Springer Proceedings in Mathematics & Statistics

Susmita Sarkar
Uma Basu
Soumen De *Editors*

Applied Mathematics

Kolkata, India, February 2014

Springer

Springer Proceedings in Mathematics & Statistics

Volume 146

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Susmita Sarkar · Uma Basu · Soumen De
Editors

Applied Mathematics

Kolkata, India, February 2014

Springer

Editors
Susmita Sarkar
Department of Applied Mathematics
University of Calcutta (UoC)
Kolkata, West Bengal
India

Soumen De
Department of Applied Mathematics
University of Calcutta (UoC)
Kolkata, West Bengal
India

Uma Basu
Department of Applied Mathematics
University of Calcutta (UoC)
Kolkata, West Bengal
India

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-81-322-2546-1 ISBN 978-81-322-2547-8 (eBook)
DOI 10.1007/978-81-322-2547-8

Library of Congress Control Number: 2015946738

Mathematics Subject Classification (2010): 53-XX, 76HXX, 74B-XX, 85AXX, 92F99, 91B80, 97MXX

Springer New Delhi Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer (India) Pvt. Ltd. is part of Springer Science+Business Media (www.springer.com)

In memory of
Sir Asutosh Mookerjee,
S.N. Bose,
M.N. Saha
and N.R. Sen

Preface

The year 2014 is the 150th birth anniversary year of the great personality Sir Asutosh Mookerjee who designed the University of Calcutta as a premier seat of education and research. Under the initiative of Sir Asutosh Mookerjee, in 1914, different postgraduate departments were opened in the University of Calcutta for carrying on advanced level teaching and research and the University of Calcutta then achieved a new dimension. The University College of Science, at 92, Upper Circular Road, was established in 1914 from the generous funding of Sir Taraknath Palit and Sir Rashbehary Ghose. The postgraduate departments in applied sciences including applied mathematics were started there in 1914. At that time, this Applied Mathematics Department was named Mixed Mathematics Department.

Professor Satyendranath Bose, Professor Meghnad Saha, and Professor Nikhil Ranjan Sen were students at the Mixed Mathematics Department during the session 1913–1915. The great visionary Sir Asutosh Mookerjee identified these budding scientists and appointed them as teachers immediately after their passing. Apart from these three legendary scientists, last 100 years of this department were also enriched by many other applied mathematicians like Prof. Ganesh Prasad (first Sir Rashbehary Ghose Professor of the University of Calcutta), Prof. B.B. Dutta, Prof. S. Banerjee, Prof. N.N. Sen, Prof. B.S. Roy, Prof. S.K. Chakraborty, and others for their illustrious contributions in different fields of Applied Mathematics.

Department of Applied Mathematics, University of Calcutta, has organized an international conference on “Emerging Trends in Applied Mathematics: Dedicated to the Memory of Sir Asutosh Mookerjee and Contributions of S.N. Bose, M.N. Saha and N.R. Sen” in collaboration with Saha Institute of Nuclear Physics, Kolkata, and Indian Association for the Cultivation of Science, Kolkata, during February 12–14, 2014, to commemorate these glorious events. This international conference was also financially supported by the National Board for Higher Mathematics (NBHM), Government of India and Department of Higher Education, Government of West Bengal.

This international conference was started with the song “subho karmo pathe dharo nirbhayo gaan” which is the theme song of the University of Calcutta written by Kabiguru Rabindranath Tagore. The program was inaugurated by Justice Chittatosh Mookerjee, former chief justice of the Bombay High Court and grandson of Sir Asutosh Mookerjee.

The keynote address was delivered by Professor Ralph Abraham, emeritus professor of mathematics, University of California, Santa Cruz, USA. The whole program was divided into sessions dedicated to Sir Asutosh Mookerjee, Prof. Satyendranath Bose, Prof. Meghnad Saha, and Prof. Nikhil Ranjan Sen. Eminent scientists from all over the world delivered lectures at different sessions. Both oral and poster sessions were arranged for contributed papers.

"Sounds of Space", harmony with the audio visual raga melodies of Indian classical music was presented by Prof. Chanchal Uberoi, former professor of mathematics, Indian Institute of Science, Bangalore.

A cultural program with violin recital followed by conference dinner was arranged in the evening of 13th February. This heritage conference was ended with the National Anthem.

The lectures of the invited speakers and the contributed papers is being published in this proceedings. All papers have been refereed by national and international referees. We are grateful to all the authors, referees, and the publisher for their kind cooperation for successful completion of this book.

We also express our grateful thanks to Prof. Dilip Kumar Sinha, former Sir Rashbehary Ghose Professor of the Department of Applied Mathematics, University of Calcutta and former Vice-Chancellor of Viswabharati for his advice and assistance in the preparation of the writeup entitled "Bedrock of Applied Mathematics in Realms of Calcutta University."

Susmita Sarkar
Uma Basu
Soumen De

Acknowledgments

It is our pleasure to present this volume consisting of invited lectures and selected papers based on oral and poster presentations at the international conference on “Emerging Trends in Applied Mathematics: Dedicated to the Memory of Sir Asutosh Mookerjee and Contributions of S.N. Bose, M.N. Saha and N.R. Sen” held during February 12–14, 2014, at the Department of Applied Mathematics, University of Calcutta.

We gratefully acknowledge the wholehearted support of Prof. Milan Kumar Sanyal, former director, Saha Institute of Nuclear Physics, Kolkata, and Prof. Debshankar Roy, director, Indian Association for the Cultivation of Science, Kolkata, to organize this conference in collaboration with their institutes. We also acknowledge the kind support of Prof. Arup Raichaudhuri, director, S.N. Bose Centre for Basic Science, Kolkata, for providing us accommodation for the delegates at the institute guest house. We are grateful to Prof. Bimal Roy, director, Indian Statistical institute, Kolkata, for his kind supports from the part of National Board for Higher Mathematics (NBHM), Government of India. We are also grateful to Mr. Bratya Basu, former minister-in-charge and Mr. Vivek Kumar, secretary, Department of Higher Education, Government of West Bengal, for providing us financial support to organize the conference.

We are extremely thankful to Prof. Bikas K. Chakraborty, former director, Saha Institute of Nuclear Physics, Kolkata, for his valuable guidance at every step of the conference. We are also grateful to him for providing us valuable pictures of Prof. M.N. Saha from SINP archive. We are indebted to our honorable vice-chancellor, Prof. Suranjan Das, for his constant encouragement and financial as well as moral support which has enabled us to organize such a commemorative conference at the international level. We express sincere gratitude to Prof. Mamata Ray, honorable former pro vice-chancellor (business affairs and finance) of the University of Calcutta for her wholehearted support in organizing this conference. We also express sincere gratitude to Prof. Dhrubajyoti Chattopadhyay, honorable pro vice-chancellor (academic affairs), Prof. Sonali Chakraborty Banerjee, honorable pro vice-chancellor (business affairs and finance), Prof. Basab Chaudhuri (registrar), Prof. Asutosh Ghosh (dean, faculty of science), Dr. Hari Sadhan Ghosh (finance officer), and Dr. Amit Ray (secretary, faculty of science) for extending all facilities in organizing this conference.

We are thankful to our following invited speakers from India and abroad for delivering lectures, which has enriched our conference.

Prof. Ralph Abraham, University of California, Santa Cruz, USA.
Prof. Peter Leach, University of Kwazulu Nata, Durban.
Prof. Eduardo Massad, University of Sao Paolo, Brazil.
Prof. Jun-Ichi-Inoue (Deceased), Hokkaido University, Japan.
Prof. Siddhartha Sen, University College of Dublin, Ireland.
Prof. Arup Mukherjee, Montclair State University, Montclair, USA.
Prof. Suhrit De, Eastern Illinoi University, Charlston, USA.
Prof. Subhas Chandra Basak, University of Minnesota, Duluth, USA.
Prof. Jayant Vishnu Narlikar, Inter University Centre for Astronomy and Astrophysics, Pune, India.
Prof. Amiya Gopal Mukherjee, Indian Statistical Institute, Kolkata, India.
Prof. Chanchal Uberoi, Indian Institute of Science, Bangalore, India.
Prof. Amit Apte, Tata Institute of Fundamental Research, Bangalore, India.
Prof. Sunil Chakraborty, National Aeronautical Laboratory, Bangalore, India.
Prof. Amita Das, Institute for Plasma research, Gujrat, India.
Prof. Sitabhra Sinha, Institute of Mathematical Sciences, Chennai, India.
Prof. Siddhartha Pratim Chakraborty, Indian Institute of Technology, Guwahati, India.
Prof. Bikas K. Chakraborty, Saha Institute of Nuclear Physics, Kolkata, India.
Prof. Sunanda Banerjee, Saha Institute of Nuclear Physics, Kolkata, India.
Prof. Birendranath Mandal, Indian Statistical Institute, Kolkata, India.
Prof. Krishnendu Sengupta, Indian Association for the Cultivation of Science, Kolkata, India.
Prof. Arabinda Roy, University of Calcutta, Kolkata, India.
Our greetings also go to all the participants who presented their paper at the conference.

We are also thankful to them who kindly agreed to chair different sessions of the conference.

Our gratitude is due to Prof. Dilip Kumar Sinha, retired Sir Rashbehary Ghose Professor of Applied Mathematics Department of Calcutta University who first sent the proposal to the head of the department for organizing this commemorative conference. We are thankful to Dr. Kausik Bal, development and planning officer of the University of Calcutta for encouraging us to organize this heritage conference at the international level.

We also express sincere thanks to Prof. Partha Guha, S.N. Bose Centre for Basic Science, Kolkata, for his constant help at every step of the conference. We are thankful to all members of the advisory committee and organizing committee for extending help and support as and when required.

We would also like to extend our gratitude to the reviewers of the papers submitted for consideration in this volume for having so generously shared their time and expertise.

We are thankful to World Book Traders, International Book Company and A.M. Book Stall for their support to this conference.

We are thankful to the officers and staffs of our engineering department, accounts department, UGC section, and also the staff of our other administrative sections (office of the vice-chancellor, pro vice-chancellors, registrar, finance officer, and secretary) for their active support and cooperation towards the conference.

Finally, we would like to thank Prof. Bikas K. Chakraborty, former director and Ms. Dipa Dasgupta, scientific officer of Saha Institute of Nuclear Physics, Kolkata, for kindly allowing us to use pictures of Meghnad Saha in the book. We wish to thank our colleagues, research scholars, students, and nonteaching staffs who contributed greatly to the organization and success of the conference.

