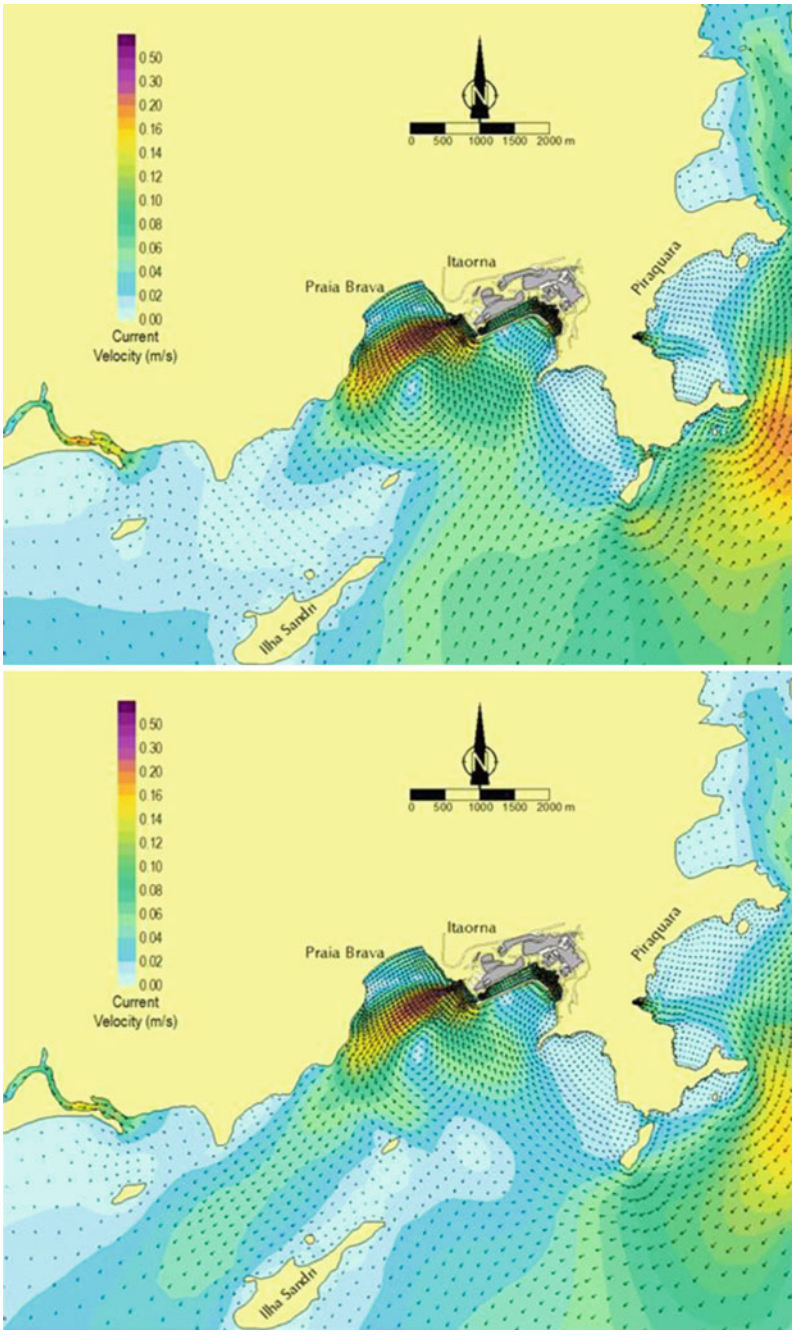

Fig. 19.5 Typical current pattern in Future Scenario for flooding (*above*) and ebbing (*below*) tides

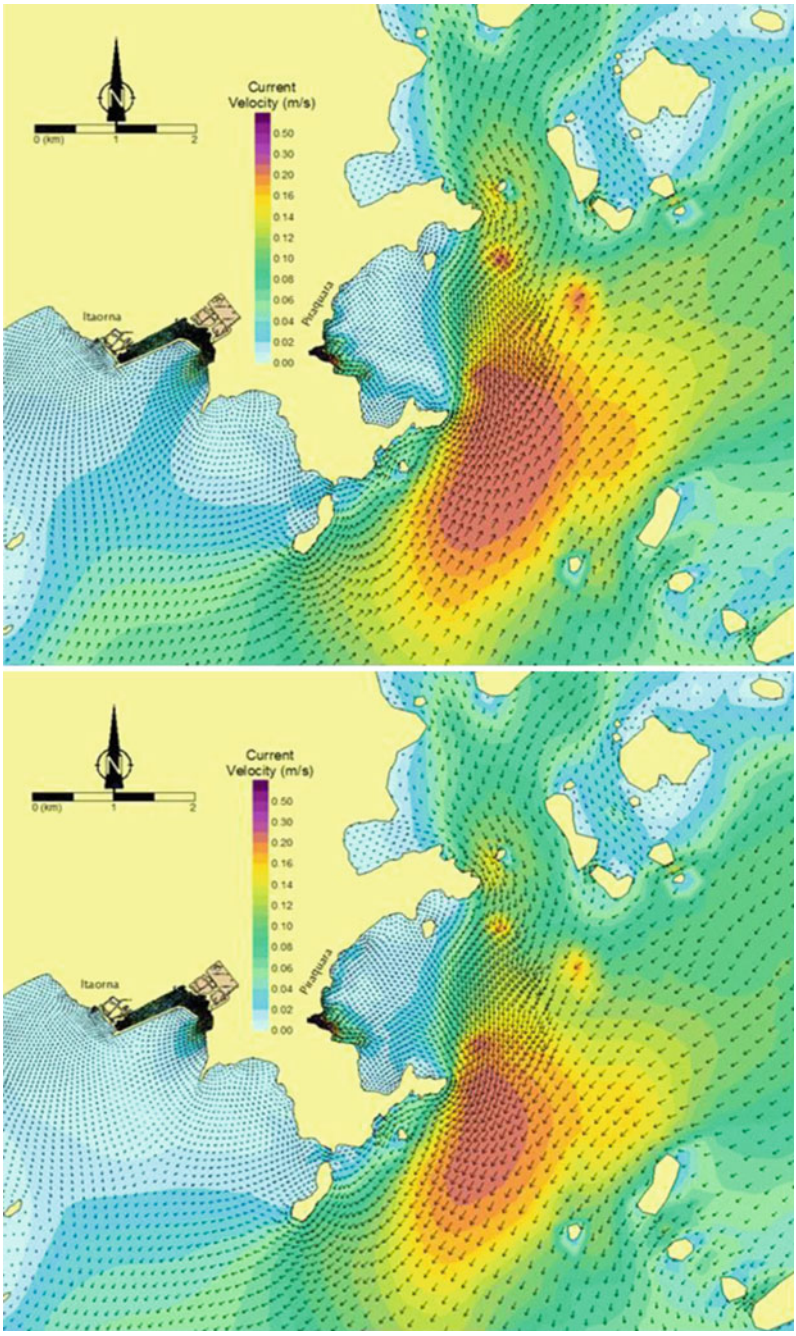

Fig. 19.6 Typical current pattern in Future Scenario for flooding (*above*) and ebbing (*below*) tides

Table 19.2 Parameters of a LOCA event causing tritium release (HTO) to Ilha Grande bay with the values of initial concentration load and requires dilution to the reference limit

Discharge = $\frac{0.0183\,m^3/s}{\text{Pollutant}}$	HTO initial concentration (Bq/m^3)	Pollutant load (Bq/s)	Concentration level ^3H-seawater (Bq/m^3)	Required dilution of source
^3H	$6.8E+10$	$3.7E+12$	$1.11E+06$	60,000

simulation of conservative pollutants. In this case, the results for tritium (HTO) were selected, because it shows the major inventory of radionuclides in the coolant systems. Loss of coolant accident could result in release of large amounts of tritium even if the leaks are initially contained in the reactor building. It was analyzed a scenario for hypothetical accidental tritium release in Ilha Grande Bay of 37 PBq of HTO in a volume of 66 m^3 of coolant after a LOCA event. The waste was released into the concrete ground around the plant producing a discharge of 0.018 m^3/s of liquid wastes to Ilha Grande Bay during 1 h after the accident (Table 19.2).

It was defined by the licensing the concentration of radioactive material dissolved or entrained noble gases released from the site shall be limited to 1.11 MBq/m^3. This specification is provided to ensure that the concentration of radioactive materials released in liquid waste effluents from the site will be less than the concentration levels specified in 10 CFR 20 (NRC, 2007). The value is applicable to the assessment and control of dose to the public. It is equivalent to the radionuclide concentrations which, if inhaled or ingested continuously over the course of a year, would produce a total effective dose equivalent of 0.5 mSv. The discharge input values are calculated taking also into account the cross-section area of discharge points. The tritium behavior was considered conservative once it forms the water molecule like its isotope hydrogen and remains in solution. Even the radioactive decay was considered negligible for the effect of modeling, once the simulation time corresponds to less than 10% of its half-live (12, 6 years). The modeling results of tritium dispersion released in Ilha Grande Bay were performed for the period between 24 h and 1 year after the accident. The reference levels explained above implicate in an intervention act if the concentration of radioactive material released from the site exceeding the above limits and immediately restore the concentration to within the above limits. The results of the 3D analytical-numerical hydrodynamic model, considering the effects of forcing input data on advection and turbulent diffusion, showed in the previous section as a velocity field that defines the water circulation in the Ilha Grande Bay, are the basis to model the transport of HTO. It should be reinforced the equations showed before applied to model advective and diffusive transport as well as in case of kinetic reactions of radionuclides occur, always considering an Eulerian referential At the end of the first day after the accident, tritium concentrations of up to 1 GBq/m^3 close to Itaorna beach could be observed, which means a dilution required of thousand times in relation to the limit. However, these values spread on a very limited surface. The tritium dispersion

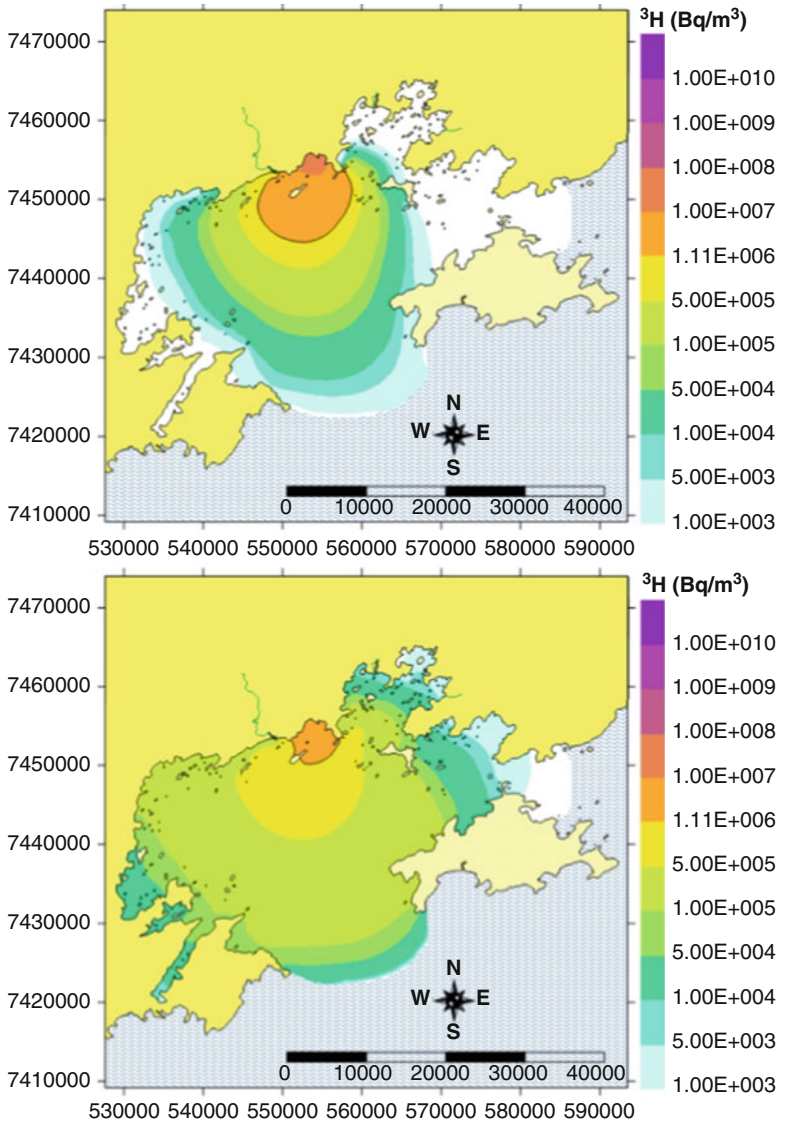

Fig. 19.7 Dispersion of HTO plume 3 days after the accident in the scenarios without (*above*) and with (*below*) seawater recirculation

for the two scenarios in the bay as a whole was similar. The differences were observed between the third and fourteenth day (Figs. 19.7–19.9), time interval in which the HTO plume with concentrations above the limit reach a maximum spread, extending in an area with more than 10 km of diameter. In the second scenario, with pumping and discharge operations, the area occupied by such concentration levels

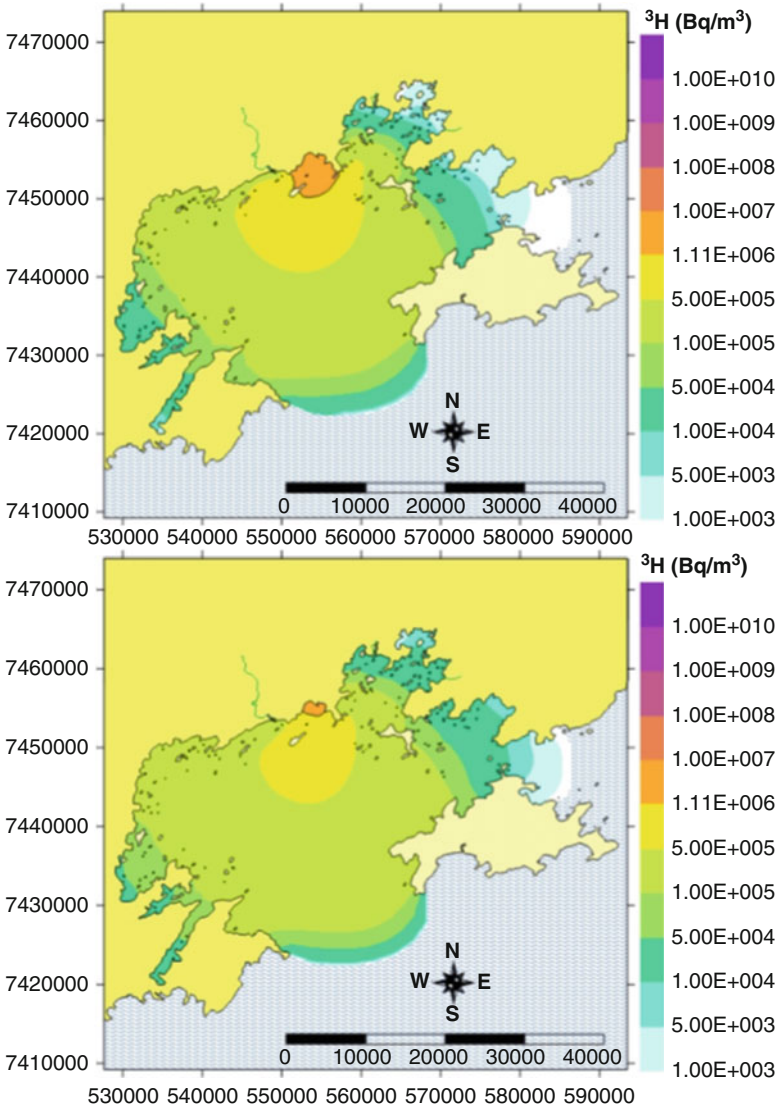

Fig. 19.8 Dispersion of HTO plume after 10 days of the accident in the scenarios without (*above*) and with (*below*) seawater recirculation

decreased more quickly in the time interval considered than in the first scenario. Such difference was caused by the removal of large volume of polluted waters from the accident site and its dilution in the discharge area, which has minor tritium concentrations. As a result of the dilution enhancement promoted by keeping the other plants operating, the tritium reference limit will not be more exceeded starting from the eleventh day and thereafter. This will occur only after the fifteenth day

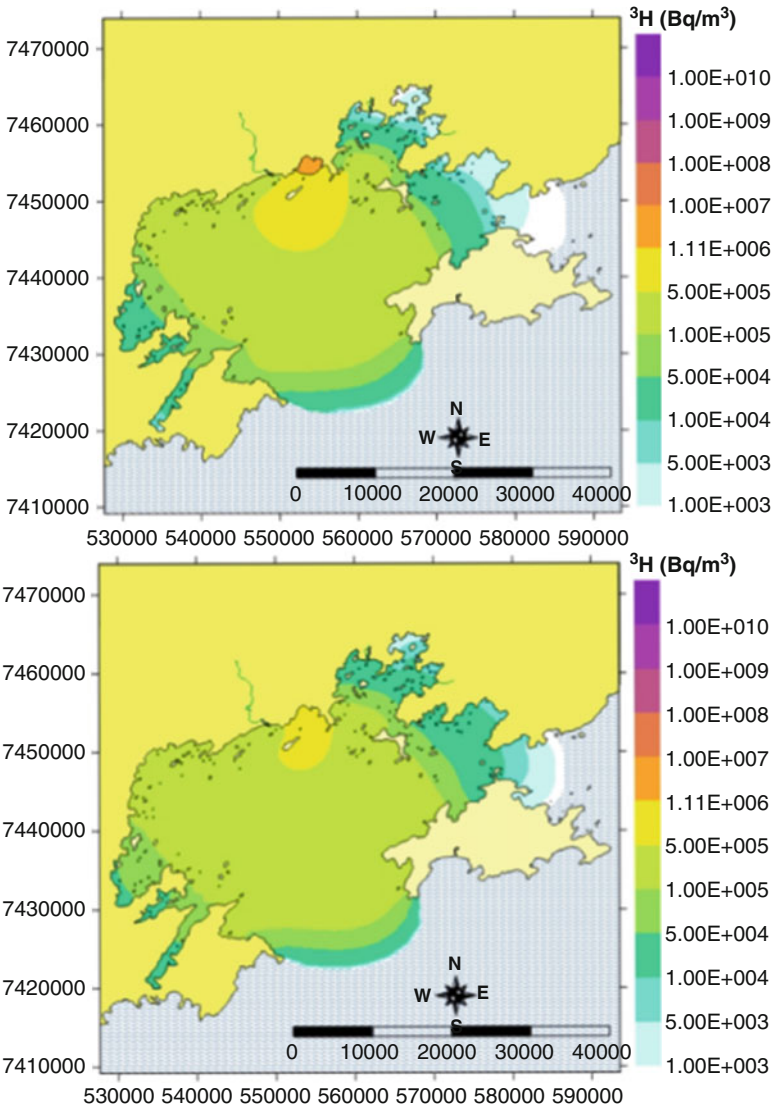

Fig. 19.9 Dispersion of HTO plume after 14 days of the accident in the scenarios without (*above*) and with (*below*) seawater recirculation

in the first scenario. Thus, increase the pumping and discharging rates could be used in the first days after the accident to accelerate the dilution. However, this operation would be effective only to manage the highest concentrations of the plume. It has shown no significant difference to the general distribution of HTO plume in the bay. The plume showed results quite similar in both scenarios for the

moments corresponding to the 1, 3, and 6 months. After one month, the plume reach maximum spread in the bay, when concentrations of the order of 50 KBq/m^3 could be observed in its major part. Such concentrations are still high and due to the uncertainties of the incurred effective dose by accumulation of organically bound tritium (OBT) in seafood it would implicate in the continuous monitoring of fish species consumed by local populations. After that time, the concentrations quickly started to decrease. After three months, the estuary would present concentrations lower than the detection limit (DL) of the technique (11 KBq/m^3) applied by the plant operator in the environmental monitoring program, with the exception of Ribeira Bay waters where the fish OBT monitoring should be sustained. After six months, the HTO plume would be completely undetectable. Finally, after one year of the accident, it is considered that the Ilha Grande Bay would return to its original condition, once the plume shows concentrations of the same order of the results obtained by a previous study of tritium routine releases using the same model, after reaching steady state conditions.

19.5 Conclusions

Model results were remarkably good in reproducing registered water level variations in all situations. Tidal components and meteorological oscillations were almost perfectly matched by computational modeling with SisBAHIA®. Results were very accurate even during the occurrence of unusual and rapid oscillations. Model results were very good in reproducing the tidal and local wind components of the registered currents. However, they were not good in representing the residual currents. This is quite an interesting fact that should to be exploited in future research. Transport modeling results showed a fast dilution of tritium in Ilha Grande Bay according to the circulation scenario, becoming more diluted in case of the recirculation of seawater promoted by the maintenance of pumping and discharging operations. However, the reference limit is exceeded at least during the first 10 days after the accident. The increase of the pumping rate during this period should be considered as an action to speed up the dilution and mitigate the impact of the accident. However, according to the linear non-threshold (LNT) paradigm, the impact of lower concentrations of tritium converted to organically bound form (OBT) and ingested by human populations and biota of tropical environments remains unknown. Some further experimental work with tropical biota is necessary to assess this issue.

References

[HuEtAl04] C-An Huh, et al. Marine environmental radioactivity near Nuclear Power Plants in Northern Taiwan, Journal of Marine Science and Technology, **12**(5), 418–23, (2004)

[Co04] COPPETEC. Environmental aspects concerning Angra 3. Final Report. PENO-4841. (2004)

[Ha87] D.R.F. Halerman, Water Quality Control. Massachusetts Institute of Technology (1987)

[NRC07] NRC, National Regulatory Commission. 10 CFR Part Appendix A to Part 50–General Design Criteria for Nuclear Power Plants. http://www.nrc.gov/reading-rm/doc-collections/cfr/part050/part050-appa.html (2007)

[Ro12] P.C.C Rosman. Referência Técnica do SisBaHiA. 2012 http://www.sisbahia.coppe.ufrj.br/SisBAHIA_RefTec_V92.pdf

[Ro87] P.C.C. Rosman, Modeling Shallow Water Bodies via Filtering Techniques. Ph.D. Thesis, Dept. of Civil Engineering, Massachusetts Institute of Technology (1987)

Chapter 20
Fractional Calculus: Application in Modeling and Control

J. Tenreiro Machado

20.1 Introduction

The generalization of the concept of derivative $D^{\alpha} f(x)$ to non-integer values of α goes back to the beginning of the theory of differential calculus. In fact, Leibniz, in his correspondence with Bernoulli, L'Hôpital (1695), had several notes about the calculation of $D^{\frac{1}{2}} f(x)$. The development of the theory of Fractional Calculus (FC) is due to the contributions of many mathematicians such as Euler, Liouville, Riemann, and Letnikov [Ol74], [SaKiMa93], [MiRo93]. In the fields of physics and engineering, FC is presently associated with the modeling of electrochemical reactions, irreversibility, and electromagnetism [Ko84], [ToBa84], [Fe96], [LeNiNi98], [Hi00], [SaAgMa07], [Ma10], [Ka11]. The adoption of the FC in control algorithms has been studied [Ro67], [An94], [WeEk94], [Ma97], [Ma01], [We02], [We03], [Za05], [Ma06], [Ta10], [CaEtAl10], [Di10], [Le11], [Or11], [BaMaLu11], [BaEtAl12] using the frequency and discrete-time domains. Nevertheless, this research is still giving its first steps and further investigation is required. This article introduces the fundamental aspects of the theory of FC and the modeling and control of dynamical systems.

The paper is organized as follows. Section 20.2 outlines the main mathematical aspects of the theory of FC. Section 20.3 introduces the main algorithms to approximate fractional-order derivatives. Sections 20.4 and 20.5 present examples of the implementation of FC-based models and controllers. Finally, Sect. 20.6 draws the main conclusions.

J.T. Machado (✉)
Institute of Engineering, Polytechnic of Porto, Department of Electrical Engineering, 4200-072 Porto, Portugal
e-mail: jtm@isep.ipp.pt

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_20,

20.2 Main Mathematical Aspects of the Theory of Fractional Calculus

Can the order of derivatives and integrals be extended to have meaning with any number irrational, fractional, or complex? Gottfried Leibniz invented that idea in 1695 and exchanged correspondence with Guillaume l'Hôpital about it. The concept motivated mathematicians, physicists, and engineers to develop the concept of FC both in theoretical aspects and in practical implementations [MaKiMa11], [Ma11]. This "new" mathematical tool is in fact as old as the standard differential calculus and has been the subject of research for more than three centuries. Figure 20.1 depicts the time line [MaKiMa10b] of the most important scientific contributions during 1695–1970. During the last decades FC was recognized to be a splendid tool to model and to analyze complex dynamical systems and we witnessed the emergence of a large number of contributions. Figure 20.2 shows the time line [MaKiMa10a] of many relevant conferences, books, and special issues held during 1966–2010 (interested readers can download the A3 posters from the web site of Journal of Fractional Calculus & Applied Analysis http://www.math.bas.bg/~fcaa/).

There are several different definitions of fractional derivatives and their comparison is outside the scope of this sub-section. The most used definitions of a fractional derivative of order α are, respectively, the Riemann-Liouville (RL), Grünwald-Letnikov (GL), and Caputo (C) formulations [Ki94], [KiSrTr06]:

$$
{}_{a}^{RL}D_{t}^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_a^t \frac{f(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau,\ t > a,\ \mathrm{Re}(\alpha) \in]n-1, n[,
$$

$$
{}_{a}^{GL}D_{t}^{\alpha} f(t) = \lim_{h \to 0} \frac{1}{h^{\alpha}} \sum_{k=0}^{\left[\frac{t-a}{h}\right]} (-1)^k \binom{\alpha}{k} f(t-kh),\ t > a,\ \alpha > 0,
$$

$$
{}_{a}^{C}D_{t}^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_a^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau,\ t > a,\ n-1 < \alpha < n,
$$

where $\Gamma(\cdot)$ is Euler's gamma function, $[x]$ means the integer part of x, and h is the step time increment.

These operators capture the history of all past events, in opposition to integer derivatives that are "local" operators. This means that fractional order systems have a memory of the dynamical evolution. This behavior has been recognized in several natural and man-made phenomena and their modeling becomes much simpler using the tools of FC, while the counterpart of building integer order models leads often to complicated expressions. The geometrical interpretation of fractional derivatives has been the subject of debate and several perspectives had been forwarded [Ta95], [Po02], [Ma03], [Ma09].

Fig. 20.1 Time line of FC during the period 1695–1970 [MaKiMa10b]

Fig. 20.2 Time line of FC during the period 1966–2010 [MaKiMa10a]

Using the Laplace transformation, we have the expressions

$$\mathscr{L}\left\{{}_0^{RL}D_t^{\alpha} f(t)\right\} = s^{\alpha}\mathscr{L}\left\{f(t)\right\} - \sum_{k=0}^{n-1} s^k \, {}_0^{RL}D_t^{\alpha-k-1} f\left(0^+\right),$$

$$\mathscr{L}\left\{{}_0^{C}D_t^{\alpha} f(t)\right\} = s^{\alpha}\mathscr{L}\left\{f(t)\right\} - \sum_{k=0}^{n-1} s^{\alpha-k-1} \, f^{(k)}(0),$$

where s and $\mathscr{L}$ denote the Laplace variable and operator, respectively.

The Mittag–Leffler (ML) function $E_{\alpha}(t)$ is defined as [Kl09], [HaMaSa11]

$$E_{\alpha}(t) = \sum_{k=0}^{\infty} \frac{t^k}{\Gamma(\alpha k + 1)}, \quad \alpha \in \mathbb{C}, \ \mathrm{Re}(\alpha) > 0. \tag{20.1}$$

The function $E_{\alpha}(t)$ was defined and studied by Mittag–Leffler in the year 1903. It is a direct generalization of the exponential series. The ML function forms a bridge between the exponential and the power laws. The first occurs in phenomena governed by integer order and the second in fractional order dynamics. In particular, when $\alpha = 1$ the ML function simplifies and we have $E_1(t) = e^t$, while, for large values of t, the asymptotic behavior yields

$$E_{\alpha}(-t) \approx \frac{1}{\Gamma(1-\alpha)} \frac{1}{t}, \quad \alpha \neq 1, \ 0 < \alpha < 2.$$

Since the Laplace transform leads to

$$\mathscr{L}\left\{E_{\alpha}(\pm a t^{\alpha})\right\} = \frac{s^{\alpha-1}}{s^{\alpha} \mp a},$$

we observe a generalization of the Laplace transform pairs from the exponential towards the ML, namely from integer up to fractional powers of s.

The more general Mittag–Leffler function, often called two-parameter ML function, is given by

$$E_{\alpha,\beta}(t) = \sum_{k=0}^{\infty} \frac{t^k}{\Gamma(\alpha k + \beta)}, \quad \alpha, \beta \in \mathbb{C}, \ \mathrm{Re}(\alpha), \mathrm{Re}(\beta) > 0. \tag{20.2}$$

The function defined by (20.1) gives a generalization of (20.2), since $E_{\alpha}(t) = E_{\alpha,1}(t)$. This generalization was studied by Wiman in 1905, Humbert and Agarwal in 1953, and others.

Based on the proposed definitions it is possible to calculate the fractional-order integrals/derivatives of several functions (Tables 20.1 and 20.2), where $H(\cdot)$, $\delta(\cdot)$,

Table 20.1 Riemann–Liouville fractional derivatives with lower terminal at 0

$f(t)$	${}_0^{RL}D_t^{\alpha} f(t), t>0, \alpha\in\mathbb{R}$
$H(t)$	$\frac{t^{-\alpha}}{\Gamma(1-\alpha)}$
$H(t-a)$	$\begin{cases} \frac{(t-a)^{-\alpha}}{\Gamma(1-\alpha)} & t>a \\ 0 & 0\le t\le a \end{cases}$
$H(t-a)f(t)$	$\begin{cases} {}_a^{RL}D_t^{\alpha} f(t) & t>a \\ 0 & 0\le t\le a \end{cases}$
$\delta(t)$	$\frac{t^{-\alpha-1}}{\Gamma(-\alpha)}$
$\delta^{(n)}(t)$	$\frac{t^{-\alpha-n-1}}{\Gamma(-\alpha-n)}, n\in\mathbb{N}$
$\delta^{(n)}(t-a)$	$\begin{cases} \frac{(t-a)^{-\alpha}}{\Gamma(1-\alpha)} & t>a \\ 0 & 0\le t\le a \end{cases}$
t^{ν}	$\frac{\Gamma(\nu+1)}{\Gamma(\nu+1-\alpha)} t^{\nu+\alpha}, \nu>-1$
$e^{\lambda t}$	$t^{-\alpha} E_{1,1-\alpha}(\lambda t)$
$\cosh\left(\sqrt{\lambda}t\right)$	$t^{-\alpha} E_{2,1-\alpha}\left(\lambda t^2\right)$
$\frac{\sinh\left(\sqrt{\lambda}t\right)}{\sqrt{\lambda}t}$	$t^{1-\alpha} E_{2,2-\alpha}\left(\lambda t^2\right)$
$\ln(t)$	$\frac{t^{-\alpha}}{\Gamma(1-\alpha)}[\ln(t)+\psi(1)-\psi(1-\alpha)]$
$t^{\beta-1}\ln(t)$	$\frac{\Gamma(\beta)t^{\beta-\alpha-1}}{\Gamma(\beta-\alpha)}[\ln(t)+\psi(\beta)-\psi(\beta-\alpha)], Re(\beta)>0$
$t^{\beta-1}E_{\mu,\beta}(\lambda t^{\mu})$	$t^{\beta-\alpha-1}E_{\mu,\beta-\alpha}(\lambda t^{\mu}), \beta,\mu>0$

Table 20.2 Riemann–Liouville fractional derivatives with lower terminal at $-\infty$

$f(t)$	${}_{-\infty}^{RL}D_t^{\alpha} f(t), t>0, \alpha\in\mathbb{R}$
$H(t-a)$	$\begin{cases} \frac{(t-a)^{-\alpha}}{\Gamma(1-\alpha)} & t>a \\ 0 & t\le a \end{cases}$
$H(t-a)f(t)$	$\begin{cases} {}_a^{RL}D_t^{\alpha} f(t) & t>a \\ 0 & t\le a \end{cases}$
$e^{\lambda t}$	$\lambda^{\alpha} e^{\lambda t}, t>0$
$e^{\lambda t+\mu}$	$\lambda^{\alpha} e^{\lambda t+\mu}, t>0$
$\sin(\lambda t)$	$\lambda^{\alpha}\sin\left(\lambda t+\alpha\frac{\pi}{2}\right), \lambda>0, \alpha>-1$
$\cos(\lambda t)$	$\lambda^{\alpha}\cos\left(\lambda t+\alpha\frac{\pi}{2}\right), \lambda>0, \alpha>-1$
$e^{\lambda t}\sin(\mu t)$	$\rho e^{\lambda t}\sin(\mu t+\alpha\phi), \lambda,\mu>0$ $\rho=\sqrt{\lambda^2+\mu^2}, \phi=\arctan\left(\frac{\mu}{\lambda}\right)$
$e^{\lambda t}\cos(\mu t)$	$\rho e^{\lambda t}\cos(\mu t+\alpha\phi), \lambda,\mu>0$ $\rho=\sqrt{\lambda^2+\mu^2}, \phi=\arctan\left(\frac{\mu}{\lambda}\right)$

and $\psi(z)=\frac{\Gamma'(z)}{\Gamma(z)}$ are the Heaviside, Dirac, and Digamma functions, respectively [Po99a]. Nevertheless, the problem of devising and implementing fractional-order algorithms is not trivial and will be the topic of the next sections.

20.3 Approximations to Fractional-Order Derivatives

In this section we analyze two methods for implementing fractional-order derivatives, namely the frequency-based and the discrete-time approaches, and its implication in control algorithms.

In order to analyze a frequency-based approach to D^{α}, $0 < \alpha < 1$, let us consider the recursive circuit [CaHa64], [Ou91], [Ou95] represented in Fig. 20.1 such that

$$i = \sum_{k=0}^{n} i_k,$$

$$R_{k+1} = \frac{1}{\varepsilon} R_k,$$

$$C_{k+1} = \frac{1}{\eta} C_k,$$

where ε and η are scale factors, i is the current due to an applied voltage v, and R_k and C_k are the resistance and capacitance elements of the kth branch of the circuit.

The admittance $Y(\imath\omega)$ is given by

$$Y(\imath\omega) = \frac{I(\imath\omega)}{V(\imath\omega)} = \sum_{k=0}^{n} \frac{\imath\omega C\varepsilon^k}{\imath\omega CR + (\varepsilon\eta)^k},$$

where $\mathscr{F}$ denotes the Fourier transform operator, ω represents the frequency, $\mathscr{F}\{i(t)\} = I(\imath\omega)$, $\mathscr{F}\{v(t)\} = V(\imath\omega)$ and $\imath = \sqrt{-1}$.

Figure 20.2 shows the asymptotic Bode diagrams of amplitude and phase of $Y(\imath\omega)$. The frequencies of the poles ω_k and zeros ω'_k obey the recursive relationships

$$\frac{\omega'_{k+1}}{\omega'_k} = \frac{\omega_{k+1}}{\omega_k} = \varepsilon\eta,$$

$$\frac{\omega_{k+1}}{\omega'_k} = \varepsilon,$$

$$\frac{\omega'_k}{\omega_k} = \eta.$$

From the Bode diagram of amplitude or of phase, the average slope m' can be calculated as

$$m' = \frac{\ln\varepsilon}{\ln\varepsilon + \ln\eta}$$

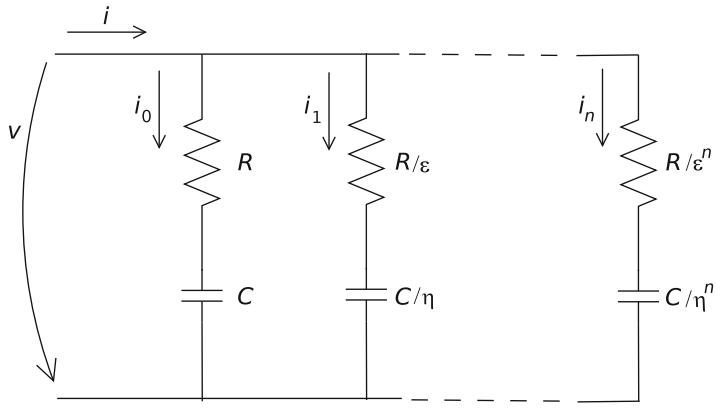

Fig. 20.3 Electrical circuit with a recursive association of resistance and capacitance elements

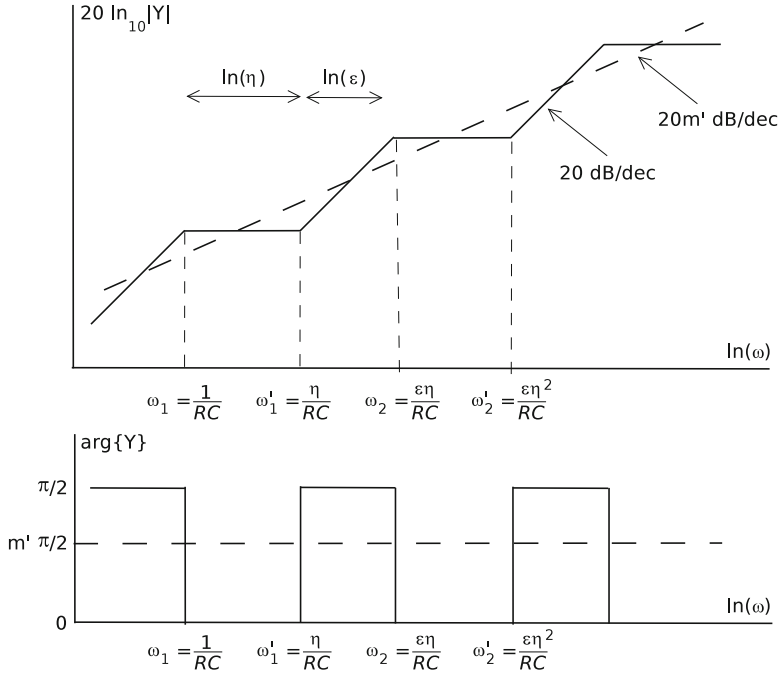

Fig. 20.4 Bode diagrams of amplitude and phase of $Y(\imath\omega)$

Consequently, the circuit of Fig. 20.3 represents an approach to D^{α}, $0 < \alpha < 1$, with $m' = \alpha$, based on a recursive pole/zero placement in the frequency domain. As mentioned in Sect. 20.2, the Laplace definition for a derivative of order $\alpha \in \mathbb{C}$ is a "direct" generalization of the classical integer-order scheme with the multiplication of the signal transform by the s operator. Therefore, in what concerns

automatic control theory this means that frequency-based analysis methods have a straightforward adaptation to their fractional-order counterparts. Nevertheless, the implementation based on the Laplace definition (adopting the frequency domain) requires an infinite number of poles and zeros obeying a recursive relationship. In a real approximation the finite number of poles and zeros yields a ripple in the frequency response and a limited bandwidth. Based on the Grünwald–Letnikov definition of a derivative of fractional order α of the signal $f(t)$, $D^{\alpha} f(t)$ leads to the expression: This formulation inspired a discrete-time calculation algorithm, based on the approximation of the time increment h by means of the sampling period T, yielding the equation in the z domain

$$\frac{\mathscr{Z}\{D^{\alpha} f(t)\}}{\mathscr{Z}\{f(t)\}} = \frac{1}{T^{\alpha}} \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(\alpha+1)}{k!\,\Gamma(\alpha-k+1)} z^{-k} = \left(\frac{1-z^{-1}}{T}\right)^{\alpha}, \tag{20.3}$$

where $\mathscr{Z}$ denotes the Z-transform operator.

An implementation of (20.3) corresponds to an r-term truncated series given by

$$\frac{\mathscr{Z}\{D^{\alpha} f(t)\}}{\mathscr{Z}\{f(t)\}} = \frac{1}{T^{\alpha}} \sum_{k=0}^{r} \frac{(-1)^k \Gamma(\alpha+1)}{k!\,\Gamma(\alpha-k+1)} z^{-k}.$$

Clearly, to have good approximations, we must have a large r and a small T.

Expression (20.3) represents the Euler, or first backward difference, approximation in the so-called $s \to z$ conversion scheme. Another possibility, often adopted in control system design, consists in the Tustin (or bilinear) rule. The Euler and Tustin rational expressions, $\psi_0\left(z^{-1}\right) = \frac{1-z^{-1}}{T}$ and $\psi_1\left(z^{-1}\right) = \frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}$, are often called generating approximants of zero and first order, respectively. Therefore, the generalization of these conversion methods leads to the non-integer order α results [Ma99]

$$s^{\alpha} \approx \left(\frac{1-z^{-1}}{T}\right)^{\alpha},$$

$$s^{\alpha} \approx \left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}\right)^{\alpha}.$$

We can obtain a family of fractional differentiators generated by $\psi_0^{\alpha}\left(z^{-1}\right) = \left[\psi_0\left(z^{-1}\right)\right]^{\alpha}$ and $\psi_1^{\alpha}\left(z^{-1}\right) = \left[\psi_1\left(z^{-1}\right)\right]^{\alpha}$ weighted by the factors p and $1-p$, yielding

$$\psi_{av}^{\alpha}\left(z^{-1}\right) = p\,\psi_0^{\alpha}\left(z^{-1}\right) + (1-p)\,\psi_1^{\alpha}\left(z^{-1}\right). \tag{20.4}$$

For example, the Al-Alaoui operator [Al93] corresponds to an interpolation of the Euler and Tustin rules with weighting factor $p = \frac{3}{4}$. In order to get a rational expression, the final approximation corresponds to a truncated Taylor series or

a rational fraction expansion. Due to its superior performance often it is used a fraction of order r; that is,

$$\psi\left(z^{-1}\right)=\frac{\sum_{k=0}^{r} a_k z^{-k}}{\sum_{k=0}^{r} b_k z^{-k}}.$$

Often it is adopted a Padé expansion in the neighborhood of $z=0$ and, since one parameter is linearly dependent, it is established $b_0=1$. The arithmetic mean (20.4)) motivates the study of an averaging method [MaGa09], [MaEtAl10] based on the generalized formula of averages (often called average of order $q\in\mathbb{R}$)

$$\psi_{av}^{\alpha}\left(z^{-1}\right)=\left\{p\left[\psi_0^{\alpha}\left(z^{-1}\right)\right]^q+(1-p)\left[\psi_1^{\alpha}\left(z^{-1}\right)\right]^q\right\}^{\frac{1}{q}}, \tag{20.5}$$

where (p,q) are two tuning degrees of freedom, corresponding q to the order of the averaging expression and p to the weighting factor. For example, when $q=\{-1,0,1\}$, in expression (20.5), we get the well-known expressions for the {harmonic, geometric, arithmetic} averages.

20.4 Fractional Modeling

In this section a classical fractional-order model is presented.

At high frequencies the electric current in a conductor distributes itself so that the current density near the surface is greater than that at its core. This phenomenon is called the skin effect (SE), or electromagnetic diffusion. As will be seen the SE shows characteristics that are well modeled by means of FC [MaGa12].

For a conductor of length l_0 and a sinusoidal field $E=\sqrt{2}\tilde{E}\sin(\omega t)$, with t and ω denoting time and frequency, the equivalent electrical complex impedance $\tilde{Z}$ is given by

$$\tilde{Z}=\frac{q l_0}{2\pi r_0\gamma}\frac{J_0(q r_0)}{J_1(q r_0)}, \tag{20.6}$$

where $q^2=-\imath\omega\gamma\mu$, ε, μ, and γ are the electrical permittivity, the magnetic permeability, and the conductivity, respectively, and J_0 and J_1 are complex-valued Bessel functions of the first kind of orders 0 and 1.

For low and high frequencies $\tilde{Z}$ yields

$$\omega\to 0 \quad\Rightarrow\quad \tilde{Z}\to\frac{l_0}{\pi r_0^2\gamma}, \tag{20.7}$$

$$\omega\to\infty \quad\Rightarrow\quad \tilde{Z}\to\frac{l_0}{2\pi r_0}\sqrt{\frac{\omega\mu}{2\gamma}}(1+\imath). \tag{20.8}$$

Expression (20.8) reveals a phenomenon of order $\alpha = \frac{1}{2}$ that is not captured by the standard integer models. Joining the two asymptotic expressions (20.7) and (20.8), we obtain the simple fractional approximation

$$\tilde{Z}_{app} = Z_0 \left(1 + \frac{\imath\omega}{a}\right)^{\alpha}, \tag{20.9}$$

where $Z_0 = \frac{l_0}{\pi r_0^2 \gamma}$, $a = \frac{4}{r_0^2 \gamma \mu}$ and $\alpha = \frac{1}{2}$. It must be noted that, while other approximations are possible, expression (20.9) has a simple analytical structure yielding

$$\tilde{Z}_{app} = \frac{l_0}{\pi r_0^2 \gamma}, \quad \tilde{Z}_{app} = \frac{l_0}{2\pi r_0} \sqrt{\frac{\omega\mu}{2\gamma}} (1 + \imath)$$

as $\omega \to 0$ and $\omega \to \infty$, respectively.

Figure 20.5 compares the Bode diagrams of amplitude and phase of $E(k_0)$ based on expressions (20.6) and (20.9) for a conductor with $\gamma = 10^7\, \Omega^{-1}$ m, $l_0 = 1$ m, $r_0 = 3.02\, 10^{-3}$ m, $\mu_0 = 1.257\, 10^{-6}\, \mathrm{Hm}^{-1}$ and $\mu_r = 10^3$. We verify this approximation leads to a very good curve fitting.

20.5 Fractional Control

In this section simple fractional-order control algorithms are presented.

Figure 20.6 illustrates an important aspect of fractional-order controllers, by using an elemental fractional system in the direct loop with transfer function $G(s) = \frac{K}{s^{\alpha}}$, $1 < \alpha < 2$. The open-loop Bode diagrams (Fig. 20.7) of amplitude and phase have a slope of -20 dB/dec and a constant phase of $-\alpha\frac{\pi}{2}$ rad, respectively. Therefore, the closed-loop system has a constant phase margin of $\pi\left(1 - \frac{\alpha}{2}\right)$ rad, that is independent of the system gain K. This important property is also revealed by means of the root-locus depicted in Fig. 20.8, illustrating the cases of $0 < \alpha < 1$ and $1 < \alpha < 2$.

Let us consider $K = 1$, so that $G(s) = \frac{1}{s^{\alpha}}$, and an unit step input $R(s) = \frac{1}{s}$ in the system represented in Fig. 20.6. The output response will be $C(s) = \frac{1}{s(s^{\alpha}+1)}$, or, in the time domain, $c(t) = 1 - E_{\alpha}(-t^{\alpha})$. Figure 20.9 depicts the responses for $\alpha = \{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$. We observe that the fractional values "interpolate" the well-known cases of $\alpha = \{1, 2\}$. Furthermore, we note the appearance of a fast initial transient, followed by an extremely slow convergence for the steady-state value, which is typical of many fractional order systems.

A popular application of FC is in the area of control [Ma97], [MoEtAl10], [Pe11], [Vada12] and corresponds to the generalization of the Proportional, Integral,

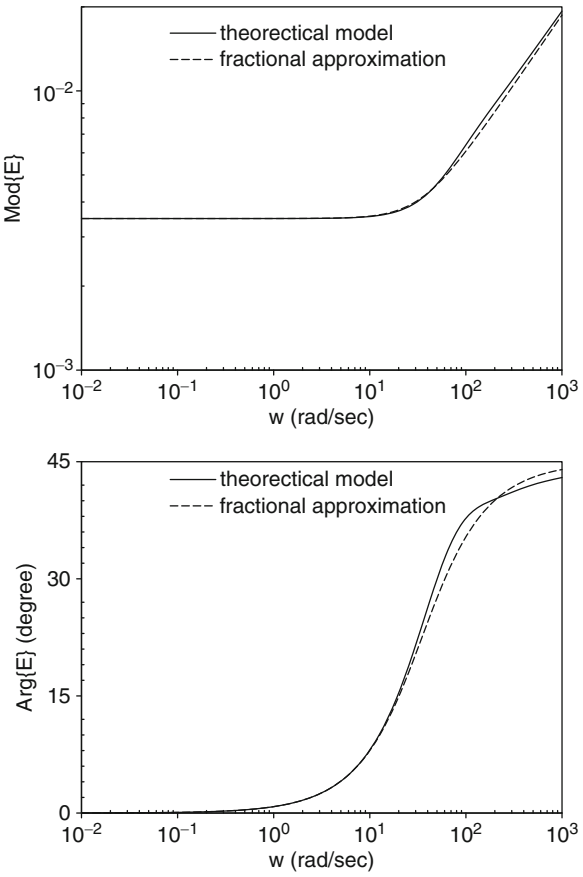

Fig. 20.5 Bode diagrams of amplitude and phase of $E(k_0)$ for the theoretical and the approximate expression with $\gamma = 10^7\,\Omega^{-1}$ m, $l_0 = 1$ m, $r_0 = 3.02\,10^{-3}$ m, $\mu_0 = 1.257\,10^{-6}\,\mathrm{Hm}^{-1}$ and $\mu_r = 10^3$

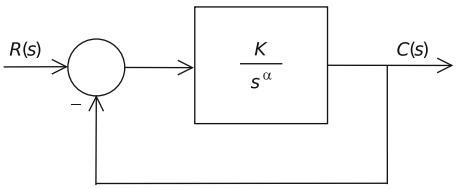

Fig. 20.6 Block diagram for an elemental feedback control system of fractional order α

and Derivative (*PID*) algorithm, namely to the fractional *PID*. The $PI^{\lambda}D^{\mu}$ by Podlubny [Po99b] control algorithm has a transfer function given by

$$G_c(s) = K_P + K_I s^{-\lambda} + K_D s^{\mu},$$

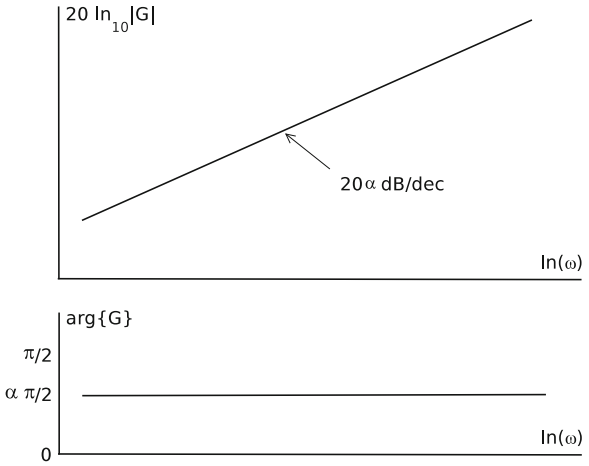

Fig. 20.7 Open-loop Bode diagrams of amplitude and phase for a system of fractional order $0 < \alpha < 1$

where K_P, K_I, and K_D are the proportional, integral, and differential gains, and λ and μ are the fractional orders of the integral and derivative actions, respectively.

The diagram of Fig. 20.10 shows that the cases $(\lambda, \mu) = \{(0,0), (1,0), \}$ $\{(0,1), (1,1)\}$, correspond to the *P*, *PI*, *PD*, and *PID*, respectively.

20.6 Conclusions

This paper presented the fundamental aspects of the FC calculus, the main approximation methods for the fractional-order derivatives calculation, and the implication of the FC concepts upon the extension of the classical automatic control theory. Bearing these ideas in mind, several approximate schemes for the calculation of fractional derivatives, and examples of FC-models and FC-controllers were described. It was shown that fractional-order models capture phenomena and properties that classical integer-order simply neglect.

References

[Al93] Al-Alaoui, M.A.: Novel digital integrator and differentiator. Electron. Lett. **29**, 376–378 (1993)

[An94] Anastasio, T.J.: The fractional-order dynamics of brainstem vestibulo-oculomotor neurons. Biol. Cybern. **72**(1), 69–79 (1994)

[BaMaLu11] Baleanu, D., Machado, J.T., Luo, A.: Fractional Dynamics and Control. Springer, New York (2011)

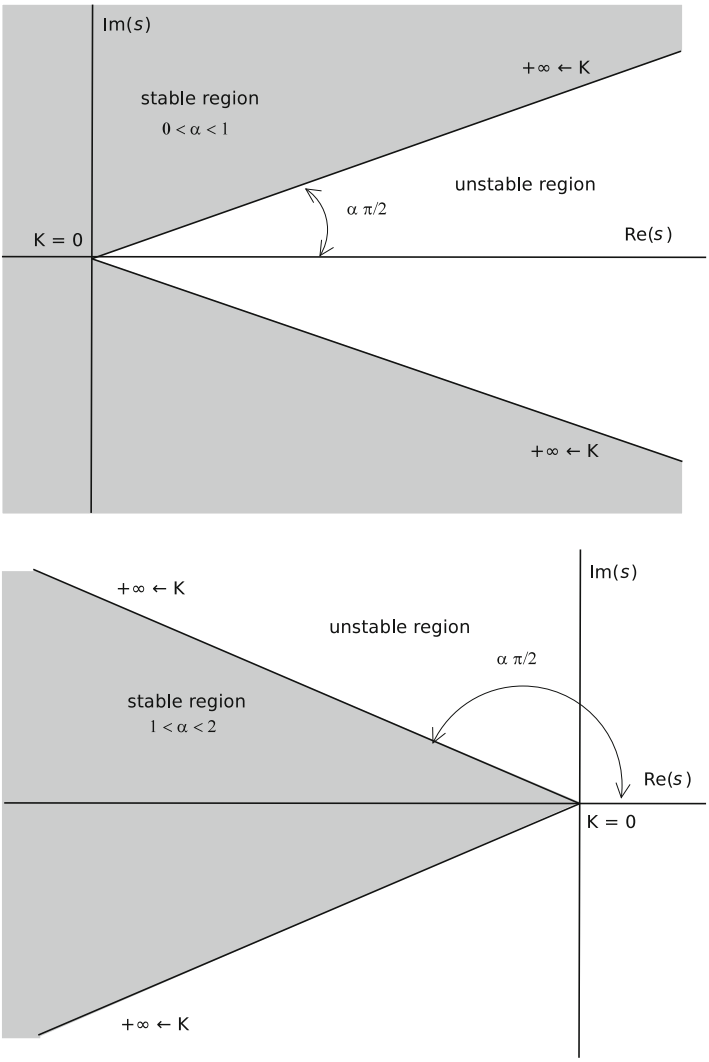

Fig. 20.8 Root locus for a feedback control system of fractional order for $0 < \alpha < 1$ and $1 < \alpha < 2$

[BaEtAl12] Baleanu, D., Diethelm, K., Scalas, E., Trujillo, J.J.: Fractional Calculus Models and Numerical Methods. World Scientific, Amsterdam (2012)

[CaEtAl10] Caponetto, R., Dongola, G., Fortuna, L., Petráš, I.: Fractional Order Systems: Modeling and Control Applications. World Scientific, Singapore (2010)

[CaHa64] Carlson, G.E., Halijak, C.A.: Approximation of fractional capacitors $(1/s)^{(1/n)}$ by a regular Newton process. IEEE Trans. Circ. Theor. **10**, 210–213 (1964)

[Di10] Diethelm, K.: The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type. Springer, Heidelberg (2010)

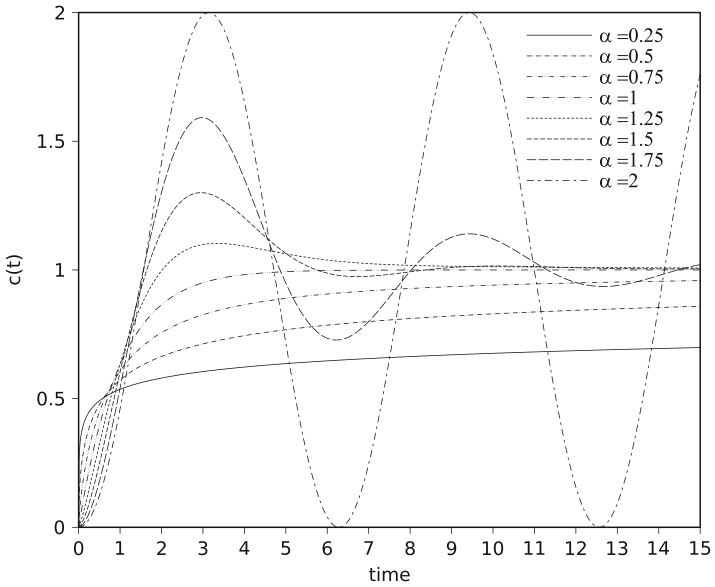

Fig. 20.9 Time response $c(t) = 1 - E_\alpha(-t^\alpha)$ of the closed-loop system represented in Fig. 20.5 for a unit step reference input and $\alpha = \{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$

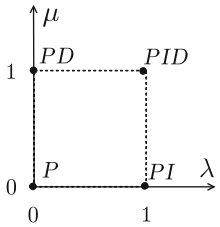

Fig. 20.10 The $PI^\lambda D^\mu$ and the four integer cases P, PI, PD, and PID

[Fe96] Fenander, Å.: Modal synthesis when modeling damping by use of fractional derivatives. AIAA J. **34**, 1051–1058 (1996)
[HaMaSa11] Haubold, H.J., Mathai, A.M., Saxena, R.K.: Mittag–Leffler functions and their applications. J. Appl. Math. **61**, 298628 (2011)
[Hi00] Hilfer, R.: Applications of Fractional Calculus in Physics. World Scientific, Singapore (2000)
[Ka11] Kaczorek, T.: Selected Problems of Fractional Systems Theory. Springer, Berlin (2011)
[KiSrTr06] Kilbas, A.A., Srivastava, H.M., Trujillo, J.J.: Theory and Applications of Fractional Differential Equations. North-Holland Mathematics Studies, vol. 204. Elsevier, Amsterdam (2006)
[Ki94] Kiryakova, V.: Generalized Fractional Calculus and Applications. Longman Scientific and Technical, Harlow (1994)
[Kl09] Klimek, M.: On Solutions of Linear Fractional Differential Equations of a Variational Type. Czestochowa University of Technology, Czestochowa (2009)

[Ko84] Koeller, R.C.: Applications of fractional calculus to the theory of viscoelasticity. ASME J. Appl. Mech. **51**(2), 299–307 (1984)
[Le11] Leszczyanski, J.S.: An Introduction to Fractional Mechanics. Czestochowa University of Technology, Czestochowa (2011)
[Ma97] Machado, J.T.: Analysis and design of fractional-order digital control systems. Syst. Anal. Model. Simulat. **27**, 107–122 (1997)
[Ma99] Machado, J.T.: Fractional-order derivative approximations in discrete-time control systems. Syst. Anal. Model. Simulat. **34**, 419–434 (1999)
[Ma01] Machado, J.T.: Discrete-time fractional-order controllers. Fractional Calculus Appl. Anal. **4**, 47–66 (2001)
[Ma03] Machado, J.T.: A probabilistic interpretation of the fractional-order differentiation. J. Fractional Calculus Appl. Anal. **6**, 73–80 (2003)
[Ma09] Machado, J.T.: Fractional derivatives: probability interpretation and frequency response of rational approximations. Comm. Nonlinear Sci. Numer. Simulat. **14**, 3492–3497 (2009)
[Ma11] Machado, J.T.: And I say to myself: "What a fractional world!". J. Fractional Calculus Appl. Anal. **14**, 635–654 (2011)
[MaGa09] Machado, J.T., Galhano, A.M.: Approximating fractional derivatives in the perspective of system control. Nonlinear Dynam. **56**, 401–407 (2009)
[MaGa12] Machado, J.T., Galhano, A.M.: Fractional order inductive phenomena based on the skin effect. Nonlinear Dynam. **68**, 107–115 (2012)
[MaKiMa10b] Machado, J.T., Kiryakova, V., Mainardi, F.: A poster about the old history of fractional calculus. J. Fractional Calculus Appl. Anal. **13**, 447–454 (2010)
[MaKiMa10a] Machado, J.T., Kiryakova, V., Mainardi, F.: A poster about the recent history of fractional calculus. J. Fractional Calculus Appl. Anal. **13**, 329–334 (2010)
[MaEtAl10] Machado, J.T., Galhano, A.M., Oliveira, A.M., Tar, J.K.: Optimal approximation of fractional derivatives through discrete-time fractions using genetic algorithms. Comm. Nonlinear Sci. Numer. Simulat. **15**, 482–490 (2010)
[MaKiMa11] Machado, J.T., Kiryakova, V., Mainardi, F.: Recent history of fractional calculus. Comm. Nonlinear Sci. Numer. Simulat. **16**, 1140–1153 (2011)
[Ma06] Magin, R.L.: Fractional Calculus in Bioengineering. Begell House, Redding (2006)
[Ma10] Mainardi, F.: Fractional Calculus and Waves in Linear Viscoelasticity: An Introduction to Mathematical Models. Imperial College Press, London (2010)
[LeNiNi98] Le Méhauté, A. Nigmatillin, R.R., Nivanen, L.: Flèches du Temps et Géométrie Fractale, 2nd edn. Hermes, Paris (1998)
[MiRo93] Miller, K.S., Ross, B.: An Introduction to the Fractional Calculus and Fractional Differential Equations. Wiley, New York (1993)
[MoEtAl10] Monje, C.A., Chen, Y., Vinagre, B.M., Xue, D., Feliu, V.: Fractional-Order Systems and Controls. Springer, London (2010)
[Ol74] Oldham, K.B., Spanier, J.: The Fractional Calculus. Academic, New York (1974)
[Or11] Ortigueira, M.D.: The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type. Springer, Berlin (2011)
[Ou91] Oustaloup, A.: La Commande CRONE: Commande Robuste d'Ordre Non Entier. Hermes, Paris (1991)
[Ou95] Oustaloup, A.: La Dérivation Non Entière: Théorie, Synthèse et Applications. Hermes, Paris (1995)
[Pe11] Petráš, I.: Fractional-Order Nonlinear Systems: Modeling, Analysis and Simulation. Springer, Berlin (2011)
[Po99a] Podlubny, I.: Fractional Differential Equations. Academic Press, San Diego (1999)
[Po99b] Podlubny, I.: Fractional-order systems and $PI^{\lambda}D^{\mu}$-controllers. IEEE Trans. Automat. Contr. **44**, 208–213 (1999)
[Po02] Podlubny, I.: Geometric and physical interpretation of fractional integration and fractional differentiation. J. Fractional Calculus Appl. Anal. **5**, 367–386 (2002)

[Ro67] Roy, S.C.: On the realization of a constant-argument immitance of fractional operator. IEEE Trans. Circ. Theor. **14**, 264–374 (1967)
[SaAgMa07] Sabatier, J., Agrawal, O.P., Machado, J.T. (eds.): Advances in Fractional Calculus: Theoretical Developments and Applications in Physics and Engineering. Springer, Dordrecht (2007)
[SaKiMa93] Samko, S.G., Kilbas, A.A., Marichev, O.I.: Fractional Integrals and Derivatives. Gordon and Breach, Yverdon (1993)
[Ta10] Tarasov, V.E.: Fractional Dynamics: Applications of Fractional Calculus to Dynamics of Particles, Fields and Media. Springer, Berlin (2010)
[Ta95] Tatom, F.B.: The relationship between fractional calculus and fractals. Fractals **3**, 217–229 (1995)
[ToBa84] Torvik, P.J., Bagley, R.L.: On the appearance of the fractional derivative in the behaviour of real materials. ASME J. Appl. Mech. **51**, 294–298 (1984)
[Vada12] Valério, D., da Costa, J.S.: An Introduction to Fractional Control. IET, Stevenage (2012)
[We03] West, B., Bologna, M., Grigolini, P.: Physics of Fractal Operators. Springer, New York (2003)
[We02] Westerlund, S.: Dead Matter Has Memory. Causal Consulting, Kalmar (2002)
[WeEk94] Westerlund, S., Ekstam, L.: Capacitor Theory. IEEE Trans. Dielectrics Electr. Insul. **1**, 826–839 (1994)
[Za05] Zaslavsky, G.M.: Hamiltonian Chaos and Fractional Dynamics. Oxford University Press, Oxford (2005)

Chapter 21
Modified Integral Equation Method for Stationary Plate Oscillations

G.R. Thomson and C. Constanda

21.1 Introduction

Each of the exterior Dirichlet, Neumann, and Robin boundary value problems associated with the high-frequency stationary oscillations of Mindlin-type elastic plates is known to have at most one solution in a certain class of functions (see [Co98] and [ThCo11]). However, in [ThCo97], [ThCo99], [ThCo09a], and [ThCo10] it was shown that, using classical integral equation techniques, it is not possible to derive single, uniquely solvable integral equations from which to construct the solution of the mathematical model. To overcome this drawback, in [ThCo12b] the solutions were sought in the form of functions satisfying a dissipative condition on some suitable curve.

Below, we propose an alternative modified method that also yields well-posed integral equations.

Techniques that resolve the uniqueness difficulties arising in exterior problems in acoustics and elastodynamics can be found in [Jo74], [Ur78], [Jo84], and [Be90].

Throughout what follows, a superscript T denotes matrix transposition and $x = (x_1, x_2)^{\mathrm{T}}$ and $y = (y_1, y_2)^{\mathrm{T}}$ are generic points in the Cartesian plane $\mathbb{R}^2$, with corresponding polar coordinates (R_x, θ_x) and (R_y, θ_y).

We denote by S^+ a domain in $\mathbb{R}^2$ bounded by a simple, closed C^2-curve ∂S and write $S^- = \mathbb{R}^2 \setminus \bar{S}^+$. The origin of coordinates is assumed to lie in S^+.

Let h_0 be the (constant) thickness of the plate, ρ the density, and λ and μ the Lamé constants of the homogeneous and isotropic plate material, which occupies the

G.R. Thomson
A.C.C.A., Glasgow, UK
e-mail: thomsongavin@gmail.com

C. Constanda (✉)
The University of Tulsa, Tulsa, OK, USA
e-mail: christian-constanda@utulsa.edu

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_21,

infinite region $\bar{S}^- \times [-h_0/2, h_0/2]$ in $\mathbb{R}^3$. The stationary oscillations of frequency ω of the plate, when transverse shear deformation is taken into account, are governed by the system [ScCo93]

$$A^{\omega}(\partial_x)u(x) = H(x), \tag{21.1}$$

where $u = (u_1, u_2, u_3)^{\mathrm{T}}$ is a vector characterizing the displacements, H is related to the averaged (across thickness) body forces and moments, and the matrix operator $A^{\omega}(\partial_x) = A^{\omega}(\partial/\partial x_1, \partial/\partial x_2)$ is defined by

$$A^{\omega}(\xi_1, \xi_2) = \begin{pmatrix} h^2\mu(\Delta + k_3^2) + h^2(\lambda+\mu)\xi_1^2 & h^2(\lambda+\mu)\xi_1\xi_2 & -\mu\xi_1 \\ h^2(\lambda+\mu)\xi_1\xi_2 & h^2\mu(\Delta + k_3^2) + h^2(\lambda+\mu)\xi_2^2 & -\mu\xi_2 \\ \mu\xi_1 & \mu\xi_2 & \mu(\Delta + k^2) \end{pmatrix};$$

here $h^2 = h_0^2/12$, $\Delta = \xi_1^2 + \xi_2^2$, and

$$k^2 = \frac{\rho\omega^2}{\mu}, \quad k_3^2 = k^2 - \frac{1}{h^2}. \tag{21.2}$$

We define constants k_1^2 and k_2^2 by

$$k_1^2 + k_2^2 = \frac{\lambda + 3\mu}{\lambda + 2\mu}k^2, \quad k_1^2 k_2^2 = \frac{\mu}{\lambda + 2\mu}k^2 k_3^2. \tag{21.3}$$

In what follows it is assumed that

$$\lambda + \mu > 0, \quad \mu > 0, \quad \rho\omega^2 h^2 > \mu. \tag{21.4}$$

These inequalities imply that k_1^2, k_2^2, and k_3^2 are real, positive, and distinct.

Without loss of generality, we restrict our attention to the homogeneous system

$$A^{\omega}(\partial_x)u(x) = 0, \tag{21.5}$$

since a particular solution of (21.1) can be constructed in terms of a Newtonian potential [ThCo98].

The boundary moment–stress operator $T(\partial_x) = T(\partial/\partial x_1, \partial/\partial x_2)$ is defined by [ScCo93]

$$T(\xi_1, \xi_2) = \begin{pmatrix} h^2((\lambda+2\mu)\nu_1\xi_1 + \mu\nu_2\xi_2) & h^2(\mu\nu_2\xi_1 + \lambda\nu_1\xi_2) & 0 \\ h^2(\lambda\nu_2\xi_1 + \mu\nu_1\xi_2) & h^2(\mu\nu_1\xi_1 + (\lambda+2\mu)\nu_2\xi_2) & 0 \\ \mu\nu_1 & \mu\nu_2 & \mu(\nu_1\xi_1 + \nu_2\xi_2) \end{pmatrix},$$

where $\nu = (\nu_1, \nu_2)^{\mathrm{T}}$ is the unit outward normal to the boundary of the middle plane of the plate.

The operators $A^{\omega}(\partial_x)$ and $T(\partial_x)$ are connected by the reciprocity relation, which states that if $u, v \in C^2(S^+) \cap C^1(\bar{S}^+)$, then [ThCo11]

$$\int\limits_{S^+} \left(u^{\mathrm{T}} A^{\omega} v - v^{\mathrm{T}} A^{\omega} u\right) da = \int\limits_{\partial S} \left(u^{\mathrm{T}} T v - v^{\mathrm{T}} T u\right) ds. \tag{21.6}$$

We denote by $\mathscr{B}^{\omega}$ the class of functions defined in S^- which satisfy the radiation conditions formulated in [Co98] (see also [ThCo11]) as $R_x \to \infty$.

Let $\mathscr{R}$, $\mathscr{S}$, and $\mathscr{G}$ be 3×1 vector functions prescribed on ∂S, and let $\sigma \in C^{1,\alpha}(\partial S)$, $\alpha \in (0,1)$, be a symmetric 3×3 matrix function. We now formulate the exterior boundary value problems ($\mathrm{D}^{\omega-}$), ($\mathrm{N}^{\omega-}$), and ($\mathrm{R}^{\omega-}$) with Dirichlet, Neumann, and Robin boundary conditions, respectively, as follows:

($\mathrm{D}^{\omega-}$) Find $u \in C^2(S^-) \cap C^1(\bar{S}^-) \cap \mathscr{B}^{\omega}$ satisfying (21.5) in S^- and

$$u|_{\partial S} = \mathscr{R}. \tag{21.7}$$

($\mathrm{N}^{\omega-}$) Find $u \in C^2(S^-) \cap C^1(\bar{S}^-) \cap \mathscr{B}^{\omega}$ satisfying (21.5) in S^- and

$$Tu|_{\partial S} = \mathscr{S}. \tag{21.8}$$

($\mathrm{R}^{\omega-}$) Find $u \in C^2(S^-) \cap C^1(\bar{S}^-) \cap \mathscr{B}^{\omega}$ satisfying (21.5) in S^- and

$$(Tu + \sigma u)|_{\partial S} = \mathscr{G}. \tag{21.9}$$

A function u is said to be regular in S^- if $u \in C^2(S^-) \cap C^1(\bar{S}^-)$.

The next assertion was proved in [Co98] and [ThCo11].

Theorem 1. *(i) Each of ($D^{\omega-}$) and ($N^{\omega-}$) has at most one regular solution.*
(ii) If $\mathrm{Im}(\sigma)$ *is positive semidefinite, then ($R^{\omega-}$) has at most one regular solution.*

21.2 A Modified Matrix of Fundamental Solutions

Let

$$\varepsilon_m = \begin{cases} 1, & m = 0, \\ 2, & m \geq 1, \end{cases} \qquad E_m^{(\sigma)}(\theta) = \begin{cases} \cos m\theta, & \sigma = 1, \\ \sin m\theta, & \sigma = 2, \end{cases}$$

where m is a nonnegative integer, and let

$$\phi_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} H_m(k_1 R_x) E_m^{(\sigma)}(\theta_x), \quad \hat{\phi}_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} J_m(k_1 R_x) E_m^{(\sigma)}(\theta_x),$$

$$\upsilon_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} H_m(k_2 R_x) E_m^{(\sigma)}(\theta_x), \quad \hat{\upsilon}_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} J_m(k_2 R_x) E_m^{(\sigma)}(\theta_x),$$
$$\psi_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} H_m(k_3 R_x) E_m^{(\sigma)}(\theta_x), \quad \hat{\psi}_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} J_m(k_3 R_x) E_m^{(\sigma)}(\theta_x),$$

where k_1, k_2, and k_3 are defined by (21.2) and (21.3), H_m is the Hankel function of the first kind and order m, and J_m is the Bessel function of the first kind and order m.

We introduce the constants

$$\alpha_1^2 = \frac{k_2^2 - \mu' k_3^2}{k_2^2 - k_1^2}, \quad \alpha_2^2 = \frac{k_1^2 - \mu' k_3^2}{k_1^2 - k_2^2},$$

where $\mu' = \mu/(\lambda + 2\mu)$. It is easily verified that these constants are strictly positive if inequalities (21.4) hold and k_1^2 is the larger of the roots defined by (21.3).

The radiating wavefunctions are [ThCo09b]

$$\boldsymbol{\Phi}_m^{(\sigma)}(x) = \left(\alpha_1 \frac{\partial}{\partial x_1} \phi_m^{(\sigma)}(x), \alpha_1 \frac{\partial}{\partial x_2} \phi_m^{(\sigma)}(x), -hk_3 \alpha_2 \phi_m^{(\sigma)}(x) \right)^{\mathrm{T}},$$
$$\boldsymbol{\Upsilon}_m^{(\sigma)}(x) = \left(\alpha_2 \frac{\partial}{\partial x_1} \upsilon_m^{(\sigma)}(x), \alpha_2 \frac{\partial}{\partial x_2} \upsilon_m^{(\sigma)}(x), hk_3 \alpha_1 \upsilon_m^{(\sigma)}(x) \right)^{\mathrm{T}},$$
$$\boldsymbol{\Psi}_m^{(\sigma)}(x) = \left(\frac{\partial}{\partial x_2} \psi_m^{(\sigma)}(x), -\frac{\partial}{\partial x_1} \psi_m^{(\sigma)}(x), 0 \right)^{\mathrm{T}}.$$

Each of these functions satisfies (21.5) in $\mathbb{R}^2 \setminus \{0\}$ and belongs to $\mathscr{B}^\omega$. The corresponding regular wavefunctions $\hat{\boldsymbol{\Phi}}_m^{(\sigma)}$, $\hat{\boldsymbol{\Upsilon}}_m^{(\sigma)}$, and $\hat{\boldsymbol{\Psi}}_m^{(\sigma)}$ are obtained by replacing $\phi_m^{(\sigma)}$, $\upsilon_m^{(\sigma)}$, and $\psi_m^{(\sigma)}$ in the above formulas by $\hat{\phi}_m^{(\sigma)}$, $\hat{\upsilon}_m^{(\sigma)}$, and $\hat{\psi}_m^{(\sigma)}$, respectively.

If $D^\omega(x,y)$ is the matrix of fundamental solutions for $A^\omega(\partial_x)$ constructed in [ThCo09b], it can be shown that for $R_x < R_y$,

$$D^\omega(x,y) = \frac{i}{4h^2 \mu k_3^2} \sum_{m=0}^{\infty} \sum_{\sigma=1}^{2} \left\{ \hat{\boldsymbol{\Phi}}_m^{(\sigma)}(x) \left[\boldsymbol{\Phi}_m^{(\sigma)}(y) \right]^{\mathrm{T}} + \hat{\boldsymbol{\Upsilon}}_m^{(\sigma)}(x) \left[\boldsymbol{\Upsilon}_m^{(\sigma)}(y) \right]^{\mathrm{T}} + \hat{\boldsymbol{\Psi}}_m^{(\sigma)}(x) \left[\boldsymbol{\Psi}_m^{(\sigma)}(y) \right]^{\mathrm{T}} \right\}. \quad (21.10)$$

We now consider a modified matrix of fundamental solutions $D_L^\omega(x,y)$ of the form

$$D_L^\omega(x,y) = D^\omega(x,y) + L^\omega(x,y), \quad (21.11)$$

where

$$L^\omega(x,y) = \frac{i}{4h^2 \mu k_3^2} \sum_{m=0}^{\infty} \sum_{\sigma=1}^{2} \left\{ a_m^{(\sigma)} \boldsymbol{\Phi}_m^{(\sigma)}(x) \left[\boldsymbol{\Phi}_m^{(\sigma)}(y) \right]^{\mathrm{T}} + b_m^{(\sigma)} \boldsymbol{\Upsilon}_m^{(\sigma)}(x) \left[\boldsymbol{\Upsilon}_m^{(\sigma)}(y) \right]^{\mathrm{T}} + c_m^{(\sigma)} \boldsymbol{\Psi}_m^{(\sigma)}(x) \left[\boldsymbol{\Psi}_m^{(\sigma)}(y) \right]^{\mathrm{T}} \right\}, \quad (21.12)$$

with arbitrary constants $a_m^{(\sigma)}$, $b_m^{(\sigma)}$, and $c_m^{(\sigma)}$. From (21.12) it is easy to see that $L^{\omega}(x,y) = [L^{\omega}(y,x)]^{\mathrm{T}}$, which, in view of (21.11), means that

$$D_L^{\omega}(x,y) = [D_L^{\omega}(y,x)]^{\mathrm{T}},$$

since the "unmodified" matrix of fundamental solutions is symmetric [ThCo09b].

Let $\varpi(b)$ be a disk centered at the origin whose radius b is sufficiently small so that $\varpi(b) \subset S^+$. We assume that, as $m \to \infty$ through real, positive values,

$$a_m^{(\sigma)} \sim \left(\frac{e^2 k_1^2 b^2}{4m^2}\right)^m, \quad b_m^{(\sigma)} \sim \left(\frac{e^2 k_2^2 b^2}{4m^2}\right)^m, \quad c_m^{(\sigma)} \sim \left(\frac{e^2 k_3^2 b^2}{4m^2}\right)^m. \tag{21.13}$$

It can be shown that if (21.13) holds, then the infinite series (21.12) is absolutely convergent in the region $R_x R_y > b^2$.

The wavefunctions satisfy certain orthogonality-type properties, established in [ThCo09b]. Let

$$\chi_m^{(\sigma 1)}(x) = \Phi_m^{(\sigma)}(x), \quad \chi_m^{(\sigma 2)}(x) = \Upsilon_m^{(\sigma)}(x), \quad \chi_m^{(\sigma 3)}(x) = \Psi_m^{(\sigma)}(x),$$

where $\sigma = 1,2$ and $m = 0,1,2,\ldots$. The functions $\hat{\chi}_m^{(\sigma j)}$, $j = 1,2,3$, are defined in the obvious way. If ∂C is any closed curve that contains the origin in its interior, then for $m \geq 1$,

$$\int_{\partial C} \left\{ \left[\hat{\chi}_m^{(\sigma j)}\right]^{\mathrm{T}} T \hat{\chi}_n^{(\nu k)} - \left[\hat{\chi}_n^{(\nu k)}\right]^{\mathrm{T}} T \hat{\chi}_m^{(\sigma j)} \right\} ds = 0, \tag{21.14}$$

$$\int_{\partial C} \left\{ \left[\hat{\chi}_m^{(\sigma j)}\right]^{\mathrm{T}} T \chi_n^{(\nu k)} - \left[\chi_n^{(\nu k)}\right]^{\mathrm{T}} T \hat{\chi}_m^{(\sigma j)} \right\} ds = 4ih^2 \mu k_3^2 \delta_{mn} \delta_{\sigma\nu} \delta_{jk}, \tag{21.15}$$

$$\int_{\partial C} \left\{ \left[\bar{\chi}_m^{(\sigma j)}\right]^{\mathrm{T}} T \chi_n^{(\nu k)} - \left[\chi_n^{(\nu k)}\right]^{\mathrm{T}} T \bar{\chi}_m^{(\sigma j)} \right\} ds = 8ih^2 \mu k_3^2 \delta_{mn} \delta_{\sigma\nu} \delta_{jk}. \tag{21.16}$$

Equalities (21.15) and (21.16) also hold for $m = 0$, $\sigma = 1$.

We introduce the modified single-layer and double-layer potentials

$$V_L^{\omega}(\varphi) = \int_{\partial S} D_L^{\omega}(x,y) \varphi(y)\, ds(y),$$

$$W_L^{\omega}(\varphi) = \int_{\partial S} \left[T(\partial_y) D_L^{\omega}(y,x)\right]^{\mathrm{T}} \varphi(y)\, ds(y),$$

where $D_L^{\omega}(x,y)$ is defined by (21.11) and (21.12). These potentials behave in the same way as the modified potentials discussed in [ThCo12b]. Their properties are gathered in the next assertion.

Theorem 2. *(i)* $V_L^{\omega}\varphi,\ W_L^{\omega}\varphi \in \mathscr{B}^{\omega}$.

(ii) If $\varphi \in C(\partial S)$, *then* $V_L^{\omega}\varphi$ *and* $W_L^{\omega}\varphi$ *are analytic and satisfy system* (21.5) *in* $\mathbb{R}^2 \setminus (\partial S \cup \{0\})$.

(iii) If $\varphi \in C^{0,\alpha}(\partial S)$, $\alpha \in (0,1)$, *then the direct values* $V_{L0}^{\omega}\varphi$ *and* $W_{L0}^{\omega}\varphi$ *of* $V_L^{\omega}\varphi$ *and* $W_L^{\omega}\varphi$ *on* ∂S *exist (the latter as principal value), the functions*

$$\mathscr{V}_L^{\omega+}(\varphi) = (V_L^{\omega}\varphi)|_{\bar{S}^+}, \quad \mathscr{V}_L^{\omega-}(\varphi) = (V_L^{\omega}\varphi)|_{\bar{S}^-}$$

are of class $C^{\infty}(S^+) \cap C^{1,\alpha}(\bar{S}^+)$ *and* $C^{\infty}(S^-) \cap C^{1,\alpha}(\bar{S}^-)$, *respectively, and*

$$T\mathscr{V}_L^{\omega+}(\varphi) = (W_{L0}^{\omega*} + \tfrac{1}{2}I)\varphi, \quad T\mathscr{V}_L^{\omega-}(\varphi) = (W_{L0}^{\omega*} - \tfrac{1}{2}I)\varphi$$

on ∂S, *where* $W_{L0}^{\omega*}$ *is the adjoint of* W_{L0}^{ω} *and* I *is the identity operator.*

(iv) If $\varphi \in C^{1,\alpha}(\partial S)$, $\alpha \in (0,1)$, *then the functions*

$$\mathscr{W}_L^{\omega+}(\varphi) = \begin{cases} (W_L^{\omega}\varphi)|_{S^+} & \text{in } S^+, \\ (W_{L0}^{\omega} - \frac{1}{2}I)\varphi & \text{on } \partial S, \end{cases} \quad \mathscr{W}_L^{\omega-}(\varphi) = \begin{cases} (W_L^{\omega}\varphi)|_{S^-} & \text{in } S^-, \\ (W_{L0}^{\omega} + \frac{1}{2}I)\varphi & \text{on } \partial S \end{cases}$$

are of class $C^{\infty}(S^+) \cap C^{1,\alpha}(\bar{S}^+)$ *and* $C^{\infty}(S^-) \cap C^{1,\alpha}(\bar{S}^-)$, *respectively, and we have* $T\mathscr{W}_L^{\omega+}(\varphi) = T\mathscr{W}_L^{\omega-}(\varphi)$ *on* ∂S.