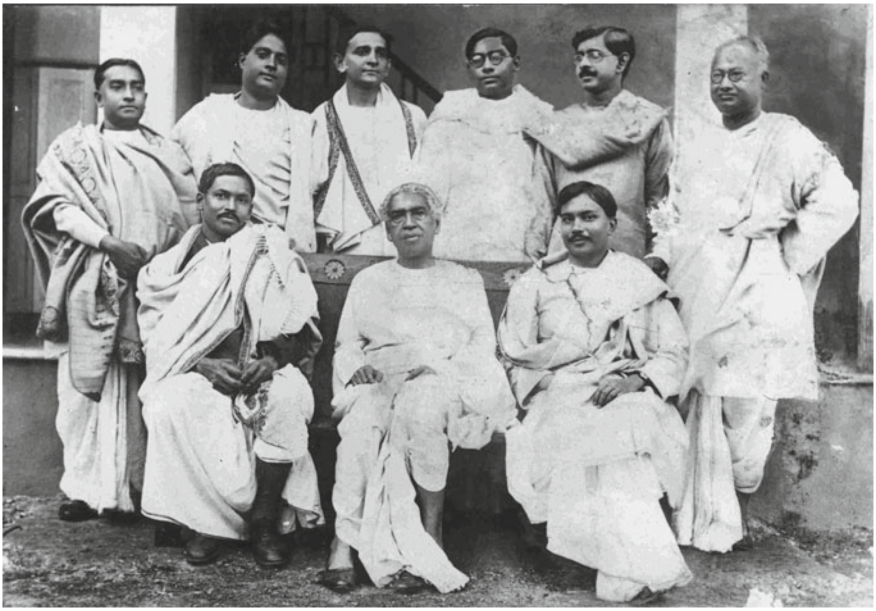
Bedrock[1] of Applied Mathematics in Realms of Calcutta University

As indicated earlier, the chief *raison d' etre* of the international conference is to recapture, in terms of contemporary times, the moorings of Applied (Mixed) Mathematics and its closely allied discipline Pure/Theoretical Physics. This happens date back to the early decades of the twentieth century. A simple hindsight shows that the relevant bedrock of such pursuits would not have been a reality, had not one of the oldest universities, University of Calcutta, developed a new trajectory, uniquely of its own. As history says, that could happen because of Sir Asutosh Mookerjee was the vice-chancellor of the University of Calcutta. Sir Asutosh Mookerjee, factually speaking, was a student and an avid researcher in Mathematics. He had his schooling in South Calcutta but, later on, moved over to the then Presidency College, Calcutta. It was mainly because of his passionate attachment to Geometry, he could make his debut through a paper published in Cambridge, the UK. He was a contemporary of many leading figures in Calcutta who played important roles in what is often called Renaissance in Bengal. He could show off his brilliance in Mathematics in both postgraduate and undergraduate examinations at the University of Calcutta. Sir Asutosh Mookerjee carried on his researches mainly on Differential Equations of Geometrical Curves, being inspired by the French Geometer G. Monge. Sir Asutosh built up linkages at the intellectual level with individual and organizations in the global level.

[1] Sir Asutosh Mookerjee, Satyendranath Bose, Meghnad Saha and Nikhil Ranjan Sen form the foundation of Applied Mathematics in Calcutta University.

Sir Asutosh Mookerjee in the *first row* (*third from the left*). *Source* Meghnad Saha Archives

The leadership of Sir Asutosh Mookerjee at the University of Calcutta was highly innovative in that he transformed the whole university into an academically vibrant institution. It is to be noted that Sir Asutosh Mookerjee, as a researcher, did not fully opt out for Geometry. Indeed, he had two papers on Hydrokinetics. Also, prior to that, he went through another postgraduate examination at the university on Physical Sciences. The Indian Association for Cultivation of Science (IACS) at Bowbazar, Calcutta, provided Sir Asutosh Mookerjee opportunities for his lectures on Mathematics and Physical Sciences. The Asiatic Society of Bengal, of course, witnessed his presentations and lectures on a variety of research and educational areas.



Seated *left to right* Meghnad Saha, Jagdish Chandra Bose, J.C. Ghosh. Standing *left to right* Snehamoy Dutta, Satyendranath Bose, D.M. Bose, N.R. Sen, J.N. Mukherjee and N.C. Nag. *Source* Meghnad Saha Archives

Sir Asutosh could build up, by then, a yearning for developing something abiding in the premises of the University of Calcutta. Some departments, reflecting his thought processes, were very much in the air of his times. Obviously, Albert Einstein's Special Theory of Relativity, in 1905, prevailed on him so much as to initiate the creation of departments and faculties, unfolding confluences of both Mathematics and Physics that accounted for his recruiting young scholars like Satyendra Nath Bose, Meghnad Saha, and Nikhil Ranjan Sen at the Department of Applied Mathematics, in close alliance with the Department of Physics. Later on, this could become all the more refueled through his encouragement of the publication, through Calcutta University Press, the translation of Einstein's classic paper on Relativity from German into English, by S.N. Bose, M.N. Saha with a preface by Prasanta Chandra Mahalanobis. It looks that Prof. N.R. Sen might have been inspired to proceed for researches in Germany and, later on, in particular after his return for further pursuits on Relativity.

Satyendranath Bose (*right*) with Albert Einstein. *Source* Meghnad Saha Archives

It is fairly known that Sir Asutosh did not have an entry as a faculty in Mathematics, even though he aspired very much. Yet, the quintessence of mathematical studies had their reflections in his agenda for restructuring the academic facets at the University of Calcutta as its vice-chancellor. Sir Asutosh had the firm conviction that postgraduate studies should necessarily include a coupling of teaching and research. As such, he was on the lookout of younger scholars, keen to engage themselves not only in respect of updated teaching but actively receptive to research pursuits. That was why Sir Asutosh did not have any ilk of hesitation in appointing S.N. Bose and M.N. Saha as lecturers in Applied (Mixed) Mathematics at the University of Calcutta. Sir Asutosh was even questioned as to using Bose and Saha for teaching as well at the Department of Physics, he could find a rebuttal in that a revolution could be much on the anvil in realms of theoretical physics because of the Special Theory of Relativity. Thus, how S.N. Bose, M.N. Saha, and N.R. Sen could be reared up in ambiences congenial to the unfolding of potential theoreticians in young researchers.

Satyendranath Bose as a student (1910–1911). *Source* Meghnad Saha Archives

It is a matter of reality and certainly, a coincidence, that all of S.N. Bose, M.N. Saha and N.R. Sen studied together at the Department of Mathematics in the erstwhile Presidency College, Calcutta. They had their master's degrees in Mixed Mathematics, S.N. Bose topping the list and M.N. Saha coming next. One should mention that the trio—S.N. Bose, M.N. Saha, and N.R. Sen—had their honors degree in Mathematics, one following the other. Both Bose and Saha stayed on together in the M.Sc. course. As history says, both of them engaged themselves initially in teaching classical areas like Elasticity, Geodesy, and Geophysics. But, later on, they reached out to the Department of Pure Physics in some of the theoretical aspects. For this, they kept on apprising themselves of ideas and methods in new domains of Mathematics and Theoretical Physics. With the founding of the Dacca University in the early twenties in the last century, S.N. Bose was invited to take over the reins of the combined Department of Mathematics and Physics at Dacca University. One finds that, prior to his joining Dacca University, he did some research work in Mathematics, jointly with M.N. Saha, and only a few alone. His simmerings with Theoretical Physics, particularly those relating to the Special Theory of Relativity kept on having somewhat unflinching attempts to seek alternatively ab initio approach to the Special Theory of Relativity. Professor S.N. Bose initially not being able to prevail upon the journals for publishing his paper, had to turn to Albert Einstein who could immediately appreciate Bose's efforts and hence, his basic paper on Planck's law could have its genesis. Professor S.N. Bose could directly bring about conceptually what is known as the "genre of Bose–Einstein statistics". Bose also felt the necessity of a wider interactive mode abroad.

Satyendranath Bose delivering a speech. *Source* Meghnad Saha Archives

Nikhil Ranjan Sen joined the University of Calcutta in 1917 and had his initial research in areas of Newtonian potential, Elasticity and Hydrodynamics. Papers on these topics were published in *Philosophical Magazine* abroad and, later on, a few others in the *Bulletin of Calcutta Mathematical Society*. N.R. Sen's identification of elastic materials was remarkable and so, about his work on hydrodynamical waves, especially on geometries of relevant shapes. N.R. Sen, having obtained his D.Sc. degree at the University of Calcutta in 1921, could not afford to be immune from the emerging influences in Mathematical Physics. Accordingly, he sought his interactive mode in Berlin, Germany, and worked with Von Laue for another doctorate degree. His work with Von Laue was about De Sitter Universe. Prof. N.R. Sen plunged himself later in the realms of the General Theory of Relativity and Cosmology. Such investigations led to his fundamental papers that came out in *Zeitschrift fur Physik* and *Proceedings of the Royal Society of London*. Needless to add that Prof. Sen came in contact with celebrities like Max Planck, Albert Einstein, Arnold Sommerfield, and Louis de Broglie.

Professor Nikhil Ranjan Sen. *Source* Meghnad Saha Archives

Satyendranath Bose inspecting a telescope. *Source* Meghnad Saha Archives

The foregoing lines ought to give an impression that S.N. Bose, M.N. Saha, and N.R. Sen, since their younger days, were keen to establish, mostly in theoretical terms, some hallmarks providing a better understanding and fruitful insight into problems of reality. True, physical reality drew their attention. None of them, appeared to be vying with one another in respect of themes of research pursuits but a common thread was strikingly discernible. While Prof. Saha could provide the prequel before leaving for Allahabad, Dacca University proved to be congenial to Bose's thought processes.

Meghnad Saha at the Department of Physics, Allahabad University. *Source* Meghnad Saha Archives

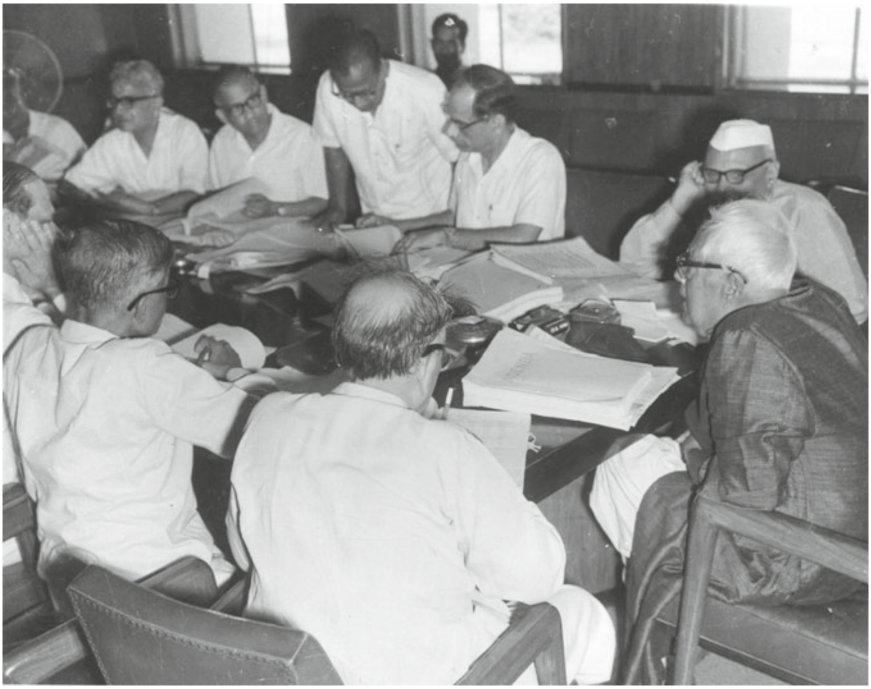
Prof. S.N. Bose did not allow any opportunity, for example, through dialogues, letters, lectures, etc., to go unattended. Bose's area of contention was to look for variants aiming at ab initio formulation and development of the concepts embedded in the well-known Planck's law. Indeed, Prof. S.N. Bose stepped in for the development of Quantum Statistical Mechanics, essentially as a conceptual breakthrough to what could be indicated by Albert Einstein. Bose's statistics was superbly distinctive in a sense that the conceptual framework could find a place in a wide diversity of statistics, founded by distinguished physicists like Maxwell, Boltzmann, Einstein, Fermi, and Dirac.