These properties are used in the next section to formulate quasi-Fredholm integral equations for each boundary value problem.

21.3 Uniquely Solvable Integral Equations

We start with the exterior Neumann problem. Seeking the solution of $(\mathrm{N}^{\omega-})$ in the form $v = \mathscr{V}_L^{\omega-}(\varphi)$ and taking the boundary condition (21.8) into account, we arrive at the integral equation

$$\left(W_{L0}^{\omega*} - \tfrac{1}{2}I\right)\varphi = \mathscr{S} \tag{21.17}$$

for the unknown density φ. In [Co90] it is shown that the Fredholm alternative is applicable to equations of this kind. Therefore, to prove that (21.17) has a unique solution, we need to show that the corresponding homogeneous equation

$$\left(W_{L0}^{\omega*} - \tfrac{1}{2}I\right)\varphi = 0 \tag{21.18}$$

has only the zero solution.

Let $u = \mathscr{V}_L^{\omega+}(\varphi)$ and $v = \mathscr{V}_L^{\omega-}(\varphi)$, where φ satisfies (21.18). Then $Tv|_{\partial S} = 0$, which means that v is a solution of the homogeneous exterior Neumann problem. By Theorem 1, $v = 0$ in $\bar{S}^-$; in particular,

$$v|_{\partial S} = V_{L0}^{\omega}\varphi = u|_{\partial S} = 0.$$

Let

$$A_m^{(\sigma)} = \frac{i}{4h^2\mu k_3^2}\int_{\partial S}\left[\Phi_m^{(\sigma)}(y)\right]^{\mathrm{T}}\varphi(y)\,ds(y),$$

$$B_m^{(\sigma)} = \frac{i}{4h^2\mu k_3^2}\int_{\partial S}\left[\Upsilon_m^{(\sigma)}(y)\right]^{\mathrm{T}}\varphi(y)\,ds(y),$$

$$C_m^{(\sigma)} = \frac{i}{4h^2\mu k_3^2}\int_{\partial S}\left[\Psi_m^{(\sigma)}(y)\right]^{\mathrm{T}}\varphi(y)\,ds(y),$$

and let $R_{\min}$ be the minimum distance from the origin to ∂S. Then, by (21.10) – (21.12), we see that for $b \le R_x < R_{\min}$,

$$u(x) = \sum_{m=0}^{\infty}\sum_{\sigma=1}^{2}\left\{A_m^{(\sigma)}\hat{\Phi}_m^{(\sigma)}(x) + B_m^{(\sigma)}\hat{\Upsilon}_m^{(\sigma)}(x) + C_m^{(\sigma)}\hat{\Psi}_m^{(\sigma)}(x)\right.$$
$$\left. + a_m^{(\sigma)}A_m^{(\sigma)}\Phi_m^{(\sigma)}(x) + b_m^{(\sigma)}B_m^{(\sigma)}\Upsilon_m^{(\sigma)}(x) + c_m^{(\sigma)}C_m^{(\sigma)}\Psi_m^{(\sigma)}(x)\right\}. \quad (21.19)$$

We apply the reciprocity relation (21.6) to u and $\bar{u}$ in $S^+ \setminus \overline{\omega}(b)$. Since $u|_{\partial S} = \bar{u}|_{\partial S} = 0$, it follows that

$$\int_{\partial\overline{\omega}(b)}(u^{\mathrm{T}}T\bar{u} - \bar{u}^{\mathrm{T}}Tu)\,ds = 0. \quad (21.20)$$

Substituting (21.19) into (21.20), taking into account that the $\hat{\chi}_m^{(\sigma j)}$, $j = 1,2,3$, are real when their arguments are real, and using (21.14)–(21.16), we find that

$$\sum_{m=0}^{\infty}\sum_{\sigma=1}^{2}\left\{\tfrac{1}{2}A_m^{(\sigma)}\bar{a}_m^{(\sigma)}\bar{A}_m^{(\sigma)} + \tfrac{1}{2}B_m^{(\sigma)}\bar{b}_m^{(\sigma)}\bar{B}_m^{(\sigma)} + \tfrac{1}{2}C_m^{(\sigma)}\bar{c}_m^{(\sigma)}\bar{C}_m^{(\sigma)}\right.$$
$$+ \tfrac{1}{2}\bar{A}_m^{(\sigma)}a_m^{(\sigma)}A_m^{(\sigma)} + \tfrac{1}{2}\bar{B}_m^{(\sigma)}b_m^{(\sigma)}B_m^{(\sigma)} + \tfrac{1}{2}\bar{C}_m^{(\sigma)}c_m^{(\sigma)}C_m^{(\sigma)} + a_m^{(\sigma)}A_m^{(\sigma)}\bar{a}_m^{(\sigma)}\bar{A}_m^{(\sigma)}$$
$$\left. + b_m^{(\sigma)}B_m^{(\sigma)}\bar{b}_m^{(\sigma)}\bar{B}_m^{(\sigma)} + c_m^{(\sigma)}C_m^{(\sigma)}\bar{c}_m^{(\sigma)}\bar{C}_m^{(\sigma)}\right\} = 0. \quad (21.21)$$

We rearrange (21.21) in two different ways. First, we write

$$\sum_{m=0}^{\infty}\sum_{\sigma=1}^{2}\left\{\left|A_m^{(\sigma)}\right|^2\left(\left|a_m^{(\sigma)} + \tfrac{1}{2}\right|^2 - \tfrac{1}{4}\right) + \left|B_m^{(\sigma)}\right|^2\left(\left|b_m^{(\sigma)} + \tfrac{1}{2}\right|^2 - \tfrac{1}{4}\right)\right.$$
$$\left. + \left|C_m^{(\sigma)}\right|^2\left(\left|c_m^{(\sigma)} + \tfrac{1}{2}\right|^2 - \tfrac{1}{4}\right)\right\} = 0.$$

If we choose the constants so that, for $\sigma = 1,2$ and $m = 0,1,2,\ldots$, either

$$\left|a_m^{(\sigma)} + \tfrac{1}{2}\right| > \tfrac{1}{2}, \quad \left|b_m^{(\sigma)} + \tfrac{1}{2}\right| > \tfrac{1}{2}, \quad \left|c_m^{(\sigma)} + \tfrac{1}{2}\right| > \tfrac{1}{2} \tag{21.22}$$

or

$$\left|a_m^{(\sigma)} + \tfrac{1}{2}\right| < \tfrac{1}{2}, \quad \left|b_m^{(\sigma)} + \tfrac{1}{2}\right| < \tfrac{1}{2}, \quad \left|c_m^{(\sigma)} + \tfrac{1}{2}\right| < \tfrac{1}{2}, \tag{21.23}$$

then

$$A_m^{(\sigma)} = B_m^{(\sigma)} = C_m^{(\sigma)} = 0, \quad \sigma = 1,2, \ m = 0,1,2,\ldots. \tag{21.24}$$

There is another set of restrictions on the constants which ensures that (21.24) hold. From (21.21) we see that

$$\sum_{m=0}^{\infty} \sum_{\sigma=1}^{2} \left\{ \left(\operatorname{Re}\left(a_m^{(\sigma)}\right) + \left|a_m^{(\sigma)}\right|^2 \right) \left|A_m^{(\sigma)}\right|^2 + \left(\operatorname{Re}\left(b_m^{(\sigma)}\right) + \left|b_m^{(\sigma)}\right|^2 \right) \left|B_m^{(\sigma)}\right|^2 \right.$$
$$\left. + \left(\operatorname{Re}\left(c_m^{(\sigma)}\right) + \left|c_m^{(\sigma)}\right|^2 \right) \left|C_m^{(\sigma)}\right|^2 \right\} = 0.$$

Consequently, if

$$\operatorname{Re}\left(a_m^{(\sigma)}\right), \ \operatorname{Re}\left(b_m^{(\sigma)}\right), \ \operatorname{Re}\left(c_m^{(\sigma)}\right) > 0, \quad \sigma = 1,2, \ m = 0,1,2,\ldots, \tag{21.25}$$

then (21.24) hold.

Now suppose that (21.24) hold. Then, by (21.19), $u = 0$ in $b \le R_x < R_{\min}$; so, from the analyticity of the potentials we deduce that $u = 0$ in $S^+ \setminus \varpi(b)$. Hence,

$$0 = Tu|_{\partial S} = \left(W_{L0}^{\omega *} + \tfrac{1}{2} I\right) \varphi,$$

which, in view of (21.18), implies that $\varphi = 0$. Thus, we have proved the following assertion.

Theorem 3. *If $\mathscr{S} \in C^{0,\alpha}(\partial S)$, $\alpha \in (0,1)$, and the constants in $L^{\omega}(x,y)$ are chosen so that equalities (21.24) hold, then the integral equation (21.17) has a unique solution $\varphi \in C^{0,\alpha}(\partial S)$.*

We now go over to the exterior Dirichlet problem ($\mathrm{D}^{\omega -}$). Seeking the solution as a modified double-layer potential $v = \mathscr{W}_L^{\omega -}(\varphi)$ and using (21.7), we arrive at the equation

$$\left(W_{L0}^{\omega} + \tfrac{1}{2} I\right) \varphi = \mathscr{R} \tag{21.26}$$

for the unknown density φ. We claim that the corresponding homogeneous equation

$$\left(W_{L0}^{\omega} + \tfrac{1}{2} I\right) \varphi = 0 \tag{21.27}$$

has only the zero solution.

Suppose that φ is a solution of (21.27), and let

$$u = \mathscr{W}_L^{\omega+}(\varphi), \quad v = \mathscr{W}_L^{\omega-}(\varphi).$$

Then $v|_{\partial S} = 0$, and, since $(\mathrm{D}^{\omega-})$ has at most one solution, we see that $v = 0$ in $\bar{S}^-$; therefore,

$$0 = Tv|_{\partial S} = T\mathscr{W}_L^{\omega-}(\varphi)|_{\partial S} = T\mathscr{W}_L^{\omega+}(\varphi)|_{\partial S} = Tu|_{\partial S}. \tag{21.28}$$

Let

$$\tilde{A}_m^{(\sigma)} = \frac{i}{4h^2\mu k_3^2} \int_{\partial S} \left[T(\partial_y)\Phi_m^{(\sigma)}(y)\right]^{\mathrm{T}} \varphi(y)\, ds(y),$$

$$\tilde{B}_m^{(\sigma)} = \frac{i}{4h^2\mu k_3^2} \int_{\partial S} \left[T(\partial_y)\Upsilon_m^{(\sigma)}(y)\right]^{\mathrm{T}} \varphi(y)\, ds(y),$$

$$\tilde{C}_m^{(\sigma)} = \frac{i}{4h^2\mu k_3^2} \int_{\partial S} \left[T(\partial_y)\Psi_m^{(\sigma)}(y)\right]^{\mathrm{T}} \varphi(y)\, ds(y).$$

Using the symmetry of $D_L^{\omega}(x,y)$, we can write $u(x)$ in the same way as (21.19) with $A_m^{(\sigma)}$, $B_m^{(\sigma)}$, and $C_m^{(\sigma)}$ replaced by $\tilde{A}_m^{(\sigma)}$, $\tilde{B}_m^{(\sigma)}$, and $\tilde{C}_m^{(\sigma)}$, respectively. The analysis then proceeds in exactly the same way as for $(\mathrm{N}^{\omega-})$, use being made of the boundary condition (21.28) when the reciprocity relation is applied.

Under the same conditions as for $(\mathrm{N}^{\omega-})$, we find that

$$\tilde{A}_m^{(\sigma)} = \tilde{B}_m^{(\sigma)} = \tilde{C}_m^{(\sigma)} = 0, \quad \sigma = 1,2,\ m = 0,1,2,\ldots. \tag{21.29}$$

Arguing as before, we see that if (21.29) holds, then $u = 0$ in $S^+ \setminus \varpi(b)$. By continuity,

$$0 = u|_{\partial S} = \left(W_{L0}^{\omega} - \tfrac{1}{2} I\right)\varphi,$$

which, in view of (21.27), implies that $\varphi = 0$. Therefore, the following assertion is true.

Theorem 4. *If $\mathscr{R} \in C^{1,\alpha}(\partial S)$, $\alpha \in (0,1)$, and the constants in $L^{\omega}(x,y)$ are chosen so that equalities (21.29) hold, then the integral equation (21.26) has a unique solution $\varphi \in C^{1,\alpha}(\partial S)$.*

Finally, we consider the exterior Robin problem. If the solution of $(\mathrm{R}^{\omega-})$ is sought in the form $u = \mathscr{V}_L^{\omega-}(\varphi)$, then, by (21.9), we need to solve the integral equation

$$(W_{L0}^{\omega*} + \sigma V_{L0}^{\omega} - \tfrac{1}{2} I)\varphi = \mathscr{G}. \tag{21.30}$$

The corresponding homogeneous equation is

$$(W_{L0}^{\omega *} + \sigma V_{L0}^{\omega} - \tfrac{1}{2} I)\varphi = 0. \tag{21.31}$$

Let φ be a solution of (21.31), and let

$$u = \mathcal{V}_L^{\omega +}(\varphi), \quad v = \mathcal{V}_L^{\omega -}(\varphi).$$

We see that

$$(Tv + \sigma v)|_{\partial S} = 0,$$

so, assuming that σ is positive semidefinite, from Theorem 1(ii) we deduce that $v = 0$ in $\bar{S}^-$; therefore,

$$v|_{\partial S} = V_{L0}^{\omega}\varphi = u|_{\partial S} = 0.$$

The argument then continues in exactly the same way as in the case of $(\mathrm{N}^{\omega -})$. If the constants satisfy the necessary conditions, we arrive at $u = 0$ in $S^+ \setminus \overline{\omega}(b)$, which then leads to

$$0 = Tu|_{\partial S} + \sigma u|_{\partial S} = (W_{L0}^{\omega *} + \tfrac{1}{2} I)\varphi + \sigma V_{L0}^{\omega}\varphi.$$

This and (21.31) imply that $\varphi = 0$. Hence, by the Fredholm alternative, (21.30) has a unique solution. The following assertion has therefore been proved.

Theorem 5. *If* $\mathrm{Im}(\sigma)$ *is positive semidefinite,* $\mathcal{G} \in C^{0,\alpha}(\partial S)$, $\alpha \in (0,1)$, *and the constants in* $L^{\omega}(x,y)$ *are chosen so that* (21.24) *are satisfied, then* (21.30) *has a unique solution* $\varphi \in C^{0,\alpha}(\partial S)$.

We remark that the modified matrix of fundamental solutions $D_L^{\omega}(x,y)$ may also be used to obtain uniquely solvable integral equations of the first kind for the solutions of $(\mathrm{D}^{\omega -})$ and $(\mathrm{N}^{\omega -})$ if the constants satisfy appropriate conditions. Such equations were derived in [ThCo12a] by means of the modified matrix $D_M^{\omega}(x,y)$ constructed in [ThCo12b].

21.4 Modification with a Finite Series

We now modify the matrix of fundamental solutions by adding to it a finite series of wavefunctions. This leads to uniquely solvable integral equations for a restricted range of oscillation frequencies.

Let

$$D_{\mathscr{L}}^{\omega}(x,y) = D^{\omega}(x,y) + \mathscr{L}^{\omega}(x,y), \tag{21.32}$$

where

$$\mathscr{L}^{\omega}(x,y)=\frac{i}{4h^2\mu k_3^2}\sum_{m=0}^{M}\sum_{\sigma=1}^{2}\left\{a_m^{(\sigma)}\Phi_m^{(\sigma)}(x)\left[\Phi_m^{(\sigma)}(y)\right]^{\mathrm{T}}+b_m^{(\sigma)}\Upsilon_m^{(\sigma)}(x)\left[\Upsilon_m^{(\sigma)}(y)\right]^{\mathrm{T}}\right.$$
$$\left.+c_m^{(\sigma)}\Psi_m^{(\sigma)}(x)\left[\Psi_m^{(\sigma)}(y)\right]^{\mathrm{T}}\right\}. \quad (21.33)$$

The nonnegative integer M and the constants $a_m^{(\sigma)}$, $b_m^{(\sigma)}$, and $c_m^{(\sigma)}$ are arbitrary.

Repeating the argument in Sect. 21.3 for ($\mathrm{N}^{\omega-}$) and using the potentials obtained by replacing (21.11) and (21.12) with (21.32) and (21.33), we find that for $R_x<R_{\min}$ [see (21.19)],

$$u(x)=\sum_{m=0}^{\infty}\sum_{\sigma=1}^{2}\left\{A_m^{(\sigma)}\hat{\Phi}_m^{(\sigma)}(x)+B_m^{(\sigma)}\hat{\Upsilon}_m^{(\sigma)}(x)+C_m^{(\sigma)}\hat{\Psi}_m^{(\sigma)}(x)\right\}$$
$$+\sum_{m=0}^{M}\sum_{\sigma=1}^{2}\left\{a_m^{(\sigma)}A_m^{(\sigma)}\Phi_m^{(\sigma)}(x)+b_m^{(\sigma)}B_m^{(\sigma)}\Upsilon_m^{(\sigma)}(x)\right.$$
$$\left.+c_m^{(\sigma)}C_m^{(\sigma)}\Psi_m^{(\sigma)}(x)\right\}. \quad (21.34)$$

If the constants are chosen so that (21.22), (21.23), or (21.25) are satisfied for $\sigma=1,2$ and $m=0,1,2,\ldots,M$, then it is easy to see that

$$A_m^{(\sigma)}=B_m^{(\sigma)}=C_m^{(\sigma)}=0,\quad \sigma=1,2,\ m=0,1,2,\ldots,M. \quad (21.35)$$

From (21.34) it follows that if (21.35) holds, then for $R_x<R_{\min}$,

$$u(x)=\sum_{m=M+1}^{\infty}\sum_{\sigma=1}^{2}\left\{A_m^{(\sigma)}\hat{\Phi}_m^{(\sigma)}(x)+B_m^{(\sigma)}\hat{\Upsilon}_m^{(\sigma)}(x)+C_m^{(\sigma)}\hat{\Psi}_m^{(\sigma)}(x)\right\}. \quad (21.36)$$

For simplicity, below we write (R,θ) instead of (R_x,θ_x).

Theorem 6. *The function u defined by (21.36) has a zero in R at the origin, of order at least $M-1$.*

Proof. By (21.36),

$$u(x)=\sum_{\sigma=1}^{2}\left\{A_{M+1}^{(\sigma)}\hat{\Phi}_{M+1}^{(\sigma)}(x)+B_{M+1}^{(\sigma)}\hat{\Upsilon}_{M+1}^{(\sigma)}(x)+C_{M+1}^{(\sigma)}\hat{\Psi}_{M+1}^{(\sigma)}(x)+\cdots\right\}.$$

It is easy to show that

$$\frac{\partial}{\partial x_1}\left\{J_{M+1}(k_1R)E_{M+1}^{(\sigma)}(\theta)\right\}=\tfrac{1}{2}k_1\left[J_M(k_1R)E_M^{(\sigma)}(\theta)-J_{M+2}(k_1R)E_{M+2}^{(\sigma)}(\theta)\right],$$

$$\frac{\partial}{\partial x_2}\left\{J_{M+1}(k_1R)E_{M+1}^{(\sigma)}(\theta)\right\} = \tfrac{1}{2}(-1)^{\sigma}k_1\left[J_M(k_1R)E_M^{(3-\sigma)}(\theta) + J_{M+2}(k_1R)E_{M+2}^{(3-\sigma)}(\theta)\right];$$

so, from the definition of the wavefunctions we deduce that

$$\hat{\Phi}_{M+1}^{(\sigma)}(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} \alpha_1 k_1\left(J_M(k_1R)E_M^{(\sigma)}(\theta) - J_{M+2}(k_1R)E_{M+2}^{(\sigma)}(\theta)\right) \\ (-1)^{\sigma}\alpha_1 k_1\left(J_M(k_1R)E_M^{(3-\sigma)}(\theta) + J_{M+2}(k_1R)E_{M+2}^{(3-\sigma)}(\theta)\right) \\ -2hk_3\alpha_2 J_{M+1}(k_1R)E_{M+1}^{(\sigma)}(\theta) \end{pmatrix}.$$

Also,

$$\frac{\partial^l}{\partial R^l}J_M(k_1R) = \left(\tfrac{1}{2}k_1\right)^l J_{M-l}(k_1R) + \sum_{m=2}^{2l}\beta_m J_{M-l+m}(k_1R),$$

where the β_m are constants. Since

$$J_0(0) = 1, \quad J_n(0) = 0,\ n \geq 1,$$

we find that

$$\left(\frac{\partial^l}{\partial R^l}J_M(k_1R)\right)\Bigg|_{R=0} = 0 \quad \text{if } M - l \geq 1.$$

Consequently, $\hat{\Phi}_{M+1}^{(\sigma)}$ has a zero in R at the origin of order at least $M-1$. This procedure can be repeated for $\hat{\Upsilon}_{M+1}^{(\sigma)}$ and $\hat{\Phi}_{M+1}^{(\sigma)}$, and the result follows.

From Theorem 6 we see that by increasing the value of M, we increase the order of the zero of u in R at the origin and, hence, strengthen the constraints on the function. It turns out that this increases (or, rather, does not decrease) the lowest value $\omega(M)$, say, of the oscillation frequency for which u can be a nonzero solution of the homogeneous interior Dirichlet problem. Unfortunately, there is no way of knowing how many terms we need to take in series (21.33) to ensure that a particular value of ω is less than $\omega(M)$. All we can claim is that by increasing M, we have a greater chance of eliminating oscillation frequencies at which nonuniqueness can occur.

Suppose that $\omega < \omega(M)$ and that (21.35) holds. Then $u = 0$ in S^+ and we can continue the argument of the previous section.

$(\mathrm{D}^{\omega-})$ and $(\mathrm{R}^{\omega-})$ are treated analogously.

References

[Be90] Bencheikh, L.: Modified fundamental solutions for the scattering of elastic waves by a cavity. Q. J. Mech. Appl. Math. **43**, 57–73 (1990)

[Co98] Constanda, C.: Radiation conditions and uniqueness for stationary oscillations in elastic plates. Proc. Am. Math. Soc. **126**, 827–834 (1988)

[Co90] Constanda, C.: A Mathematical Analysis of Bending of Plates with Transverse Shear Deformation. Longman, Harlow (1990)
[Jo74] Jones, D.S.: Integral equations for the exterior acoustic problem. Q. J. Mech. Appl. Math. **27**, 129–142 (1974)
[Jo84] Jones, D.S.: An exterior problem in elastodynamics. Math. Proc. Camb. Philos. Soc. **96**, 173–182 (1984)
[ScCo93] Schiavone, P., Constanda, C.: Oscillation problems in thin plates with transverse shear deformation. SIAM J. Appl. Math. **53**, 1253–1263 (1993)
[ThCo97] Thomson, G.R., Constanda, C.: On stationary oscillations in bending of plates. In: Constanda, C., Saranen, J., Seikkala, S. (eds.) Integral Methods in Science and Engineering, Vol. 1: Analytic Methods, pp. 190–194. Longman, Harlow (1997)
[ThCo98] Thomson, G.R., Constanda, C.: Area potentials for thin plates. An. Stiint. Al.I. Cuza Univ. Iasi Sect. Ia Mat. **44**, 235–244 (1998)
[ThCo99] Thomson, G.R., Constanda, C.: Scattering of high frequency flexural waves in thin plates. Math. Mech. Solids **4**, 461–479 (1999)
[ThCo09b] Thomson, G.R., Constanda, C.: A matrix of fundamental solutions in the theory of plate oscillations. Appl. Math. Lett. **22**, 707–711 (2009)
[ThCo09a] Thomson, G.R., Constanda, C.: Integral equation methods for the Robin problem in stationary oscillations of elastic plates. IMA J. Appl. Math. **74**, 548–558 (2009)
[ThCo10] Thomson, G.R., Constanda, C.: The direct method for harmonic oscillations of elastic plates with Robin boundary conditions. Math. Mech. Solids **16**, 200–207 (2010)
[ThCo11] Thomson, G.R., Constanda, C.: Uniqueness of solution for the Robin problem in high-frequency vibrations of elastic plates. Appl. Math. Lett. **24**, 577–581 (2011)
[ThCo12a] Thomson, G.R., Constanda, C.: Integral equations of the first kind in the theory of oscillating plates. Appl. Anal. **91**, 2235–2244 (2012)
[ThCo12b] Thomson, G.R., Constanda, C.: Nonstandard integral equations for the harmonic oscillations of thin plates. This volume, pp. 311–328
[Ur78] Ursell, F.: On the exterior problems of acoustics: II. Math. Proc. Camb. Philos. Soc. **84**, 545–548 (1978)

Chapter 22
Nonstandard Integral Equations for the Harmonic Oscillations of Thin Plates

G.R. Thomson, C. Constanda, and D.R. Doty

22.1 Prerequisites

In [ThCo97] and [ThCo09a] the problems of high frequency harmonic oscillations of thin elastic plates with Dirichlet, Neumann, and Robin boundary conditions were investigated by means of a classical indirect boundary integral equation method. This method was not entirely satisfactory since, for the exterior problems, it produced integral equations with nonunique solutions for certain values of the oscillation frequency, although the actual boundary value problems always had at most one solution. When a direct method was employed (see [ThCo99] and [ThCo10]), it was found that uniqueness could be guaranteed only if a *pair* of integral equations was derived for each exterior problem. The classical techniques did not seem to offer any answer to the question of whether the solutions could be obtained from single, uniquely solvable equations. Below we propose a modified indirect boundary integral equation method, based on constructing a matrix of fundamental solutions satisfying a dissipative (or Robin-type) condition on a curve interior to the scatterer, which answers the above question in the affirmative.

Problems of this nature in two-dimensional acoustics and plane deformation were addressed in [Ur73] and [Be90].

In the sequel, a superscript T denotes matrix transposition and $x = (x_1, x_2)^{\mathrm{T}}$ and $y = (y_1, y_2)^{\mathrm{T}}$ are generic points in the Cartesian plane $\mathbb{R}^2$, with polar radii R_x and R_y, respectively.

Let S^+ be a domain in $\mathbb{R}^2$ bounded by a simple, closed C^2-curve ∂S, and let $S^- = \mathbb{R}^2 \setminus \bar{S}^+$. We assume that the origin of coordinates lies in S^+.

G.R. Thomson
A.C.C.A., Glasgow, UK
e-mail: thomsongavin@gmail.com

C. Constanda (✉) • D.R. Doty
The University of Tulsa, Tulsa, OK, USA
e-mail: christian-constanda@utulsa.edu; dale-doty@utulsa.edu

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_22,

We consider a homogeneous and isotropic elastic plate of density ρ and Lamé constants λ and μ, which occupies the infinite region $\bar{S}^- \times [-h_0/2, h_0/2]$ in $\mathbb{R}^3$, where $h_0 =$ const is the plate thickness. The harmonic oscillations of frequency ω of the plate, when transverse shear deformation is taken into account, are governed by the system [ScCo93]

$$A^{\omega}(\partial_x)u(x) = H(x), \tag{22.1}$$

where $u = (u_1, u_2, u_3)^{\mathrm{T}}$ is a vector characterizing the displacements, H is related to the averaged body forces and moments, and the matrix operator $A^{\omega}(\partial_x) = A^{\omega}(\partial/\partial x_1, \partial/\partial x_2)$ is defined by

$$A^{\omega}(\xi_1, \xi_2) = \begin{pmatrix} h^2\mu(\Delta + k_3^2) + h^2(\lambda+\mu)\xi_1^2 & h^2(\lambda+\mu)\xi_1\xi_2 & -\mu\xi_1 \\ h^2(\lambda+\mu)\xi_1\xi_2 & h^2\mu(\Delta + k_3^2) + h^2(\lambda+\mu)\xi_2^2 & -\mu\xi_2 \\ \mu\xi_1 & \mu\xi_2 & \mu(\Delta + k^2) \end{pmatrix};$$

here $h^2 = h_0^2/12$, $\Delta = \xi_1^2 + \xi_2^2$, and

$$k^2 = \frac{\rho\omega^2}{\mu}, \quad k_3^2 = k^2 - \frac{1}{h^2}. \tag{22.2}$$

It is assumed throughout that

$$\lambda + \mu > 0, \quad \mu > 0, \quad \rho\omega^2 h^2 > \mu. \tag{22.3}$$

Since a particular solution of (22.1) is readily constructed (see [ThCo98]), without loss of generality we consider the homogeneous system

$$A^{\omega}(\partial_x)u(x) = 0. \tag{22.4}$$

The boundary moment–stress operator $T(\partial_x) = T(\partial/\partial x_1, \partial/\partial x_2)$ is defined by [ScCo93]

$$T(\xi_1, \xi_2) = \begin{pmatrix} h^2((\lambda+2\mu)\nu_1\xi_1 + \mu\nu_2\xi_2) & h^2(\mu\nu_2\xi_1 + \lambda\nu_1\xi_2) & 0 \\ h^2(\lambda\nu_2\xi_1 + \mu\nu_1\xi_2) & h^2(\mu\nu_1\xi_1 + (\lambda+2\mu)\nu_2\xi_2) & 0 \\ \mu\nu_1 & \mu\nu_2 & \mu(\nu_1\xi_1 + \nu_2\xi_2) \end{pmatrix},$$

where $\nu = (\nu_1, \nu_2)^{\mathrm{T}}$ is the unit outward normal to the boundary of the middle plane of the plate.

We denote by $\mathscr{B}^{\omega}$ the class of functions defined in S^- which satisfy the radiation conditions formulated in [Co98] (see also [ThCo11]) as $R_x \to \infty$.

Let $\mathscr{R}$, $\mathscr{S}$, and $\mathscr{G}$ be 3×1 vector functions defined on the curve ∂S, and let $\sigma \in C^{1,\alpha}(\partial S)$, $\alpha \in (0,1)$ be a symmetric 3×3 matrix function. We state the exterior boundary value problems ($D^{\omega-}$), ($N^{\omega-}$), and ($R^{\omega-}$) with Dirichlet, Neumann, and Robin boundary data, respectively, as follows:

($D^{\omega-}$) Find $u \in C^2(S^-)\cap C^1(\bar{S}^-)\cap \mathscr{B}^\omega$ that satisfies (22.4) in S^- and

$$u|_{\partial S} = \mathscr{R}.$$

($N^{\omega-}$) Find $u \in C^2(S^-)\cap C^1(\bar{S}^-)\cap \mathscr{B}^\omega$ that satisfies (22.4) in S^- and

$$Tu|_{\partial S} = \mathscr{S}.$$

($R^{\omega-}$) Find $u \in C^2(S^-)\cap C^1(\bar{S}^-)\cap \mathscr{B}^\omega$ that satisfies (22.4) in S^- and

$$(Tu+\sigma u)|_{\partial S} = \mathscr{G}.$$

A function u is said to be *regular* in S^- if $u \in C^2(S^-)\cap C^1(\bar{S}^-)$.

Theorem 1. *(i) Each of ($D^{\omega-}$) and ($N^{\omega-}$) has at most one regular solution.*
(ii) If $Im(\sigma)$ is positive semidefinite, then ($R^{\omega-}$) has at most one regular solution.

The first assertion is proved in [Co98] and the second one in [ThCo11].

22.2 Fundamental Solutions

We make certain suitable modifications to the matrix of fundamental solutions constructed in [ThCo09b] for the operator $A^\omega(\partial_x)$.

The wavenumbers k_1^2, k_2^2, and k_3^2, where

$$k_1^2+k_2^2 = \frac{\lambda+3\mu}{\lambda+2\mu}k^2, \quad k_1^2k_2^2 = \frac{\mu}{\lambda+2\mu}k^2k_3^2 \tag{22.5}$$

and k^2 and k_3^2 are defined by (22.2), arise naturally in system (22.4) (see [Co98] and [ThCo09b]). In [ThCo09b] it is shown that all the wavenumbers are real, positive, and distinct if conditions (22.3) hold.

We introduce the symbols ε_m and $E_m^{(\sigma)}(\theta)$ by

$$\varepsilon_m = \begin{cases} 1, & m=0,\\ 2, & m\geq 1,\end{cases} \quad E_m^{(\sigma)}(\theta) = \begin{cases} \cos m\theta, & \sigma=1,\\ \sin m\theta, & \sigma=2,\end{cases}$$

where m is a nonnegative integer. Let

$$\phi_m^{(\sigma)}(x) = \sqrt{\varepsilon_m}H_m(k_1R_x)E_m^{(\sigma)}(\theta_x), \quad \hat{\phi}_m^{(\sigma)}(x) = \sqrt{\varepsilon_m}J_m(k_1R_x)E_m^{(\sigma)}(\theta_x),$$

$$\upsilon_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} H_m(k_2 R_x) E_m^{(\sigma)}(\theta_x), \quad \hat{\upsilon}_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} J_m(k_2 R_x) E_m^{(\sigma)}(\theta_x),$$

$$\psi_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} H_m(k_3 R_x) E_m^{(\sigma)}(\theta_x), \quad \hat{\psi}_m^{(\sigma)}(x) = \sqrt{\varepsilon_m} J_m(k_3 R_x) E_m^{(\sigma)}(\theta_x),$$

where θ_x and θ_y are the polar angles of x and y, respectively, H_m is the Hankel function of the first kind and order m, and J_m is the Bessel function of the first kind and order m. We also write

$$\alpha_1^2 = \frac{k_2^2 - \mu' k_3^2}{k_2^2 - k_1^2}, \quad \alpha_2^2 = \frac{k_1^2 - \mu' k_3^2}{k_1^2 - k_2^2},$$

where $\mu' = \mu/(\lambda + 2\mu)$. These constants are also strictly positive if (22.3) holds, provided that k_1^2 is the larger root of (22.5).

The *radiating wavefunctions* are [ThCo09b]

$$\Phi_m^{(\sigma)}(x) = \left(\alpha_1 \frac{\partial}{\partial x_1} \phi_m^{(\sigma)}(x), \alpha_1 \frac{\partial}{\partial x_2} \phi_m^{(\sigma)}(x), -hk_3 \alpha_2 \phi_m^{(\sigma)}(x) \right)^{\mathrm{T}}, \tag{22.6}$$

$$\Upsilon_m^{(\sigma)}(x) = \left(\alpha_2 \frac{\partial}{\partial x_1} \upsilon_m^{(\sigma)}(x), \alpha_2 \frac{\partial}{\partial x_2} \upsilon_m^{(\sigma)}(x), hk_3 \alpha_1 \upsilon_m^{(\sigma)}(x) \right)^{\mathrm{T}}, \tag{22.7}$$

$$\Psi_m^{(\sigma)}(x) = \left(\frac{\partial}{\partial x_2} \psi_m^{(\sigma)}(x), -\frac{\partial}{\partial x_1} \psi_m^{(\sigma)}(x), 0 \right)^{\mathrm{T}}. \tag{22.8}$$

The corresponding *regular wavefunctions* $\hat{\Phi}_m^{(\sigma)}$, $\hat{\Upsilon}_m^{(\sigma)}$, and $\hat{\Psi}_m^{(\sigma)}$ are defined in the same way, with $\phi_m^{(\sigma)}$, $\upsilon_m^{(\sigma)}$, and $\psi_m^{(\sigma)}$ replaced by $\hat{\phi}_m^{(\sigma)}$, $\hat{\upsilon}_m^{(\sigma)}$, and $\hat{\psi}_m^{(\sigma)}$, respectively.

Let $D^\omega(x,y)$ be the matrix of fundamental solutions for $A^\omega(\partial_x)$ constructed in [ThCo09b], where it was shown that, for $R_x < R_y$,

$$D^\omega(x,y) = \frac{i}{4h^2 \mu k_3^2} \sum_{m=0}^{\infty} \sum_{\sigma=1}^{2} \left\{ \hat{\Phi}_m^{(\sigma)}(x) \left[\Phi_m^{(\sigma)}(y) \right]^{\mathrm{T}} + \hat{\Upsilon}_m^{(\sigma)}(x) \left[\Upsilon_m^{(\sigma)}(y) \right]^{\mathrm{T}} + \hat{\Psi}_m^{(\sigma)}(x) \left[\Psi_m^{(\sigma)}(y) \right]^{\mathrm{T}} \right\}. \tag{22.9}$$

This decomposition in terms of wavefunctions is used below to modify the matrix $D^\omega(x,y)$.

22.3 Modified Fundamental Solutions

Let $\mathscr{D}$ be the annular region in $\mathbb{R}^2$ bounded externally and internally by simple, closed curves ∂S_1 and ∂S_2, respectively, and let K be a 3×3 matrix whose elements are such that

$$K_{ij} = \bar{K}_{ji} \quad \text{for } i \neq j \tag{22.10}$$

and either

$$\mathrm{Im}(K_{11}),\ \mathrm{Im}(K_{22}),\ \mathrm{Im}(K_{33}) > 0 \tag{22.11}$$

or

$$\mathrm{Im}(K_{11}),\ \mathrm{Im}(K_{22}),\ \mathrm{Im}(K_{33}) < 0. \tag{22.12}$$

The following assertion, proved in [ThCo12], is the main ingredient that informs our choice of modified matrix of fundamental solutions.

Theorem 2. *If u is an analytic solution of (22.4) in $\mathscr{D} \cup \partial S_2$ such that*

$$u = 0 \ \text{ or } \ Tu = 0 \quad \text{on } \partial S_1, \quad Tu + Ku = 0 \quad \text{on } \partial S_2,$$

where K is such that (22.10) and either (22.11) or (22.12) are satisfied, then $u = 0$ in $\mathscr{D}$.

We denote by $\partial\varpi(a)$ the circle centered at the origin whose radius a is chosen so that $\partial\varpi(a)$ lies entirely within S^+, and by S_a^- the infinite region exterior to this circle. The modified matrix of fundamental solutions $D_M^\omega(x,y)$ that we envisage is of the form

$$D_M^\omega(x,y) = D^\omega(x,y) + M^\omega(x,y), \tag{22.13}$$

where the columns of $M^\omega(x,y)$ are regular solutions of (22.4) in $S_a^- \cup \partial\varpi(a)$ with respect to x and satisfy the radiation conditions as $R_x \to \infty$. We also require that

$$T(\partial_x) D_M^\omega(x,y) + K D_M^\omega(x,y) = 0, \quad x \in \partial\varpi(a), \tag{22.14}$$

where

$$K = \begin{pmatrix} h^2\mu\kappa & 0 & 0 \\ 0 & h^2\mu\kappa & 0 \\ 0 & 0 & \mu\kappa \end{pmatrix}, \tag{22.15}$$

with

$$\kappa = |\kappa| e^{i\delta}, \quad 0 < \delta < \pi. \tag{22.16}$$

Obviously, this choice of K satisfies the conditions in Theorem 2.

Using arguments similar to those employed in [ThCo11] in connection with the symmetry of the Green's tensor for the interior Robin problem, it can be shown that

$$D_M^\omega(x,y) = [D_M^\omega(y,x)]^{\mathrm{T}}. \tag{22.17}$$

We assume that $M^{\omega}(x,y)$ is of the form

$$M^{\omega}(x,y) = \frac{i}{4h^2\mu k_3^2} \sum_{m=0}^{\infty} \sum_{\sigma=1}^{2} \left\{ \Phi_m^{(\sigma)}(x) \left[A_m^{(\sigma)}(y)\right]^{\mathrm{T}} + \Upsilon_m^{(\sigma)}(x) \left[B_m^{(\sigma)}(y)\right]^{\mathrm{T}} + \Psi_m^{(\sigma)}(x) \left[C_m^{(\sigma)}(y)\right]^{\mathrm{T}} \right\}, \quad (22.18)$$

where $A_m^{(\sigma)}$, $B_m^{(\sigma)}$, and $C_m^{(\sigma)}$ are 3×1 vector functions to be determined via (22.13) and (22.14). It is easily verified that the columns of $M^{\omega}(x,y)$ satisfy (22.4) in $\mathbb{R}^2 \setminus \{0\}$ and the radiation conditions as $R_x \to \infty$.

For simplicity, from now on we write (R,θ) instead of (R_x,θ_x). Using (22.9), (22.13), and (22.18), we see that (22.14) is satisfied if

$$\begin{aligned} &\sum_{\sigma=1}^{2} \left\{ \left[T\Phi_m^{(\sigma)} + K\Phi_m^{(\sigma)}\right]\big|_{R=a} \left[A_m^{(\sigma)}(y)\right]^{\mathrm{T}} + \left[T\Upsilon_m^{(\sigma)} + K\Upsilon_m^{(\sigma)}\right]\big|_{R=a} \left[B_m^{(\sigma)}(y)\right]^{\mathrm{T}} \right. \\ &\qquad \left. + \left[T\Psi_m^{(\sigma)} + K\Psi_m^{(\sigma)}\right]\big|_{R=a} \left[C_m^{(\sigma)}(y)\right]^{\mathrm{T}} \right\} \\ &= -\sum_{\sigma=1}^{2} \left\{ \left[T\hat{\Phi}_m^{(\sigma)} + K\hat{\Phi}_m^{(\sigma)}\right]\big|_{R=a} \left[\Phi_m^{(\sigma)}(y)\right]^{\mathrm{T}} + \left[T\hat{\Upsilon}_m^{(\sigma)} + K\hat{\Upsilon}_m^{(\sigma)}\right]\big|_{R=a} \left[\Upsilon_m^{(\sigma)}(y)\right]^{\mathrm{T}} \right. \\ &\qquad \left. + \left[T\hat{\Psi}_m^{(\sigma)} + K\hat{\Psi}_m^{(\sigma)}\right]\big|_{R=a} \left[\Psi_m^{(\sigma)}(y)\right]^{\mathrm{T}} \right\} \end{aligned}$$

for $m = 0,1,2,\ldots$. This can be written as the 3×3 system of equations

$$\begin{aligned} &\sum_{\sigma=1}^{2} U_m^{(\sigma)} \left(A_m^{(\sigma)}(y), B_m^{(\sigma)}(y), C_m^{(\sigma)}(y)\right)^{\mathrm{T}} \\ &\qquad = -\sum_{\sigma=1}^{2} \hat{U}_m^{(\sigma)} \left(\Phi_m^{(\sigma)}(y), \Upsilon_m^{(\sigma)}(y), \Psi_m^{(\sigma)}(y)\right)^{\mathrm{T}}, \quad (22.19) \end{aligned}$$

where

$$U_m^{(\sigma)} = \begin{pmatrix} \left[T\Phi_m^{(\sigma)} + K\Phi_m^{(\sigma)}\right]_1 & \left[T\Upsilon_m^{(\sigma)} + K\Upsilon_m^{(\sigma)}\right]_1 & \left[T\Psi_m^{(\sigma)} + K\Psi_m^{(\sigma)}\right]_1 \\ \left[T\Phi_m^{(\sigma)} + K\Phi_m^{(\sigma)}\right]_2 & \left[T\Upsilon_m^{(\sigma)} + K\Upsilon_m^{(\sigma)}\right]_2 & \left[T\Psi_m^{(\sigma)} + K\Psi_m^{(\sigma)}\right]_2 \\ \left[T\Phi_m^{(\sigma)} + K\Phi_m^{(\sigma)}\right]_3 & \left[T\Upsilon_m^{(\sigma)} + K\Upsilon_m^{(\sigma)}\right]_3 & \left[T\Psi_m^{(\sigma)} + K\Psi_m^{(\sigma)}\right]_3 \end{pmatrix}\Bigg|_{R=a}$$

and $\hat{U}_m^{(\sigma)}$ is defined similarly with $\Phi_m^{(\sigma)}$, $\Upsilon_m^{(\sigma)}$, and $\Psi_m^{(\sigma)}$ replaced by $\hat{\Phi}_m^{(\sigma)}$, $\hat{\Upsilon}_m^{(\sigma)}$, and $\hat{\Psi}_m^{(\sigma)}$, respectively.

Let

$$a_m = \alpha_1 \left[-(\lambda/\mu)k_1^2 a^2 H_m(k_1 a) + 2k_1^2 a^2 H_m''(k_1 a) + \kappa k_1 a^2 H_m'(k_1 a)\right],$$

$$b_m = \alpha_2 \left[-(\lambda/\mu)k_2^2 a^2 H_m(k_2 a) + 2k_2^2 a^2 H_m''(k_2 a) + \kappa k_2 a^2 H_m'(k_2 a)\right],$$

$$
\begin{aligned}
c_m &= -2mk_3aH'_m(k_3a) - m(\kappa a - 2)H_m(k_3a),\\
d_m &= \alpha_1\left[2mk_1aH'_m(k_1a) + m(\kappa a - 2)H_m(k_1a)\right],\\
e_m &= \alpha_2\left[2mk_2aH'_m(k_2a) + m(\kappa a - 2)H_m(k_2a)\right],\\
f_m &= -k_3^2a^2H''_m(k_3a) - k_3a(\kappa - a)H'_m(k_3a) - m^2H_m(k_3a),\\
g_m &= (\alpha_1 - hk_3\alpha_2)k_1aH'_m(k_1a) - \kappa hk_3\alpha_2aH_m(k_1a),\\
h_m &= (\alpha_2 + hk_3\alpha_1)k_2aH'_m(k_2a) + \kappa hk_3\alpha_1aH_m(k_2a),\\
i_m &= -mH_m(k_3a),
\end{aligned}
$$

where the "prime" symbol denotes differentiation with respect to the argument. The constants $\hat{a}_m, \hat{b}_m, \ldots, \hat{\imath}_m$ are defined analogously, with J_m in place of H_m.

After a lengthy calculation, from (22.6) to (22.8) we conclude that system (22.19) is equivalent to

$$
\begin{aligned}
&R_m(\theta)P_m(A_m^{(1)}, B_m^{(1)}, -C_m^{(2)})^{\mathrm{T}} + S_m(\theta)P_m(A_m^{(2)}, B_m^{(2)}, C_m^{(1)})^{\mathrm{T}}\\
&\qquad = -R_m(\theta)\hat{P}_m(\Phi_m^{(1)}, \Upsilon_m^{(1)}, -\Psi_m^{(2)})^{\mathrm{T}} - S_m(\theta)\hat{P}_m(\Phi_m^{(2)}, \Upsilon_m^{(2)}, \Psi_m^{(1)})^{\mathrm{T}},
\end{aligned}
$$

where

$$
P_m = \begin{pmatrix} a_m & b_m & c_m\\ d_m & e_m & f_m\\ g_m & h_m & i_m \end{pmatrix},
$$

$$
\hat{P}_m = \begin{pmatrix} \hat{a}_m & \hat{b}_m & \hat{c}_m\\ \hat{d}_m & \hat{e}_m & \hat{f}_m\\ \hat{g}_m & \hat{h}_m & \hat{\imath}_m \end{pmatrix},
$$

and

$$
R_m(\theta) = \frac{\sqrt{\varepsilon_m}h^2\mu}{a^2}\begin{pmatrix} \cos\theta\cos(m\theta) & \sin\theta\sin(m\theta) & 0\\ \sin\theta\cos(m\theta) & -\cos\theta\sin(m\theta) & 0\\ 0 & 0 & (a/h^2)\cos(m\theta) \end{pmatrix},
$$

$$
S_m(\theta) = \frac{\sqrt{\varepsilon_m}h^2\mu}{a^2}\begin{pmatrix} \cos\theta\sin(m\theta) & -\sin\theta\cos(m\theta) & 0\\ \sin\theta\sin(m\theta) & \cos\theta\cos(m\theta) & 0\\ 0 & 0 & (a/h^2)\sin(m\theta) \end{pmatrix}.
$$

By the orthogonality of the trigonometric functions, it suffices to choose the vectors $A_m^{(\sigma)}$, $B_m^{(\sigma)}$, and $C_m^{(\sigma)}$ so that

$$\left(A_m^{(\sigma)}, B_m^{(\sigma)}, (-1)^{\sigma} C_m^{(3-\sigma)}\right)^{\mathrm{T}} = -P_m^{-1} \hat{P}_m \left(\Phi_m^{(\sigma)}, \Upsilon_m^{(\sigma)}, (-1)^{\sigma} \Psi_m^{(3-\sigma)}\right)^{\mathrm{T}} \tag{22.20}$$

for $\sigma = 1,2$ and $m = 0,1,2,\ldots$. Thus, to construct the modified matrix $D_M^{\omega}(x,y)$ we need to show that P_m is invertible.

We recall some orthogonality-type relations established in [ThCo09b]. If ∂C is any simple, closed curve containing the origin in its interior, then for $m \geq 1$,

$$\int_{\partial C} \left\{ \left[\bar{\chi}_m^{(\sigma j)}\right]^{\mathrm{T}} T \chi_n^{(\nu k)} - \left[\chi_n^{(\nu k)}\right]^{\mathrm{T}} T \bar{\chi}_m^{(\sigma j)} \right\} ds = 8ih^2 \mu k_3^2 \delta_{mn} \delta_{\sigma\nu} \delta_{jk}, \tag{22.21}$$

where

$$\chi_m^{(\sigma 1)}(x) = \Phi_m^{(\sigma)}(x), \quad \chi_m^{(\sigma 2)}(x) = \Upsilon_m^{(\sigma)}(x), \quad \chi_m^{(\sigma 3)}(x) = \Psi_m^{(\sigma)}(x).$$

Equality (22.21) also holds for $m = 0$, $\sigma = 1$.

Theorem 3. *P_m is invertible for all $m \geq 0$.*

Proof. First, we deal with the case $m \geq 1$. Consider the system of equations

$$\left[T\Phi_m^{(1)} + K\Phi_m^{(1)}\right]\Big|_{R=a} z_1 + \left[T\Upsilon_m^{(1)} + K\Upsilon_m^{(1)}\right]\Big|_{R=a} z_2 + \left[T\Psi_m^{(2)} + K\Psi_m^{(2)}\right]\Big|_{R=a} z_3 = 0. \tag{22.22}$$

This can be rewritten as

$$TU + KU = 0 \quad \text{on } R = a, \tag{22.23}$$

where

$$U = \Phi_m^{(1)} z_1 + \Upsilon_m^{(1)} z_2 + \Psi_m^{(2)} z_3. \tag{22.24}$$

By (22.23), (22.15), and (22.16), on $R = a$ we have

$$\bar{U}^{\mathrm{T}} T U - U^{\mathrm{T}} T \bar{U} = -2i\mu |\kappa| (\sin\delta) \left(h^2 |U_1|^2 + h^2 |U_2|^2 + |U_3|^2 \right);$$

so, taking (22.24) into account, we find that

$$\begin{aligned} &\left[\bar{\Phi}_m^{(1)} \bar{z}_1 + \bar{\Upsilon}_m^{(1)} \bar{z}_2 + \bar{\Psi}_m^{(2)} \bar{z}_3\right]^{\mathrm{T}} \left[T\Phi_m^{(1)} z_1 + T\Upsilon_m^{(1)} z_2 + T\Psi_m^{(2)} z_3\right] \\ &\quad - \left[\Phi_m^{(1)} z_1 + \Upsilon_m^{(1)} z_2 + \Psi_m^{(2)} z_3\right]^{\mathrm{T}} \left[T\bar{\Phi}_m^{(1)} \bar{z}_1 + T\bar{\Upsilon}_m^{(1)} \bar{z}_2 + T\bar{\Psi}_m^{(2)} \bar{z}_3\right] \\ &\quad = -2i\mu |\kappa| (\sin\delta) \left(h^2 |U_1|^2 + h^2 |U_2|^2 + |U_3|^2 \right). \end{aligned} \tag{22.25}$$

Integrating (22.25) around $\partial\varpi(a)$ and using (22.21), we arrive at

$$8ih^2\mu k_3^2\left(|z_1|^2+|z_2|^2+|z_3|^2\right) = -2i\mu|\kappa|(\sin\delta)\int\limits_{\partial\varpi(a)}\left(h^2|U_1|^2+h^2|U_2|^2+|U_3|^2\right)ds,$$

which, since $0<\delta<\pi$, implies that $z_1=z_2=z_3=0$. We have shown that system (22.22) has only the zero solution, or, equivalently, that the system

$$\Theta_m Z=0,$$

where

$$\Theta_m=\left(\begin{array}{ccc}\left[T\Phi_m^{(1)}+K\Phi_m^{(1)}\right]_1 & \left[T\Upsilon_m^{(1)}+K\Upsilon_m^{(1)}\right]_1 & \left[T\Psi_m^{(2)}+K\Psi_m^{(2)}\right]_1\\ \left[T\Phi_m^{(1)}+K\Phi_m^{(1)}\right]_2 & \left[T\Upsilon_m^{(1)}+K\Upsilon_m^{(1)}\right]_2 & \left[T\Psi_m^{(2)}+K\Psi_m^{(2)}\right]_2\\ \left[T\Phi_m^{(1)}+K\Phi_m^{(1)}\right]_3 & \left[T\Upsilon_m^{(1)}+K\Upsilon_m^{(1)}\right]_3 & \left[T\Psi_m^{(2)}+K\Psi_m^{(2)}\right]_3\end{array}\right)\Bigg|_{R=a}$$

and $Z=(z_1,z_2,z_3)^{\mathrm{T}}$, has no nonzero solutions; consequently,

$$\begin{aligned}0\neq{}&\det\Theta_m\\ ={}&\left(\frac{\sqrt{2}h^2\mu}{a^2}\right)^3\left(\frac{a}{h^2}\right)\begin{vmatrix}\cos\theta\cos(m\theta) & \sin\theta\sin(m\theta) & 0\\ \sin\theta\cos(m\theta) & -\cos\theta\sin(m\theta) & 0\\ 0 & 0 & \cos(m\theta)\end{vmatrix}\begin{vmatrix}a_m & b_m & -c_m\\ d_m & e_m & -f_m\\ g_m & h_m & -i_m\end{vmatrix}\\ ={}&2\sqrt{2}\,\frac{h^4\mu^3}{a^5}\cos^2(m\theta)\sin(m\theta)(\det P_m).\end{aligned}$$

Since $m\neq0$, we deduce that $\det P_m\neq0$, so P_m is invertible for $m\geq1$.

When $m=0$, the matrix reduces to

$$P_0=\begin{pmatrix}a_0 & b_0 & 0\\ 0 & 0 & f_0\\ g_0 & h_0 & 0\end{pmatrix},$$

from which we see that $\det P_0=-f_0(a_0h_0-b_0g_0)$. Consider the system

$$\left[T\Phi_0^{(1)}+K\Phi_0^{(1)}\right]\Big|_{R=a}z_1+\left[T\Upsilon_0^{(1)}+K\Upsilon_0^{(1)}\right]\Big|_{R=a}z_2=0. \tag{22.26}$$

This is not an overdetermined system since the second row is equal to the first row multiplied by $\tan\theta$. Consequently, (22.26) is equivalent to

$$\begin{pmatrix} \left[T\Phi_0^{(1)}+K\Phi_0^{(1)}\right]_1\big|_{R=a} & \left[T\Upsilon_0^{(1)}+K\Upsilon_0^{(1)}\right]_1\big|_{R=a} \\ \left[T\Phi_0^{(1)}+K\Phi_0^{(1)}\right]_3\big|_{R=a} & \left[T\Upsilon_0^{(1)}+K\Upsilon_0^{(1)}\right]_3\big|_{R=a} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0. \tag{22.27}$$

System (22.26) can be written as

$$TU + KU = 0 \quad \text{on } R = a,$$

where

$$U = \Phi_0^{(1)} z_1 + \Upsilon_0^{(1)} z_2.$$

Following the argument for the case $m \geq 1$, we arrive at

$$8ih^2\mu k_3^2 \left(|z_1|^2 + |z_2|^2\right) = -2i\mu|\kappa|(\sin\delta) \int\limits_{\partial\varpi(a)} \left(h^2|U_1|^2 + h^2|U_2|^2 + |U_3|^2\right) ds,$$

from which we conclude that $z_1 = z_2 = 0$. Hence, (22.27) has only the zero solution, so

$$0 \neq \left(\frac{h^2\mu}{a^2}\right)^2 \left(\frac{a}{h^2}\right) \begin{vmatrix} \cos\theta & 0 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} a_0 & b_0 \\ g_0 & h_0 \end{vmatrix} = \frac{h^2\mu^2}{a^3} \cos\theta \left(a_0 h_0 - b_0 g_0\right),$$

which implies that

$$a_0 h_0 - b_0 g_0 \neq 0.$$

It remains to show that $f_0 \neq 0$. First, we note that

$$\begin{aligned} H_0'(k_3 a) &= -H_1(k_3 a), \\ H_0''(k_3 a) &= -H_1'(k_3 a) = -H_0(k_3 a) + (1/(k_3 a))H_1(k_3 a) \\ &= -H_0(k_3 a) - (1/(k_3 a))H_0'(k_3 a), \end{aligned}$$

which means that

$$\begin{aligned} f_0 &= -k_3^2 a^2 \left[-H_0(k_3 a) - (1/(k_3 a))H_0'(k_3 a)\right] - k_3 a(\kappa - a)H_0'(k_3 a) \\ &= k_3^2 a^2 H_0(k_3 a) - k_3 a(\kappa - a - 1)H_0'(k_3 a). \end{aligned}$$

Suppose that $f_0 = 0$; that is,

$$k_3 a H_0(k_3 a) = (\kappa - a - 1)H_0'(k_3 a). \tag{22.28}$$

Then we also have $\bar{f}_0 = 0$, which implies that

$$k_3 a \bar{H}_0(k_3 a) = (\bar{\kappa} - a - 1)\bar{H}_0'(k_3 a). \tag{22.29}$$

By (22.28), (22.29), and (22.16),

$$\begin{aligned} k_3 a \left[H_0'(k_3 a)\bar{H}_0(k_3 a) - \bar{H}_0'(k_3 a)H_0(k_3 a)\right] & \\ = (\bar{\kappa} - a - 1)\left|H_0'(k_3 a)\right|^2 - (\kappa - a - 1)\left|H_0'(k_3 a)\right|^2 & \\ = -2i|\kappa|(\sin\delta)\left|H_0'(k_3 a)\right|^2. & \end{aligned}$$

Also, as shown in [Ur73],

$$H_0'(k_3 a)\bar{H}_0(k_3 a) - \bar{H}_0'(k_3 a)H_0(k_3 a) = \frac{4i}{\pi k_3 a},$$

so

$$\frac{4i}{\pi} = -2i|\kappa|(\sin\delta)\left|H_0'(k_3 a)\right|^2;$$

that is,

$$2 + \pi|\kappa|(\sin\delta)\left|H_0'(k_3 a)\right|^2 = 0.$$

But this is impossible since, given that $0 < \delta < \pi$, the left-hand side is strictly positive. Therefore, $f_0 \neq 0$, which implies that $\det P_0 \neq 0$ and, hence, that P_0 is invertible.

Thus, $M^\omega(x,y)$ can be constructed in the form (22.18) by means of (22.20). As in [Be90], it can be shown that the infinite series defining $M^\omega(x,y)$ is absolutely convergent in the region $R_x R_y > a^2$. Since $y \in \partial S$ wherever we intend to apply this modification, we have $R_y > a$, which also means that $R_x \geq a$, so the region of interest to us is a subset of $R_x R_y > a^2$.

22.4 Modified Integral Equations

Consider the modified single-layer and double-layer plate potentials

$$\begin{aligned} \left(V_M^\omega \varphi\right)(x) &= \int_{\partial S} D_M^\omega(x,y)\varphi(y)\,ds(y), \\ \left(W_M^\omega \varphi\right)(x) &= \int_{\partial S} \left[T(\partial_y)D_M^\omega(y,x)\right]^{\mathrm{T}} \varphi(y)\,ds(y), \end{aligned}$$

where $D_M^\omega(x,y)$ is as defined in the preceding section. Since $M^\omega(x,y)$ is regular in S_a^-, these potentials behave in the same way as the corresponding "unmodified" potentials when $x \in \partial S$.

Theorem 4. *(i)* $V_M^\omega \varphi, W_M^\omega \varphi \in \mathscr{B}^\omega$.

(ii) If $\varphi \in C(\partial S)$, *then* $V_M^\omega \varphi$ *and* $W_M^\omega \varphi$ *are analytic and satisfy system* (22.4) *in* $\mathbb{R}^2 \setminus (\partial S \cup \{0\})$.

(iii) If $\varphi \in C^{0,\alpha}(\partial S)$, $\alpha \in (0,1)$, *then the direct values* $V_{M0}^\omega \varphi$ *and* $W_{M0}^\omega \varphi$ *of* $V_M^\omega \varphi$ *and* $W_M^\omega \varphi$ *on* ∂S *exist (the latter as principal value), the functions*

$$\mathscr{V}_M^{\omega+}(\varphi) = (V_M^\omega \varphi)|_{\bar{S}^+}, \quad \mathscr{V}_M^{\omega-}(\varphi) = (V_M^\omega \varphi)|_{\bar{S}^-}$$

are of class $C^\infty(S^+) \cap C^{1,\alpha}(\bar{S}^+)$ *and* $C^\infty(S^-) \cap C^{1,\alpha}(\bar{S}^-)$, *respectively, and*

$$T\mathscr{V}_M^{\omega+}(\varphi) = (W_{M0}^{\omega *} + \tfrac{1}{2} I)\varphi, \quad T\mathscr{V}_M^{\omega-}(\varphi) = (W_{M0}^{\omega *} - \tfrac{1}{2} I)\varphi$$

on ∂S, *where* $W_{M0}^{\omega *}$ *is the adjoint of* W_{M0}^ω *and* I *is the identity operator.*

(iv) If $\varphi \in C^{1,\alpha}(\partial S)$, $\alpha \in (0,1)$, *then the functions*

$$\mathscr{W}_M^{\omega+}(\varphi) = \begin{cases} (W_M^\omega \varphi)|_{S^+} & \text{in } S^+, \\ (W_{M0}^\omega - \frac{1}{2} I)\varphi & \text{on } \partial S, \end{cases} \quad \mathscr{W}_M^{\omega-}(\varphi) = \begin{cases} (W_M^\omega \varphi)|_{S^-} & \text{in } S^-, \\ (W_{M0}^\omega + \frac{1}{2} I)\varphi & \text{on } \partial S \end{cases}$$

are of class $C^\infty(S^+) \cap C^{1,\alpha}(\bar{S}^+)$ *and* $C^\infty(S^-) \cap C^{1,\alpha}(\bar{S}^-)$, *respectively, and we have* $T\mathscr{W}_M^{\omega+}(\varphi) = T\mathscr{W}_M^{\omega-}(\varphi)$ *on* ∂S.

These properties are typical of potentials (see, for example, [Co90]). In the proof of (ii) we take into account the fact that $M^\omega(x,y)$ is singular at the origin. Also, symmetry (22.17) ensures that the modified double-layer potential satisfies (22.4).

Using Theorem 2 and the modified potentials, we are able to formulate boundary integral equations of the second kind representing ($D^{\omega-}$), ($N^{\omega-}$), and ($R^{\omega-}$), each of which is uniquely solvable for every value of the oscillation frequency ω.

Theorem 5. *If* $\mathscr{R} \in C^{1,\alpha}(\partial S)$, $\alpha \in (0,1)$, *then the unique solution of* ($D^{\omega-}$) *is given by*

$$u = \mathscr{W}_M^{\omega-}(\varphi), \tag{22.30}$$

where $\varphi \in C^{1,\alpha}(\partial S)$ *is the unique solution of the integral equation*

$$\left(W_{M0}^\omega + \tfrac{1}{2} I\right) \varphi = \mathscr{R}. \tag{22.31}$$

Proof. Seeking the solution of ($D^{\omega-}$) in the form (22.30) leads to the integral equation (22.31) for the unknown density φ. Consider the homogeneous adjoint equation

$$\left(W_{M0}^{\omega *} + \tfrac{1}{2} I\right) \psi = 0, \tag{22.32}$$

and let $v_1 = \mathscr{V}_M^{\omega +}(\psi)$, where ψ satisfies (22.32). Then $Tv_1|_{\partial S} = 0$; also, by (22.14), we have $Tv_1 + Kv_1 = 0$ on $\partial\varpi(a)$, so, by Theorem 2, $v_1 = 0$ in $S_a^- \cap S^+$. By continuity, $v_1|_{\partial S} = V_{M0}^{\omega}\psi = 0$, which implies that $v_2 = \mathscr{V}_M^{\omega -}(\psi)$ is a solution of the homogeneous exterior Dirichlet problem. By Theorem 1(i), $v_2 = 0$ in $\bar{S}^-$, so

$$Tv_2|_{\partial S} = T\mathscr{V}_M^{\omega -}(\psi)|_{\partial S} = \left(W_{M0}^{\omega *} - \tfrac{1}{2} I\right) \psi = 0. \tag{22.33}$$

Combining (22.32) and (22.33) yields $\psi = 0$. Therefore, according to the Fredholm alternative (the applicability of which was established in [Co90]), equation (22.31) has a unique solution $\varphi \in C^{1,\alpha}(\partial S)$.

Theorem 6. *If $\mathscr{S} \in C^{0,\alpha}(\partial S)$, $\alpha \in (0,1)$, then the unique solution of $(N^{\omega -})$ is given by*

$$u = \mathscr{V}_M^{\omega -}(\varphi), \tag{22.34}$$

where $\varphi \in C^{0,\alpha}(\partial S)$ is the unique solution of the integral equation

$$\left(W_{M0}^{\omega *} - \tfrac{1}{2} I\right) \varphi = \mathscr{S}. \tag{22.35}$$

Proof. If the solution of $(\mathrm{N}^{\omega -})$ is sought in the form (22.34), then we arrive at the integral equation (22.35). As above, we consider the homogeneous adjoint equation

$$\left(W_{M0}^{\omega} - \tfrac{1}{2} I\right) \psi = 0. \tag{22.36}$$

We define $v_1 = \mathscr{W}_M^{\omega +}(\psi)$, where ψ is a solution of (22.36). Then $v_1|_{\partial S} = 0$, so, by (22.14), (22.15), and Theorem 2, $v_1 = 0$ in $S_a^- \cap S^+$; consequently,

$$Tv_1|_{\partial S} = T\mathscr{W}_M^{\omega +}(\psi)|_{\partial S} = T\mathscr{W}_M^{\omega -}(\psi)|_{\partial S} = 0.$$

Hence, $v_2 = \mathscr{W}_M^{\omega -}(\psi)$ is a solution of the homogeneous exterior Neumann problem, and Theorem 1(i) implies that $v_2 = 0$ in $\bar{S}^-$. In particular,

$$v_2|_{\partial S} = \left(W_{M0}^{\omega} + \tfrac{1}{2} I\right) \psi = 0, \tag{22.37}$$

so from (22.36) and (22.37) it follows that $\psi = 0$. As in the proof of Theorem 5, the assertion now follows from the Fredholm alternative.

Theorem 7. *If $Im(\sigma)$ is positive semidefinite and $\mathscr{G} \in C^{0,\alpha}(\partial S)$, $\alpha \in (0,1)$, then the unique regular solution of $(R^{\omega -})$ is given by*

$$u = \mathscr{V}_M^{\omega -}(\varphi), \tag{22.38}$$

where $\varphi \in C^{0,\alpha}(\partial S)$ is the unique solution of

$$(W_{M0}^{\omega *} + \sigma V_{M0}^{\omega} - \tfrac{1}{2} I)\varphi = \mathscr{G}. \tag{22.39}$$

Proof. Assuming that the solution of $(\mathrm{R}^{\omega -})$ is of the form (22.38) leads to the integral equation (22.39) for the unknown density. Once again, we consider the homogeneous adjoint equation

$$(W_{M0}^{\omega} + V_{M0}^{\omega}\sigma - \tfrac{1}{2} I)\psi = 0 \tag{22.40}$$

and introduce the function

$$v_1 = \mathscr{W}_M^{\omega +}(\psi) + \mathscr{V}_M^{\omega +}(\sigma\psi),$$

where ψ is a solution of (22.40). Then $v_1|_{\partial S} = 0$, so, by (22.14), (22.15), and Theorem 2, $v_1 = 0$ in $S_a^- \cap S^+$; therefore,

$$Tv_1|_{\partial S} = N_{M0}^{\omega}\psi + (W_{M0}^{\omega *} + \tfrac{1}{2} I)(\sigma\psi) = 0. \tag{22.41}$$

Consider the function

$$v_2 = \mathscr{W}_M^{\omega -}(\psi) + \mathscr{V}_M^{\omega -}(\sigma\psi).$$

Taking (22.40) and (22.41) into account, we find that

$$\begin{aligned} Tv_2|_{\partial S} + \sigma v_2|_{\partial S} &= N_{M0}^{\omega}\psi + (W_{M0}^{\omega *} - \tfrac{1}{2} I)(\sigma\psi) + \sigma\left[(W_{M0}^{\omega} + \tfrac{1}{2} I)\psi + V_{M0}^{\omega}(\sigma\psi)\right] \\ &= -\sigma\psi + \sigma\psi = 0; \end{aligned}$$

hence, v_2 is a solution of the homogeneous exterior Robin problem. By Theorem 1(ii), $v_2 = 0$ in $\bar{S}^-$. In particular,

$$v_2|_{\partial S} = (W_{M0}^{\omega} + \tfrac{1}{2} I)\psi + V_{M0}^{\omega}(\sigma\psi) = 0,$$

which, combined with (22.40), yields $\psi = 0$. Thus, by the Fredholm alternative, (22.39) has a unique $C^{0,\alpha}$-solution.

22.5 Numerical Example

In this section, we illustrate the power and efficiency of the boundary integral equation method by computing the solution of two interior Robin problems for a material with scaled parameters

$$\lambda = 2, \quad \mu = 3, \quad h = 1, \quad \rho = \frac{1}{300}, \quad \omega = 150,$$

and boundary conditions described by

$$\mathscr{G}(x) = (10, 10, 10)^{\mathrm{T}}, \quad \sigma(x) = 10E_3,$$

where E_3 is the identity 3×3 matrix. As shown in [ThCo10], the unique solution of this type of problem is

$$u = \mathscr{V}^{\omega+}(\mathscr{G} - \sigma\varphi) - \mathscr{W}^{\omega+}(\varphi),$$

where the density φ satisfies the boundary integral equation

$$\left(W_0^{\omega} + V_0^{\omega}\sigma + \tfrac{1}{2}I\right)\varphi = V_0^{\omega}\mathscr{G}$$

and V_0^{ω}, W_0^{ω}, $\mathscr{V}^{\omega+}$, and $\mathscr{W}^{\omega+}$ are defined in [ThCo10].

All integrals (including those with weakly singular integrands or defined as Cauchy principal values) have been computed with Mathematica's internal numerical integration schemes. The boundary basis functions (elements) are piecewise cubic Hermite splines (cubic polynomials joined with C^1 continuity). However, no continuity has been provided at any of the domain corners. The computation makes use of a total of 28 collocation points for the full boundary.

Figures 22.1–22.3 show the graphs of the functions u_i when the domain is a half circle.

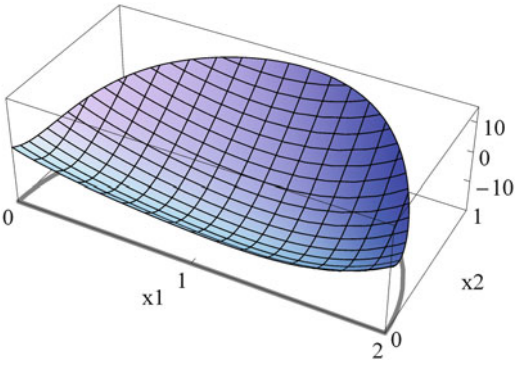

Fig. 22.1 Graph of u_1 for a half circle

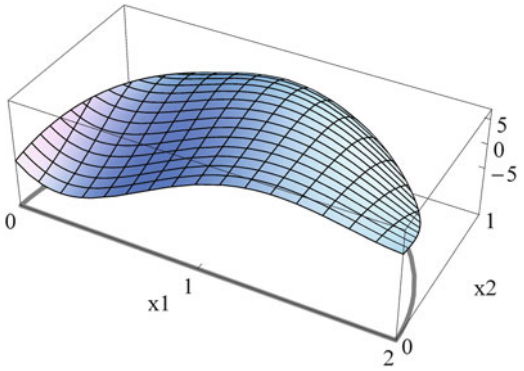

Fig. 22.2 Graph of u_2 for a half circle

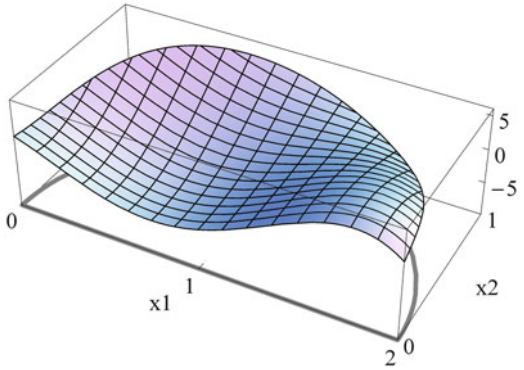

Fig. 22.3 Graph of u_3 for a half circle

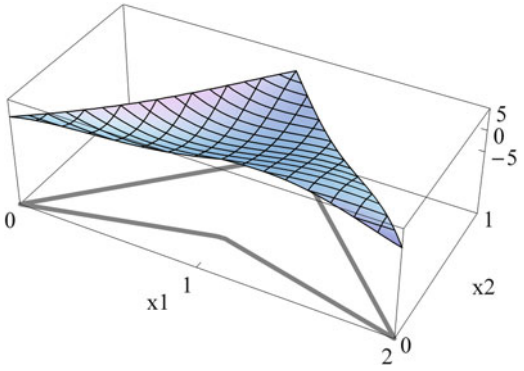

Fig. 22.4 Graph of u_1 for a wing-shaped domain

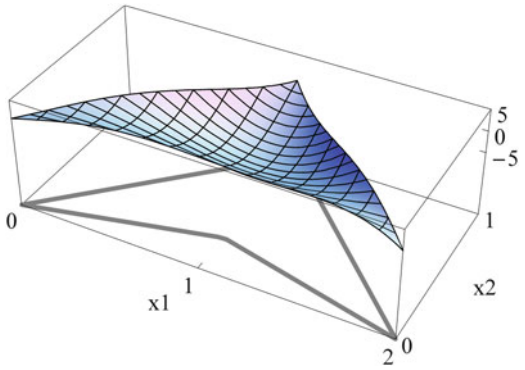

Fig. 22.5 Graph of u_2 for a wing-shaped domain

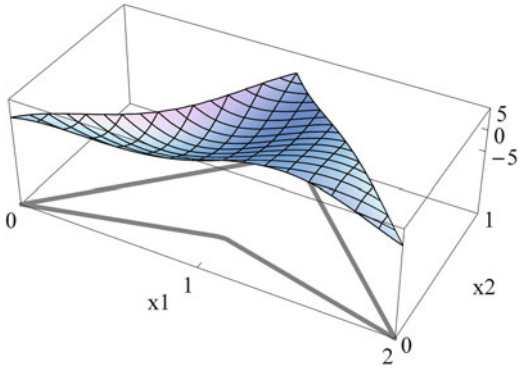

Fig. 22.6 Graph of u_3 for a wing-shaped domain

Figures 22.4–22.6 show the graphs of the u_i for a wing-shaped domain.

References

[Be90] Bencheikh, L.: Modified fundamental solutions for the scattering of elastic waves by a cavity. Q. J. Mech. Appl. Math. **43**, 57–73 (1990)

[Co90] Constanda, C.: A Mathematical Analysis of Bending of Plates with Transverse Shear Deformation. Longman, Harlow (1990)

[Co98] Constanda, C.: Radiation conditions and uniqueness for stationary oscillations in elastic plates. Proc. Am. Math. Soc. **126**, 827–834 (1998)

[ScCo93] Schiavone, P., Constanda, C.: Oscillation problems in thin plates with transverse shear deformation. SIAM J. Appl. Math. **53**, 1253–1263 (1993)

[ThCo97] Thomson, G.R., Constanda, C.: On stationary oscillations in bending of plates. In: Constanda, C., Saranen, J., Seikkala, S. (eds.) Integral Methods in Science and Engineering Vol. 1: Analytic Methods, pp. 190–194. Longman, Harlow (1997)

[ThCo98] Thomson, G.R., Constanda, C.: Area potentials for thin plates. An. Stiint. Al.I. Cuza Univ. Iasi Sect. Ia Mat. **44**, 235–244 (1998)
[ThCo99] Thomson, G.R., Constanda, C.: Scattering of high frequency flexural waves in thin plates. Math. Mech. Solids **4**, 461–479 (1999)
[ThCo09b] Thomson, G.R., Constanda, C.: A matrix of fundamental solutions in the theory of plate oscillations. Appl. Math. Lett. **22**, 707–711 (2009)
[ThCo09a] Thomson, G.R., Constanda, C.: Integral equation methods for the Robin problem in stationary oscillations of elastic plates. IMA J. Appl. Math. **74**, 548–558 (2009)
[ThCo10] Thomson, G.R., Constanda, C.: The direct method for harmonic oscillations of elastic plates with Robin boundary conditions. Math. Mech. Solids **16**, 200–207 (2010)
[ThCo11] Thomson, G.R., Constanda, C.: Uniqueness of solution for the Robin problem in high-frequency vibrations of elastic plates. Appl. Math. Lett. **24**, 577–581 (2011)
[ThCo12] Thomson, G.R., Constanda, C.: Uniqueness of analytic solutions for stationary plate oscillations in an annulus. Appl. Math. Lett. **25**, 1050–1055 (2012)
[Ur73] Ursell, F.: On the exterior problems of acoustics. Math. Proc. Camb. Philos. Soc. **74**, 117–125 (1973)

Chapter 23
A Genuine Analytical Solution for the SN Multi-group Neutron Equation in Planar Geometry

F.K. Tomaschewski, C.F. Segatto, and M.T. Vilhena

23.1 Introduction

The analytical solution program for the time-dependent neutron transport equation has undergone a significant evolution since the work of Case [CaZw67], where the one-dimensional stationary problem in a slab was solved analytically. There exists a relevant literature concerning the issue of solving the time-dependent neutron equation in a planar geometry for an unbounded domain. We mention the works of Ganapol and Filippone [GaFi82] , Ganapol and Pomraning [GaPo83], Ganapol [Ga86], Ganapol and Matsumoto [GaMa86], and Abdul [Ab06]. On the other hand, regarding the literature for bounded domains, we cite the works of Windhofer and Pucker [WiPu85], Warsa and Prinja [WaPr98], Oliveira et al. [OlCaVi02], [OlEtAl02], El-Wakil et al. [ElDeSa05], [ElDeSa06], Türeci et al. [TuGuTe07], Türeci and Türeci [TuTu07], Hadad et al. [HaPiAy08], Coppa et al. [CoEtAl08], [CoDuRa10], and Cargo and Samba [Ca10].

Recently the double Laplace transform S_N method ($DLTS_N$), which solves the S_N time-dependent transport equation for mono-energetic neutrons, either for bounded and unbounded planar geometry domain [SeViGo08], [SeViGo10] was developed. In the present work we extend this solution for this sort of problem, now assuming a neutron multi-group model. To this end, we apply the double Laplace transform techniques in the multi-group S_N equation in the time and spatial variable. Next, we solve the resulting algebraic equation for the transformed angular flux and finally, we determine the solution by Laplace transform inversion of the transformed solution by the LTS_N technique in the spatial variable and by the Laplace transform inversion theorem for the time variable. Due to the analytical character of the solution, expressed in matrix integral form, we obtain the integration constants of the

F.K. Tomaschewski • C.F. Segatto (✉) • M.T. Vilhena
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil,
e-mail: fernandasls_89@hotmail.com; csegatto@pq.cnpq.br; vilhena@pq.cnpq.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_23,

solution for the bounded domain, upon applying the slab boundary condition. For the unbounded domain, we replace this condition by the boundedness of the angular flux at infinity. At this point, we mention that this procedure leads to an analytical representation of the solution in line matrix integral form, which is then evaluated by the Stehfest numerical scheme [St70] and appears from a computational point of view a suitable approach, to work out numerically this kind of problem with a prescribed accuracy either for large or small times [SeViGo08], [SeViGo10]. Finally, we report on numerical results attained by this methodology specialized for the two-group energy as well as asymptotic behavior of the scalar flux for large times.

23.2 Time-Dependent Multi-group Transport Equation for Heterogeneous Domain

In order to construct the general solution, let us initially consider the following multi-group, time-dependent S_N neutron transport problem in a multi-layered slab, depicted in Fig. 23.1

$$\frac{1}{v_g}\frac{\partial}{\partial t}\psi^k_{n,g}(t,x)+\mu_n\frac{\partial}{\partial x}\psi^k_{n,g}(t,x)+\sigma^k_{tg}\psi^k_{n,g}(t,x) = \sum_{g'=1}^{G}\frac{\sigma^k_{sg'g}}{2}\sum_{i=1}^{N}\psi^k_{i,g'}(t,x)w_i+S^k_{n,g}(t,x), \quad (23.1)$$

for $g=1:G$ and $k=1:K$, with the initial condition

$$\psi^k_{n,g}(0,x)=\phi^k_{n,g}(x),$$

subject to the boundary conditions

$$\psi^1_{n,g}(t,x_0)=f_{n,g}(t,x), \quad t>0, \quad n=1:N/2,$$
$$\psi^K_{n,g}(t,x_K)=g_{n,g}(t,x), \quad t>0, \quad n=(N/2+1):N,$$

and the interface angular flux continuity condition

$$\psi^k_{n,g}(t,x_k)=\psi^{k+1}_{n,g}(t,x_k), \quad k=1:K-1.$$

Here, in standard notation, $\psi^k_{n,g}(t,x,)$ is the angular flux for the gth group in the kth region at position x travelling in the discrete direction μ_n at time t; v_g is the mean velocity for the gth group; σ^k_{tg} is the total differential cross section for the gth group in the kth region and $\sigma^k_{s,g'g}$ is the differential scattering cross section (from group g'

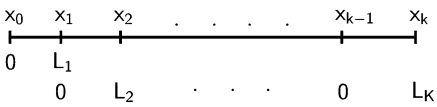

Fig. 23.1 The multilayered slab domain

into group g) in the kth region, μ_n and w_n are, respectively, the roots and weights of the Gaussian quadrature scheme.