It looks to be a congenial turn of events that Prof. S.N. Bose and Prof. M.N. Saha returned to the University of Calcutta, holding the chairs of Khaira Professor and Palit Professor of Physics, respectively, in the middle of the 1940s and late 1930s. Professor N.R. Sen was continuing as the Sir Rash Behary Ghose Professor

Satyendranath Bose with P.A.M. Dirac (*left*). *Source* Meghnad Saha Archives of Applied Mathematics at the University of Calcutta. But the trio—S.N. Bose, M.N. Saha, and N.R. Sen—kept on relentlessly on dimensionalizing their areas of academic engagements. Professor M.N. Saha strove hard to establish studies on Nuclear Science, Bio-Physics, and the like at the Institute of Nuclear Physics under the aegis of the University of Calcutta.

History says that Calcutta would not have the heydays of nuclear studies, had not there been any active support from the then Vice-Chancellor Dr. Syama Prasad Mookerjee (son of Sir Asutosh Mookerjee) at the University of Calcutta. Professor S.N. Bose became a national professor and could rally around him a good number of mathematicians, physicists, chemists, and others. As already mentioned, his interest and involvement in the General Theory of Relativity did not suffer any sort of withering away. Remarkably, Prof. N.R. Sen never shied away from any frontier area of Applied Mathematics vis-à-vis Theoretical Physics.

In front of the Magnet of the Cyclotron, Institute of Nuclear Physics, Kolkata, 1948–49; *front row* (*left to right*): Dr. A.P. Patro, Dr. B.D. Nagchowdhury, Mr. B.M. Banerjee (only part of his face visible), and Prof. Meghnad Saha. *Source* Meghnad Saha Archives



Satyendranath Bose (*extreme right*) chairing the meeting of the executive council of the National Physical Laboratory, New Delhi. *Source* Meghnad Saha Archives

A.N. Whitehead's commentary on Sen's work on Relativity still continues to be cited. Under his leadership, the Department of Applied Mathematics rose to remarkable heights so that the first report of the newly established University Grant Commission was highly effusive about the department's activities and programmes. The report went to the extent of situating the first Centre of Advanced Study in Applied Mathematics at this Department, in the independent India. Prof. N.R. Sen's close linkages with higher seats of learning and laboratories used to be a talk of the day in all academic circles relating to the development of Applied Mathematics. Indeed, any specter of Applied Mathematics could hardly go unheeded, if it did not come within Prof. N.R. Sen's mindset.

One may be failing in the total appraisal of Bose, Saha, and Sen, if their roles, standing and programmes in shaping few professional organizations are not considered. All of them happened to be Presidents of Calcutta Mathematical Society in its different phases. Their linkages with Indian Science Congress Association were highly exemplary. In particular, Saha and Bose's general presidential addresses need to be focussed, even now. Also, one cannot miss Prof. M.N. Saha building up, on the eve of his return to Calcutta University, Indian Science News Association with Sir P.C. Ray as its president, Prof. N.R. Sen, treasurer and himself as secretary. Its journal *Science and Culture* had some valuable issues in its early days, dealing with wider concerns and obligations to sciences and culturally-minded scientific communities. Professor S.N. Bose established, with some of its close cohorts, *Bangiya Bijnan Parishad* (Bengal Science Society) as a part of his unstinted conviction on dissemination of science through mother tongue. *Jnan and Bijnan* (knowledge and science), the organ of the parishad continues to be a torchbearer of science, technology, and literature with their growing ambits.

Any simple account of all the trends described above show, in no way mistaken terms, that some establishments should have been in the anvil so that relevant pursuits assume rigor and vitality. It is, therefore, a part of history to be reckoned with that Saha Institute of Nuclear Physics and S.N. Bose Institute of Physical Sciences, could be brought into existence, both being in the precincts of the University College of Science (at 92, APC Road Kolkata 700009). The auditorium of the Saha Institute of Nuclear Physics has since then kept on witnessing a rich variety of programmes of Applied Mathematics and Mathematical/Theoretical Sciences. Saha Institute of Nuclear Physics could move out for an identity of its own elsewhere but intellectual kinships could hardly disappear.

One can say the simmerings within what used to happen under the auspices of S.N. Bose Institute of Physical Sciences seldom confined themselves wholly with the walls of the erstwhile sitting room of Prof. S.N. Bose. There used to be unfettered and consistent efforts through interdisciplinary, if not, multidisciplinary exposures and dialogues through lectures, seminars, etc., in various institutions/departments with the collaboration of S.N. Bose Institute of Physical Sciences. Truly, S.N. Bose National Centre for Basic Sciences could be established as upshots of earlier adventures, the culture and the temper continue, even now through a plethora of collaborative activities and programmes under the auspices of the Department of Applied Mathematics, University of Calcutta.

The International Conference on Emerging Trends in Applied Mathematics stands as a positive accrual of a long array of events having similar but distinctive complexions and flavors, too. In sum, this conference stands out as an exemplar of connectivities, in mathematically agile terms, between forward looking researchers belonging to institutions with distinct genres.

Susmita Sarkar
Uma Basu
Soumen De

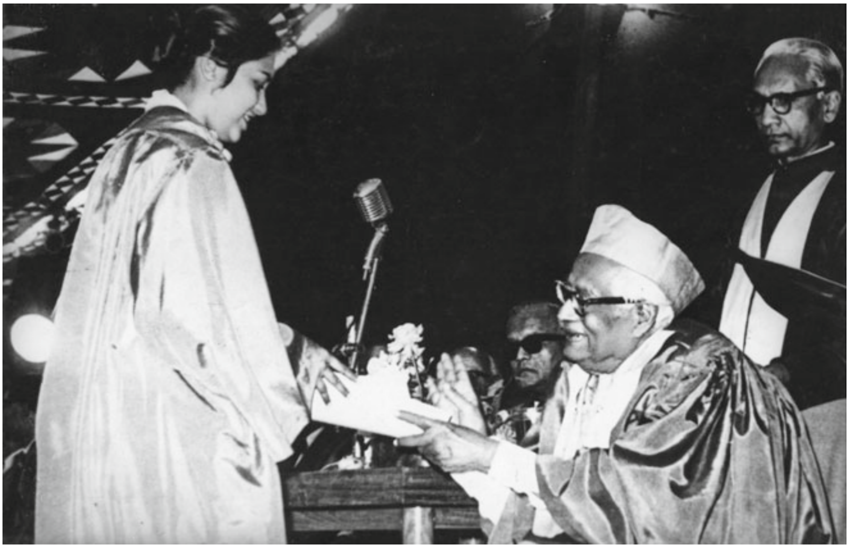
Some Unforgettable Pictures

Meghnad Saha elected FRS. *Source* Meghnad Saha Archives

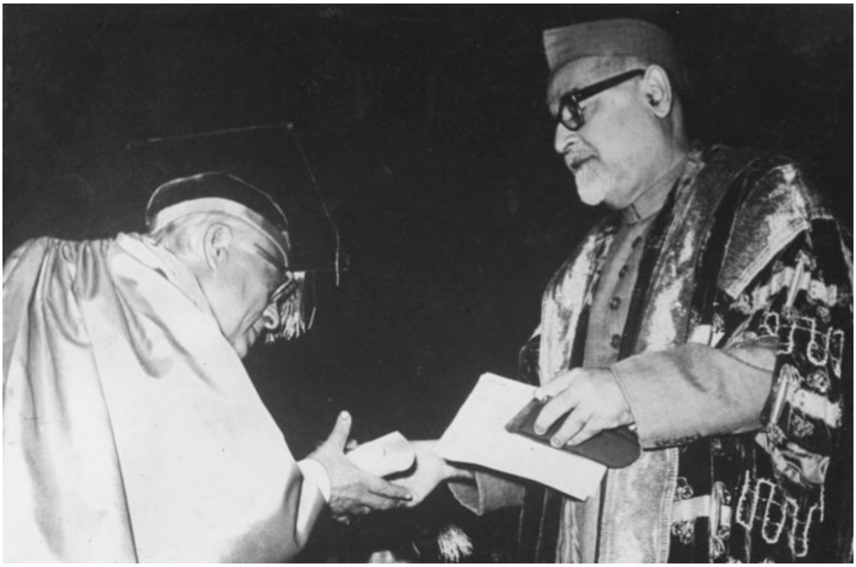
Satyendranath Bose (*left*) with J.V. Narlikar. *Source* Meghnad Saha Archives

Satyendranath Bose with Niels Bohr. *Source* Meghnad Saha Archives

Satyendranath Bose with K.N. Kolmogorov, P.C. Mahalanobis, attending the convocation at Indian Statistical Institute, Kolkata. *Source* Meghnad Saha Archives



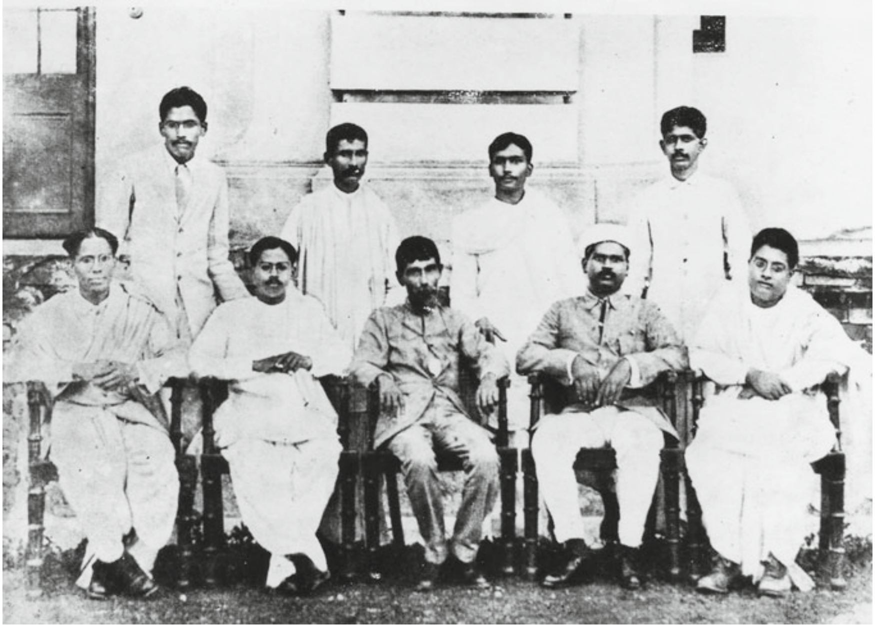
Satyendranath Bose, as the vice-chancellor of Viswabharati. *Source* Meghnad Saha Archives



Satyendranath Bose receiving D.Sc. (Hon.) from Dr. Zakir Hussain, the then President of India, at the University of Delhi in 1964. *Source* Meghnad Saha Archives

Satyendranath Bose with R.A. Fisher (*left*) and P.C. Mahalanabis (*right*). *Source* Meghnad Saha Archives

Satyendranath Bose as upacharya with Jawaharlal Nehru and with staffs and students at Viswabharati. *Source* Meghnad Saha Archives



Sitting (*left to right*): P.B. Sarkar, Amaresh Chakravarti, Acharya P.C. Ray, T.S. Muttu, Satyendranath Bose. *Source* Meghnad Saha Archives

Satyendranath Bose delivering the Meghnad Saha Memorial Lecture at SINP, Kolkata. *Source* Meghnad Saha Archives

Prof. Satyendranath Bose (*third from left*) at a seminar (29 December 1973) with Prof. S.N. Sen (*second from left*), the then vice-chancellor of the University of Calcutta and Prof. J.N. Kapur (*third from right*), the then vice-chancellor of Meerut University. *Source* Meghnad Saha Archives

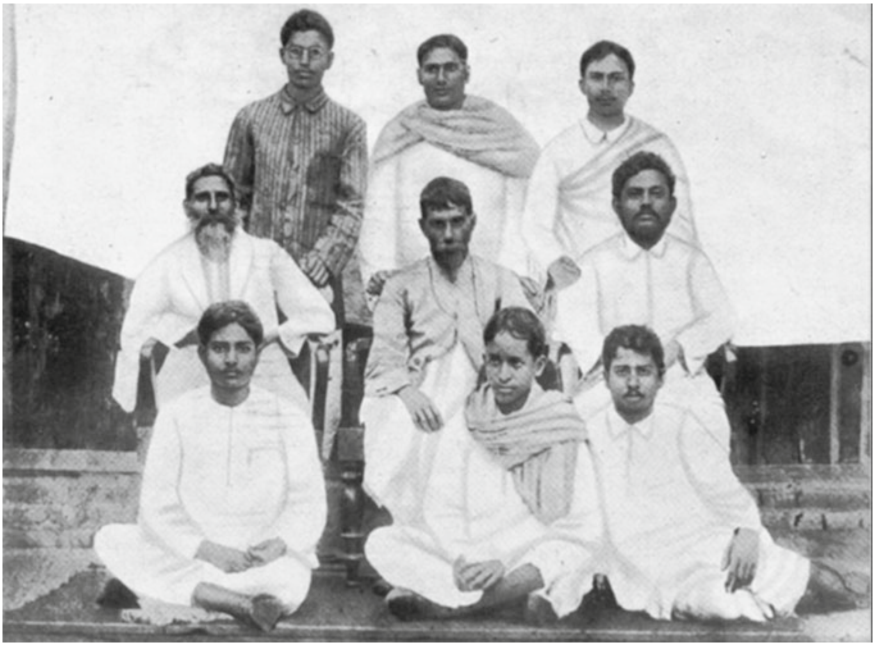
On the 70th birth celebration of Satyendranath Bose. *Source* Meghnad Saha Archives

Satyendranath Bose with his family members. *Source* Meghnad Saha Archives

Satyendranath Bose playing Esraj. *Source* Meghnad Saha Archives

Satyendranath Bose married with Ushabati Devi in 1914 at the age of 20. *Source* Meghnad Saha Archives

Satyendranath Bose and Ushabati Devi in older days. *Source* Meghnad Saha Archives



Source Meghnad Saha Archives

Contents

About the Editors

Susmita Sarkar is Professor in the Department of Applied Mathematics, University of Calcutta. Earlier, she served as a faculty member at postgraduate mathematics departments of Vidyasagar University, Burdwan University, and Jadavpur University in West Bengal. An active researcher since 1987, Dr. Sarkar's major field of research is plasma dynamics. Currently, she is working on fractional calculus and running a project in this field funded by the Department of Atomic Energy, Government of India, in collaboration with the Reactor Control Division of BARC, Mumbai. She has 44 research publications to her credit published in several international journals of repute. A national scholar and UGC-NET qualified, Dr. Sarkar received the Jnanendra Bhusan Memorial Book Prize in 1983 for securing the highest marks in mathematics honors from Presidency College. She received the INSA's visiting fellowship and was awarded the Associateship of the Third World Academy of Sciences and Regular Associateship of Abdus Salam International Centre for Theoretical Physics, Trieste, Italy. Dr. Sarkar has organized various academic activities including international and national seminars, workshops, refresher courses, and lecture courses at the applied mathematics department and invited internationally reputed resource persons for these activities. Under her initiative, the Department of Applied Mathematics, University of Calcutta, published the book *Glimpsing Through One Hundred Years of Applied Mathematics Department* during the 150th birth anniversary of Sir Asutosh Mookerjee. She was Head of the Department of Applied Mathematics, University of Calcutta, for the period September 2012–August 2014 when this international conference was held.

Uma Basu former Professor at the Department of Applied Mathematics, University of Calcutta. Professor Basu, after completing her Ph.D. from the University of Calcutta, did her post-doctoral work at the University of Central Florida, USA. She was awarded the UGC Career Award in 1994 and engaged in a number of research projects funded by UGC, CSIR, and the Department of Atomic Energy (DAE). She organized three Department of Science and Technology (DST) sponsored workshops on techniques in Applied Mathematics and edited two books containing some lectures delivered at the workshop by experts from various institutions (published by Narosa). Professor Basu was elected as fellow of National Academy of Sciences, India, in 2004. She is a lifetime member of many learned societies such as Calcutta Mathematical Society, Indian Statistical Institute, Asiatic Society, Ramanujan Mathematical Society, Indian Society of Theoretical and Applied Mechanics and many others. She acted as the secretary, treasurer, and council member for many years at the Calcutta Mathematical Society. She visited several international institutes and participated in international conferences and seminars. Areas of her research work are Fluid Mechanics, Water Waves, Thermo-Elasticity, and Mathematical Techniques in Applied Mathematics. Now, she is engaged in a research project sponsored by BRNS (DAE) at the Department of Applied Mathematics, University of Calcutta

Soumen De is an Assistant Professor at the Department of Applied Mathematics, University of Calcutta. Dr. De, a Ph.D. from Indian Statistical Institute, Kolkata, has research interests in water waves theory and integral equations. He has 15 research publications to his credit published in several international journals of repute. He is the coauthor of the book *Water Wave Scattering* (published by CRC Press).

Chapter 1
Emergent Periodicity in a Field of Chaos

Ralph Abraham and Michael Nivala

Abstract The synchronization of nonlinear oscillators is well-known and is a traditional topic in complex dynamical system theory. The synchronization of chaotic attractors is less well-known, but is of obvious interest in many applications to the sciences: physical, biological, and social. In a recent experimental study of coupled lattices of Rössler attractors, we were surprised to discover global periodic behavior in large regimes of the parameter space. This emergent periodicity in a field of chaos may be of significance in the origin of life and in many life processes. In this talk, we will explore the emergence of global periodicity and also the periodic windows in the bifurcation diagram of the Rössler attractor, which may be the local cause of this global behavior.

Keywords Nonlinear oscillator · Rössler attractor · Emergent periodicity · Synchronization

1.1 Introduction: Periodicity and Life

Living organisms are complex systems and have been modeled by complex dynamical systems. A biological organ, for example, may be simplified as a two- or three-dimensional lattice of cells or nodes, each modeled by identical dynamical schemes, with each node coupled mutually to its nearest neighbors.[1] Lattices of oscillators, for example, abound in the literature of biological modeling.

However, biological cells frequently exhibit chaotic behavior, so we have been motivated to explore two-dimensional lattices of Rössler schemes. An amazing natural phenomenon, crucial to life, is the emergence of global periodicity in such a

[1] A dynamical *scheme* is a family of dynamical systems (in this case, flows) parameterized by control variables.

R. Abraham (✉) · M. Nivala
Mathematics Department, University of California, Santa Cruz, CA 95064, USA
e-mail: rha@ucsc.edu

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_1

complex system that we call a field of chaos. For instance, we may cite swimming bacteria, respiration, the beating heart, and the regular rhythms of the brain.

In this talk, I will demonstrate the ubiquity of periodic behavior in three-related contexts.

1.2 One Rössler

This system is known as the simplest chaotic flow (continuous dynamical system) and exhibits an oscillation in the plane, together with spiking behavior in a third dimension.

1.2.1 The Basic Scheme

The scheme is defined by the equations,

$$\begin{aligned} x' &= -y - z \\ y' &= x + Ay \\ z' &= B - Cz + Mxz \end{aligned} \tag{1.1}$$

The usual values of the control parameters are $A = B = 0.2$, $C = 5.7$, and $M = 1.0$. The attractor is shown in Fig. 1.1, in which the speed along the trajectory is indicated by the colors, from blue (slowest) to red (fastest).