Applying the Laplace transform technique in (23.1) in the time variable for a generic slab (kth slab), we come out with

$$\mu_n \frac{d}{dx}\Psi^k_{n,g}(p,x) + \sigma^{pk}_{tg}\Psi^k_{n,g}(p,x) = \sum_{g'=1}^{G} \frac{\sigma^k_{sg'g}}{2} \sum_{i=1}^{N} \Psi^k_{i,g'}(p,x)w_i + Q^k_{n,g}(p,x), \quad (23.2)$$

subject to the boundary conditions

$$\Psi^1_{n,g}(p,0) = F_{n,g}(p), \quad \Psi^K_{\frac{n}{2}+1,g}(p,L_K) = G_{n,g}(p), \quad n = 1 : \frac{N}{2}.$$

Here, $\Psi^k_{n,g}(p,x)$ denotes the Laplace transform in the time variable ($t \to s$) of $\psi^k_{n,g}(t,x)$ with σ^{pk}_{tg}, and $Q^k_{n,g}(p,x)$ is

$$\sigma^{pk}_{tg} = \sigma^k_{tg} + \frac{p}{v_g}, \quad Q^k_{n,g}(p,x) = \frac{1}{v_g}\phi^k_{n,g}(x) + \bar{S}^k_{n,g}(p,x).$$

Casting (23.2) in matrix form, we arrive at

$$\frac{d}{dx}\Psi^k(p,x) - \mathbf{A}(p)\Psi^k(p,x) = \mathbf{Q}^k(p,x), \quad (23.3)$$

where $\mathbf{A}(p)$ is a matrix of order NG with entries

$$a_{i+N(g-1),j+N(g'-1)} = \begin{cases} \dfrac{\sigma^k_{sg'g}w_j}{2\mu_i} - \dfrac{\sigma^{pk}_{tg}}{\mu_i} & \text{if } i = j \text{ and } g = g', \\ \dfrac{\sigma^k_{sg'g}w_j}{2\mu_i} & \text{if } i \neq j \text{ and } g \neq g', \end{cases}$$

for $i = 1 : N$ and $j = 1 : N$ and for g and g' ranging from 1 to G. Also, the NG-dimensional vectors $\mathbf{Q}^k(p,x)$ and $\Psi^k(p,x)$ are defined by

$$\mathbf{Q}^k(p,x) = \begin{bmatrix} \mathbf{Q}^k_{1,g}(p,x) \\ \mathbf{Q}^k_{2,g}(p,x) \end{bmatrix} = (\frac{Q^k_1}{\mu_1}, \ldots, \frac{Q^k_1}{\mu_k}, \ldots, \ldots, \frac{Q^k_G}{\mu_1}, \ldots, \frac{Q^k_G}{\mu_N})^T$$

and

$$\Psi^k(p,x) = \begin{bmatrix} \Psi^k_{1,g}(p,x) \\ \Psi^k_{2,g}(p,x) \end{bmatrix} = \left(\Psi^k_{1,1},\ldots,\Psi^k_{N,1},\ldots,\Psi^k_{1,G},\ldots,\Psi^k_{N,G}\right)^T,$$

subject to the boundary conditions

$$\Psi^1_{1,g}(p,0) = \begin{bmatrix} F_{1,g} \\ \vdots \\ F_{N/2,g} \end{bmatrix}, \quad \Psi^K_{2,g}(p,L_K) = \begin{bmatrix} G_{N/2+1,g} \\ \vdots \\ G_{N,g} \end{bmatrix}.$$

Applying the $\mathrm{LTS_N}$ method [GoSeVi00], [Se99] in (23.3), we find the solution

$$\Psi^k(p,x) = \mathbf{X}[\mathbf{E}^{k+}(p,x-L_k) + \mathbf{E}^{k-}(p,x)]\zeta^k + \mathbf{H}^k(p,x), \tag{23.4}$$

where $\mathbf{E}^{k+}(p,x)$ and $\mathbf{E}^{k-}(p,x)$ are the diagonal matrix functions

$$\mathbf{E}^{k+}(p,x) = \begin{cases} e^{d_i(p)x} & \text{if } i = 1:N/2,\ (d_i > 0), \\ 0 & \text{if } i = N/2+1:N,\ (d_i < 0), \end{cases}$$

$$\mathbf{E}^{k-}(p,x) = \begin{cases} 0 & \text{if } i = 1:N/2,\ (d_i > 0), \\ e^{d_i(p)x} & \text{if } i = N/2+1:N,\ (d_i < 0), \end{cases}$$

and $d_i(p)$, $i = 1:N$ are the eigenvalues in decreasing order and $\mathbf{X}(p)$ is the eigenvector matrix of $\mathbf{A}(p)$. Furthermore, the particular solution $\mathbf{H}^k(p,x)$ has the form

$$\begin{aligned} \mathbf{H}^k(p,x) &= \mathbf{H}^{k+}(p,x) \ + \ \mathbf{H}^{k-}(p,x) \\ &= \int_{L_k}^{x} \mathbf{B}^{k+}(p,x-\xi)\mathbf{Q}(p,\xi)d\xi + \int_0^x \mathbf{B}^{k-}(p,x-\xi)\mathbf{Q}(p,\xi)d\xi, \end{aligned}$$

where

$$\mathbf{B}^{k+}(p,x) = \mathbf{X}\mathbf{E}^{k+}(p,x)\mathbf{X}^{-1},$$
$$\mathbf{B}^{k-}(p,x) = \mathbf{X}\mathbf{E}^{k-}(p,x)\mathbf{X}^{-1}.$$

To find the solution for a generic slab (kth), we perform the Laplace inversion in the time coordinate of the solution given by (23.4), using the usual definition of Laplace transform inversion; that is,

$$\psi^k_g(t,x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \Psi^k_g(p,x)e^{pt}dp. \tag{23.5}$$

To determine the solution for the multi-layered slab we evaluate the integration constants by solving an algebraic linear system resulting from the application of the boundary condition and continuity of angular flux at interface. To this point, we emphasize that based on the good results achieved for the monoenergetic problem for small and large times [SeViGo08], [SeViGo10], we evaluate the line integral solution (23.5) and (23.4) by the Gaver–Stehfest numerical scheme [St70], defined by

$$f(t) = \frac{\ln 2}{t} \sum_{i=1}^{N} V_i f\left(\frac{\ln 2}{t}\right),$$

where N is an even number and V_i is of the form

$$V_i = (-1)^{N/2+i} \sum_{k=\left[\frac{i+1}{2}\right]}^{Min(i,N/2)} \frac{k^{N/2}(2k)!}{(N/2-k)!(k)!(k-1)!(i-k)!(2k-i)!}.$$

To complete the analysis of a solution for bounded and unbounded domains, we replace the far end boundary condition of the slab by the boundedness condition of the angular flux at infinity.

$$\lim_{L_K \to \infty} \Psi_{n,g}^{K}(p, L_K) = 0.$$

From this assumption, it turns out that the $\mathrm{LTS_N}$ solution for unbounded domains for $0 \leq x \leq \infty$, simplifies to

$$\Psi_g^k(p,x) = \mathbf{X}\mathbf{E}^{k-}(p,x)\zeta^k + \mathbf{H}^{k-}(p,x).$$

Therefore, a similar procedure leads to the following solution for a generic slab (kth slab) of the problem for an unbounded domain.

$$\psi_g^k(t,x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Psi_g^k(p,x) e^{pt}\, dp. \tag{23.6}$$

The angular flux $\psi_g^k(t,x)$ is now given by (23.6). At this point, it is noteworthy that the proposed methodology can be applied to other types of boundary conditions, for instance, reflexive ones.

23.3 Numerical Results

To show the aptness of the proposed solution for the multi-group neutron transport equation considering bounded as well as unbounded planar geometry domains, let

Table 23.1 Nuclear parameters

		1th region $L_1 = 1.2549$ cm		2th region $L_2 = 7.8663$ cm	
		Group 1	Group 2	Group 1	Group 2
v_g	(cm/s)	10^7	2×10^5	10^7	2×10^5
σ_{s1g}	(cm^{-1})	0.75	0.30	0.90	0.20
σ_{s2g}	(cm^{-1})	0.10	0.99	0.05	0.08
σ_{tg}	(cm^{-1})	0.90	1.50	1.00	1.20
S_g	(cm^{-1})	1	0	0	0

Table 23.2 Asymptotic behavior of the fast and the thermal scalar flux for times ranging from 10^{-6} to 100 s

	$x = 0$		$x = 1.2549$		$x = 9.1212$	
t	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
10^{-6}	4.294645	1.094369	3.210277	1.030643	6.760881(−1)	9.542958(−1)
10^{-5}	5.159098	2.142995	3.962320	1.621230	4.228770(−1)	5.903738(−1)
10^{-4}	5.489941	2.878981	4.229715	2.184906	2.267534(−1)	1.331771(−1)
10^{-3}	5.490147	2.879403	4.229878	2.180869	2.263873(−1)	1.323116(−1)
10^{-2}	5.490067	2.879351	4.229798	2.180824	2.263641(−1)	1.322894(−1)
10^{-1}	5.490092	2.879373	4.229715	2.180049	2.267534(−1)	1.323030(−1)
1	5.490138	2.879414	4.229895	2.180883	2.263849(−1)	1.323030(−1)
10	5.490162	2.879411	4.229898	2.180880	2.263866(−1)	1.323017(−1)
100	5.490153	2.879413	4.229877	2.879414	2.263916(−1)	1.323047(−1)
LTS_{100}	5.490138	2.879397	4.229869	2.180865	2.263851(−1)	1.323009(−1)

us consider a time-dependent, two-group, two-slab problem with constant source of unitary intensity emitting fast neutrons in the first slab. In this problem we consider reflexive boundary conditions either for $x = 0$ and $x = L_2$.

Problem 1. To show the expected asymptotic behavior of the fast and thermal scalar fluxes for large times, we consider the problem with the parameters displayed in Table 23.1. The numerical results achieved for the fast and thermal fluxes are displayed in Table 23.2.

Analyzing the results attained for the fast and thermal scalar flux depicted in Table 23.2, we observe that the obtained solution gets closer to the stationary one, when we increase the time from 10^{-6} to 10^2 s. We reinforce this argument by noticing that the stationary solution was obtained by the exact LTS_N method, which converges to the exact solution when N goes to infinity. Therefore, the LTS_{100} results adopted for comparison can be considered almost exact.

Problem 2. Now, we show the expected asymptotic behavior of the fast and thermal scalar flux for the same problem with parameters displayed in Table 23.1, by increasing the slab thickness from 40 to 70 cm. The results encountered are shown in Table 23.3 for times of $t = 10^{-2}$ and $t = 50$ s.

Table 23.3 Asymptotic behavior of the fast and the thermal scalar flux obtained by increasing the slab thickness L_2 from 40 to 70 cm

	$t = 10^{-2}$ s					
	$x = 0$		$x = 1.2549$		$x = 30.$	
L	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
40	5.486544	2.877067	4.225630	2.178195	9.892873(−6)	5.783470(−6)
50	5.486545	2.877067	4.225631	2.178195	9.892871(−6)	5.780980(−6)
60	5.486545	2.877067	4.225631	2.178195	9.890461(−6)	5.783470(−6)
70	5.486545	2.877067	4.225631	2.178195	9.891659(−6)	5.784084(−6)
$TLTS_{100}$	5.486546	2.877070	4.225631	2.178195	9.889937(−6)	5.781673(−6)
	$t = 50$ s					
	$x = 0$		$x = 1.2549$		$x = 30.$	
L	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
50	5.486593	2.877099	4.225674	2.178222	9.893950(−6)	5.783779(−6)
60	5.486593	2.877099	4.225674	2.178222	9.893950(−6)	5.783781(−6)
70	5.486592	2.877099	4.225674	2.178222	9.893949(−6)	5.783782(−6)
$TLTS_{100}$	5.487100	2.877097	4.225672	2.178222	9.893951(−6)	5.783782(−6)

By similar reasoning and confirmed by the results in Table 23.3, one approaches the asymptotic behavior of the solutions for the fast and thermal scalar fluxes, respectively, upon increasing the slab thickness from 40 to 70 cm. Therefore, from the previous observed asymptotic behavior combined with the proved convergence of the LTS_N solution in the limit N goes to infinity, we present the results for the fast and the thermal scalar flux for the discussed problem. In Tables 23.4 and 23.5, we show the fast and thermal scalar fluxes, as a function of position and time, which are also consistent with the asymptotic behaviors discussed for the fast and thermal scalar fluxes.

23.4 Conclusion

From the presented analysis, we are confident that the novel contribution of this work is the determination of an analytical solution for a multi-group time-dependent S_N neutron transport equation in planar geometry for bounded and unbounded domains. To the best of our knowledge this type of solution is new in literature. It is worth mentioning that this solution is unique for bounded and unbounded planar geometry domains. Notice that we need to evaluate the integration constants for the bounded domain problem only, applying the far end slab boundary condition, whereas for the unbounded domain problem, we assume the boundedness of the angular flux at infinity.

Further, we emphasize that increasing the number of energy groups for this kind of problem does not impose restrictions on the generality of the discussed solution.

Table 23.4 $TLTS_{100}$ solution for the fast and the thermal scalar fluxes obtained for the slab

Fast group

t \ x (s) (cm)	0.	0.62745	1.2549	3.877	6.4991	9.1212
10^{-6}	4.294644	4.074190	3.210277	9.524878(−1)	6.909226(−1)	6.760880(−1)
2×10^{-6}	4.705693	4.478853	3.598998	1.114940	6.393075(−1)	5.647808(−1)
5×10^{-6}	4.964300	4.726437	3.815582	1.185832	6.103102(−1)	4.997250(−1)
10^{-5}	5.159104	4.908164	3.962325	1.187851	5.496722(−1)	4.228770(−1)
2×10^{-5}	5.352904	5.091379	4.115456	1.190888	4.739069(−1)	3.256572(−1)
5×10^{-5}	5.478663	5.211895	4.220056	1.198422	4.096384(−1)	2.385719(−1)
10^{-4}	5.489940	5.222791	4.229714	1.199245	4.014490(−1)	2.267533(−1)
5×10^{-4}	5.490123	5.222961	4.229854	1.199193	4.011776(−1)	2.263863(−1)
10^{-3}	5.490146	5.222984	4.229878	1.199199	4.011795(−1)	2.263872(−1)
10^{-2}	5.490066	5.222903	4.229797	1.199139	4.011470(−1)	2.263641(−1)
10^{-1}	5.490091	5.222936	4.229842	1.199194	4.011793(−1)	2.263887(−1)
1	5.490163	5.223001	4.229895	1.199198	4.011767(−1)	2.263848(−1)

Thermal group

t \ x (s) (cm)	0.	0.62745	1.2549	3.877	6.4991	9.1212
10^{-6}	1.094369	1.084537	1.030642	9.576160(−1)	9.543845(−1)	9.542957(−1)
2(−6)	1.247224	1.225301	1.107157	9.243359(−1)	9.062088(−1)	9.042936(−1)
5(−6)	1.664906	1.606986	1.335072	8.529935(−1)	7.797904(−1)	7.674408(−1)
10^{-5}	2.142995	2.029739	1.621230	7.738424(−1)	6.191981(−1)	5.903739(−1)
2×10^{-5}	2.588794	2.439552	1.938144	7.041024(−1)	4.233843(−1)	3.683176(−1)
5×10^{-5}	2.856116	2.693800	2.159705	6.915387(−1)	2.554751(−1)	1.623655(−1)
10^{-4}	2.878981	2.716038	2.180490	6.934071(−1)	2.347754(−1)	1.331771(−1)
5(−4)	2.879387	2.716431	2.180854	6.933627(−1)	2.341477(−1)	1.323032(−1)
10^{-3}	2.879403	2.716448	2.180868	6.933665(−1)	2.341474(−1)	1.323020(−1)
10^{-2}	2.879350	2.716398	2.180823	6.933345(−1)	2.341297(−1)	1.322893(−1)
10^{-1}	2.879373	2.716420	2.180849	6.933633(−1)	2.341473(−1)	1.323029(−1)
1	2.879414	2.716460	2.180882	6.933677(−1)	2.341465(−1)	1.323009(−1)

In fact, the LTS_N method has shown a good performance to work with S_N neutron transport problems in planar geometry with N as large as $1,500$. Moreover, it is possible to solve this type of problem in unbounded domains, for $-\infty < x < \infty$, by the use of the Placzek lemma [CaHoPl53].

Finally, we emphasize that the character of the analytical representation of this solution for the neutron S_N transport equation is not restricted in the sense that no approximation is made along its derivation and has proven convergence, which guarantees that this solution converges to the exact one when N goes to infinity. Thus we attained an exact solution to an approximate problem and focus our future attention on solving, the multi-group S_N neutron kinetic transport equation in planar geometry.

Table 23.5 $TLTS_{100}$ solution for the fast and the thermal scalar fluxes obtained for unbounded domain

Fast Group							
t \ x (s) (cm)	0.	0.62745	1.2549	3.877	6.4991	9.1212	30.
10^{-6}	4.199570	3.965531	3.047731	6.646077(−1)	3.889566(−1)	3.731818(−1)	3.730109(−1)
2×10^{-6}	4.552406	4.310592	3.375669	7.530343(−1)	2.492995(−1)	1.590168(−1)	1.464716(−1)
5×10^{-6}	4.785360	4.531383	3.581930	8.383309(−1)	2.256507(−1)	7.152805(−2)	2.236927(−2)
10^{-5}	4.993262	4.734398	3.770805	9.251912(−1)	2.537540(−1)	7.564771(−2)	1.232125(−2)
2×10^{-5}	5.238615	4.975316	3.995680	1.043153	2.974433(−1)	8.779042(−2)	6.810620(−3)
5×10^{-5}	5.452953	5.186036	4.193910	1.165047	3.524395(−1)	1.067383(−1)	1.177180(−3)
10^{-4}	5.485226	5.217925	4.224381	1.187052	3.653462(−1)	1.127204(−1)	7.061426(−5)
5×10^{-4}	5.486603	5.219287	4.225689	1.188099	3.660901(−1)	1.131677(−1)	9.926119(−6)
10^{-3}	5.486626	5.219315	4.225713	1.188106	3.660926(−1)	1.131682(−1)	9.899780(−6)
10^{-2}	5.486546	5.219233	4.225631	1.188046	3.660640(−1)	1.131563(−1)	9.889937(−6)
10^{-1}	5.486567	5.219258	4.225673	1.188098	3.660913(−1)	1.131681(−1)	9.893540(−6)
1	5.486639	5.219326	4.225724	1.188104	3.660901(−1)	1.131669(−1)	9.892806(−6)
Thermal group							
t \ x (s) (cm)	0.	0.62745	1.2549	3.877	6.4991	9.1212	30.
10^{-6}	1.092013	1.081596	5.608741(−1)	3.765065(−1)	3.731052(−1)	3.730110(−1)	3.730109(−1)
2×10^{-6}	1.237652	1.214253	6.616690(−1)	1.680036(−1)	1.488620(−1)	1.466612(−1)	1.464716(−1)
5×10^{-6}	1.632932	1.492842	9.519494(−1)	1.204215(−1)	4.266364(−2)	2.637025(−2)	2.236943(−2)
10^{-5}	2.016400	1.857003	1.325700	2.279368(−1)	6.450320(−2)	2.481403(−2)	1.232043(−2)
2×10^{-5}	2.449617	2.289866	1.757533	4.142705(−1)	1.115362(−1)	3.443126(−2)	6.808957(−3)
5×10^{-5}	2.821428	2.658430	2.121823	6.412823(−1)	1.891409(−1)	5.597077(−2)	1.173758(−3)
10^{-4}	2.874890	2.711834	2.175944	6.847609(−1)	2.121852(−1)	6.531001(−2)	6.647717(−5)
5×10^{-4}	2.877104	2.714057	2.178227	6.868696(−1)	2.136359(−1)	6.613637(−2)	5.820284(−6)
10^{-3}	2.877121	2.714073	2.178241	6.868740(−1)	2.136364(−1)	6.613591(−2)	5.792388(−6)
10^{-2}	2.877070	2.714024	2.178195	6.868430(−1)	2.136214(−1)	6.612958(−2)	5.781673(−6)
10^{-1}	2.877087	2.714042	2.178219	6.868702(−1)	2.136361(−1)	6.613603(−2)	5.783529(−6)
1	2.877128	2.714083	2.178253	6.868754(−1)	2.136361(−1)	6.613553(−2)	5.783135(−6)

References

[Ab06] Abdul, M.A.: Chapman–Enskog-maximum entropy method on time-dependent neutron transport equation. J. Quant. Spectros. Radiat. Tran. **101**, 210–225 (2006)

[Ca10] Cargo, P., Samba, G.: Resolution of the time-dependent P_N equations by Godunov type scheme having the diffusion limit. Math. Model. Numer. Anal. **44**, 1193–1224 (2010)

[CaZw67] Case, K.M., Zweifel, P.F.: Linear Transport Theory. Addison-Wesley, Reading (1967)

[CaHoPl53] Case, K.M., Hoffmann, F., Placzek, G.: Introduction to the Theory of Neutron Diffusion, vol. 1. US Government Printing Office, Washington (1953)

[CoEtAl08] Coppa, G.G.M., Dulla, S., Peano, F., Ravetto, P.: Alternative forms of the time-dependent neutron transport equation. Progr. Nucl. Energ. **50**, 934–938 (2008)

[CoDuRa10] Coppa, G.G.M., Dulla, S., Ravetto, P.: The time-dependent P-1 model for the neutronics of multiplying systems: a review. Kerntechnik **75**, 200–205 (2010)

[ElDeSa05] El-Wakil, S.A., Degheidy, A.R., Sallah, M.: Time-dependent neutron transport in finite media using Pomraning–Eddington approximation. Ann. Nucl. Energ. **32**, 343–353 (2005)

[ElDeSa06] El-Wakil, S.A., Degheidy, A.R., Sallah, M.: Time-dependent radiation transfer with Rayleigh scattering in finite slab media. J. Quant. Radiat. Tran. **102**, 152–161 (2006)

[Ga86] Ganapol, B.D.: Solution of the one-group time dependent neutron transport equation in an infinite medium by polynomial reconstruction. Nucl. Sci. Eng. **92**, 272–279 (1986)

[GaFi82] Ganapol, B.D., Filippone W.L.: Time dependent emergent intensity from an anisotropically-scattering semi-infinite atmosphere. J. Quant. Spectros. Radiat. Tran. **27**, 15–21 (1982)

[GaMa86] Ganapol, B.D., Matsumoto, M.: Numerical evaluation of time-dependent reflected intensity from an anisotropically scattering semi-infinite atmosphere. J. Quant. Spectros. Radiat. Tran. **35**, 71–78 (1986)

[GaPo83] Ganapol, B.D., Pomraning, G.C.: The non-equilibrium Marshak wave problem: a transport theory solution. J. Quant. Spectros. Radiat. Tran. **24**, 311–320 (1983)

[GoSeVi00] Goncalves, G.A., Segatto, C.F., Vilhena, M.T.: The LTS_N particular solution in a slab for an arbitrary source and large order of quadrature. J. Quant. Radiat. Tran. **66**, 271–276 (2000)

[HaPiAy08] Hadad, K., Pirouzmand. A., Ayoobian, N.: Cellular neural networks (CNN) simulation for the TN approximation of the time dependent neutron transport equation in slab geometry. Ann. Nucl. Energ. **35**, 2313–2320 (2008)

[OlEtAl02] Oliveira, J.V.P., Cardona, A.V., Vilhena, M.T, Barros, R.C.: A semi-analytical numerical method for time-dependent radiative transfer problems in slab geometry with coherent isotropic scattering. J. Quant. Radiat. Trans. **73**, 55–62 (2002)

[OlCaVi02] Oliveira, J.V.P., Cardona A.V., Vilhena, M.T.: Solution of the one-dimensional time-dependent discrete ordinates problem in a slab by the spectral and LTS_N methods. Ann. Nucl. Energ. **29**, 13–20 (2002)

[Se99] Segatto, C.F., Vilhena, M.T., Gomes, M.G.: The one-dimensional LTS_N solution in a slab with high degree of quadrature. Ann. Nucl. Energ. **26**, 925–934 (1999)

[SeViGo08] Segatto, C.F., Vilhena, M.T., Goncalez, T.T.: An analytical solution for the time-dependent S_N transport equation in a slab. Kerntechnik **73**, 176–178 (2008)

[SeViGo10] Segatto, C.F., Vilhena, M.T., Goncalez, T.T.: An analytical solution for the one-dimensional time-dependent S_N transport equation for bounded and unbounded domains in cartesian geometry. Kerntechnik **75**, 53–57 (2010)

[St70] Stehfest, H.: Numerical inversion of Laplace transforms. Comm. ACM **13**, 47 (1970)

[TuGuTe07] Türeci, G., Güleçyüz, C., Tezcan, C.: H_N solutions of the time dependent linear neutron transport equation for a slab transport equation for slab and a sphere. Kerntechnik **72**, 66–73 (2007)

[TuTu07] Türeci, R.G., Türeci, D.: Time-dependent albedo problem for quadratic anisotropic scattering. Kerntechnik **72**, 59–65 (2007)
[WaPr98] Warsa, J.S., Prinja, A.K.: Bilinear-discontinuous numerical solution of the time-dependent transport equation in slab geometry. Ann. Nucl. Energ. **26**, 195–215 (1998)
[WiPu85] Windhofer, P.F., Pucker, N.: Multiple-collision solutions for time-dependent neutron-transport in slabs of finite thickness. Nucl. Sci. Eng. **91**, 223–233 (1985)

Chapter 24
Single-Phase Flow Instabilities: Effect of Pressure Waves in a Pump–Pipe–Plenum–Choke System

R.A.M. Vieira and M.G. Prado

24.1 Introduction

Stability is a very important topic in several sciences because it refers to real observable conditions. In classical mechanics, if a system is described by a set of differential equations, an equilibrium solution may be determined by setting all time derivatives equal to zero. This equilibrium solution is also known as the steady-state solution, fixed point, critical point, and equilibrium point, to name a few.

It is very important to distinguish between mathematical derivation and actual physical existence of a steady-state solution. An equilibrium solution may be obtained mathematically but physically may not exist or may never be achieved. Usually, analytical criteria can be obtained to establish the stability around equilibrium solutions without the necessity of solving the set of differential equations.

Several steady-state codes are used to calculate "equilibrium" solutions for different processes. This is a shortcut to obtain the "expected" steady-state solution, since the dynamics of the system are neglected.

Taking, for instance, a one-dimensional (1D) problem

$$\dot{x} = f(x), \tag{24.1}$$

the steady-state solution ($\bar{x}$) is obtained from the equation

$$f(x) = 0.$$

R.A.M. Vieira (✉)
Petrobras, Rio de Janeiro, Brazil
e-mail: rinaldo_vieira@petrobras.com.br

M.G. Prado
The University of Tulsa, Tulsa, OK, USA
e-mail: mauricio-prado@utulsa.edu

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_24,

One can say that around the equilibrium solution, the function x can be expressed as the summation of the equilibrium solution plus a small disturbance $\delta x(t)$, therefore,

$$x = \bar{x} + \delta x. \tag{24.2}$$

Substitution of (24.2) into (24.1) yields

$$\frac{d(\bar{x} + \delta x)}{dt} = f(\bar{x} + \delta x). \tag{24.3}$$

The right-hand side can be expanded using Taylor series

$$f(\bar{x} + \delta x) = f(\bar{x}) + f'(\bar{x})\,\delta x + O(\delta x^2). \tag{24.4}$$

Substituting (24.4) into (24.3), canceling out the terms that are zero at the equilibrium solution and neglecting the higher order terms yields to a linearized ordinary differential equations (ODE) which represents the propagation of small disturbances around the equilibrium solution. Integration of this equation leads to

$$\delta x = \delta x(0)\, e^{f'(\bar{x})t}, \tag{24.5}$$

where $\delta x(0)$ is the value of the initial infinitesimal disturbance. It is clear from (24.5) that the disturbance will increase, resulting in an unstable equilibrium, if

$$f'(\bar{x}) > 0.$$

For a two-dimensional (2D) homogeneous linear system, represented by the following system of ODE

$$\begin{cases} \dot{x}_1 = a_1 x_1 + a_2 x_2, \\ \dot{x}_2 = a_3 x_1 + a_4 x_2, \end{cases} \tag{24.6}$$

which can be written in matrix notation as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x},$$

where

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix},$$

the stability of the equilibrium solution (the origin for this particular case) is based on the eigenvalues of $\mathbf{A}$, which are provided by

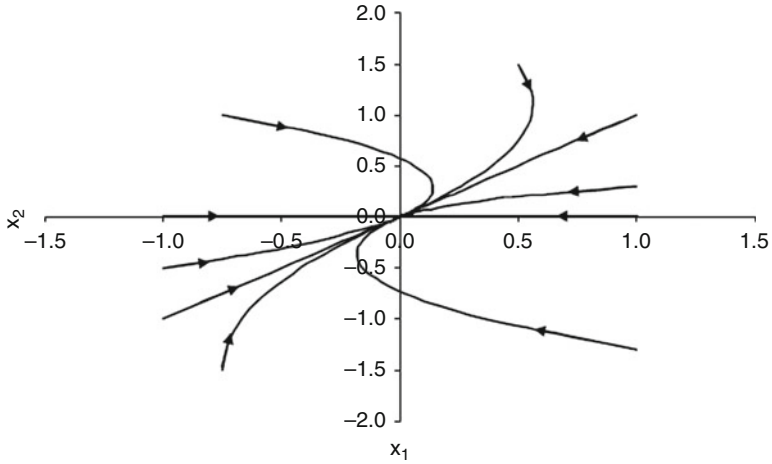

Fig. 24.1 Phase portrait: stable node

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0,$$

where det is the determinant, λ are the eigenvalues, and $\mathbf{I}$ is the identity matrix.

Defining the trace (P) and the determinant (q) of the coefficient matrix $\mathbf{A}$ as

$$\begin{cases} P = a_1 + a_4, \\ q = a_1 a_4 - a_2 a_3, \end{cases} \tag{24.7}$$

the eigenvalues may be written as

$$\lambda = \frac{P \pm \sqrt{P^2 - 4q}}{2}.$$

The equilibrium solution is asymptotically stable if, and only if, the eigenvalues have negative real parts [JL06]. A very useful graph that helps in the understanding of stability concepts is the phase portrait. This graph illustrates the relationship between solutions x_1 and x_2 as time evolves for several different initial conditions.

Figure 24.1 shows a generic phase portrait, which represents a stable equilibrium. Each path corresponds to a different initial condition and the arrows provide a visual interpretation of the stability.

A useful tool that can be constructed using the definitions given in (24.7) is the graph of P versus q. Figure 24.2 shows such graph which is divided into regions according to the eigenvalues characteristics (real positive, real negative, null or complex) and repeatability.

If an equilibrium solution attracts all nearby initial conditions, it is said to be an attractor. If an equilibrium solution repels all nearby initial conditions, it is

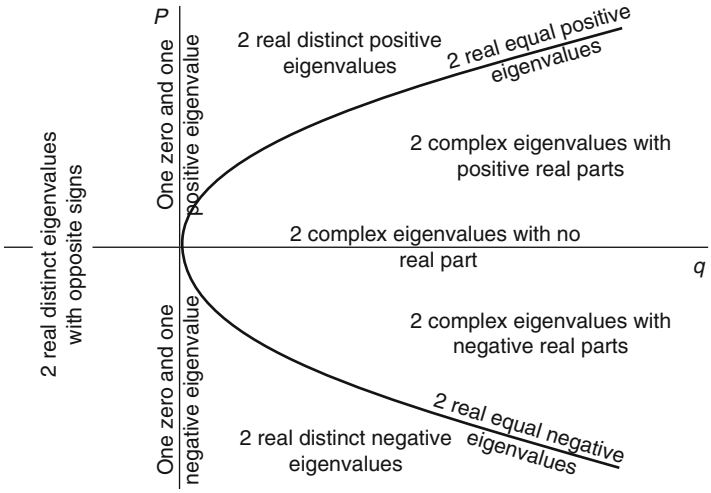

Fig. 24.2 Eigenvalue characteristics

called a repeller. Equilibrium solutions located in the fourth quadrant of Fig. 24.2 are attractors.

The linear system given by (24.6) can be related to the nonlinear case

$$\begin{cases} \dot{x_1} = f_1(x_1, x_2) \\ \dot{x_2} = f_2(x_1, x_2) \end{cases}, \tag{24.8}$$

where f_1 and f_2 are nonlinear functions. If small disturbances $\delta x_i(t)$ are applied to the equilibrium solutions $(\bar{x}_1, \bar{x}_2)$, that is,

$$\begin{gathered} x_i(t) = \bar{x}_i + \delta x_i(t) \\ i = 1, 2 \end{gathered} \tag{24.9}$$

and then (24.9) is substituted in (24.8), the equation of how the propagation of small disturbances *around* the equilibrium solution evolves appears. Proceeding with Taylor expansions, neglecting second and high order terms, the final linearized system is obtained. In terms of matrix notation, it is given by

$$\dot{\mathbf{x}} = \mathbf{J}|_{\bar{\mathbf{x}}}\mathbf{x}, \tag{24.10}$$

where $\mathbf{J}$ is the Jacobian matrix. Similar to the linear case, the stability of the steady-state solution *would* be given based on the eigenvalues of $\mathbf{J}$, evaluated at each equilibrium solution:

$$\mathbf{J}|_{\bar{\mathbf{x}}} = \left[\begin{matrix} a_1 = \dfrac{\partial f_1}{\partial x_1} & a_2 = \dfrac{\partial f_1}{\partial x_2} \\ a_3 = \dfrac{\partial f_2}{\partial x_1} & a_4 = \dfrac{\partial f_2}{\partial x_2} \end{matrix}\right]\Bigg|_{\bar{\mathbf{x}}}.$$

It *would* be asymptotically stable if, and only if, the eigenvalues had negative real parts. In other words, as (24.10) represents how disturbances are propagated (in a "linearized" way), if they die-out—meaning they are attracted to the equilibrium solution—the equilibrium solution of the original nonlinear system (24.8) also exists and is stable.

For real systems, this linearization process usually leads to easy inequalities that determine whether or not a solution is stable, which are based on steady-state parameters. Because of nonlinearities, usually these criteria are only valid in a *very small* vicinity of the equilibrium solution.

In addition, another mathematical entity called "limit cycle" exists in the phase portrait of 2D nonlinear systems and is very important in determining if a steady state solution exists and if it can be achieved. A limit cycle is an isolated closed trajectory, meaning that its neighboring trajectories are not closed—they spiral either towards (stable) or away (unstable) from the limit cycle. If one of the variables of a limit cycle is plotted against time, a periodic waveform is obtained. It only exists in nonlinear systems and cannot be determined through LLA. Usually transient numerical simulation is the only way to confirm the presence or not of such entity.

Figure 24.3 shows a very interesting situation that may occur in systems described by (24.8). It represents a phase portrait containing a "locally" stable equilibrium solution that is surrounded by two limit cycles. The inner one is unstable while the outer, stable. The internal area of the unstable limit cycle represents the "basin of attraction" of this equilibrium solution. The equilibrium solution will only exist if the initial condition is placed inside its basin of attraction. In addition, the magnitude of any perturbation needs to be small enough to maintain the system inside this area. If these conditions are not satisfied, the limit cycle, which represents

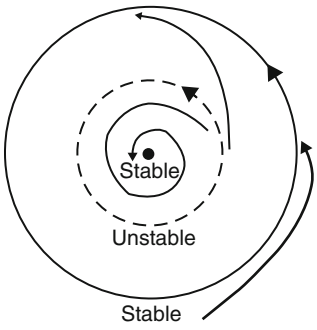

Fig. 24.3 Possible phase portrait in a 2D nonlinear system

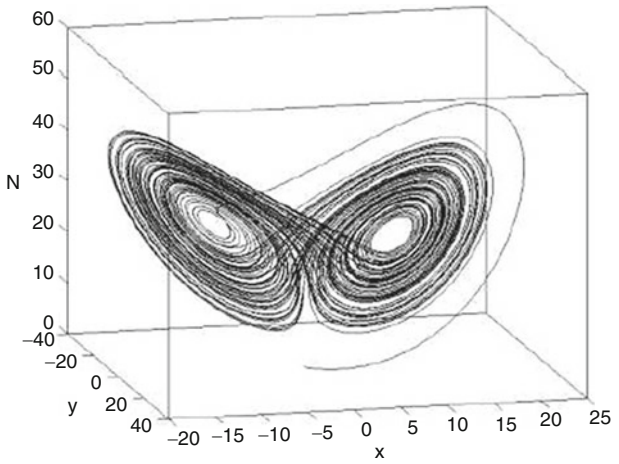

Fig. 24.4 Strange attractor: Lorenz attractor

a cyclical behavior, will be the final state of the system. This example clearly shows that criteria based on LLA may be useless.

3D and higher-order nonlinear systems also have a different entity named "strange attractor." It represents waveforms that do not have any periodicity and remain bounded within a definite volume. This particular situation is usually called *chaos*. Figure 24.4 shows a well-known strange attractor named Lorenz attractor—the path never repeats itself and it remains bounded indefinitely [St01].

Another very important aspect of chaotic systems, in fact a hallmark of chaos, is the sensitive dependence on initial condition. In non-chaotic systems two very close initial conditions are expected to evolve similarly in time. This is not what happens in chaotic systems—at a certain time the two trajectories diverge from each other and follow different courses of evolution. For practical purposes, this property of chaotic systems implies something of a grave consequence. Errors in specifying accurate initial conditions make future predictions inaccurate. Prediction is impossible beyond a certain time frame. After a specific time, predictions are unreliable.

Oscillatory behavior is also observed in fluid flow systems. Two phase flow system instability is a well-known problem in nuclear industry [BoBeTo73], [LaPo89]. It may cause flow oscillations which can induce boiling crises, disturb control systems, or cause mechanical damage in nuclear equipment devices. Oil wells also face production instabilities that usually lead to operational problems to surface and subsurface equipment. Most importantly, they also cause production losses [HuGo03].

LLA may also be applied to fluid flow systems to determine analytical stability criteria. It is not trivial to derive such equations as the governing equations are partial differential equations (PDE). To obtain easy practical criteria, several simplifying

assumptions must be made. Most of them may end up reducing the system from PDE to ODE, to allow the use of LLA based on the eigenvalues of the Jacobian matrix. There exist other methods based on Laplace transformation and frequency domain but the resulting criteria are somehow equivalent. It should be noted that the number of criteria is related to the size of the Jacobian matrix.

The simplifying assumptions combined with the nonlinearities effects may cause these criteria to fail in several cases, including some very simple systems [Vi11]. The combination of steady-state simulators and LLA criteria may not be a good choice in real case situations. Transient simulation seems to be the most adequate method to determine if a fluid flow system will exhibit or not an unstable behavior.

24.2 Single-Phase Flow Instabilities Criteria

24.2.1 Static Instability

For fluid flow problems it is useful to describe this "instability" with the concept of nodal analysis or required pressure versus available pressure. The nodal analysis, *which is done considering steady state equations*, consists of selecting a division point (or node) in a flowing system. This division "creates" two sections: an inflow section (or available pressure section) which represents the available pressure at the node that this section can deliver for a certain mass flow rate and the outflow section representing the required pressure to flow the same mass flow rate. Different values of flow rates are simulated and the required and available pressures are plotted in a graph as a function of the flow rate.

Figure 24.5 shows an example of a horizontal pumping system. The two tanks (T_1 and T_2) have perfectly constant pressures P_1 and P_2, respectively. The tanks are connected through a horizontal pipeline of length (ΔL), internal diameter (d) and flow cross section area (A_p). The fluid is water, treated as incompressible and therefore with a constant density (ρ). Assuming that the hypothetical nodal point is located just after the pump discharge, the nodal analysis curves are shown in Fig. 24.6. The equilibrium mass flow rate ($\bar{m}$) is the one where the required pressure is equal to the available pressure at the nodal point.