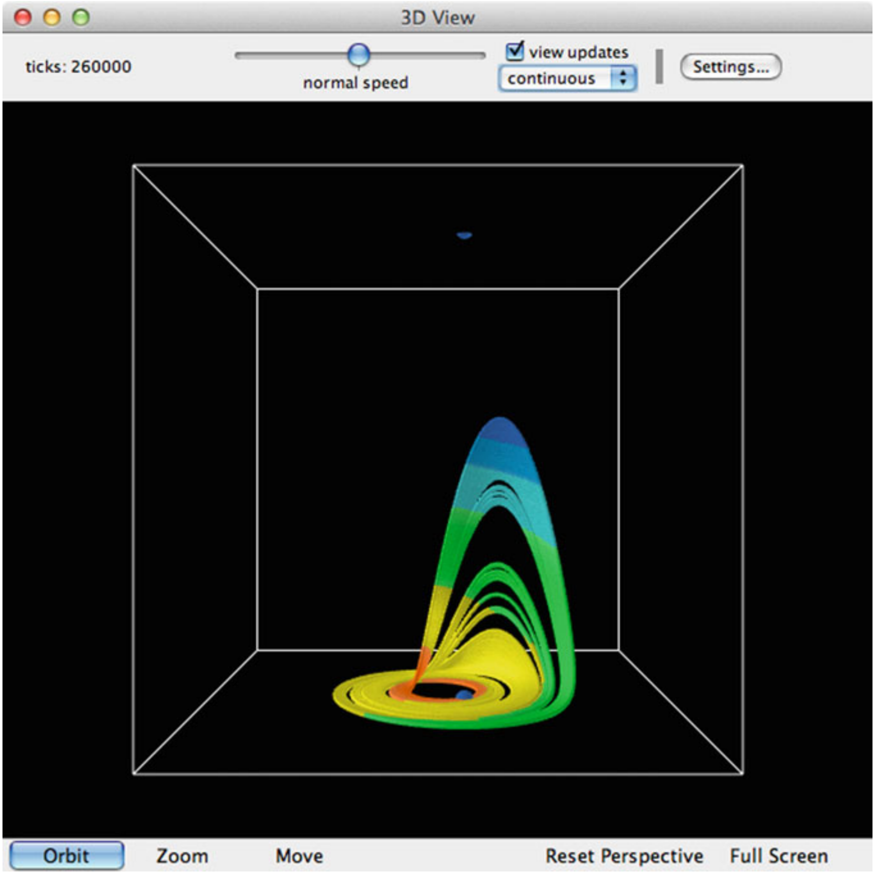
Note the simple rotation in the (x, y) plane, and the spike in the z direction.

1.2.2 Bifurcations and Periodic Windows

As one of the four control parameters is varied, while the other three are held constant, the behavior of the attractor changes through slightly different chaotic states, with occasional windows of periodic behavior. These are shown in abbreviated form in Fig. 1.2. All four exhibit periodic windows, but note that they are most conspicuous in the B plot. There, as B is decreased from the right, a unit periodic attractor undergoes a period doubling bifurcation, and then another and another, as in the familiar route to chaos of the logistic family.

1.3 Two Rösslers

There are several ways of coupling two identical dynamical systems. We will be mostly interested in the direct coupling method, due to a geometric theory of syn-



ticks: 260000

normal speed

☑ view updates

continuous

Fig. 1.1 The usual Rössler

chronization. In this method, a proportion of the z-value of each trajectory is added to the z-component of the other vectorfield. This is expressed precisely in these equations, in which we have assumed identical values of the four control parameters in each of the coupled systems.

1.3.1 Synchronization

The 0-system now includes a z_1-dependent perturbation in the third component of its vectorfield, with coupling coefficient, D_0,

$$\begin{aligned} x_0' &= -y_0 - z_0 \\ y_0' &= x_0 + Ay_0 \\ z_0' &= (B + D_0 z_1) - Cz_0 + Mx_0 z_0. \end{aligned} \tag{1.2}$$

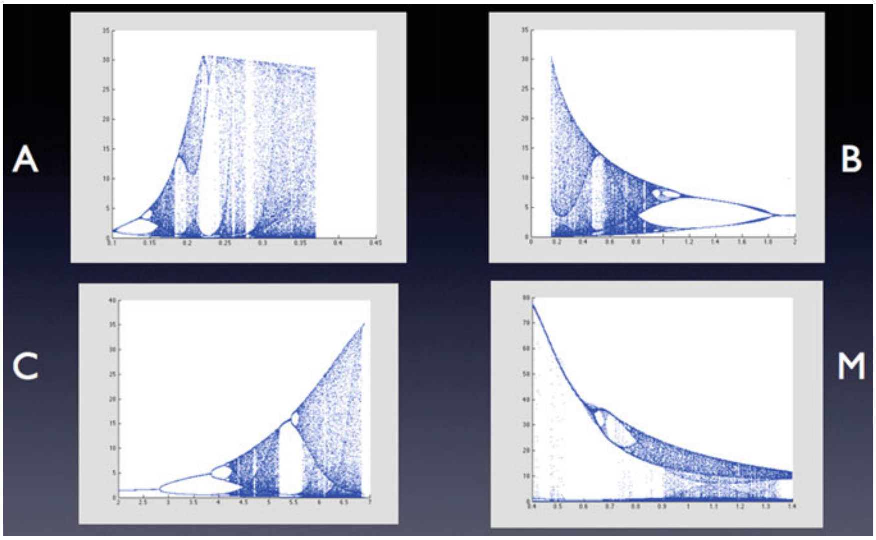


Fig. 1.2 The four bifurcation plots: Poincaré-z versus *A, B, C, M*

Similarly, the 1-system now includes a z_0-dependent perturbation in the third component of its vectorfield, with coupling coefficient, D_1,

$$\begin{aligned} x_1' &= -y_1 - z_1 \\ y_1' &= {}_1x + Ay_1 \\ z_1' &= (B + D_1 z_0) - Cz_1 + Mx_1 z_1 \end{aligned} \tag{1.3}$$

Note the addition of two additional control parameters, D_0 and D_1, the coupling coefficients. Also, we have grouped together the terms $(B + D_0 z_1)$ and $(B + D_1 z_0)$ to foreground the fact that the forcing terms effectively modify the B coefficients of each system. We call these terms *effective* B_0 and *effective* B_1 for the coupled Rössler systems.

We are interested in two special cases.

In the case $D_0 = 0$, the 0-system is called the master, and the 1-system is the slave. The master system forces the slave, while the master behaves as if the slave did not exist.

In the case $D_0 = D_1 \geq 0$, we say the two systems are mutually and symmetrically coupled.

In both of these special cases, increasing the coupling coefficients produces synchronization of the z-spikes, even though these spikes occur chaotically in time and in strength as well, as shown in Fig. 1.3, for the master and slave.

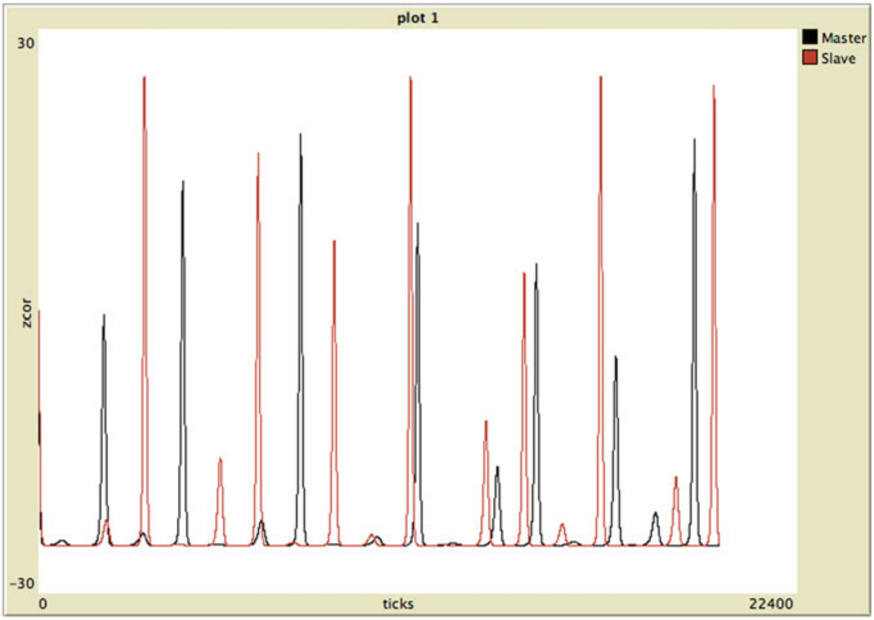


Fig. 1.3 The master (*black*) and slave (*red*) z's versus time

1.3.2 Bifurcations and Periodic Windows

We now consider the double Rössler system in the second case of symmetrical coupling. A simulation with NetLogo 3D, showing the two trajectories side-by-side, one green, the other blue, reveals the chaotic synchronization. The trajectories may be clarified by indicating a Poincaré cross-section. For this, we have chosen the positive half of the X–Z plane. When the green trajectory pierces this half-plane, it leaves a red drop. And when the blue transits the section, it leaves a yellow drop. A bifurcation movie of this system, as $D = D_0 = D_1$ increases from 0.0 to 4.0, reveals an extensive periodic window around $D = 2.0$. This emergent periodicity, seen from the positive Y-axis, is illustrated in Fig. 1.4.

1.4 Many Rösslers

Finally we consider a regular lattice of 160,000 usual Rösslers, in a 400 by 400 square grid, on a two-dimensional torus. Each node is mutually and symmetrically coupled to each of its four nearest neighbors. Thus, at each node, we have,

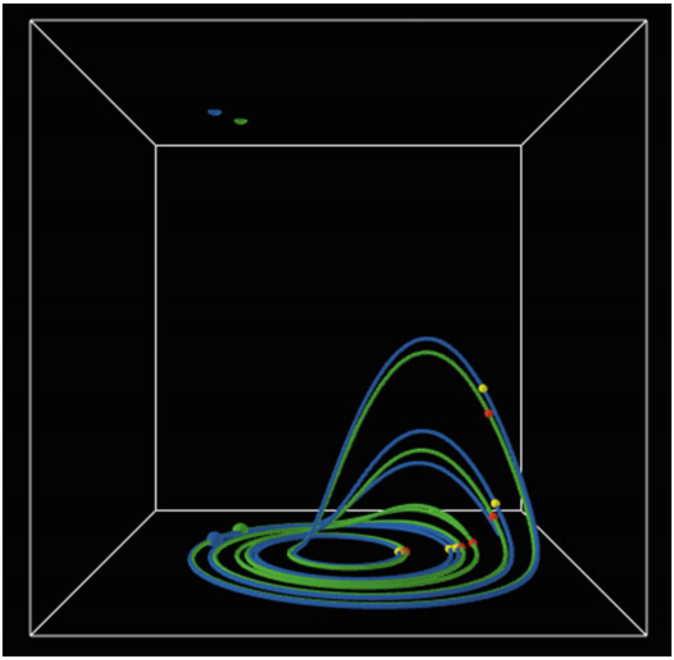


Fig. 1.4 Visualizing the Poincaré cross-section

$$
\begin{aligned}
x' &= -y - z \\
y' &= x + Ay \\
z' &= (B + Dz_s) - Cz + Mxz
\end{aligned}
\tag{1.4}
$$

where D is the common coupling constant and z_s denotes the sum of the z-coordinates of the four neighbors.

1.4.1 Synchronization

Extensive simulations of this system, the $2D$-toral Rössler lattice, by Michael Nivala of the UCLA have been recorded as movies, with each frame representing an instantaneous state revealing the z of each node as a color, from blue (0) to red (25). Three frames of such a movie[2] are shown in Fig. 1.5. The colors, especially in the third frame, reveal islands of z-synchronization, which move around with advancing simulation time.

[2] Nivala's a14.

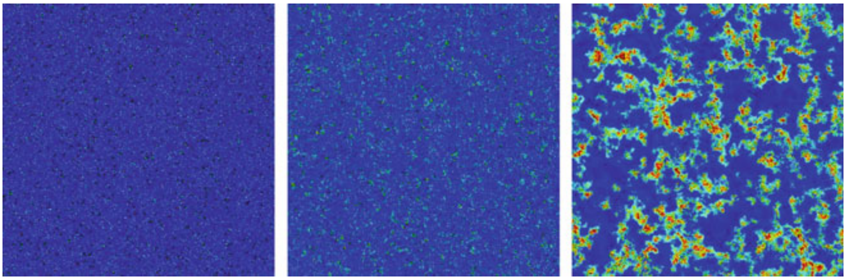


Fig. 1.5 Three frames of the Rössler lattice

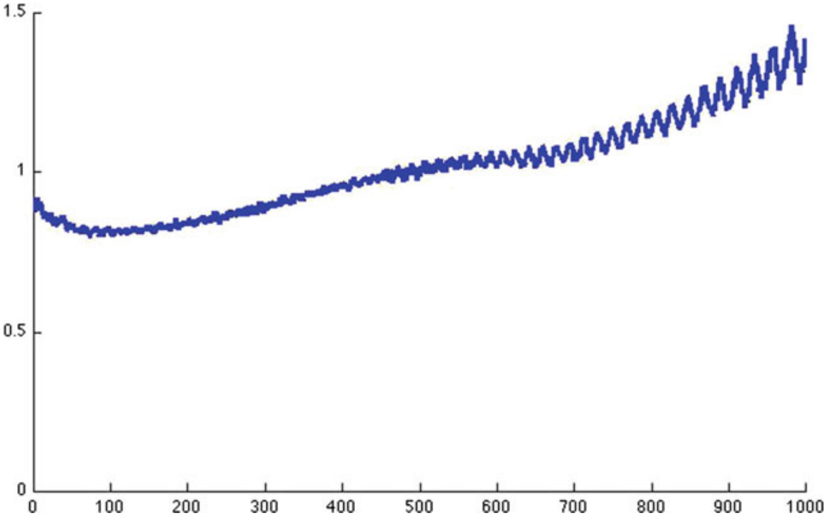


Fig. 1.6 Periodic temporal fluctuations in the z average

1.4.2 Global Periodicity

A surprising feature of these simulations is a robust global periodicity. This may be observed by averaging the z-values of all the nodes and plotting as a function of time. As we see in Fig. 1.6, there is a periodic fluctuation in this average value.

1.5 Conclusion: Future Work

Our interest in the emergence of global periodicity in a field of chaos is heightened by the crucial role of periodicity in life processes, and we feel justified in thinking that nature has selected for attractors with shapes that facilitate synchronization and bifurcation diagrams with periodic windows. And yet, these periodic windows are quite surprising from the point of view of pure mathematics. We began our investigation with the idea of observing patterns of chaotic synchronization and were astonished to discover global periodicity by accident.

We have by now a large number of related simulations and will be filing more progress reports as time goes on. But at this point, we may say that global periodicity is ubiquitous for Rössler lattices. In the future, we plan to explore other fields of chaos, such as Lorenz and Ueda lattices, to discover their secrets as well.

Chapter 2
The Globalisation of Applied Mathematics

Peter Leach

Abstract Applied mathematics has a history going back several thousands of years at least to the time of the Babylonians. In a sense, (pure) mathematics evolved from applied mathematics. Over the centuries, applied mathematics became closely associated with mechanics and new developments were called applicable mathematics. This terminology has faded in recent decades due to the widespread use of mathematics in the solution of problems in any fields which can be treated quantitatively. One recent field is that of financial mathematics, and we illustrate a few of the problems which have been solved using the techniques of applied mathematics.

Keywords Mathematical modelling · Financial mathematics · Lie algebra · Lie point symmetry

2.1 Background

We commence with a story motivated by an observation of Professor KM Tamizhmani concerning modelling. It concerns a beautiful young woman and two suitors, both of them handsome young men named Krishnakumar and Sinuvasan. The young woman takes a pile of pebbles and puts one to the left for each ewe owned by Krishnakumar and one to the right for each ewe owned by Sinuvasan. When all of the ewes have been counted, there are 15 pebbles to the left and 10 to the right. Which suit will the woman accept?

The modelling behind this question is to use a pebble to represent an ewe, but there is a further consideration. What criterion does the woman use to reach her decision? Is it the number of ewes or is it the social convention as to who does the milking?

Here we have a very simple instance of mathematical modelling in application. All that was required for such a model to be considered is the social value of goats.

P. Leach (✉)
School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Private Bag X54001,
Durban 4000, Republic of South Africa
e-mail: leach@ucy.ac.cy

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_2

A more sophisticated application of mathematics can be found in the times of the Babylonians who used linear interpolation to approximate the sine function as an aid to the determination of the position of the Moon which was an important matter for a society which used the lunar month. A millennium or so later Pythagoras travelled to the Land of Egypt and observed that the surveyors were making use of the 3-4-5 rule to ensure that the farmers' fields were quite rectangular. He returned to his native Samos and did what philosophers do best which is to take something practical and construct a Theorem from it.

Another example of applied mathematics from ancient times is found in the work of Claudios Ptolemaios who devised a method for the calculation of orbits by the simple expedient of having circles move on circles. Ptolemaios made a mistake in his modelling in two respects. Firstly he assumed that the Earth was the centre of the Universe and secondly that the basic component of an orbit was a circle. Nevertheless his method provided accurate predictions for roughly one and a half millennia until after the physically acceptable model was introduced in the seventeenth century. This illustrates an interesting point about modelling and applied mathematics. The model may be physically incorrect and yet provide correct answers.

With the advent of the seventeenth century mathematical modelling developed remarkably well in the various branches of Mechanics. As the years turned into centuries, Mechanics became subdivided into specific areas—Classical, Continuum, Quantum, Relativistic—to such an extent that Applied Mathematics became synonymous with Mechanics. When other areas of application and modelling were developed, it was fashionable to call these areas Applicable Mathematics to avoid the obvious taint of Mechanics. Fortunately such a distinction appears to have faded in recent decades.

One of the reasons for the loss of the distinction can be found in the universality of differential equations. The same equation appears in various diverse applications. An example of the proliferation of models can be seen in a recent paper [17] devoted to solutions of the Fisher Equation and some of its generalisations. Some of the fields of application mentioned are logistic models of population growth, flame propagation, neurophysiology, autocatalytic chemical reactions and branching processes based on Brownian motion. According to [26], the original problem modelled the propagation of a gene in a population. The classical Fisher Equation

$$u_t = bu_{xx} + au(1-u), \quad ab \neq 0, \tag{2.1}$$

first appeared seventy-five years ago in [12]. It is remarkable how disparate processes can be modelled by what is essentially the same equation. Only the names of the labels and possibly boundary/initial conditions have been changed.

In 1828, Robert Brown [4] reported his observations of 1827 concerning the motion of particles, such as small pieces of broken pollen, suspended in a fluid. The irregular motion subsequently became called Brownian motion and became an important concept in the modelling of a wide variety of phenomena ranging from

Statistical Physics to Financial Mathematics with quite a few stops in between. The essential point is that the mechanisms in each of these phenomena are based on some form of random, or stochastic, motion. Our particular interest today comprises some equations which have arisen in the field of Financial Mathematics. The literature is quite vast, but the seminal papers can be counted on the fingers of one hand.

2.2 An Algebraic Diversion

Before we begin our examination of some of the equations which arise in finance, we should recall a little of the algebraic theory of differential equations.

A differential equation,

$$E\left(x,\, u,\, u_x,\, u_{xx},\, \ldots\right) = 0, \tag{2.2}$$

in which all symbols can be multisymbols, is invariant under the infinitesimal transformation generated by the operator

$$\Gamma = \xi \partial_x + \eta \partial_u \tag{2.3}$$

if

$$\Gamma^{[n]} E_{|E=0} = 0, \tag{2.4}$$

where $\Gamma^{[n]}$ is the extension of Γ to account for all of the derivatives occurring in E. The invariance contained in (2.4) occurs when the Eq. (2.2) is taken into account. Usually the coefficient functions, ξ and η, are taken to be functions of x and u only, i.e., the infinitesimal transformation is a point transformation, but it is also possible to include derivatives.

The number of symmetries, (2.3), can range from zero to infinity, depending upon the equation being studied. Under the operation of taking the Lie Bracket

$$\left[\Gamma_i,\, \Gamma_j\right]_{LB} = \Gamma_i \Gamma_j - \Gamma_j \Gamma_i \tag{2.5}$$

one obtains a Lie algebra. Different equations can have the same algebra even if their provenances are quite disparate.

The symmetries, if sufficient in number and type, can be used to reduce the equation and even find its solution. The calculation of the symmetries is usually an exercise in advanced tedium and is best left to a computer algebra code on some computer. Various packages are available and they are of variable quality. We use Sym [1, 7–9] which operates in Mathematica. For the identification of the algebra, we make use of the classification scheme of Mubarakzyanov [20–23].

2.3 The Black–Scholes Equation

The Black–Scholes–Merton equation [2, 3, 19],

$$u_t + \tfrac{1}{2}\sigma^2 x^2 u_{xx} + rxu_x - ru = 0, \tag{2.6}$$

is the precursor of the many evolution partial differential equations which have been derived in the modelling of various financial processes. Basically it has to do with the pricing of options, but anything vaguely connected such as corporate debt is equally grist for its mill. The symmetry analysis of (2.6) was first undertaken by Gasizov and Ibragimov [13]. After determining the symmetries they obtained of the solution for the initial condition being a delta function which is a typical initial condition for the heat equation.[1] A more typical problem is the solution of (2.6) subject to what is known as a terminal condition, i.e. $u(T, x) = U$ when $t = T$, and it is this problem which we solve to give a demonstration of the methodology.

The Lie point symmetries of (2.6) are

$$\begin{aligned}
\Gamma_1 &= x\partial_x \\
\Gamma_2 &= 2tx\partial_x + \left\{t - \frac{2}{\sigma^2}(rt - \log x)\right\} u\partial_u \\
\Gamma_3 &= u\partial_u \\
\Gamma_4 &= \partial_t \\
\Gamma_5 &= 8t\partial_t + 4x\log x\partial_x + \left\{4tr + \sigma^2 t + 2\log x + \frac{4r}{\sigma^2}(rt - \log x)\right\} u\partial_u \\
\Gamma_6 &= 8t^2\partial_t + 8tx\log x\partial_x + \left\{-4t + 4t^2 r + \sigma^2 t^2 + 4t\log x + \frac{4}{\sigma^2}(rt - \log x)^2\right\} u\partial_u \\
\Gamma_\infty &= f(t, x)\partial_u,
\end{aligned}$$

where Γ_∞ is the infinite subset of solutions to (2.6). The algebra of the finite subset is $sl(2, R) \oplus_s W_3$, where W_3 is the three-dimensional Heisenberg-Weyl algebra.