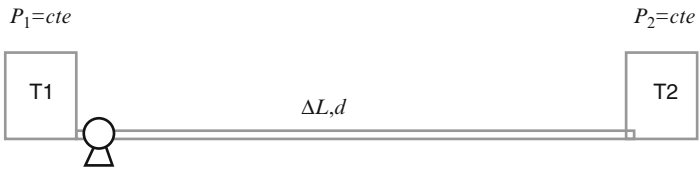

Fig. 24.5 Horizontal pumping unit

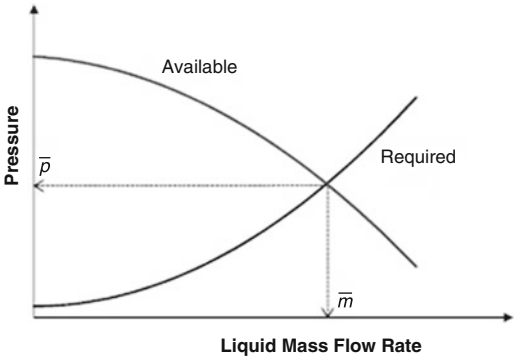

Fig. 24.6 Typical nodal analysis graph

The available pressure and the required pressure are given, respectively, by

$$P_{avail}(\dot{m}) = P_1 + \Delta P_P(\dot{m}),$$
$$P_{req}(\dot{m}) = P_2 + \Delta P_F(\dot{m}),$$

where $\Delta P_P(\dot{m})$ is the pump increment pressure and $\Delta P_F(\dot{m})$ is the friction loss in the pipe. At the steady-state condition the inflow and outflow pressures must be equal, yielding

$$P_1 + \Delta P_P(\bar{\dot{m}}) = P_2 + \Delta P_F(\bar{\dot{m}}). \tag{24.11}$$

To determine whether the equilibrium solution is locally stable or not, one can apply LLA. For this condition, the differential equation that governs this system may be written as (assuming the pump to have no inertia) [DeGiRi81]:

$$P_2 - P_1 - \Delta P_P(\dot{m}) + \Delta P_F(\dot{m}) = -\frac{\Delta L}{A_P}\frac{d\dot{m}}{dt}. \tag{24.12}$$

Applying a perturbation $\dot{m} = \bar{\dot{m}} + \delta\dot{m}$ in (24.12), performing linearization of pump and friction terms using Taylor series:

$$\Delta P_P(\bar{\dot{m}} + \delta\dot{m}) = \Delta P_P(\bar{\dot{m}}) + \left.\frac{d\Delta P_P(\dot{m})}{d\dot{m}}\right|_{\bar{\dot{m}}} \delta\dot{m},$$

$$\Delta P_F(\bar{\dot{m}} + \delta\dot{m}) = \Delta P_F(\bar{\dot{m}}) + \left.\frac{d\Delta P_F(\dot{m})}{d\dot{m}}\right|_{\bar{\dot{m}}} \delta\dot{m},$$

and using the relation giving by (24.11), we obtain the linearized differential equation

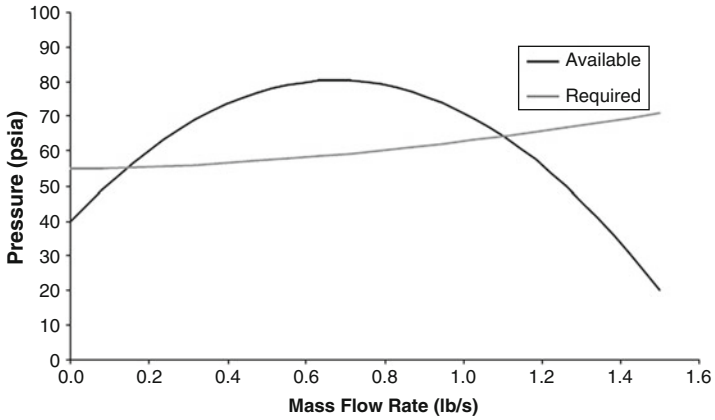

Fig. 24.7 Nodal analysis graph: unstable and stable points

$$\left(\left. \frac{d\Delta P_F(\dot{m})}{d\dot{m}} \right|_{\bar{\dot{m}}} - \left. \frac{d\Delta P_P(\dot{m})}{d\dot{m}} \right|_{\bar{\dot{m}}} \right) \delta\dot{m} = -\frac{\Delta L}{A_p} \frac{d(\delta\dot{m})}{dt}. \tag{24.13}$$

Integration of (24.13) leads to

$$\delta\dot{m} = \delta\dot{m}(0)\, e^{B},$$

where

$$B = -\frac{A_p}{\Delta L} \left(\left. \frac{d\Delta P_F(\dot{m})}{d\dot{m}} \right|_{\bar{\dot{m}}} - \left. \frac{d\Delta P_P(\dot{m})}{d\dot{m}} \right|_{\bar{\dot{m}}} \right) t. \tag{24.14}$$

Equation 24.14 reveals that the disturbance will grow (resulting in a local unstable equilibrium) if

$$\left. \frac{d\Delta P_P(\dot{m})}{d\dot{m}} \right|_{\bar{\dot{m}}} > \left. \frac{d\Delta P_F(\dot{m})}{d\dot{m}} \right|_{\bar{\dot{m}}}.$$

As friction is the only parameter that affects the slope of the required pressure and the pump curve is the one responsible for the slope of the available pressure, the above instability criterion may be rewritten as

$$\left. \frac{dP_{avail.}}{d\dot{m}} \right|_{eq.} > \left. \frac{dP_{req.}}{d\dot{m}} \right|_{eq.}. \tag{24.15}$$

This type of instability is named "static instability" because one may determine if an equilibrium solution exists based only on the nodal analysis graph—(24.15) is not even necessary. Take, for instance, the nodal analysis shown in Fig. 24.7.

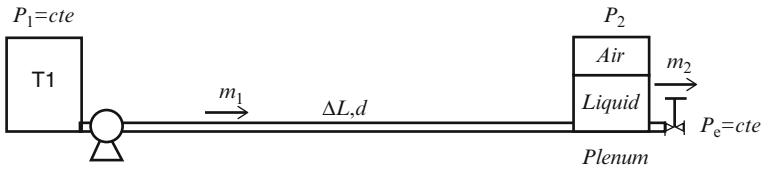

Fig. 24.8 Horizontal pumping unit with a plenum

Applying (24.15), we see that the lower equilibrium solution is unstable. One may come up with this same conclusion without using the criterion. If a positive disturbance is applied to the lower steady-state solution, as the available pressure is higher than the required pressure, the system accelerates toward the higher flow rate (which is not unstable). If a negative perturbation is applied to the same equilibrium solution, the system dies out, as the required pressure is higher than the available.

On the other hand, statically speaking, the higher flow rate should be stable. It has a self-control characteristic, meaning that the equilibrium is restored if small perturbations are applied around this equilibrium solution.

24.2.2 Dynamic Instability

These types of instabilities cannot be determined based on the nodal analysis graph. Didactically, they are usually classified into categories like pressure–drop instabilities, density–waves, etc., but from a skeptical point of view, they are based on the real part of the eigenvalues as previously described.

Figure 24.8 shows a modification in Fig. 24.5, where a plenum with pressurized air is located at the outlet of pumping system [Gr81], [RoRu78]. The plenum acts as a buffer tank. A choke is installed downstream of the plenum and the external pressure downstream of the choke is P_e.

The mass flow rate at the pipe is $\dot{m}_1$ while the one that leaves the plenum is $\dot{m}_2$. Assuming the liquid to be incompressible, the same equation for the system without plenum (Fig. 24.5) still holds for the liquid flow inside the pipe. The global force balance equation is

$$P_2 - P_1 - \Delta P_P(\dot{m}_1) + \Delta P_F(\dot{m}_1) = -\frac{\Delta L}{A_p}\frac{d\dot{m}_1}{dt}, \tag{24.16}$$

where P_2 is the plenum pressure (hydrostatic and liquid and gas are neglected). The choke at the outlet is assumed to have no inertia and modeled in [MuYoOkHu09]:

$$\dot{m}_2|\dot{m}_2| = \left(C\frac{\pi d_c^2}{4}\right)^2 2\rho_l(P_2 - P_e), \tag{24.17}$$

where C and d_c are the choke flow coefficient and the choke diameter. The plenum is considered as a passage tank without pressure losses. A mass balance applied to the liquid phase in the plenum yields

$$\dot{m}_1 - \dot{m}_2 = \frac{d\rho_l V_l}{dt} = -\rho_l \frac{dV_g}{dt}. \tag{24.18}$$

If the gas behavior is assumed to be isentropic, then

$$dV_g = -\frac{V_g}{kP_2} dP_2, \tag{24.19}$$

$$V_g = V_g^R \left(\frac{P_g^R}{P_2} \right)^{\frac{1}{k}},$$

where V_g^R and P_g^R represent some reference state for the gas in the plenum and k is the gas adiabatic index.

Combining (24.16)–(24.19), applying perturbations around the equilibrium solution, and linearizing the nonlinear terms, we find the following linearized matrix of coefficients $\mathbf{A}$, as described in [Vi11]:

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix},$$

where

$$a_1 = \frac{A_p}{\Delta L} \left(\left. \frac{d\Delta P_P(\dot{m}_1)}{d\dot{m}_1} \right|_{\bar{m}} - \left. \frac{d\Delta P_F(\dot{m}_1)}{d\dot{m}_1} \right|_{\bar{m}} \right),$$

$$a_2 = -\frac{A_p}{\Delta L},$$

$$a_3 = \frac{k\bar{P}_2^{\frac{k+1}{k}}}{\rho_l V_g^R \left(P_g^R\right)^{\frac{1}{k}}},$$

$$a_4 = -\frac{k\bar{P}_2^{\frac{k+1}{k}}}{\rho_l V_g^R \left(P_g^R\right)^{\frac{1}{k}}} \left(\left. \frac{d\Delta P_C(\dot{m}_2)}{d\dot{m}_2} \right|_{\bar{m}} \right)^{-1}.$$

Based on this matrix, we establish the instability criteria [Vi11]

$$\left. \frac{dP_{avail.}}{d\dot{m}} \right|_{eq.} > \left. \frac{dP_{req.}}{d\dot{m}} \right|_{eq.}, \tag{24.20}$$

$$\left.\frac{d\Delta P_P(\dot{m}_1)}{d\dot{m}_1}\right|_{\bar{m}} > \left.\frac{d\Delta P_F(\dot{m}_1)}{d\dot{m}_1}\right|_{\bar{m}} + \frac{\Delta L k \bar{P}_2^{\frac{k+1}{k}}}{A_p \rho_L V_g^R \left(P_g^R\right)^{\frac{1}{k}}} \left(\left.\frac{d\Delta P_C(\dot{m}_2)}{d\dot{m}_2}\right|_{\bar{m}} \right)^{-1}, \quad (24.21)$$

where $\Delta P_C(\dot{m}_2)$ is the choke pressure drop.

As this system is two dimensional, two criteria are obtained. It should be noted that the static instability criterion (24.20) appears once more. The second criterion is given by (24.21). It is clear that the second instability criterion cannot be replaced by an inspection in the nodal analysis graph. That is why it is named "dynamic instability." Being a 2D system, it may exhibit oscillatory behavior (limit cycles).

24.3 Single-Phase Flow Models

24.3.1 Incompressible Model

If one assumes water as incompressible (as it was considered to derive the instability criteria), the following set of equations needs to be solved, which represents the pump–plenum–choke dynamics (see Fig. 24.8):

$$P_2 - P_1 - \Delta P_P(\dot{m}_1) + \Delta P_F(\dot{m}_1) = -\frac{\Delta L}{A_p}\frac{d\dot{m}_1}{dt},$$

$$\dot{m}_2|\dot{m}_2| = \left(C\frac{\pi d_c^2}{4}\right)^2 2\rho_l(P_2 - P_e),$$

$$\dot{m}_1 - \dot{m}_2 = \rho_l \frac{V_g^R\left(P_g^R\right)^{\frac{1}{k}}}{kP_2^{\frac{k+1}{k}}}\frac{dP_2}{dt},$$

This system was solved using the second-order Runge–Kutta method. A generic pump curve (for didactic purpose) was used in the simulations. The pump curve was divided into four segments as shown in Fig. 24.9.

24.3.2 Compressible Model

Under isothermal conditions, the single phase conservation laws are reduced to mass and momentum balance equations

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho V)}{\partial z} = 0, \quad (24.22)$$

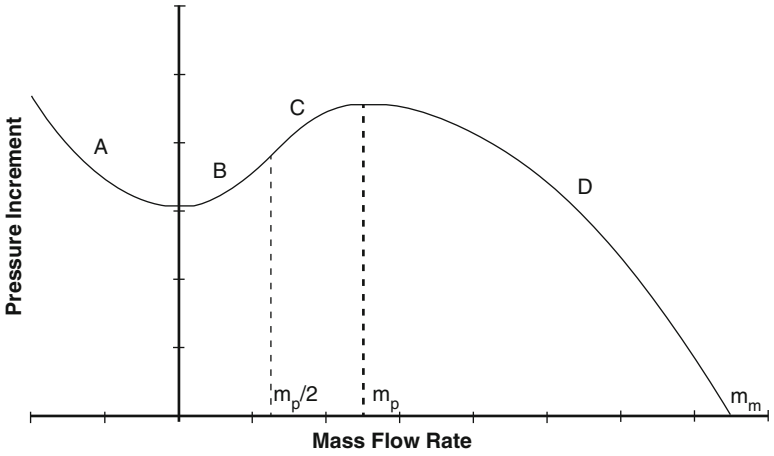

Fig. 24.9 Generic pump performance curve

$$\frac{\partial \rho V}{\partial t} + \frac{\partial (\rho VV)}{\partial z} + \frac{\partial P}{\partial z} = -\rho g \sin(\theta) - \varphi_w, \tag{24.23}$$

where θ is the angle with the horizontal and φ_w the friction loss term, given by

$$\varphi_w = \frac{1}{2} f \frac{\rho V |V|}{d},$$

where f is the Darcy–Weisbash friction factor. The relation between pressure and density was assumed to be of the type

$$\rho = \rho^R e^{c\left(P - P^R\right)}.$$

The discretization of (24.22) and (24.23) was done using a fully implicit first-order finite difference method on a staggered grid [Pa80] using pressure and fluxes as variables. Pressure was defined at the cell center, while mass fluxes at the cell faces [Vi11].

One of the advantages of implicit codes is their ability to use either large or small time-steps. If it is desired to capture pressure waves, small time steps should be used, like the ones calculated by the CFL criterion

$$\Delta t \leq \min_i \left\{ \frac{\Delta z_i}{|a_i \pm V_i|} \right\},$$

where a is the speed of sound in the liquid phase.

24.4 Application and Discussion

24.4.1 Example 1: Phase Portrait, Incompressible Model

This example considers the incompressible model assuming that the pump behaves as a fully opened valve for reverse flow. Figure 24.10 shows the nodal analysis (at plenum position), which presents three equilibrium solutions.

The table below presents the summary of the local linearization analysis (Table 24.1).

Figure 24.11 shows the transient simulation of a small disturbance applied to solution 3 (or an initial condition very close to solution 3). As the disturbance is close enough to the steady-state solution, the perturbation dies out and the system returns to equilibrium.

Increasing the disturbance, the equilibrium solution is no longer capable of attracting the system to equilibrium state number 3. The system tends to an oscillatory behavior, and a limit cycle is obtained. Figure 24.12 shows this situation.

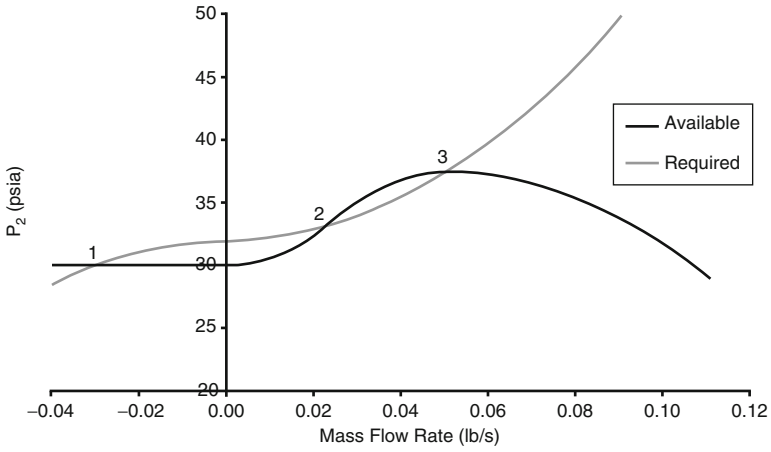

Fig. 24.10 Nodal analysis: example 1

Table 24.1 Local linearization analysis: example 1

Solution	Flow Rate (lb/s)	Eigenvalues	Stability
1	–0.03027	$\lambda_{1,2} = -0.026 \pm 0.183i$	Stable
2	0.02288	$\lambda_1 = 1.362$, $\lambda_2 = -0.048$	Unstable
3	0.05021	$\lambda_{1,2} = -0.022 \pm 0.206i$	Stable

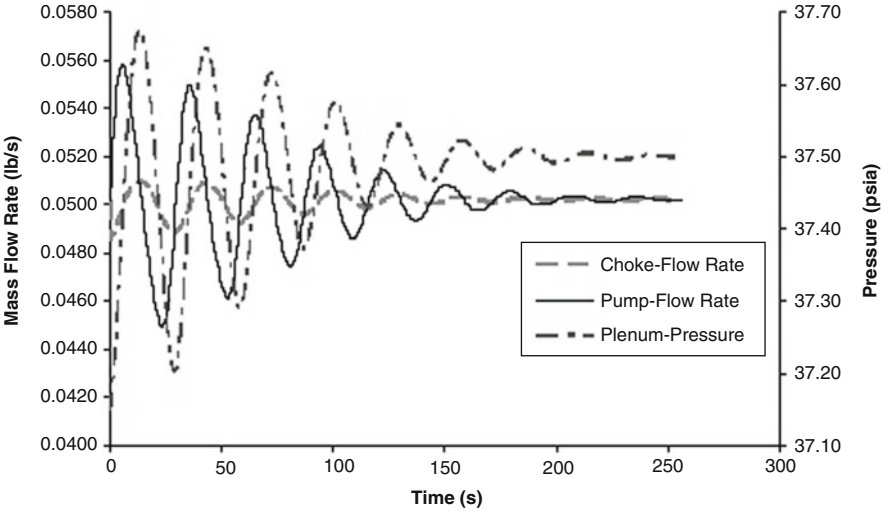

Fig. 24.11 Time plot: disturbance 1, example 1

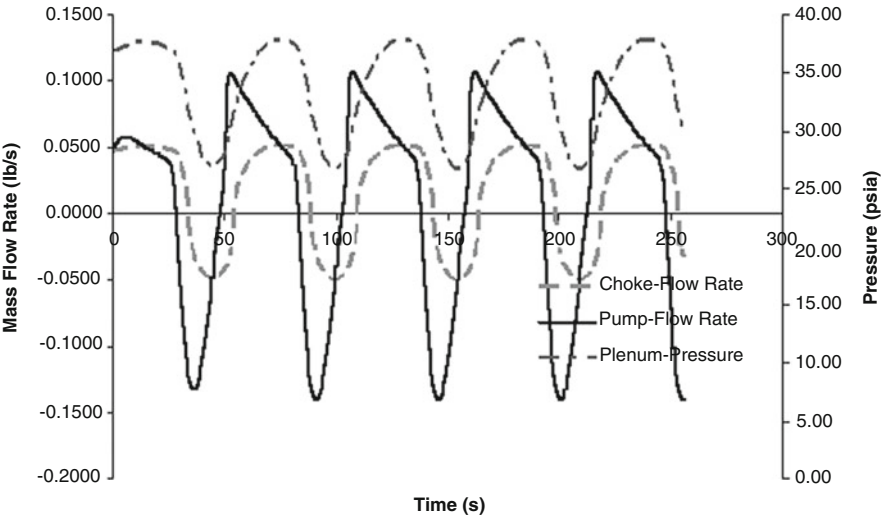

Fig. 24.12 Time plot: disturbance 2, example 1

Two locally stable equilibrium solutions and a stable limit cycle co-exist in this system. Figure 24.13 shows the phase portrait for several different initial conditions.

It is easily seen that the only way to reach solution 3 is to set the initial condition close to it. Solution 1 has a larger basin of attraction when compared to solution 3. Several initial conditions will converge to the external stable limit.

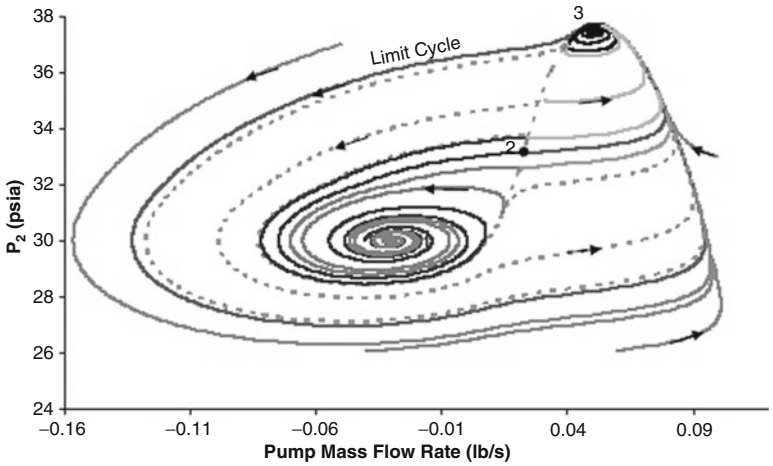

Fig. 24.13 Phase portrait: example 1

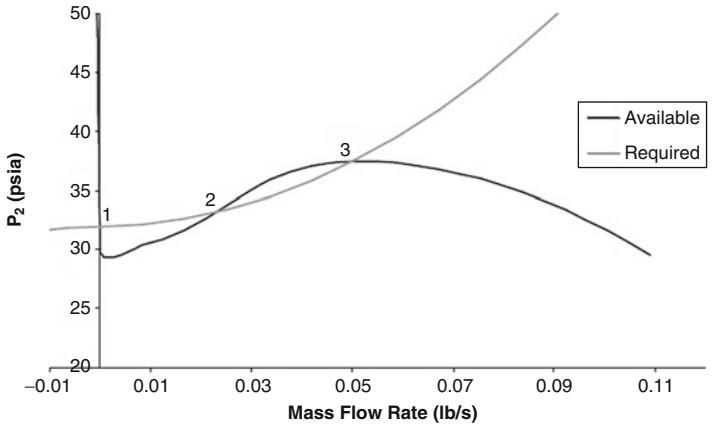

Fig. 24.14 Nodal analysis: example 2

24.4.2 Example 2: Phase Portrait, Incompressible Model with Check-Valve

In this example, a single modification is made to Example 1, by adding a check valve at the pump. Unlike the last example, no reverse flow is allowed through the pump now. Figure 24.14 shows the nodal analysis for this example. The table following this figure presents the summary of the local linearization analysis (Table 24.2).

The resulting phase portrait is shown below. Equilibrium solution 1 has changed to a stable node instead of a stable spiral (Fig. 24.15).

Table 24.2 Local linearization analysis: example 2

Solution	Flow Rate (lb/s)	Eigenvalues	Stability
1	0.0000	$\lambda_1 = 3.68$, $\lambda_2 = -45.19$	Stable
2	0.02288	$\lambda_1 = 1.362$, $\lambda_2 = -0.048$	Unstable
3	0.05021	$\lambda_{1,2} = -0.022 \pm 0.206i$	Stable

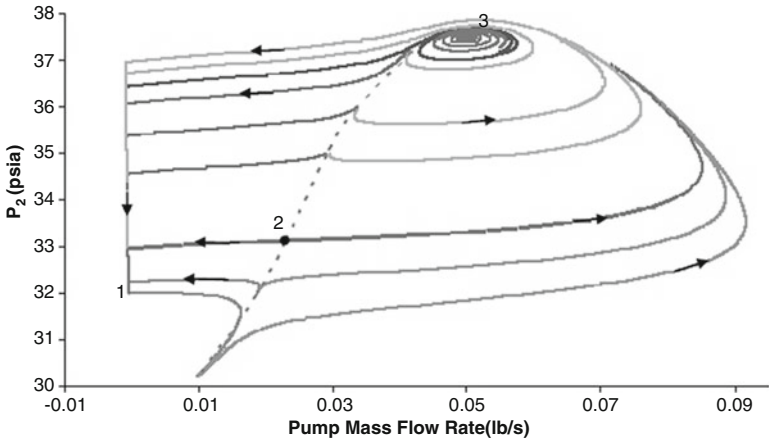

Fig. 24.15 Phase portrait: example 2

The stable node (solution 1) attracts all initial conditions that are not sufficiently close to equilibrium solution 3, which still has an unstable limit cycle surrounding it. If one compares Examples 1 and 2, the only difference between them is the presence of a check-valve, which avoids the appearance of the limit cycle.

It should be noted that a dynamic model for the check-valve could lead to different results but this detailed analysis is beyond the scope of this work.

24.4.3 Example 3: Incompressible Versus Compressible Model

The objective of this example is to compare the two models previously described. The nodal analysis graph considering the plenum as nodal section is shown in Fig. 24.16.

Incompressible Model. The instability criterion developed for incompressible liquid indicates that the equilibrium solution in Fig. 24.16 is an unstable spiral.

The transient solution is shown in Fig. 24.17. A limit cycle takes place (see Fig. 24.18).

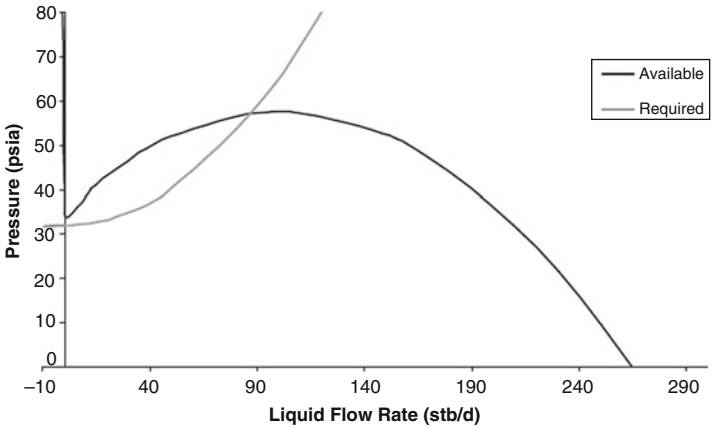

Fig. 24.16 Nodal analysis: example 3

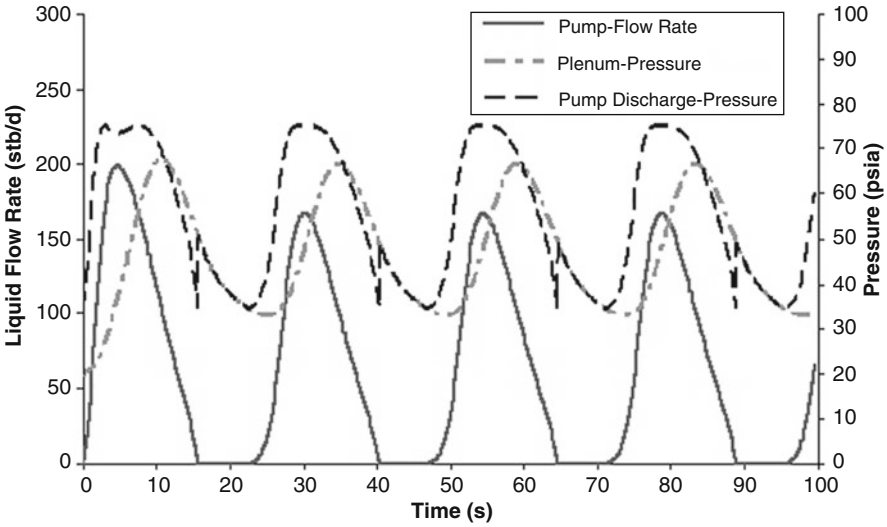

Fig. 24.17 Transient solution: example 3, incompressible model

Compressible Model. Using this model, which assumes the liquid to have some compressibility ($c = 3\ 10^{-6}\ \text{psi}^{-1}$) slightly different results are obtained.

Figure 24.19 shows the result of the simulation using a time step (0.01 s) less than the *CFL* criterion (in order to capture pressure-waves type propagation).

The transient solution seems to exhibit chaotic characteristics. This oscillation has the contribution of acoustic instability since traveling pressure waves contribute to destabilization of the system.

One diagnostic to infer whether the system is chaotic is the sensitivity to initial condition. Figure 24.20 shows the results for two very close initial conditions for the

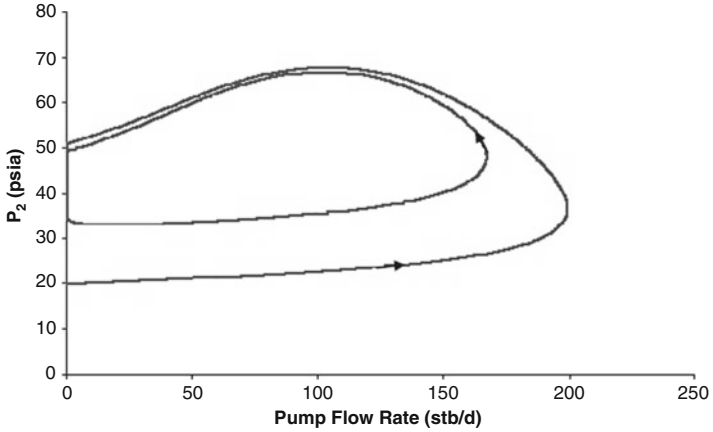

Fig. 24.18 Limit cycle: example 3, incompressible model

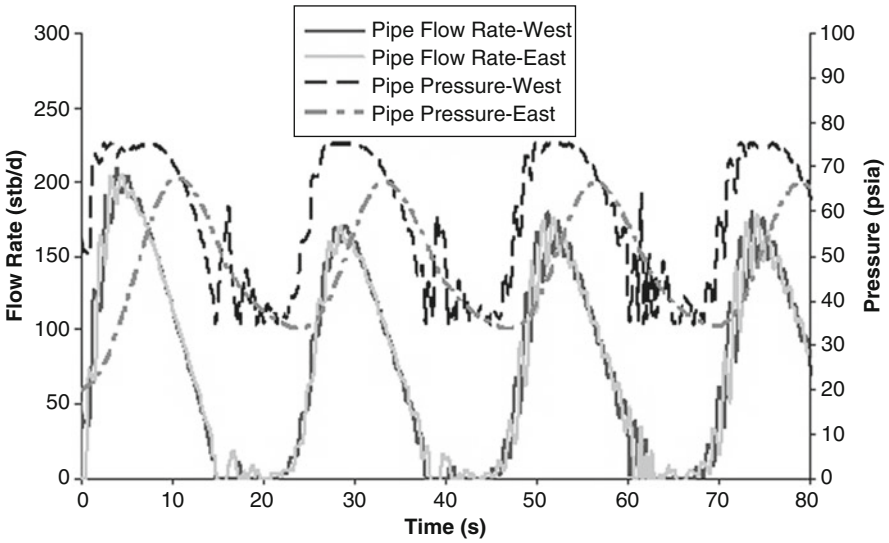

Fig. 24.19 Transient solution: example 3, compressible model

first 40 s of simulation. Only one variable is compared (pressure at pump discharge) and they are practically following the same path.

Figure 24.21 shows the divergence point around simulation time of 40 s. From that time on, the paths are different.

If a larger time step is used (0.25 s) and as a consequence the pressure-waves are missed, the result is similar to the incompressible model given in Fig. 24.17. Figure 24.22 shows this condition. For such large time step, the pressure waves are no longer captured. Under this situation, the chaotic behavior and tubing dynamics disappear and, the dynamic is basically represented by the plenum.

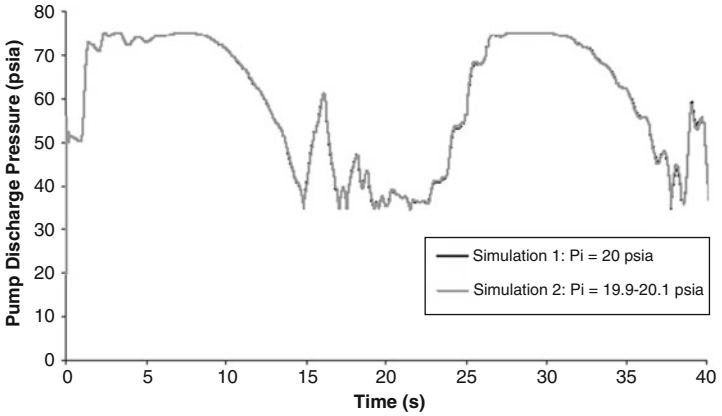

Fig. 24.20 Sensitivity to initial condition: example 3, first 40 s

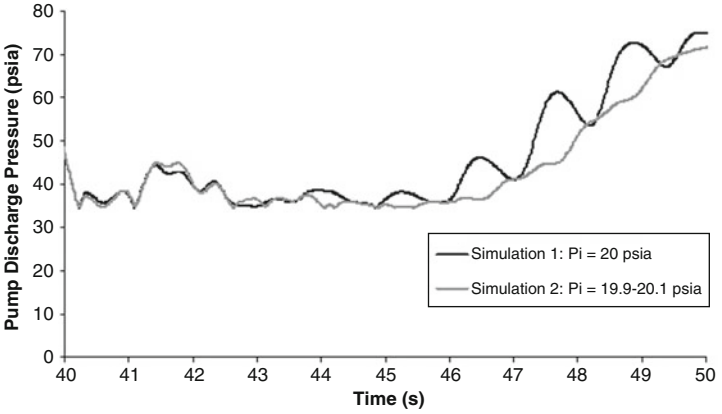

Fig. 24.21 Sensitivity to initial condition: example 3, divergence

If an even larger time step is used (1 s) wrong results are obtained. Even the plenum dynamic is lost and the unstable solution is reached as if it were stable. Figure 24.23 shows this condition. This may be a pitfall of implicit codes. Some software developers of implicit codes claim that the use of large time steps is one of the big advantages of their codes. The use of large time steps, for some conditions, may cause the physics to be lost, and wrong results may be obtained.

24.4.4 Example 4: Incompressible Versus Compressible Model

This example considers the same data as Example 3. The only difference is the use of a larger choke opening, leading to a higher equilibrium flow rate.

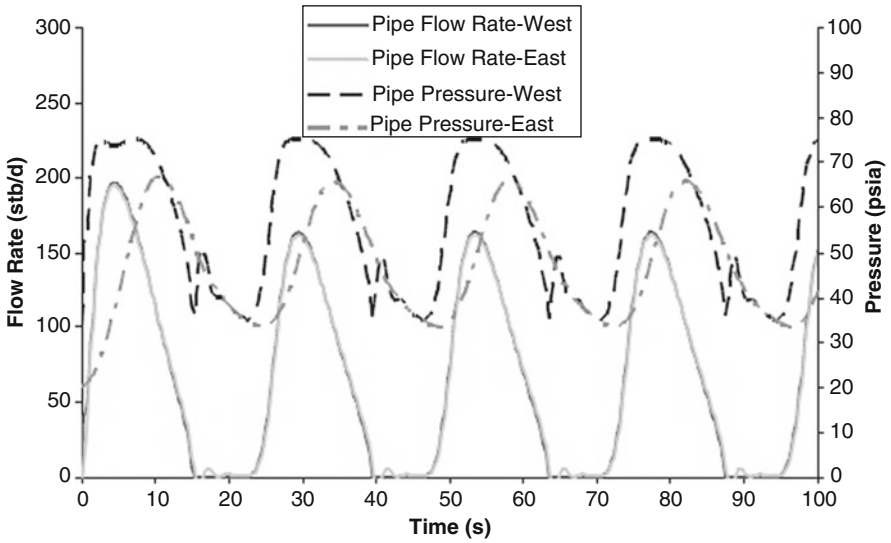

Fig. 24.22 Compressible model: example 3, $\Delta t = 0.25$ s

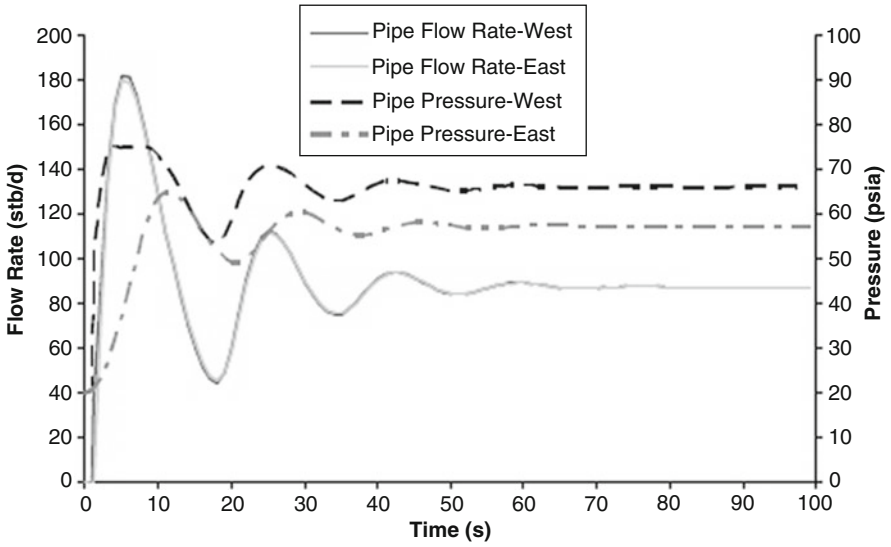

Fig. 24.23 Compressible model: example 3, $\Delta t = 1$ s

Incompressible Model. The analytical criteria establish that the equilibrium solution is stable (stable spiral). Transient simulation using the incompressible model "confirms" the stability as shown in Fig. 24.24. No limit cycle is observed and the only attractor is the equilibrium solution.

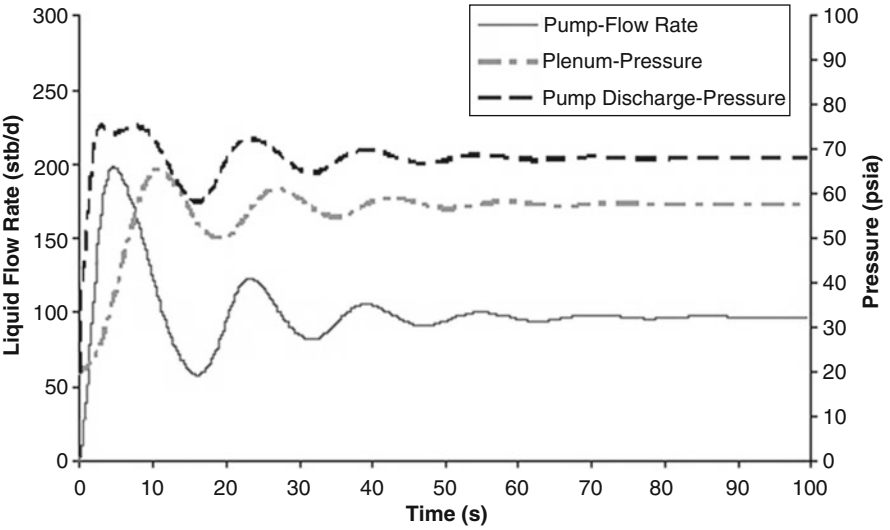

Fig. 24.24 Transient solution: example 4, incompressible model

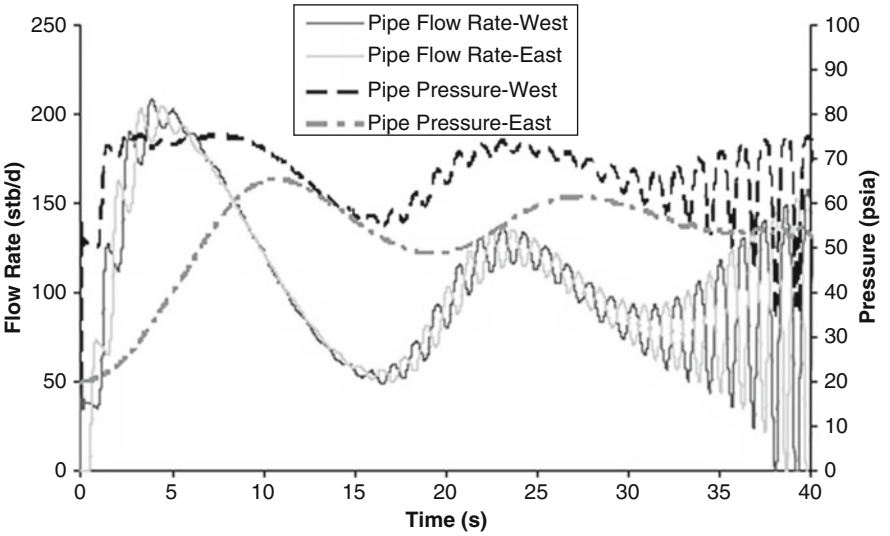

Fig. 24.25 Transient solution: example 4, compressible model

Compressible Model. A different response is obtained using the compressible model. The dynamics are revealed to be unstable and chaotic (see Fig. 24.25). This example shows that simplifying assumptions to develop analytical models can lead to a misrepresentation of reality leading to failure of the criteria.

24.5 Conclusions

1. There are different attractors in multidimensional nonlinear systems, such as equilibrium solutions, limit cycles, and strange attractors. LLA provides limited information regarding a tiny piece of a big puzzle and depending on the initial condition, the equilibrium solution may never be reached, even though it is stable. Only numerical simulations can really determine whether or not a dynamic system is stable.
2. The equations that govern fluid flow are in reality a system of PDEs. Several of the analytical criteria were obtained by simplification of these PDEs into ODEs in addition to other simplifying assumptions. The transient simulations showed that these analytical criteria may fail in some particular situations.
3. A single-phase transient code was developed and several instability examples were shown. Comparisons relating incompressible vs. compressible models were also made.
4. Depending on the system being simulated, the selection of the correct time step is very important. Large time steps may misrepresent reality, since the transient solution under such circumstances is not able to capture all dynamics. Wrong results such as stable conditions that should be unstable can be obtained.

24.6 Nomenclature

a	Speed of sound
a_i	Constants in matrix **A**
A	Area
A	Coefficient matrix
c	Isothermal compressibility
C	Choke flow coefficient
d	Diameter
D	Dimension
f	Darcy–Weisbash friction factor
$f(x)$	Generic function of x
$f'(x)$	First derivative of generic function of x
g	Gravitational acceleration
I	Identity matrix
J	Jacobian matrix
k	Gas adiabatic index
L	Length
$\dot{m}$	Mass flow rate
$\bar{\dot{m}}$	Equilibrium mass flow rate

P Pressure or trace of a 2 by 2 matrix
q 2 by 2 matrix determinant
Q Volumetric flow rate
t Time
V Velocity or volume
$\dot{x}_i$ First derivative of x_i
$\mathbf{x}$ Column vector of variables x_i
$\dot{\mathbf{x}}$ First derivative of column vector $\mathbf{x}$
$\bar{x}$ Equilibrium solution
z Position

Greek Letters:

δ Disturbance
φ_w Friction loss gradient
λ Eigenvalues
θ Angle with horizontal
ρ Density

Subscripts:

$avail.$ Available
c Choke
$eq.$ Equilibrium
f Friction
g Gas
l Liquid
P Pump or pipe
$req.$ Required

Superscripts:

R Reference

Acknowledgments The authors appreciate the technical and financial support of Tulsa University Artificial Lift Projects' member companies. The progress on this work is the result of the support of Baker–Hughes Centrilift, Chevron, ENI, Kuwait Oil Company, PEMEX, Petrobras, Shell International, Total and Wood Group ESP.