To solve the problem of the terminal condition we take a linear combination of the finite set of symmetries, $\Gamma = \sum_{i=1}^{i=6} \alpha_i \Gamma_i$, and apply it to the two conditions given above.[2] In the case of $t = T$ we obtain

$$\alpha_4 + 8T\alpha_5 + 8T^2\alpha_6 = 0 \tag{2.7}$$

[1] One would hope that this initial condition would not apply in financial matters! Unfortunately there are some instances of financial instability in which such an initial condition is far too accurate a model. Note that the paper [16] with more realistic conditions appeared earlier, but the content of [13] had already been presented at a seminar in the Department of Physics, The University of the Witwatersrand, in 1996.

[2] The solution symmetries, Γ_∞, can play no role in this as their action on $u(T, x) = U$ produces a linear combination of linearly independent solutions.

in which we have replaced t by its specified value. When we turn to the condition $u(T,x)=U$ and make the appropriate substitutions, we obtain

$$\begin{aligned}&\alpha_2\left\{T-\frac{2}{\sigma^2}\left(rT-\log x\right)\right\}U+\alpha_3 U\\&+\alpha_5\left\{4Tr+\sigma^2 T+2\log x+\frac{4r}{\sigma^2}\left(Tr-\log x\right)\right\}U\\&+\alpha_6\left\{-4T+4T^2r+\sigma^2t^2+4T\log x+\frac{4}{\sigma^2}\left(rT-\log x\right)^2\right\}U=0.\end{aligned}\quad(2.8)$$

The coefficient of $(\log x)^2$ in (2.8) means that $\alpha_6=0$ and hence from (2.7) that $\alpha_4=-8T\alpha_5$. Returning to (2.8) the coefficient of $\log x$ leads to $\alpha_2\sigma^2U/2-4rU\alpha_5/\sigma^2=0$ and the remaining terms give $\alpha_2(1-r\sigma^2/2)U+\alpha_3U+\alpha_5(4Tr+\sigma^2T+4r^2T)U=0$. Consequently we have

$$\begin{aligned}&\alpha_1 \text{ is arbitrary}\\&\alpha_2=(2r-\sigma^2)\alpha_5\\&\alpha_3=-8rT\alpha_5\\&\alpha_4=-8T\alpha_5.\end{aligned}\quad(2.9)$$

As is common with $(1+1)$ evolution partial differential equations of maximal symmetry, there are two symmetries which are compatible with the terminal condition. They are

$$\begin{aligned}&\Lambda_1=x\partial_x\quad\text{and}\\&\Lambda_2=8(t-T)\partial_t+(4rt-2\sigma^2t+4\log x)x\partial_x+8r(t-T)u\partial_u\end{aligned}$$

with the Lie Bracket $[\Lambda_1,\ \Lambda_2]_{LB}=4\Lambda_1$ so that reduction by the normal subgroup, represented by Λ_1, is to be preferred. The invariants of the associated Lagrange's system,

$$\frac{\mathrm{d}t}{0}=\frac{\mathrm{d}x}{x}=\frac{\mathrm{d}u}{0},$$

are t and u so that we introduce the change of variables $y=t$ and $v=u$ into (2.6) to obtain the ordinary differential equation

$$v'-rv=0$$

with solution

$$v=K\mathrm{e}^{ry}.$$

In terms of the original variables the solution obtained using Λ_1 is

$$u = K\mathrm{e}^{rt}$$

and on the substitution of the terminal conditions to evaluate the constant of integration, we find that the solution of the terminal problem for (2.6) is

$$u(t,x) = U\exp[r(t-T)]. \tag{2.10}$$

As the solution of this problem is unique, there is no need to make use of the second symmetry.

2.4 The Cox–Ingersoll–Ross equation

The Cox–Ingersoll–Ross equation [6] (see also [5, 11, 14, 25] for studies of similar equations),

$$u_t + \tfrac{1}{2}\sigma^2 x u_{xx} - (\kappa - \lambda x)u_x - xu = 0, \tag{2.11}$$

is an example of an equation for which the number of Lie point symmetries depends upon a relationship between the parameters in the equation.

For unconstrained values of the parameters (2.11) possesses the symmetries [10]

$$\begin{aligned}
\Gamma_1 &= u\partial_u \\
\Gamma_{2\pm} &= \exp[\pm\beta t]\left\{\pm\partial_t + \beta x\partial_x - \frac{1}{\sigma^2}(-\beta \pm \lambda)(\kappa \pm \beta x)u\partial_u\right\} \\
\Gamma_3 &= \partial_t \\
\Gamma_\infty &= f(t,x)\partial_u,
\end{aligned}$$

where, as above, Γ_∞ represents the solution symmetries of the linear evolution partial differential equation. The finite subalgebra is $sl(2,R) \oplus A_1$. Although there does not exist a point transformation which takes (2.11) to the classical heat equation, the algebraic structure is that of a heat equation with a source/sink term proportional to U/X^2 in the transformed variables [10, 18].

Despite the diminution in the number of symmetries compared to (2.6), we can still investigate to see if there are sufficient symmetries to solve the problem with a terminal condition. As we did above, we take a linear combination of the elements of the finite subalgebra and apply it to the conditions $u(T,x) = U$ when $t = T$. The latter gives

$$\alpha_{2+}\exp[\beta T] - \alpha_{2-}\exp[-\beta T] + \alpha_3 = 0$$

and the former

$$\begin{aligned}
&\alpha_1 U - \alpha_{2+} \exp[\beta T] \frac{1}{\sigma^2}(-\beta + \lambda)(\kappa + \beta x)U \\
&\qquad - \alpha_{2-} \exp[-\beta T] \frac{1}{\sigma^2}(-\beta - \lambda)(\kappa - \beta x)U = 0.
\end{aligned}$$

It is necessary to separate the coefficient of x from the constant term. This gives a relationship between α_{2+} and α_{2-}. When this is substituted into the remaining terms, we obtain the relationships

$$\begin{aligned}
\alpha_1 &= -\frac{2\kappa(\beta - \lambda)}{\sigma^2} \exp[\beta T]\alpha_{2+}, \\
\alpha_{2-} &= \frac{\beta - \lambda}{\beta + \lambda} \exp[2\beta T]\alpha_{2+}, \\
\alpha_3 &= -\frac{2\lambda\lambda}{\beta + \lambda} \exp[\beta T]\alpha_{2+}.
\end{aligned} \tag{2.12}$$

Even with the reduced number of symmetries we have been able to obtain a symmetry which is compatible with the terminal condition and this may be used to reduce (2.11) to an ordinary differential equation to be solved.

2.5 The Heath Equation

The evolution partial differential equations which arise in Financial Mathematics are not confined to linear equations. As a simple example we consider the equation treated in Heath [15], namely

$$2u_t + 2au_x + b^2 u_{xx} - u_x^2 + 2\nu(x) = 0. \tag{2.13}$$

For a general function $\nu(x)$ (2.13) possesses the Lie point symmetries [24]

$$\begin{aligned}
\Gamma_1 &= \partial_t, \\
\Gamma_2 &= \partial_u, \\
\Gamma_\infty &= b^2 f(t, x) \exp[u/b^2]\partial_u,
\end{aligned}$$

where $f(t, x)$ is any solution of the linear equation

$$2u_t + 2au_x + b^2 u_{xx} + 2\nu(x)u = 0. \tag{2.14}$$

Due to the presence of the arbitrary function $\nu(x)$ in (2.13) one would not expect any symmetries apart from the obvious Γ_1 and Γ_2. Due to the nonlinearity of (2.13) one would certainly not expect the presence of Γ_∞ as this is a characteristic of linear equations. As Γ_∞ is present, it is evident that a linearising transformation exists and it is easily inferred from the other terms in the symmetry to be given by $U(t, x, u) = -\exp[-u/b^2]$. The transformation is of the same form as the Cole-Hopf transformation well-known from its linearising effect upon the Burgers equation.

It is known (*cf.* [24]) that (2.13) possesses additional symmetries if $\nu(x)$ has certain specific forms. In the symmetry analysis of the equivalent equation, (2.14), using SYM two special cases arise naturally. They are

$$\begin{aligned}\nu_1(x) &= a_1 + a_2x + a_3x^2 \quad \text{and} \\ \nu_2(x) &= a_1 + a_2x + a_3x^2 + \frac{a_4}{(a_2 + 2a_3x)^2}.\end{aligned}$$

In the case of $\nu_1(x)$ the number of symmetries and their algebra are the same as for the classical heat equation and consequently there exists a point transformation connecting (2.14), hence (2.13), to the heat equation. This is not the case with $\nu_2(x)$. The number of symmetries corresponds to the heat equation with a source/sink term proportional to u/x^2. Obviously the algebra is $\{sl(2, R) \oplus A_1\} \oplus_s \infty$ which is characteristic of evolution equations derived from the Ermakov–Pinney equation [18].

2.6 A Really Nonlinear One!

What is essentially a variant of the Black–Scholes equation

$$2V_t + 2(r - q)SV_S + \Sigma^2 S^2 V_{SS} - 2rV = 0$$

and readily reducible to the heat equation is rendered more than moderately nonlinear if Σ is assumed to be proportional to V_{SS} to become the differential equation,

$$2V_t + 2(r - q)SV_S + \sigma^2 S^2 (V_{SS})^3 - 2rV = 0, \tag{2.15}$$

which possesses five Lie point symmetries, namely

$$\begin{aligned}\Gamma_1 &= \exp[rt]\partial_V, \\ \Gamma_2 &= S\exp[qt]\partial_V, \\ \Gamma_3 &= \partial_t, \\ \Gamma_4 &= \exp[(2r - 4q)t]\{\partial_t + (r - q)S\partial_S + rV\partial_V\}, \\ \Gamma_5 &= S\partial_S + 2V\partial_V.\end{aligned}$$

The five-dimensional algebra is $\{A_1 \oplus A_2\} \oplus_s 2A_1$.