References

[BoBeTo73] Boure, J.A., Bergles, A.E., Tong, L.S.: Review of Two-Phase Flow Instabilities. Nucl. Eng. Des. **25**, 165–192 (1973)

[DeGiRi81] Delhaye, J.M, Giot, M., Riethmuller, M.L.: Thermohydraulics of Two-Phase System for Industrial Design and Nuclear Engineering. Hemisphere, Washington (1981)

[Gr81] Greitzer, E.M.: The stability of pumping systems. J. Fluid. Eng. **103**, 193–242 (1981)

[HuGo03] Hu, B., Golan, M.: Gas-lift instability resulted production loss and its remedy by feedback control: dynamical simulation results. In: SPE International Improved Oil Recovery Conference in Asia Pacific, 84917, Kuala Lumpur (2003)

[LaPo89] Lahey, R.T. Jr., Podowski, M.Z.: On the analysis of various instabilities in two-phase flows. In: Multiphase Science and Technology, vol. 4. Hemisphere, Washington (1989)

[JL06] Logan, J.D.: Applied Mathematics, 3rd edn. Wiley-Interscience, New York (2006)

[MuYoOkHu09] Munson B.R., Young, D.F., Okiishi, T.H., Huebsch, W.W.: Fundamentals of Fluid Mechanics, 6th edn. Wiley, New York (2009)

[Pa80] Patankar, S.V.: Numerical Heat Transfer and Fluid Flow. Hemisphere, Washington (1980)

[RoRu78] Rothe, P.H., Runstadler, P.W. Jr.: First-order pump surge behavior. J. Fluid. Eng. **100**, 459–466 (1978)

[St01] Strogatz, S.H.: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Westview, Boulder (2001)

[Vi11] Vieira, R.A.M.: Flow dynamics in oil wells. Ph.D. dissertation, The University of Tulsa (2011)

Chapter 25
Two-Phase Flow Instabilities in Oil Wells: ESP Oscillatory Behavior and Casing-Heading

R.A.M. Vieira and M.G. Prado

25.1 Introduction

If a system is described by a set of differential equations, an equilibrium solution may be determined by setting all time derivatives equal to zero. This equilibrium solution is also known as steady-state solution, fixed point, critical point, and equilibrium point, to name a few.

Several commercial steady-state two-phase flow codes are used by petroleum engineers to calculate the "equilibrium" flow rate for oil wells. This is a shortcut to obtain the "expected" steady-state solution, since the dynamics of the system are neglected. Steady-state simulators are widely used because they are cheaper and easy to use when compared to sophisticated transient simulators.

It is very important to distinguish between mathematical calculation and actual physical existence of a steady-state solution. A steady-state solution may be mathematically determined but physically may not exist or may never be achieved.

Using as example a two-dimensional (2D) homogeneous linear system, represented by the system of ordinary differential equations (ODE)

$$\begin{cases} \dot{x}_1 = a_1 x_1 + a_2 x_2 \\ \dot{x}_2 = a_3 x_1 + a_4 x_2 \end{cases} \tag{25.1}$$

which can be written in matrix notation as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x},$$

R.A.M. Vieira (✉)
Petrobras, Rio de Janeiro, Brazil
e-mail: rinaldo_vieira@petrobras.com.br

M.G. Prado
The University of Tulsa, Tulsa, OK, USA
e-mail: mauricio-prado@utulsa.edu

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_25,

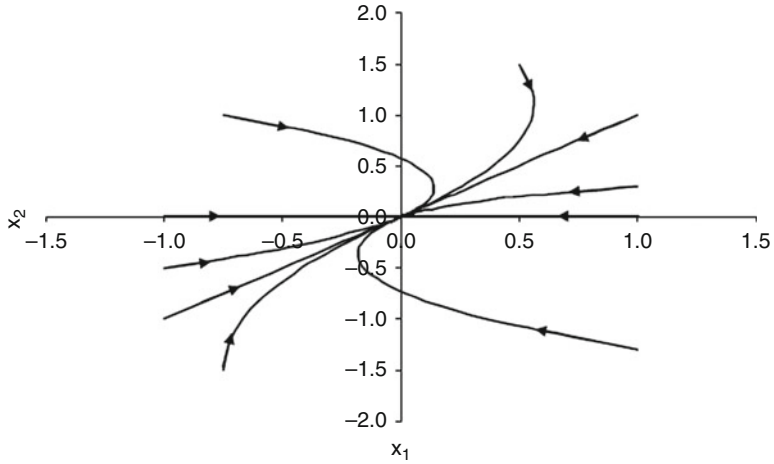

Fig. 25.1 Phase portrait: degenerate stable node

where

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}.$$

The stability of the equilibrium solution (the origin for this particular case) is based on the eigenvalues of A, which are provided by

$$\det(A - \lambda I) = 0,$$

where *det* is the determinant, λ are the eigenvalues, and **I** is the identity matrix.

If one defines the trace (P) and the determinant (q) of the coefficient matrix **A** as

$$\begin{cases} P = a_1 + a_4 \\ q = a_1 a_4 - a_2 a_3 \end{cases},$$

the eigenvalues may be written as

$$\lambda = \frac{P \pm \sqrt{P^2 - 4q}}{2},$$

The steady-state solution is asymptotically stable if, and only if, the eigenvalues have negative real parts [Lo06]. A graph called phase portrait is very useful to help the understanding of stability. This graph illustrates the relationship between solutions x_1 and x_2 as time evolves for different initial conditions.

Figure 25.1 shows a generic phase portrait, which represents a stable equilibrium solution. Each path corresponds to a different initial condition and the arrows provide a visual interpretation of the stability.

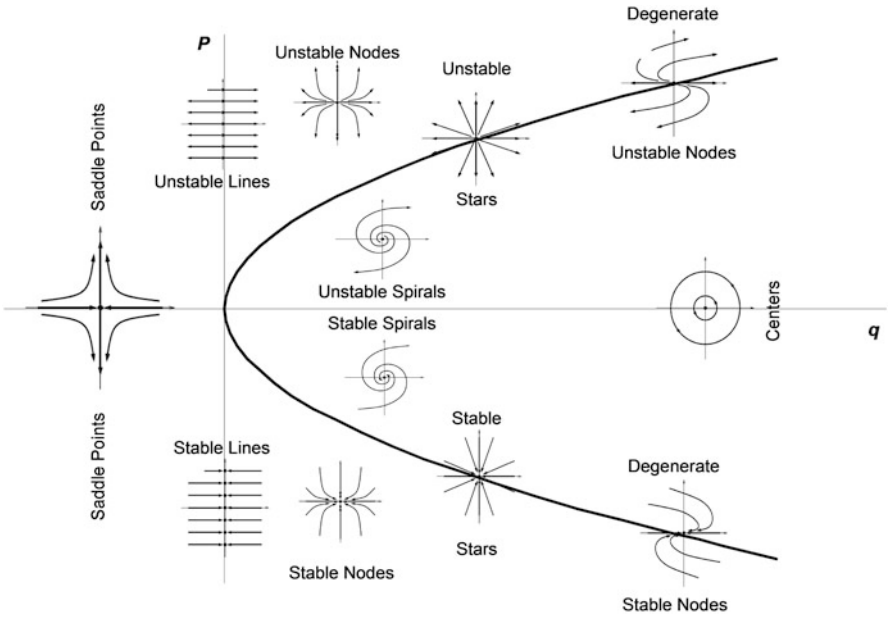

Fig. 25.2 Equilibrium solutions stability: linear 2D problems

Figure 25.2, adapted from [Wi09], shows all possible phase portraits for 2D linear systems. The axes are given by P and q. The fourth quadrant of Fig. 25.2 comprises 2D linear systems where the real part of the eigenvalues is negative (negative P), representing asymptotically stable solutions.

A quick analysis of the graph reveals the presence of "neutrally stable" entities named centers. Each different initial condition generates a different center which is neither "attracted" nor "repelled" by the equilibrium solution. For all other situations, the paths are attracted or repelled by the equilibrium solution following lines or spirals.

The most simple and used procedure to check the stability of equilibrium solutions in nonlinear systems is known as local linearization analysis (LLA). The linear system given by (25.1) can be related to the nonlinear case

$$\begin{cases} \dot{x}_1 = f_1(x_1, x_2) \\ \dot{x}_2 = f_2(x_1, x_2) \end{cases}, \tag{25.2}$$

where f_1 and f_2 are nonlinear functions. If small disturbances $\delta x_i(t)$ are applied to the equilibrium solutions $(\bar{x}_1, \bar{x}_2)$:

$$x_i(t) = \bar{x}_i + \delta x_i(t), \quad i = 1, 2, \tag{25.3}$$

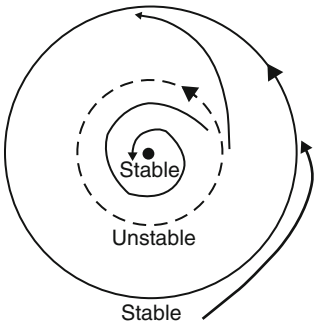

Fig. 25.3 Possible phase portrait of a 2D nonlinear system

and then (25.3) is substituted into (25.2), the equation of how the propagation of small disturbances *around* the equilibrium solution evolves appears. Proceeding with Taylor expansions, neglecting second and high order terms, the final linearized system is obtained. In terms of matrix notation, it is given by

$$\dot{\mathbf{x}} = \mathbf{J}|_{\bar{\mathbf{x}}}\,\mathbf{x}, \tag{25.4}$$

where $\mathbf{J}$ is the Jacobian matrix. Similar to the linear case, the stability of the steady-state solution *would* be given based on the eigenvalues of $\mathbf{J}$, evaluated at each equilibrium solution.

It *would* be asymptotically stable if, and only if, the eigenvalues had negative real parts. In other words, as (25.4) represents how disturbances are propagated (in a "linearized" way), if they die-out—meaning they are attracted to the equilibrium solution—the equilibrium solution of the original nonlinear system 25.2 does exist and is also stable.

For real systems, this linearization process usually leads to easy inequalities that determine whether or not a solution is stable, which are based on steady-state parameters. Because of nonlinearities, usually these criteria are only valid in a *very small* vicinity of the equilibrium solution. In addition, another mathematical entity called "limit cycle" exists in the phase portrait of 2D nonlinear systems and is very important in determining if a steady state solution exists and if it can be achieved.

A limit cycle is an isolated closed trajectory, meaning that its neighboring trajectories are not closed—they spiral either towards (stable) or away (unstable) from the limit cycle. If one of the variables of a limit cycle is plotted against time, a periodic waveform is obtained. It only exists in nonlinear systems and cannot be determined through LLA. Transient numerical simulation is usually the best way to confirm the presence or not of such entity.

Figure 25.3 shows a very interesting situation that may occur in systems described by (25.2). It represents a phase portrait containing a "locally" stable equilibrium solution that is surrounded by two limit cycles. The inner one is unstable while the outer, stable. The internal area of the unstable limit cycle represents

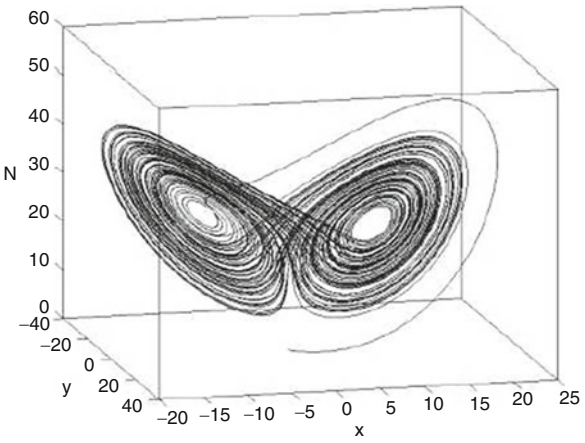

Fig. 25.4 Strange attractor: Lorenz attractor

the "basin of attraction" of this equilibrium solution. The equilibrium solution will only exist if the initial condition is placed inside its basin of attraction. In addition, the magnitude of any perturbation needs to be small enough to maintain the system inside this area. If these conditions are not satisfied, the limit cycle, which represents a cyclical behavior, will be the final state of the system. This example clearly shows that criteria based on LLA may be useless.

3D and higher-order nonlinear systems also have a different entity named "strange attractor." It represents waveforms that do not have any periodicity and remain bounded within a definite volume. This particular situation is usually called *chaos*. Figure 25.4 shows a well-known strange attractor named Lorenz attractor—the path never repeats itself and it remains bounded indefinitely [St01].

Oscillatory behavior is also observed in fluid flow systems. Two phase flow system instability is a well-known problem in the nuclear industry [BoBeTo73], [LaPo89]. It may cause flow oscillations which can induce boiling crises, disturb control systems, or cause mechanical damage in nuclear equipment devices. Oil wells also face production instabilities that usually lead to operational problems to surface and subsurface equipment. Most importantly, they also cause production losses [HuGo03].

LLA may, in addition, be applied to fluid flow systems to determine analytical stability criteria. It is not trivial to derive such equations as the governing equations are partial differential equations (PDE). To obtain easy practical criteria, several simplifying assumptions must be made. Most of them may end up reducing the system from PDE to ODE, to allow the use of LLA based on the eigenvalues of the Jacobian matrix. There exist other methods based on Laplace transformation and frequency domain but the resulting criteria are somehow equivalent. It should be noted that the number of criteria is related to the size of the Jacobian matrix.

The simplifying assumptions combined with the nonlinearities effects may cause these criteria to fail in several cases, including some very simple systems [Vi11].

The combination of steady-state simulators and LLA criteria may not be a good choice in real case situations. Transient simulation seems to be the most adequate method to determine if a well will exhibit or not an unstable behavior.

25.2 Two-Phase Flow Modeling Overview

Figure 25.5 shows the schematic of a production well. There are basically three domains in the system: casing, tubing, and annular space. One of the extremities of each domain forms a shared interface called "junction" in this work with the other domains. The casing domain comprehends the volume between the reservoir and the junction, while the tubing and the annular space are bounded by the junction and each respective surface choke.

Three variations are possible: (1) The electrical submersible pumps (ESP) may be located in front of the perforations (casing not included in the solution domain), (2) ESP in front of the perforations and the assumption that only gas is separated—all liquid from reservoir goes inside the tubing and the gas separated to the annular

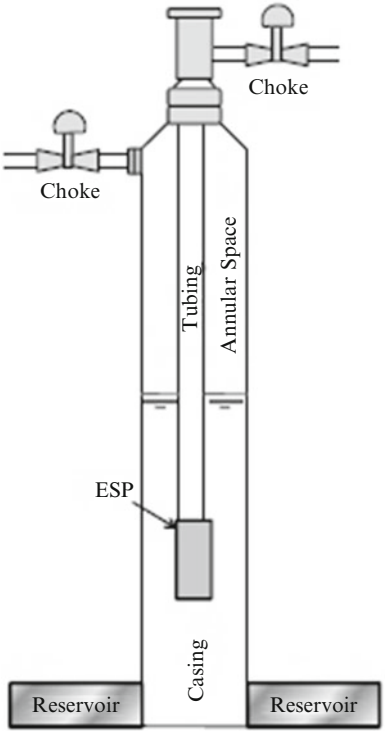

Fig. 25.5 Reservoir–casing–tubing–annular space model

space vanishes—casing and annular space not included in the solution domain—and (3) no ESP installed—which represents natural flowing wells.

For the ESP case, it is interesting to note that during well start-up, both annular space and casing feed the pump with liquid. Gas is separated at pump's intake (which can be a regular intake or a rotary separator), meaning that at this moment counter-current flow takes place at annular space. If a steady state situation is reached, the liquid level at the annular space reaches a constant depth, which is called "dynamic level."

Independent of the scenario, each domain must obey the conservation laws and the junction must receive an appropriate treatment to correctly model the problem, including the consideration of gas and liquid mass conservation.

Equations and Numerical Solution. The model is based on the drift-flux approach [ZuFi65], assuming isothermal flow and no mass transfer between phases:

$$\frac{\partial\left(\alpha_g \rho_g\right)}{\partial t}+\frac{\partial\left(\alpha_g \rho_g V_g\right)}{\partial z}=0, \tag{25.5}$$

$$\frac{\partial\left(\alpha_l \rho_l\right)}{\partial t}+\frac{\partial\left(\alpha_l \rho_l V_l\right)}{\partial z}=0, \tag{25.6}$$

$$\frac{\partial\left(\alpha_g \rho_g V_g+\alpha_l \rho_l V_l\right)}{\partial t}+\frac{\partial P}{\partial z}=-\rho_m g sin(\theta)+\varphi_{wt}, \tag{25.7}$$

$$V_g-V_l=V_S, \tag{25.8}$$

where (25.5) and (25.6) represent, respectively, gas and liquid mass conservation, (25.7) is the mixture momentum conservation equation (convective terms were neglected), and (25.8) the slip velocity closure relationship.

Closure relationships must be provided for the slip velocity (V_S) and for the two-phase friction term (ϕ_{wt}). The slip velocity is obtained using the traditional drift-flux model.

$$\alpha=\frac{V_{sg}}{C_0\left(V_{sg}+V_{sl}\right)+V_d},$$

where C_0 is the distribution parameter and V_d the drift velocity. These two parameters are obtained from published correlations. For co-current upward flow, a modification in the correlation stated in [WoGh07] was proposed, while for co-current downward flow the result in [IsHi05] was used. Because of the lack of correlations to model counter-current flow, a linear interpolation procedure between co-current upward/downward has been developed in [Vi11].

The discretization of the equations was done using a fully implicit first-order finite difference method on a staggered grid [Pa80], with pressure and void fractions defined at the cell centers and velocities at cell edges, using an upwind scheme.

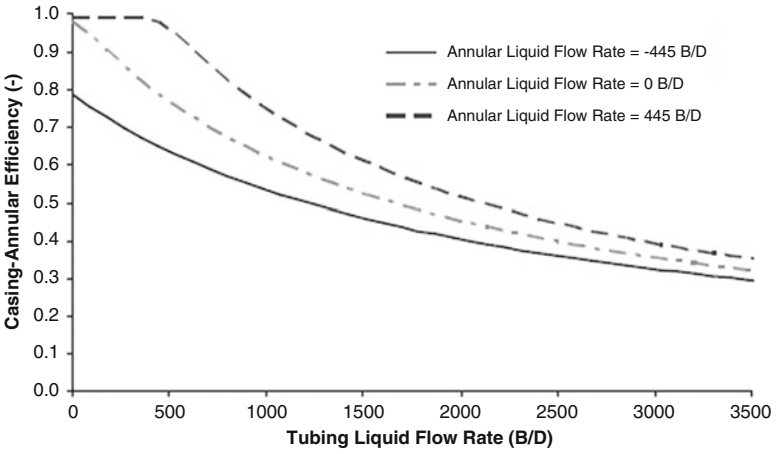

Fig. 25.6 Natural casing–annulus separation efficiency: pumped well

(25.5)–(25.8) were discretized in each domain, with some particular adaptations such as the use of equivalent and hydraulic diameters for annular geometry.

The reservoir was modeled as a source of liquid and gas, following linear relationships. As the chokes may be under single or two-phase flow conditions, the model in [Sa84] was used. A proper description of the "junction" was done, assuring gas and liquid mass conservation as well as pressure continuity [Vi11].

Gas Separation Models. The bottomhole natural gas separation efficiency was calculated using the model in [Al93]. This model assumes that all liquid coming from the casing goes through the pump and the liquid inside the annular space is static. These premises do not satisfy the reality of this work since the liquid within the annular space may be flowing upward or downward. Because of the lack of correlations, a modification was proposed in [Vi11].

Figure 25.6 shows the results of the proposed modification for some arbitrary condition. For the case when the annular liquid flow rate is zero, the correlation represents Alhanati's original model itself. If the liquid is getting into the annular space (positive flow rate) the efficiency is higher since it drags more gas. On the other hand, if the annular liquid is going inside the intake (negative flow rate) more gas is dragged into the pump, reducing the separation efficiency.

A simplified rotary separator based on Alhanati's work [Al93] rotary separator model can also be used. Figure 25.7 shows a typical curve for the global efficiency of this equipment, for some arbitrary conditions. As suggested by Alhanati, the existence of operational conditions in which rotary separators are not effective was considered in this simplified model.

Pump Model. Electrical submersible pumps are widely utilized in the oil industry. They are multistage vertical pumps with a diffuser casing that can handle large liquid volumes. In an artificial lift system, these pumps are installed within a cased hole well and produce the reservoir while staying "submersed" in the fluid.

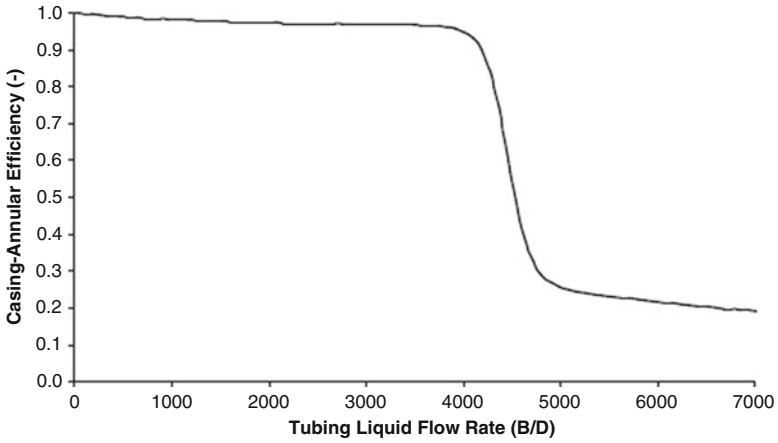

Fig. 25.7 Generic rotary separator efficiency curve

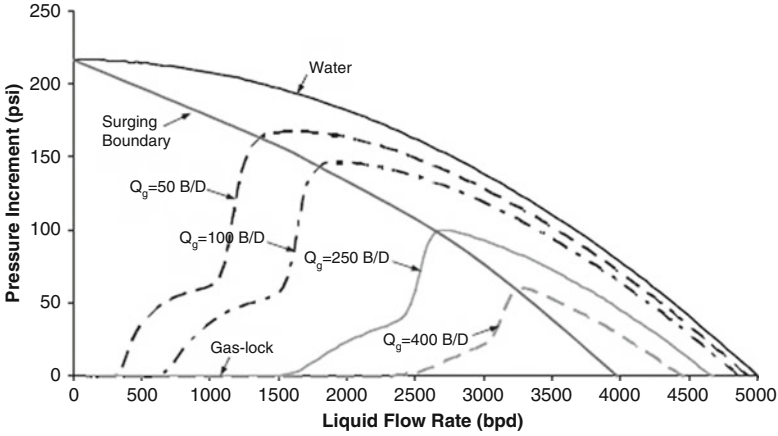

Fig. 25.8 Two-phase pump performance curve

Free gas directly impacts the pump curve performance deteriorating its ability to lift liquids. The degree of head deterioration varies from a simple reduction in performance to more severe problems such as surging and gas-lock.

The pump model used in this work is proposed based on the work presented in [Du03] and [CaEtAl09]. The calculation is not done stage by stage; instead, it is considered an average total pressure increment. Figure 25.8 shows a typical curve performance for some arbitrary intake conditions for different constant gas flow rates as described in [Vi11]. The stable operational envelope of the pump is assumed to be the region limited by the surging boundary, the water curve performance and the no-pressure increment horizontal line.

25.3 Application and Discussion

25.3.1 Example 1. ESP: Tubing and Annular Space Included in the Solution Domain. Stability Example

This example considers a pump equipped with a rotary separator. The pump maximum flow rate is 8,640 B/D and it is located in front of the perforations. The separator is under-sized as its maximum operational liquid flow rate is about 1,250 B/D. For this scenario, all liquid from reservoir goes to the pump and the gas separated to the annular space disappears.

The gas split is determined through Alhanati's model previously described. The fluids considered in this simulation are air and water. Figure 25.9 shows the nodal analysis under these premises.

A widely used instability criterion derived from LLA, given in [DeGiRi81], is

$$\left.\frac{dP_{avail.}}{dQ}\right|_{eq.} > \left.\frac{dP_{req.}}{dQ}\right|_{eq.}. \tag{25.9}$$

According to this criterion, the solution is unstable if, at the equilibrium flow rate, the derivative of the available pressure is greater than the required pressure. Inequality (25.9) does not guarantee that the equilibrium solution in Fig. 25.9 is unstable. It should be noted that this is just one criterion among several others that may exist. For complex systems like this example, the other criteria are very difficult to obtain, even using simplifying assumptions.

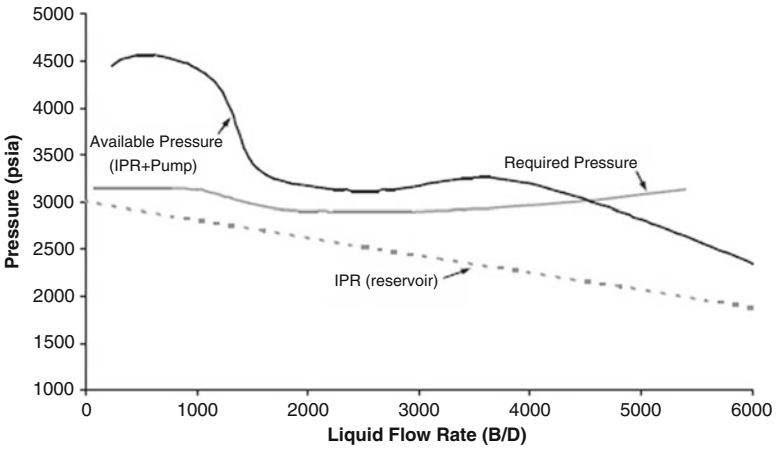

Fig. 25.9 Nodal analysis: example 1

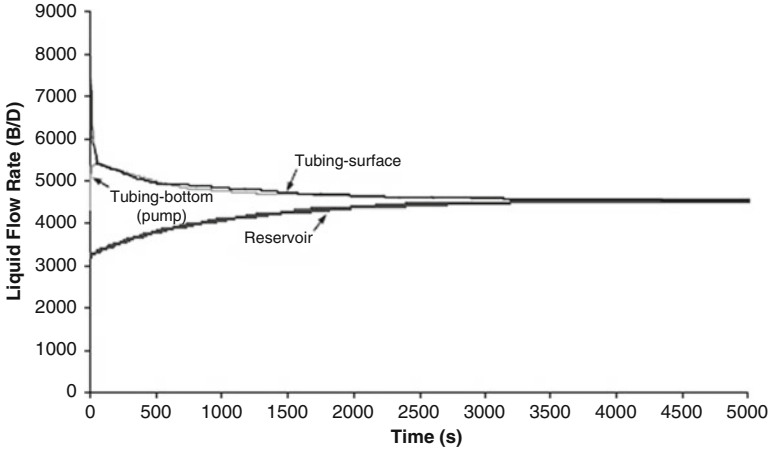

Fig. 25.10 Transient solution: tubing and reservoir liquid flow rates

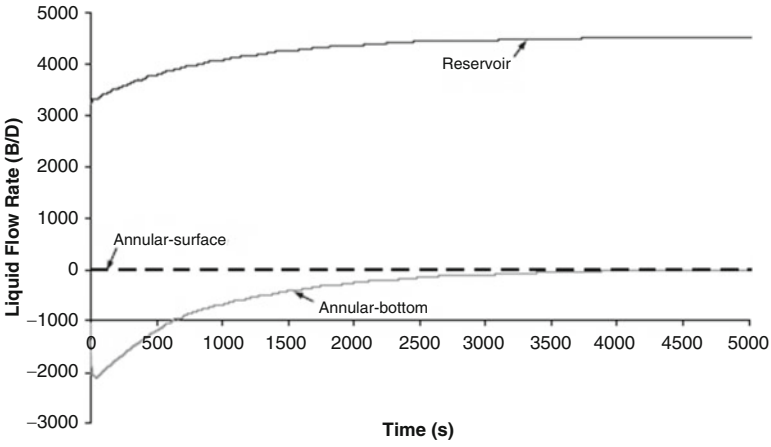

Fig. 25.11 Transient solution: annular space and reservoir liquid flow rates

Figures 25.10 and 25.11 show the results of the transient simulations. The steady-state condition is reached with a constant dynamic level in the annular space, since no liquid flows in this domain after 5,000 s.

25.3.2 Example 2. ESP: Neither Casing nor Annular Space Included in the Solution Domain. Instability Example

The only difference between this example and the last one is the tubing diameter. In this example, the tubing diameter is smaller than the one used in Example 1—which

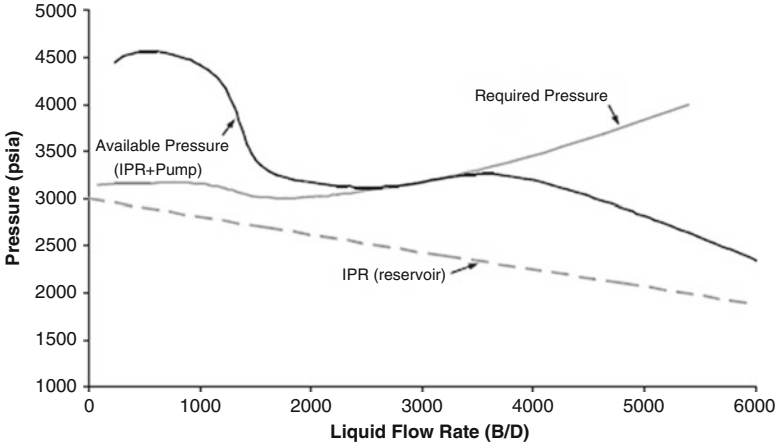

Fig. 25.12 Nodal analysis: example 2

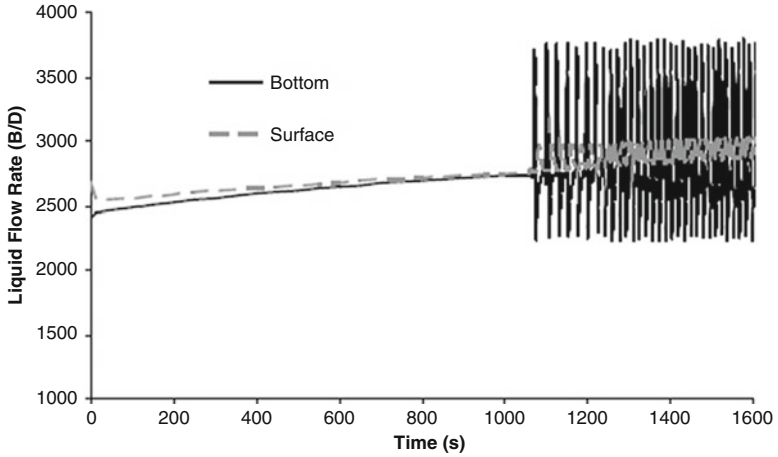

Fig. 25.13 Transient solution: liquid flow rates

increases the required pressure—leading to a smaller liquid equilibrium flow rate. Figure 25.12 shows the nodal analysis.

One more time (25.9) does not guarantee that the equilibrium solution in Fig. 25.12 is unstable. Figure 25.13 shows the result of transient simulation. The equilibrium solution is unstable and no steady-state is obtained. The pump presents a high frequency oscillatory behavior.

Figure 25.14 shows the oscillatory behavior with a different time scale range, while Fig. 25.15 shows the oscillatory behavior reached around the equilibrium solution.

An interesting observation is that the surface flow rates show small amplitudes while at the pump they are in the order of 1,500 B/D. This is not a desirable

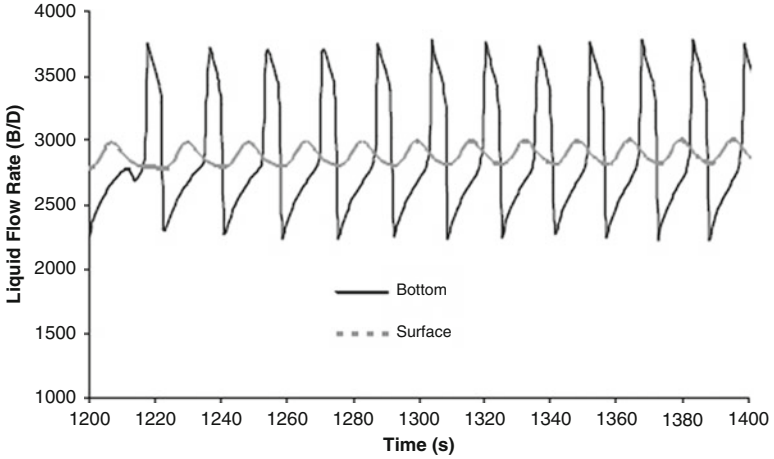

Fig. 25.14 Transient solution: close-up

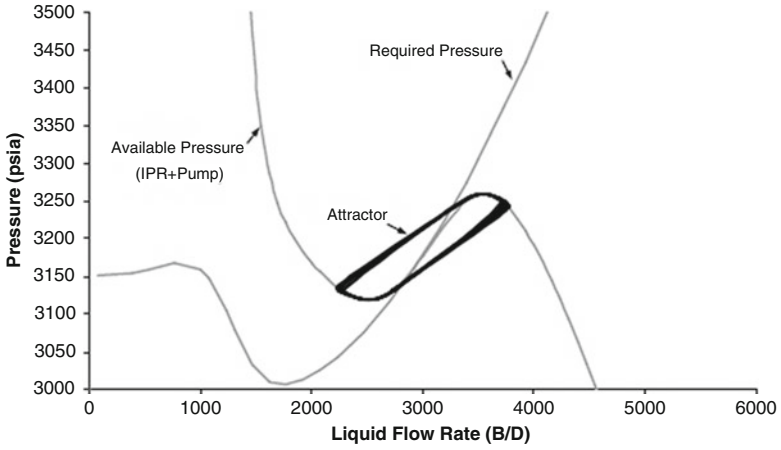

Fig. 25.15 Oscillatory behavior

operational condition for an ESP, especially for a pump with floating impellers. In a real well, probably the protective relay would shut down the equipment.

25.3.3 Example 3. ESP: Tubing and Annular Space Included in the Solution Domain. Instability Example

The objective of this example is to determine the influence of the annular space dynamics in the unstable behavior of Example 2. To solve this problem under

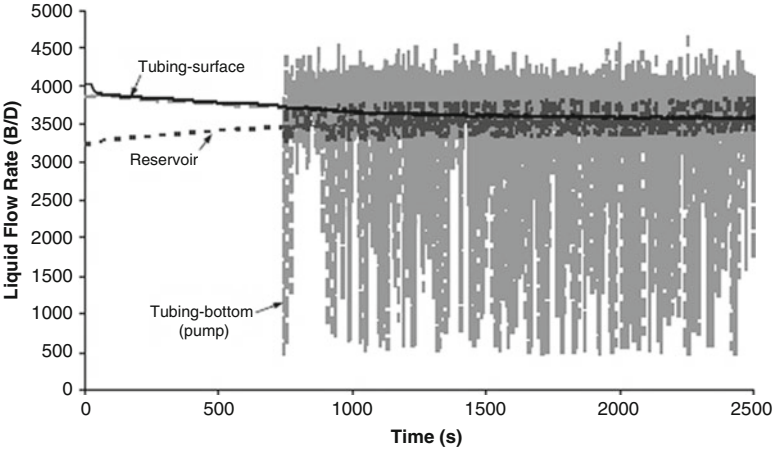

Fig. 25.16 Transient solution: tubing and reservoir liquid flow rates

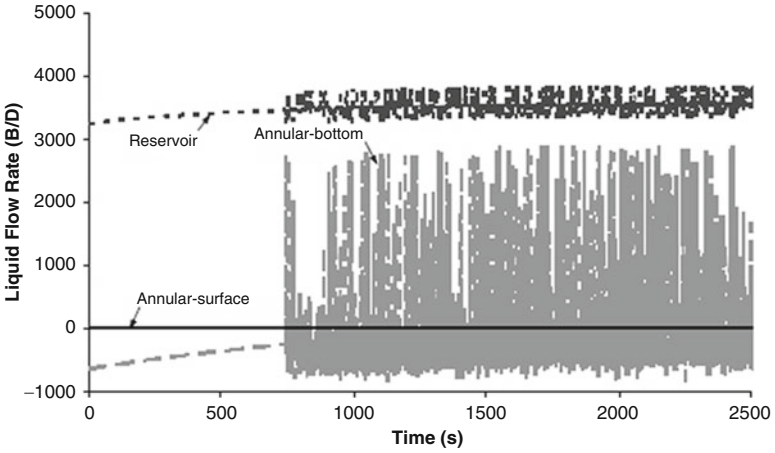

Fig. 25.17 Transient solution: annular space and reservoir liquid flow rates

steady-state conditions, the first thing assumed is that the annular space has reached a constant dynamic level (no liquid moving inside the annular space) and thus all liquid coming from reservoir goes into the pump.

The nodal analysis is the same as the one shown in Fig. 25.12. Figures 25.16 and 25.17 show the result of the transient simulation.

The well exhibits an oscillatory behavior, confirming the previous result. It should be noted that the fluctuations at surface are even smaller while the amplitudes downhole have increased. In addition, the solution shows more evidence of a chaotic behavior.

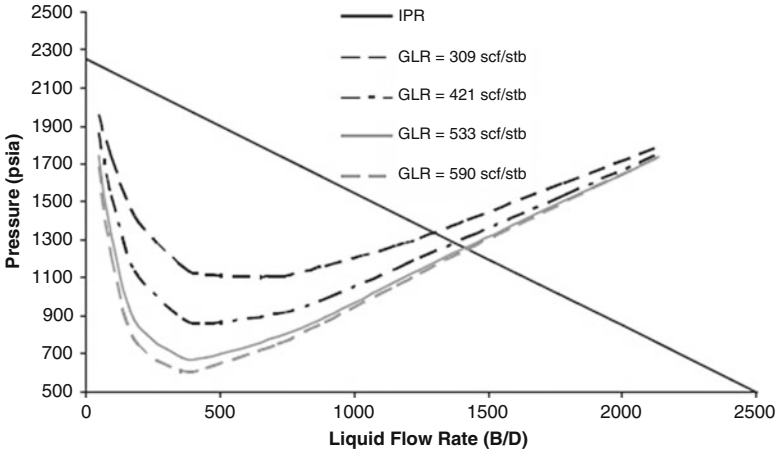

Fig. 25.18 Nodal analysis: example 3

25.3.4 Example 4: Natural Flowing Well. Casing Heading

If one assumes no ESP installed in the well shown in Fig. 25.5 and that the annular surface choke is closed, gas accumulates in the annular space due to natural separation. This gas accumulation increases the pressure at the liquid level, reducing production and pushing the liquid in annular space into the tubing.

At some point the liquid level reaches the bottom of the tubing and the gas in the annulus is produced, causing a "gas-lift" effect. When this occurs, the annular space is depressurized increasing the liquid production from reservoir.

When the annular space pressure is no longer enough to maintain the "gas-lift", liquid and gas will start to accumulate in the annular space, increasing its pressure. At some point, the "gas-lift" process starts over again. This cyclical behavior is common in naturally flowing wells without packers.

It should be noted that casing heading does not occur in every well without packer. Figure 25.18 shows the nodal analysis for several gas–liquid–ratios (GLRs), assuming that all gas goes into the tubing.

One can see that for GLRs over 533 scf/stb, the required pressure curve in the region around the equilibrium solution starts to present a reverse behavior. Although the mixture-density reduces as the amount of gas increases (reducing hydrostatic), friction and acceleration becomes higher, overcoming the hydrostatic reduction. A well producing under this condition should not exhibit heading, since more gas into tubing stabilizes the system. Figure 25.19 shows the liquid production at surface for different GLRs.

As expected, the well does not exhibit casing heading for GLRs equal or greater than 533 scf/stb. For this situation, friction acts as a stabilizing mechanism (reverse pressure gradient behavior).

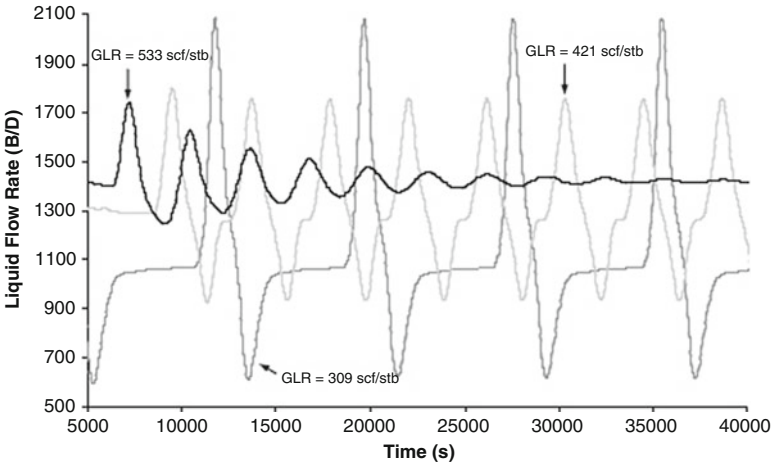

Fig. 25.19 GLRs comparison

25.4 Conclusions

1. There are different attractors in multi-dimensional nonlinear systems, such as: equilibrium solutions, limit cycles, and strange attractors. LLA provides limited information regarding a tiny piece of a big puzzle and depending on the initial condition, the equilibrium solution may never be reached, even though it is stable. Only numerical simulations can really determine whether or not a dynamic system is stable.
2. A two-phase flow code based on the drift flux approach was developed in order to simulate well configurations without packers. Under this condition, bottom-hole gas segregation and storage effects were considered. For wells equipped with ESP, the two-phase flow pump performance as well as separation models were used. Due to the nonexistence of models for some conditions, some modifications in similar models were proposed.
3. Examples of casing heading and ESP oscillatory behavior were shown. ESP oscillatory behavior can only be captured if the performance curves of the devices are correctly modeled.

25.5 Nomenclature

a_i	Constants in matrix **A**
A	Coefficient matrix
C_0	Distribution parameter
D	Dimension

$f_i(x_1, x_2)$	Generic nonlinear function
$\mathbf{J}$	Jacobian matrix
P	Pressure or trace of a 2 by 2 matrix
q	2 by 2 matrix determinant
Q	Volumetric flow rate
t	Time
V	Velocity
V_d	Drift velocity
V_S	Slip velocity
V_{sg}	Superficial gas velocity
V_{sl}	Superficial liquid velocity
$\dot{x}_i$	First derivative of x_i
$\mathbf{x}$	Column vector of variables x_i
$\dot{\mathbf{x}}$	First derivative of column vector $\mathbf{x}$
$\bar{x}$	Equilibrium solution
z	Position

Greek Letters:

α	Gas void fraction
δ	Disturbance
θ	Angle with horizontal
ρ	Density
φ_{wt}	Two-phase friction loss gradient

Subscript:

$avail.$	Available
$eq.$	Equilibrium
g	Gas
l	Liquid
$req.$	Required

Acknowledgments The authors appreciate the technical and financial support of Tulsa University Artificial Lift Projects' member companies. The progress on this work is the results of the support of Baker-Hughes Centrilift, Chevron, ENI, Kuwait Oil Company, PEMEX, Petrobras, Shell International, Total and Wood Group ESP.

References

[Al93] Alhanati, F.J.S.: Bottomhole gas separation efficiency in electrical submersible pump installations. Ph.D. dissertation, The University of Tulsa (1993)

[BoBeTo73] Boure, J.A., Bergles, A.E., Tong, L.S.: Review of two-phase flow instabilities. Nucl. Eng. Des. **25**, 165–192 (1973)

[CaEtAl09] Carvalho, P.C.G., Prado, M.G., Blanco, J.G., Morooka, C., Estevam, V.: Multiphase performance of ESP stage part II: case study. Technical Report, The University of Tulsa (2009)

[DeGiRi81] Delhaye, J.M., Giot, M., Riethmuller, M.L.: Thermohydraulics of Two-Phase System for Industrial Design and Nuclear Engineering. Hemisphere, Washington (1981)

[Du03] Duran, J.: Pressure effects on esp stages air-water performance. MS thesis, The University of Tulsa (2003)

[HuGo03] Hu, B., Golan, M.: Gas-lift instability resulted production loss and its remedy by feedback control: dynamical simulation results. In: SPE International Improved Oil Recovery Conference in Asia Pacific, no. SPE 84917, Kuala Lumpur (2003)

[IsHi05] Ishii, M., Hibiki, T.: One-dimensional drift–flux model for various flow conditions. In: The Eleventh International Topical Meeting on Nuclear Reactor Thermal-Hydraulics. Avignon, France (2005)

[LaPo89] Lahey, R.T. Jr., Podowski, M.Z.: On the analysis of various instabilities in two-phase flows. In: Matar, O.K., Delhaye, J.M. (eds.) Multiphase Science and Technology, vol. 4, pp. 183–370. Hemisphere, Washington (1989)

[Lo06] Logan, J.D.: Applied Mathematics, 3rd edn. Wiley, New York (2006)

[Pa80] Patankar, S.V.: Numerical Heat Transfer and Fluid Flow. Hemisphere, Washington (1980)

[Sa84] Sachdeva, R.: Two-phase flow through chokes. MS thesis, The University of Tulsa (1984)

[St01] Strogatz, S.H.: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Westview Press, Boulder (2001)

[Vi11] Vieira, R.A.M.: Flow dynamics in oil wells. Ph.D. dissertation, The University of Tulsa (2011)

[Wi09] Wiens, E.G.: Two dimensional flows and phase diagrams. www.egwald.ca/nonlineardynamics/twodimensionaldynamics.php (2009)

[WoGh07] Woldesemayat, M.A., Ghajar, A.J.: Comparison of void fraction correlations for different flow patterns in horizontal and upward inclined pipes. Int. J. Multiphas. Flow **33** (2007)

[ZuFi65] Zuber, N., Findlay, J.A.: Average volumetric concentration in two-phase flow systems. J. Heat Tran. **87**, 453–468 (1965)

Chapter 26
Validating a Closed Form Advection–Diffusion Solution by Experiments: Tritium Dispersion after Emission from the Brazilian Angra Dos Reis Nuclear Power Plant

G.J. Weymar, D. Buske, M.T. Vilhena, and B.E.J. Bodmann

26.1 Introduction

As one of the consequences of the last two nuclear accidents (Chernobyl in 1986 and Fukushima in 2011), nuclear safety regulations have progressively improved. One crucial issue for safety control, emergency plans, and related actions is the knowledge of dispersion of radioactive substances in the planetary boundary layer. While monitoring procedures are a standard routine by the controlled release of Tritium, predicting dispersion of this substance is still a challenge, especially if a rugged orography is present, such as the environment around the Brazilian nuclear power plant Angra dos Reis. Although there are available program platforms that allow to simulate dispersion processes, their underlying models are frequently too simple, frequently based on simple Gaussian models, so that distributions for specific scenarios may only be attained by tuning the simulations according to certain experimental findings instead of predicting them.

The present work is one contribution in a larger program that has the intention to determine general closed form solutions that allow to match a variety of meteorological conditions based on phenomenological approaches for turbulence. A generally accepted deterministic model makes use of Fickian closure and leads thus to an advection–diffusion model for dispersion processes. A well-established method that solves the equation in closed form is based on spectral theory and integral transform, also known as GILTT (see [BuEtAl11], [BuEtAl12], [MoEtAl09], [MoViBu09], [ViEtAl12]). The equation has to be complemented by a known wind profile,

G.J. Weymar • M.T. Vilhena • B.E.J. Bodmann (✉)
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
e-mail: guicefetrs@gmail.com; vilhena@math.ufrgs.br; bardo.bodmann@ufrgs.br

D. Buske
Federal University of Pelotas, Pelotas, Rio Grande do Sul, Brazil
e-mail: daniela.buske@ufpel.edu.br

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7_26,

which is usually determined using experimental meteorological data and the micro-meteorological parameters are calculated from empirical equations established in the literature.

The closed form solution is then applied to the complete set of experiments of the Angra campaign using the associated meteorological conditions. From the comparison of expectation values and measured values the solution is validated and checked for adequacy. Also some comparisons to other approaches are presented.

26.2 The Advection–Diffusion Approach

Upon developing a mathematical dispersion model one typically faces various problems. First one has to identify a differential equation that shall represent the model or the underlying physical law. Once the law/model is accepted as the fundamental equation one challenges the task of solving the equation in many cases approximately and analyze the error of approximation and numerical errors in order to validate its prediction against experimental data. Experimental data of a stochastic process typically spread around average values, i.e. are distributed according to probability distributions. Hence, the model shall within certain limits reproduce the experimental findings. However, the fundamental equation is already a simplification so that deviations may occur which in general have their origin in a model error superimposed by numerical or approximation-based errors. In case of a genuine convergence criterion one may pin down the error analysis essentially to a model validation. Since in general convergence is handled by heuristic convergence criteria, a model validation is not obvious.

For a time-dependent regime considered in the present work, we assume that the associated advection–diffusion equation adequately describes such a dispersion process, which we test by comparison with other methods in order to pin down computational errors and finally analyze for model adequacy. In this line we show with the present discussion that our analytical approach does not only yield a solution for the three-dimensional advection–diffusion equation but predicts tracer concentrations closer to observed values compared to other approaches from the literature, which is also manifest in better statistical coefficients.

Approaches to the advection–diffusion problem are not new in the literature; they are either based on numerical schemes, stochastic simulations or (semi-)analytical methods, as shown in a selection of articles (see [ScFi75], [De78], [NiDe81], [Ti89], [ShSiYa96], [LiHi97]). Note that in these works all solutions are valid for scenarios with strong restrictions with respect to their specific wind and vertical eddy diffusivity profiles. A more general approach, the advection diffusion multilayer method (ADMM) approach solves the two-dimensional advection–diffusion equation with variable wind profile and eddy diffusivity coefficient [MoEtAl06]. The main idea here relies on the discretization of the atmospheric boundary layer in a multilayer domain, assuming in each layer that the eddy diffusivity and wind profile take averaged values. The resulting advection–diffusion equation in each layer is then

solved by the Laplace transform technique. The generalized integral advection–diffusion multilayer technique (GIADMT) method [CoEtAl06] is a dimensional extension to the previous work, but again assuming the stepwise approximation for the eddy diffusivity coefficient and wind profile. In this work we improve the solutions of the aforementioned articles and report on a general analytical solution for the advection–diffusion problem, assuming that eddy diffusivity and wind profiles are arbitrary functions having a continuous dependence on the vertical and longitudinal spatial variables.

Our starting equation is the advection–diffusion equation for the simulation of contaminant or tracer release in the atmospheric boundary layer assuming a Fickian closure for the turbulence. Here, $\bar{c}$ represents the mean concentration of a contaminant (in units of g/m^3) and $\bar{\mathbf{V}} = (\bar{u}, \bar{v}, \bar{w})$ is the mean wind velocity (in m/s) and the domain of interest is a cuboid with $\mathbf{0} \leq \mathbf{r} \leq \mathbf{L}$. Here, the shorthand notation signifies $\mathbf{0} = (0,0,0)$ and $\mathbf{L} = (L_x, L_y, h)$, with h is the height of the atmospheric boundary layer in units of m. The emission source is approximated by a point source (hot spot) with constant emission rate Q (in g/s) at position $\mathbf{r}_s = (0, y_0, H_s)$.

$$\frac{\partial \bar{c}}{\partial t} + \bar{\mathbf{V}} \cdot \nabla \bar{c} = \nabla \cdot (\mathbb{K} \cdot \nabla)\, \bar{c} + S \tag{26.1}$$

In the most general case the diffusion term contains a local and anisotropic (3×3) diffusion coefficient matrix $\mathbb{K}$, which in the present case we assume to be diagonal $\mathbb{K} = \mathrm{diag}(K_x, K_y, K_z)$. The problem is subject to zero flux Neumann-type boundary conditions on the cuboid bounding surface Γ

$$\mathbb{K} \cdot \nabla \bar{c}\big|_{\mathbf{r} \in \Gamma} = \mathbf{0}$$

and initial condition (at $t = 0$)

$$\bar{c} = 0 \quad \forall \mathbf{r} = (x, y, z) \neq \mathbf{r}_s.$$

Instead of including an explicit source term into the advection–diffusion equation, a further constant source flux ($\forall t$) constraint is added to the boundary conditions,

$$\left(\bar{\mathbf{V}} \cdot \hat{\mathbf{x}}\right) \bar{c}\big|_{\mathbf{r}=\mathbf{r}_0} \hat{\mathbf{x}} = Q \delta(y - y_0) \delta(z - H_s) \hat{\mathbf{x}},$$

with the unit vector $\hat{\mathbf{x}} = (1,0,0)$ and $\mathbf{r}_0 = (0, y, z)$.

26.3 A Closed Form Solution

In this section we first introduce the general formalism to solve a general problem and subsequently reduce the problem to a more specific one, that is solved and compared to experimental findings.

26.3.1 General Procedure

In order to solve the problem (26.1) we reduce the dimensionality by one and thus cast the problem into a form already solved in reference [MoEtAl09]. To this end we apply the integral transform technique in the y variable and expand the pollutant concentration as

$$\bar{c}(x,y,z,t) = \mathbf{R}^T(x,z,t)\mathbf{Y}(y), \tag{26.2}$$

where $\mathbf{R} = (R_1, R_2, \ldots)^T$ and $\mathbf{Y} = (Y_1, Y_2, \ldots)^T$ is a vector in the space of orthogonal eigenfunctions, given by $Y_m(y) = \cos(\lambda_m y)$ with eigenvalues $\lambda_m = m\frac{\pi}{L_y}$ for $m = 0, 1, 2, \ldots$. For convenience we introduce some shorthand notations, $\nabla_2 = (\partial_x, 0, \partial_y)^T$ and $\hat{\partial}_y = (0, \partial_y, 0)^T$, so that (26.1) now reads

$$\begin{aligned}
&(\partial_t \mathbf{R}^T)\mathbf{Y} + \bar{\mathbf{U}}\left(\nabla_2 \mathbf{R}^T\mathbf{Y} + \mathbf{R}^T\hat{\partial}_y\mathbf{Y}\right) \\
&\quad = \left(\nabla^T\mathbf{K} + (\mathbf{K}\nabla)^T\right)\left(\nabla_2\mathbf{R}^T\mathbf{Y} + \mathbf{R}^T\hat{\partial}_y\mathbf{Y}\right) \\
&\quad = \left(\nabla_2^T\mathbf{K} + (\mathbf{K}\nabla_2)^T\right)(\nabla_2\mathbf{R}^T\mathbf{Y}) + \left(\hat{\partial}_y^T\mathbf{K} + (\mathbf{K}\hat{\partial}_y)^T\right)(\mathbf{R}^T\hat{\partial}_y\mathbf{Y}).
\end{aligned}$$

Applying the integral operator

$$\int_0^{L_y} dy\mathbf{Y}[\mathbf{F}] = \int_0^{L_y} \mathbf{F}^T \wedge \mathbf{Y}\, dy, \tag{26.3}$$

where $\mathbf{F}$ is an arbitrary function and $\wedge$ signifies the dyadic product, and making use of orthogonality, we rewrite (26.1) as a matrix equation in which the integral terms are

$$\begin{aligned}
\mathbf{B}_0 &= \int_0^{L_y} dy\mathbf{Y}[\mathbf{Y}] = \int_0^{L_y} \mathbf{Y}^T \wedge \mathbf{Y}\, dy, \\
\mathbf{Z} &= \int_0^{L_y} dy\mathbf{Y}[\hat{\partial}_y\mathbf{Y}] = \int_0^{L_y} \hat{\partial}_y\mathbf{Y}^T \wedge \mathbf{Y}\, dy, \\
\mathbf{W}_1 &= \int_0^{L_y} dy\mathbf{Y}[(\nabla_2^T\mathbf{K})(\nabla_2\mathbf{R}^T\mathbf{Y})] = \int_0^{L_y} \left((\nabla_2^T\mathbf{K})(\nabla_2\mathbf{R}^T\mathbf{Y})\right)^T \wedge \mathbf{Y}\, dy, \\
\mathbf{W}_2 &= \int_0^{L_y} dy\mathbf{Y}[(\mathbf{K}\nabla_2)^T(\nabla_2\mathbf{R}^T\mathbf{Y})] = \int_0^{L_y} \left((\mathbf{K}\nabla_2)^T(\nabla_2\mathbf{R}^T\mathbf{Y})\right) \wedge \mathbf{Y}\, dy, \\
\mathbf{T}_1 &= \int_0^{L_y} dy\mathbf{Y}[((\hat{\partial}_y^T\mathbf{K})(\hat{\partial}_y\mathbf{Y})] = \int_0^{L_y} \left(((\hat{\partial}_y^T\mathbf{K})(\hat{\partial}_y\mathbf{Y})\right)^T \wedge \mathbf{Y}\, dy, \\
\mathbf{T}_2 &= \int_0^{L_y} dy\mathbf{Y}[(\mathbf{K}\hat{\partial}_y)^T(\hat{\partial}_y\mathbf{Y})] = \int_0^{L_y} \left((\mathbf{K}\hat{\partial}_y)^T(\hat{\partial}_y\mathbf{Y})\right)^T \wedge \mathbf{Y}\, dy.
\end{aligned}$$

Here, $\mathbf{B}_0 = \frac{L_y}{2}\mathbf{I}$, where $\mathbf{I}$ is the identity, the elements $(\mathbf{Z})_{mn} = \frac{2}{1-n^2/m^2}\delta_{1,j}$ with $\delta_{i,j}$ the Kronecker symbol and $j = (m+n)\text{mod}2$ is the remainder of an integer division (i.e., one for $m+n$ odd and zero else). Note that the integrals $\mathbf{W}_i$ and $\mathbf{T}_i$ depend on the specific form of the eddy diffusivity $\mathbf{K}$. The above integrals are general, but for practical purposes and for application to a case study we truncate the eigenfunction space and consider M components in $\mathbf{R}$ and $\mathbf{Y}$ only, though continue using the general nomenclature that remains valid. The obtained matrix equation determines now together with initial and boundary condition uniquely the components R_i for $i = 1,\ldots,M$ following the procedure introduced in reference [MoEtAl09]:

$$(\partial_t \mathbf{R}^T)\mathbf{B} + \bar{\mathbf{U}}\left(\nabla_2 \mathbf{R}^T\mathbf{B} + \mathbf{R}^T\mathbf{Z}\right) = \mathbf{W}_1(\mathbf{R}) + \mathbf{W}_2(\mathbf{R}) + \mathbf{R}^T(\mathbf{T}_1 + \mathbf{T}_2).$$

26.3.2 A Specific Case for Application

In order to discuss a specific case we introduce a convention and consider the average wind velocity $\bar{\mathbf{U}} = (\bar{u},0,0)^T$ aligned with the x-axis. We superimpose the solution after rotation in the $x-y$-plane in order to transform every instantaneous solution into the same coordinate frame, i.e. the coordinate frame for $t = 0$. By comparison of physically meaningful cases, one finds for the operator norm $||\partial_x K_x \partial_x|| << |\bar{u}|$, which can be understood intuitively because eddy diffusion is observable predominantly perpendicular to the mean wind propagation. As a consequence we neglect the terms with K_x and $\partial_x K_x$.

The principal aspect of interest in pollution dispersion is the vertical concentration profile that responds strongly to the atmospheric boundary layer stratification, so that the simplified eddy diffusivity depends in leading order approximation $\mathbf{K} \to \mathbf{K}_1 = \text{diag}(0, K_y, K_z)$, only on the vertical coordinate $\mathbf{K}_1 = \mathbf{K}_1(z)$. For this specific case the integrals $\mathbf{W}_i$ reduce to

$$\begin{aligned}
\mathbf{W}_1 &\to (\partial_z K_z)(\partial_z \mathbf{R}^T)\mathbf{B},\\
\mathbf{W}_2 &\to K_z(\partial_z^2 \mathbf{R}^T)\mathbf{B},\\
\mathbf{T}_1 &\to \mathbf{0},\\
\mathbf{T}_2 &\to -K_y \Lambda \mathbf{B},
\end{aligned}$$

where $\Lambda = \text{diag}(\lambda_1^2, \lambda_2^2, \ldots)$. Then the simplified equation system to be solved is

$$\partial_t \mathbf{R}^T\mathbf{B} + \bar{u}\partial_x \mathbf{R}^T\mathbf{B} = (\partial_z K_z)\partial_z \mathbf{R}^T\mathbf{B} + K_z\partial_z^2 \mathbf{R}^T\mathbf{B} - K_y \mathbf{R}^T \Lambda \mathbf{B},$$

which is equivalent to the problem

$$\partial_t \mathbf{R} + \bar{u}\partial_x \mathbf{R} = (\partial_z K_z)\partial_z \mathbf{R} + K_z\partial_z^2 \mathbf{R} - K_y \Lambda \mathbf{R}, \tag{26.4}$$

by virtue of $\mathbf{B}$ being a diagonal matrix.

Once the problem (26.4) is solved by the GILTT method, the solution of problem (26.1) is well determined. In reference [MoEtAl09] a two-dimensional problem with advection in the x direction in stationary regime was solved which has the same formal structure than (26.4) except for the time dependence. We apply the Laplace Transform in the t variable, $(t \to r)$ obtaining the following pseudo-steady-state problem:

$$r\tilde{\mathbf{R}}_0 + \overline{u}\partial_x \tilde{\mathbf{R}}_0 = \partial_z \left(K_z \partial_z \tilde{\mathbf{R}}_0\right) - \Lambda\, K_y \tilde{\mathbf{R}}_0. \tag{26.5}$$

The x and z dependence may be separated using the same reasoning as already introduced in (26.2). To this end we pose the solution of problem (26.5) in the form

$$\tilde{\mathbf{R}}_0 = \mathbf{PC},$$

where $\mathbf{C} = (\zeta_1(z), \zeta_2(z), \ldots)^T$ are a set of orthogonal eigenfunctions, given by $\zeta_i(z) = \cos(\gamma_i z)$, and $\gamma_i = i\pi/h$ (for $i = 0, 1, 2, \ldots$) are the set of eigenvalues.

Replacing this in (26.5) and using (26.3) with respect to the z-dependent degrees of freedom, that is,

$$\int_0^h dz \mathbf{C}[\mathbf{F}] = \int_0^h \mathbf{F}^T \wedge \mathbf{C}\, dz,$$

we arrive at the first-order differential equation system

$$\partial_x \mathbf{P} + \mathbf{HP} = \mathbf{0}, \tag{26.6}$$

where $\mathbf{P} = \mathbf{P}(x, r)$ and $\mathbf{H} = \mathbf{B}_1^{-1}\mathbf{B}_2$. The entries of matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ are

$$\begin{aligned}
(\mathbf{B}_1)_{i,j} &= -\int_0^h \overline{u}\zeta_i(z)\zeta_j(z)\, dz \\
(\mathbf{B}_2)_{i,j} &= \int_0^h \partial_z K_z \partial_z \zeta_i(z)\zeta_j(z)\, dz - \gamma_i^2 \int_0^h K_z \zeta_i(z)\zeta_j(z)\, dz \\
&\quad -r\int_0^h \zeta_i(z)\zeta_j(z)\, dz - \lambda_i^2 K_y \int_0^h \zeta_i(z)\zeta_j(z)\, dz.
\end{aligned}$$

Following the reasoning in [MoEtAl09], we solve (26.6) applying Laplace transform and diagonalization of the matrix $\mathbf{H} = \mathbf{XDX}^{-1}$, which results in

$$\tilde{\mathbf{P}}(s, r) = \mathbf{X}(s\mathbf{I} + \mathbf{D})^{-1}\mathbf{X}^{-1}\mathbf{P}(0, r), \tag{26.7}$$

where $\tilde{\mathbf{P}}(s, r)$ denotes the Laplace Transform of $\mathbf{P}(x, r)$. Here $\mathbf{X}^{(-1)}$ is the (inverse) matrix of the eigenvectors of matrix $\mathbf{B}_1^{-1}\mathbf{B}_2$ with diagonal eigenvalue matrix $\mathbf{D}$ and the entries of matrix $(s\mathbf{I} + \mathbf{D})_{ii} = s + d_i$. After performing the Laplace transform inversion of (26.7), we get

$$\mathbf{P}(x,r) = \mathbf{X}\mathbf{G}(x,r)\mathbf{X}^{-1}\xi,$$

where $\mathbf{G}(x,r)$ is the diagonal matrix with components $(\mathbf{G})_{ii} = e^{-d_i x}$. In addition, the still unknown arbitrary constant matrix is given by $\xi = \mathbf{X}^{-1}\mathbf{P}(0,r)$.

The time dependence is obtained upon applying the inverse Laplace transform definition

$$\mathbf{R}_0(x,z,t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mathbf{P}(x,r)\mathbf{C}(z)e^{rt}\,dr\,.$$

To overcome the drawback of evaluating a line integral, we perform the calculation of this integral by the Gaussian quadrature scheme, which is exact if the integrand is a polynomial of degree $2M-1$ in the $\frac{1}{r}$ variable

$$\mathbf{R}_0(x,z,t) = \frac{1}{t}\mathbf{a}^T\left(\mathbf{p}\mathbf{R}_0(x,z,\frac{\mathbf{p}}{t})\right)\,, \tag{26.8}$$

where **a** and **p** are, respectively, vectors with the weights and roots of the Gaussian quadrature scheme [StSe66].

26.4 Experimental Data and Turbulent Parametrization

For model validation we chose a controlled release of radioactive material performed in 1985 at the Itaorna Beach, close to the nuclear reactor site Angra dos Reis in the Rio de Janeiro state, Brazil. Details of the dispersion experiment are described elsewhere [BiEtAl85]. The experiment consisted in the controlled releases of radioactive tritium loaded water vapor from the meteorological tower at 100 m height during 5 days (November 28 to December 4, 1984). During the whole experiment, four meteorological towers collected the relevant meteorological data. Wind speed and direction were measured at three levels (10, 60, and 100 m) together with the temperature gradients between 10 and 100 m. Some additional data of relative humidity were available in some of the sampling sites, and were used to calculate the concentration of radioactive tritium loaded water in the air (after measuring the radioactivity of the collected samples). All relevant details, as well as the synoptic meteorological conditions during the dispersion campaign are described in [BiEtAl85]. The data from the 5 experiments were used to obtain the numerical results and are presented in Table 26.1.

The micro-meteorological parameters shown in Table 26.1 are calculated from equations obtained in the literature. The roughness length utilized was 1 m and the Monin–Obukhov length for convective conditions can be written as [Za90]

$$L = -h/k\,(u_*/w_*)^3\,,$$

Table 26.1 Micro-meteorological parameters and emission rate for the Angra dos Reis experiments

Experiment	Period	$\overline{u}(10)$ ($m\,s^{-1}$)	u_* ($m\,s^{-1}$)	w_* ($m\,s^{-1}$)	L (m)	h (m)	Q ($MBq\,s^{-1}$)
1	1	1.83	0.32	0.46	−809.51	965.09	20.46
	2	2.43	0.42	0.60	−1056.86	1259.98	20.46
	3	2.76	0.48	0.69	−1214.26	1447.64	20.46
2	1	2.59	0.44	0.63	−1108.58	1321.64	25.34
	2	2.21	0.38	0.55	−966.91	1152.75	25.34
	3	2.18	0.38	0.54	−951.17	1133.98	25.34
3	1	2.21	0.38	0.55	−966.91	1152.75	20.46
	2	1.97	0.34	0.49	−861.23	1026.75	20.46
	3	2.61	0.46	0.66	−1146.81	1367.21	20.46
4	1	1.23	0.21	0.31	−539.67	643.40	24.34
	2	1.01	0.18	0.25	−440.73	525.44	24.34
	3	1.05	0.18	0.26	−456.47	544.21	24.34
5	1	1.95	0.34	0.49	−854.48	1018.71	31.32
	2	1.54	0.27	0.39	−674.59	804.24	31.32
	3	2.61	0.45	0.65	−1137.81	1356.49	31.32

where k is the von Karman constant ($k = 0.4$), w_* is the convective velocity scale with wind speed U, $u_* = kU/\ln(z_r/z_0)$ is the friction velocity, where U is the wind velocity at the reference height $z_r = 10\,\text{m}$, and $h = 0.3u_*/f_c$ is the height of the boundary layer with the Coriolis coefficient $f_c = 10^{-4}$.

In the atmospheric diffusion problems the choice of a turbulent parametrization represents a fundamental aspect for contaminant dispersion modeling. From the physical point of view a turbulence parametrization is an approximation for the natural phenomenon, where details are hidden in the parameters that are being used and have to be adjusted in order to reproduce experimental findings. The reliability of each model strongly depends on the way the turbulent parameters are calculated and related to the current understanding of the planetary boundary layer. In terms of the convective scaling parameters the vertical and lateral eddy diffusivity can be formulated as in [DeCVCa97], namely

$$K_z = 0.22 w_* h \left(\frac{z}{h}\right)^{\frac{1}{3}} \left(1 - \frac{z}{h}\right)^{\frac{1}{3}} \left(1 - e^{\frac{4z}{h}} - 0.0003 e^{\frac{8z}{h}}\right), \tag{26.9}$$

$$K_y = \frac{\sqrt{\pi}\sigma_v}{16 (f_m)_v q_v}, \tag{26.10}$$

where

$$\sigma_v^2 = \frac{0.98 c_v}{(f_m)_v^{\frac{2}{3}}} \left(\frac{\psi_\varepsilon}{q_v}\right)^{\frac{2}{3}} \left(\frac{z}{h}\right)^{\frac{2}{3}} w_*^2,$$

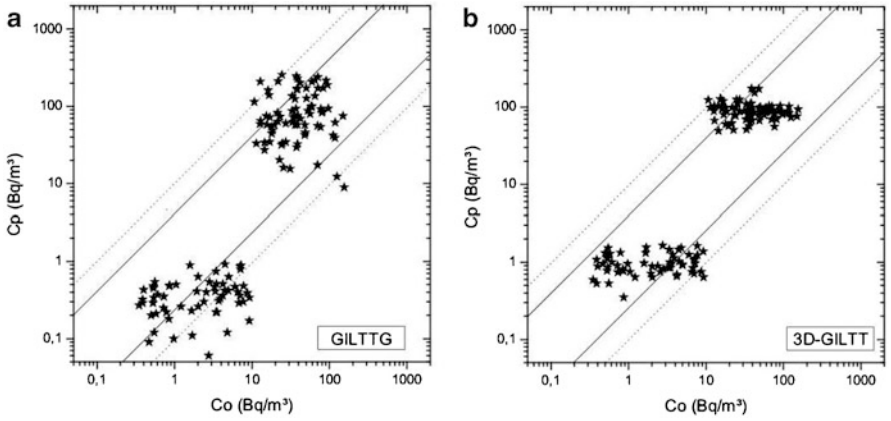

Fig. 26.1 Scatter diagram of the observed versus predicted maximum ground level concentrations. Data between *lines* correspond to a factor of two and five

$$q_v = 4.16\frac{z}{h}, \quad \psi_\varepsilon^{\frac{1}{3}} = \left(\left(1-\frac{z}{h}\right)^2\left(-\frac{z}{L}\right)^{-\frac{2}{3}} + 0.75\right)^{\frac{1}{2}}, \quad (f_m)_v = 0.16,$$

σ_v is the standard deviation of the longitudinal turbulent velocity component, q_v is the stability function, ψ_ε is the dimensionless molecular dissipation rate, and $(f_m)_v$ is the transverse wave peak.

The wind speed profile can be described by a power law $u_z/u_1 = (z/z_1)^n$ [PaDu88], where u_z and u_1 are the horizontal mean wind speeds at heights z and z_1, and n is an exponent related to the intensity of turbulence [Ir79].

26.5 Numerical Results

In this study we introduce the vertical and lateral eddy diffusivities ((26.9) and (26.10)) and the power law wind profile in the 3D-GILTT model to calculate the ground-level concentration of emissions released from an elevated continuous source point in an unstable/neutral atmospheric boundary layer.

The validation of the 3D-GILTT model predictions against experimental data from the Angra site together with a two-dimensional model (GILTTG) are shown in Fig. 26.1. While the present approach (3D-GILTT) is based on a genuine three-dimensional description an earlier analytical approach (GILTTG) uses a Gaussian assumption for the horizontal transverse direction [MoEtAl09]. The 3D-GILTT approach reproduces acceptably the observed concentrations, although this simulation did not make use of the terrain's realistic complexity.

In the further we use the standard statistical indices in order to compare the quality of the two approaches. Note that we present the two analytical model

Table 26.2 Statistical comparisons between GILTTG and 3D-GILTT results

Statistical indices	GILTTG	3D-GILTT
NMSE = $\overline{(C_o - C_p)^2}/\overline{C_p}\,\overline{C_o}$	2.82	1.44
COR = $\overline{(C_o - \overline{C_o})(C_p - \overline{C_p})}/\sigma_o\sigma_p$	0.46	0.59
FA2 = $0.5 \leq (C_p/C_o) \leq 2$	0.32	0.38
FA5 = $0.2 \leq (C_p/C_o) \leq 5$	0.67	0.80
FB = $\overline{C_o} - \overline{C_p}/0.5(\overline{C_o} + \overline{C_p})$	−0.62	−0.59
FS = $(\sigma_o - \sigma_p)/0.5(\sigma_o + \sigma_p)$	−0.69	−0.37

approaches, since the earlier one was found to be acceptable in comparison with other approaches found in the literature and both give a solution in closed form. The standard statistical indices are NMSE, the normalized mean square error; COR, the correlation coefficient; FA2 and FA5, the fraction of data (in %) in the cones determined by a factor of two and five, respectively; FB, the fractional bias and FS, the fractional standard deviation. The subscripts o and p refer to observed and predicted quantities, respectively, and $\bar{C}$ indicates the averaged values. Table 26.2 presents the results of the statistical indices used to evaluate the model performance [Ha89] and further compare our model to the GILTTG approach. The statistical index FB indicates whether the predicted quantities (C_p) under- or overestimates the observed ones (C_o). The statistical index NMSE represents the quadratic error of the predicted quantities in relation to the observed ones. Best results are indicated by values compatible with zero for NMSE, FB, and FS, and compatible with unity for COR, FA2, and FA5. The statistical indices point out that a reasonable agreement is obtained between experimental data and the 3D-GILTT model.

In order to validate the two models we fit the predicted versus observed values by a linear regression, where the closer their intersect to the origin and the closer the slope is to unity the better is the approach. The GILTTG approach results in $\bar{C}_p = 0.95\bar{C}_o + 26.53$ with $R^2 = 0.46$ and $\kappa = 0.95$, whereas the 3D-GILTT obeys the result $\bar{C}_p = 0.86\bar{C}_o + 27.61$ with $R^2 = 0.59$ and $\kappa = 0.99$. In order to perform a model validation we introduced an index

$$\kappa = \sqrt{(a-1)^2 + \left(b/\bar{C}_o\right)^2},$$

where

$$\bar{C}_o = \frac{1}{n}\sum_{i=1}^{n} C_{oi},$$

which if identical zero indicates a perfect match between the model and the experimental findings. Here a is the slope, b the intersection, C_{oi} of the experimental data, and $\bar{C}_o$ its arithmetic mean. Since the experiment is of stochastic character whereas the stochastic properties are hidden in the model parameters, considerable

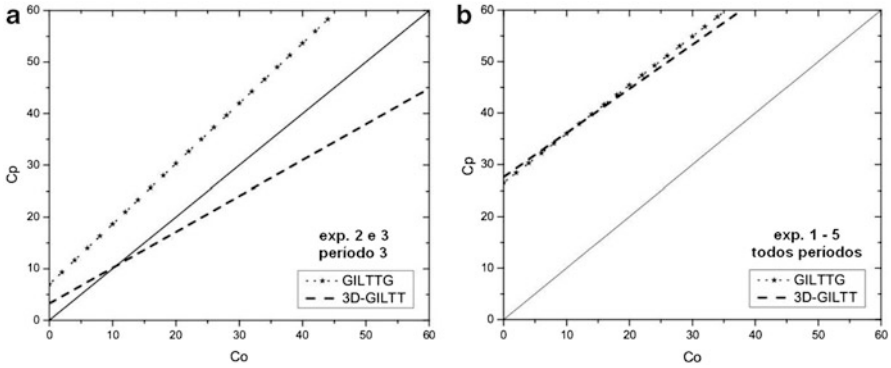

Fig. 26.2 Linear regression for the GILTTG and 3D-GILTT. The bisector was added as an eye guide

fluctuations are present. Nevertheless, by comparison (see Fig. 26.2) one observes that the present approach yields the better description of the data.

26.6 Conclusions

The present work was based on an Eulerian approach to determine dispersion of radioactive contaminants in the planetary boundary layer. To this end the diffusion equation for the cross-wind integrated concentrations was closed by the relation of the turbulent fluxes to the gradient of the mean concentration by means of eddy diffusivity (K-theory). We are completely aware of the fact that K-closure has its intrinsic limits so that one would like to remove these inconsistencies. However, comparisons of predictions by this approach to experimental data have shown that there are scenarios where this lack is not significantly manifest, which we use as a justification together with its computational simplicity to perform our simulations based on this approach. Moreover, the present work may be understood as one tile in a larger program development that simulates radioactive material dispersion using analytical resources. In a longer term we intend to build a library that allows to predict radioactive material transport in the planetary boundary layer that extends from the micro- to the meso-scale. In this sense this contribution is a step into this direction.

In the present discussion we restricted our comparison with the two-dimensional and GILTT approach only, since its usefulness was already proven [BuEtAl10] and the specification of diffusion and wind profile are identical. Other approaches like ADMM among others make use of step-wise approximations for $\mathbb{K}$ and $\mathbf{V}$ or determine the velocity field from large eddy simulations, in other words they are not self-contained. Although the measurements were at ground level one could think that a two-dimensional approach would suffice, the present comparison clearly

shows the influence of the additional dimension. While in the two-dimensional approach the tendency of the predicted concentrations is to overestimate the observed values in the experiments 2 and 3, this is not the case for the results of the three-dimensional description, mainly because it does not assume turbulence to be homogeneous. In the remainder of the runs (1, 4, and 5) considerably larger variations of the mean wind velocity as well as lower wind velocities among others are not compatible with some of the simplifications that were made in order to obtain the solution for the studied case (compare the model validation in Fig. 26.2, left and right). However, the solution method of the advection diffusion equation discussed here is more general than shown in the present context, so that in principle a wider range of applications is possible. Especially, other assumptions for the velocity field and the diffusion matrix are possible and also necessary. In a future work we will focus on a variety of applications and introduce a rigorous proof of convergence of the method from a mathematical point of view.

References

[BiEtAl85] Biagio, R., Godoy, G., Nicoli, I., Nicolli, D., Thomas, P.: First atmospheric diffusion experiment campaign at the Angra site, KfK 3936, Karlsruhe, and CNEN 1201, Rio de Janeiro (1985)

[BuEtAl10] Buske, D., Vilhena, M.T., Moreira, D.M., Tirabassi, T.: An analytical solution for the transient two-dimensional advection–diffusion equation with non-Fickian closure in Cartesian geometry by integral transform technique. In: Constanda, C., Perez, M.E. (eds.) Integral Methods in Science and Engineering. Computational Methods, pp. 33–40. Birkhäuser, Boston (2010)

[BuEtAl11] Buske, D., Vilhena, M.T., Segatto, C.F., Quadros, R.S.: A general analytical solution of the advection–diffusion equation for Fickian closure. In: Constanda, C., Harris, P.J. (eds.) Integral Methods in Science and Engineering. Computational and Analytic Aspects, pp. 25–34. Birkhäuser, Boston (2011)

[BuEtAl12] Buske, D., Vilhena, M.T., Bodmann, B., Tirabassi, T.: Analytical model for air pollution in the atmospheric boundary layer. In: Khare, M. (ed.) Air Pollution-Monitoring, Modelling and Heath, pp. 39–58. InTech (2012)

[CoEtAl06] Costa, C.P., Vilhena, M.T., Moreira, D.M., Tirabassi, T.: Semi-analytical solution of the steady three-dimensional advection–diffusion equation in the planetary boundary layer. Atmos. Environ. **40**, 5659–5669 (2006)

[DeCVCa97] Degrazia, G.A., Campos Velho, H.F., Carvalho, J.C.: Nonlocal exchange coefficients for the convective boundary layer derived from spectral properties. Contrib. Atmos. Phys. **70**, 57–64 (1997)

[De78] Demuth, C.: A contribution to the analytical steady solution of the diffusion equation for line sources. Atmos. Environ. **12**, 1255–1258 (1978)

[Ha89] Hanna, S.R.: Confidence limit for air quality models as estimated by bootstrap and jackknife resampling methods. Atmos. Environ. **23**, 1385–1395 (1989)

[Ir79] Irwin, J.S.: A theoretical variation of the wind profile power-low exponent as a function of surface roughness and stability. Atmos. Environ. **13**, 191–194 (1979)

[LiHi97] Lin, J.S., Hildemann, L.M.: A generalised mathematical scheme to analytically solve the atmospheric diffusion equation with dry deposition. Atmos. Environ. **31**, 59–71 (1997)

[MoEtAl06] Moreira, D.M, Vilhena, M.T., Tirabassi, T., Costa, C., Bodmann, B.: Simulation of pollutant dispersion in atmosphere by the Laplace transform: the ADMM approach. Water Air Soil Pollut. **177**, 411–439 (2006)

[MoViBu09] Moreira, D.M., Vilhena, M.T., Buske, D.: On the GILTT formulation for pollutant dispersion simulation in the atmospheric boundary layer. In: Air Pollution and Turbulence: Modeling and Applications, pp. 179–202. CRC Press, Boca Raton (2009)

[MoEtAl09] Moreira, D.M., Vilhena, M.T., Buske, D., Tirabassi, T.: The state-of-art of the GILTT method to simulate pollutant dispersion in the atmosphere. Atmos. Res. **92**, 1–17 (2009)

[NiDe81] Nieuwstadt, F.T.M., de Haan, B.J.: An analytical solution of one-dimensional diffusion equation in a non-stationary boundary layer with an application to inversion rise fumigation. Atmos. Environ. **15**, 845–851 (1981)

[PaDu88] Panofsky, A.H., Dutton, J.A.: Atmospheric Turbulence. Wiley, New York (1988)

[ScFi75] Scriven, R.A., Fisher, B.A.: The long range transport of airborne material and its removal by deposition and washout. II. The effect of turbulent diffusion. Atmos. Environ. **9**, 59–69 (1975)

[ShSiYa96] Sharan, M., Singh, M.P., Yadav, A.K.: A mathematical model for the atmospheric dispersion in low winds with eddy diffusivities as linear functions of downwind distance. Atmos. Environ. **30**, 1137–1145 (1996)

[StSe66] Stroud, A.H., Secrest, D.: Gaussian Quadrature Formulas. Prentice Hall, Englewood Cliffs (1966)

[Ti89] Tirabassi, T.: Analytical air pollution and diffusion models. Water Air Soil Pollut. **47**, 19–24 (1989)

[ViEtAl12] Vilhena, M.T., Buske, D., Degrazia, G.A., Quadros, R.S.: An analytical model with temporal variable eddy diffusivity applied to contaminant dispersion in the atmospheric boundary layer. Phys. A **391**, 2576–2584 (2012)

[Za90] Zanetti, P.: Air Pollution Modeling. Computational Mechanics Publications, Southampton (1990)

Index

C. Constanda et al. (eds.), *Integral Methods in Science and Engineering: Progress in Numerical and Analytic Techniques*, DOI 10.1007/978-1-4614-7828-7,