The symmetries Γ_1 and Γ_2 satisfy (2.15) and as solution symmetries are of no use in giving a symmetry which is compatible with any other conditions. Fortunately the remaining three symmetries are sufficient for the purpose of satisfying the requirement that $V(T, S) = G(S)$ when $t = T$ provided that $G(S)$ takes a specific form. The application of $\Gamma = \alpha_3\Gamma_3 + \alpha_4\Gamma_4 + \alpha_5\Gamma_5$ to the terminal condition above leads to the conditions

$$\begin{aligned}&\alpha_3 = -\alpha_4 \exp[(2r - 4q)T] \quad \text{and}\\&\alpha_4 \exp[(2r - 4q)T](rG(S) - (r - q)SG'(S)) + \alpha_5(2G(S) - SG'(S)) = 0.\end{aligned}$$

One possibility for the second condition is that $r = 2q$ in which case the conditions become

$$\begin{aligned}&\alpha_3 = -\alpha_4 \quad \text{and}\\&(q\alpha_4 + \alpha_5)(2G - SG') = 0\end{aligned}$$

so that either $\alpha_5 = -q\alpha_4$ or $G(S) = KS^2$ for some constant, K. In the case of the former possibility Γ is zero. In the case of the latter α_4 and α_5 are arbitrary, but we have only the single symmetry

$$\Gamma = S\partial_S + 2V\partial_V \tag{2.16}$$

for which the invariants are t and VS^{-2}. We substitute $V = S^2 f(t)$ into (2.15) and easily find that

$$V = \frac{S^2}{\sqrt{8\sigma^2(t + c)}}, \tag{2.17}$$

where c is the constant of integration. The value of this constant is determined by imposing the terminal condition which gives

$$c = \frac{1}{8\sigma^2 K^2} - T.$$

If $r \neq 2q$, the second condition gives two possibilities. If $G(S)$ is still given by KS^2, $\alpha_4 = 0$ and so α_3 is also zero. The solution (2.17) still applies. On the other hand α_4 is arbitrary and $\alpha_5 = 0$ if $G(S) = KS^{r/(r-q)}$. Now

$$\Gamma = \{\exp[(2r - 4q)t] - \exp[(2r - 4q)T]\}\,\partial_t + \exp[(2r - 4q)t]\,\{(r - q)S\partial_S + rV\partial_V\}\,. \tag{2.18}$$

For other functions GS all the α_i are zero and so there is no symmetry compatible with the terminal condition.

2.7 Conclusion

Given the constraints of time we have been able only to explore one aspect of Applied Mathematics. This is the quite recent application of Lie's Theory of Continuous Groups to problems which arise in Financial Mathematics. We noted a recent paper which mentioned a few applications of the Fisher Equation—originally formulated in a biological context—to divers fields. The mechanisms of the various problems have a certain similarity and so we find the same equations, maybe *mutatis mutandis*, recurring. One of the important features is that methods developed in one field can find application in many other fields.

As the decades progress, the quantification of all manner of phenomena increases in number and diversity. The quantification is the gist of Applied Mathematics and so we have Globalisation.

Acknowledgments I thank Professor KM Tamizhmani and the Department of Mathematics for generous provision of facilities and support and furthermore thank the University of KwaZulu-Natal and the National Research Foundation of South Africa for their continued support. The opinions expressed in this paper should not be construed as being those of either institution.

References

1. K. Andriopoulos, S. Dimas, P.G.L. Leach, D. Tsoubelis, On the systematic approach to the classification of differential equations by group theoretical methods. J. Comput. Appl. Math. **230**, 224–232 (2009). doi:10.1016/j.cam.2008.11.002
2. F. Black, M. Scholes, The valuation of option contracts and a test of market efficiency. J. Financ. **27**, 399–417 (1972)
3. F. Black, M. Scholes, The pricing of options and corporate liabilities. J. Polit. Econ. **81**, 637–654 (1973)
4. R. Brown, A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Philos. Mag. **4**, 161–173 (1828)
5. K. Chan, A. Karolyi, F. Longstaff, A. Sanders, An empirical comparison of alternate models of the short-term interest rate. J. Financ. **47**, 1209–1227 (1992)
6. J.C. Cox, J.E. Ingersoll, S.A. Ross, An intertemporal general equilibrium model of asset prices. Econometrica **53**, 363–384 (1985)
7. S. Dimas, D. Tsoubelis, *SYM: A New Symmetry-Finding Package for Mathematica*, ed. by N.H. Ibragimov, C. Sophocleous, P.A. Damianou. Group Analysis of Differential Equations (University of Cyprus, Nicosia, 2005), pp. 64–70. http://www.math.upatras.gr/~spawn
8. S. Dimas, D. Tsoubelis, *A New Mathematica-Based Program for Solving Overdetermined Systems of PDEs.* 8th International Mathematica Symposium, Avignon, France, 2006
9. S. Dimas, Partial differential equations, algebraic computing and nonlinear systems, Thesis, University of Patras, Patras, Greece, 2008
10. S. Dimas, K. Andriopoulos, D. Tsoubelis, P.G.L. Leach, Complete specification of some partial differential equations that arise in financial mathematics. J. Nonlinear Math. Phys. (2009) (submitted)
11. U.L. Dothan, On the term structure of interest rates. J. Financ. Econ. **6**, 59–69 (1978)
12. R.A. Fisher, The wave of advance of advantageous genes. Ann. Eugenics **7**, 353–369 (1937)

13. R. Gasizov, N.H. Ibragimov, Lie symmetry analysis of differential equations in finance. Nonlinear Dyn. **17**, 387–407 (1998)
14. J. Goard, New solutions to the bond-pricing equation via Lie's classical method. Math. Comput. Model. **32**, 299–313 (2000)
15. D. Heath, E. Platin, M. Schweizer, *Numerical Comparison of Local Risk-minimisation and Mean-variance Hedging*, ed. by E. Jouini, J. Cvitanić, M. Musiela. Option Pricing, Interest Rates and Risk Management (Cambridge University Press, Cambridge, 2001), pp. 509–537
16. N.H. Ibragimov, C. Wafo Soh, *Solution of the Cauchy Problem for the Black-Scholes Equation Using Its Symmetries* Modern Group Analysis, ed. by N.H. Ibragimov, K.R. Naqvi, E. Straume. International Conference at the Sophus Lie Conference Centre (MARS Publishers, Norway, 1997)
17. R. Jiwari, A. Verma, Analytic, Power series and numerical solutions of nonlinear diffusion equations via symmetry reductions, preprint school of mathematics and computer applications, Thapar University, Patiala—147004, India (2014)
18. R.L. Lemmer, P.G.L. Leach, A classical viewpoint on quantum chaos. Arab. J. Math. Sci. **5**, 1–17 (1999)
19. R.C. Merton, On the pricing of corporate data: the risk structure of interest rates. J. Financ. **29**, 449–470 (1974)
20. V.V. Morozov, Classification of six-dimensional nilpotent Lie algebras. Izv. Vyssh. Uchebn. Zavad. Mat. **5**, 161–171 (1958)
21. G.M. Mubarakzyanov, On solvable Lie algebras. Izv. Vyssh. Uchebn. Zavad. Mat. **32**, 114–123 (1963)
22. G.M. Mubarakzyanov, Classification of real structures of five-dimensional Lie algebras. Izv. Vyssh. Uchebn. Zavad. Mat. **34**, 99–106 (1963)
23. G.M. Mubarakzyanov, Classification of solvable six-dimensional Lie algebras with one nilpotent base element. Izv. Vyssh. Uchebn. Zavad. Mat. **35**, 104–116 (1963)
24. V. Naicker, K. Andriopoulos, P.G.L. Leach, Symmetry reductions of a Hamilton-Jacobi-Bellman equation arising in financial mathematics. J. Nonlinear Math. Phys. **12**, 268–283 (2005)
25. C.A. Pooe, F.M. Mahomed, S.C. Wafo, Fundamental solutions for zero-coupon bond-pricing models. Nonlinear Dyn. **36**, 69–76 (2004)
26. O.O. Vaneeva, R.O. Popovych, C. Sophocleous, *Group Classification of the Fisher Equation with Time-dependent Coefficients*, ed. by O.O. Vaneeva, C. Sophocleous, R.O. Popovych, P.G.L. Leach, V.M. Boyko, P.A. Damianou. Group Analysis of Differential Equations and Integrable Systems VI (University of Cyprus, Lefkosia, 2013), pp. 225–236

Chapter 3
The Ricci Flow Equation and Poincaré Conjecture

Amiya Mukherjee

Abstract The Poincaré conjecture was formulated by the French mathematician Henri Poincaré more than hundred years ago. The conjecture states, when reformulated in modern language, that any simply connected closed 3-manifold is diffeomorphic to the standard 3-sphere S^3. This was the most famous open problem, and its solution turned out to be extraordinarily difficult. It had eluded all attempts at solution for more than hundred years. During 2002 and 2003, Grigoriy Perelman posted a proof of the conjecture on the Internet in three instalments, completing a program initiated in the 1980s by Richard Hamilton to solve a more general conjecture, called the geometrization conjecture of William Thurston. The key tool of Hamilton's program is the Ricci flow, a differential equation on the space of Riemannian metrics of a 3-manifold. The equation is designed after the mathematical model for heat flow. As heat gradually flows from hotter to cooler parts of a metallic body until a uniform temperature is achieved throughout the body, it was expected that in Ricci flow, regions of higher curvature will tend to diffuse into regions of lower curvature to produce an equilibrium geometry for the 3-manifold for which Ricci curvature is uniform over the entire manifold. Thus in principle, a 3-manifold when subject to Ricci flow will produce a kind of normal form which will ultimately solve the geometrization conjecture. Although Hamilton established a number of beautiful geometric results using the Ricci flow equation, the progress in applying this program to the conjecture eventually came to a standstill mainly because of the formation of singularities, which defied solution of the problem. In his proof, Perelman constructed a program for getting around to these obstacles. He modified the Ricci flow used by Hamilton with "Ricci flow with surgery". This expunges the singular regions as they develop in a controlled way and eventually solves the geometrization conjecture.

Keywords Poincaré conjecture · Ricci curvature

Perhaps the greatest breakthrough of the new millennium is the solution of a century-old problem, called the Poincaré conjecture, by the Russian mathematician Grigoriy

A. Mukherjee (✉)
Indian Statistical Institute, Kolkata 700108, India
e-mail: amiya@isical.ac.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_3

Perelman. This research is the completion of a program initiated in the 1980s by Richard Hamilton to solve a more general conjecture, called the geometrization conjecture, by William Thurston. The main tool of this program is the Ricci flow, a differential equation on the space of Riemannian metrics. The present lecture provides a brief description of the Ricci flow program for the solution of the geometrization conjecture.

During 2002–2003, Perelman posted a proof of the conjecture on the Internet in three instalments [15, 16] and [17]. The proof was unconventional, having no direct mention of the Poincaré or the Thurston conjecture. It was extremely brief at crucial points which could have been elaborated to many pages, and it contained many elegant results which were irrelevant to the central argument. The proof was really a challenge, and only a few mathematicians had the expertise necessary to evaluate and defend it. At least two groups of experts examined the proof for four years and found no significant errors or gaps. It had also been scrutinized in various seminars around the world. By 2006, detailed exposition of Perelman's work had appeared in three separate manuscripts [1, 13], and [14], each more than 300 pages in length. In 2007, a committee of prominent mathematicians of the International Mathematical Union nominated Perelman for a Fields Medal, traditionally considered the highest honour in mathematics. But Perelman had declined the prize, even after long persuasion by the President Sir John Ball of the Union. The reclusive mathematician said that if everybody understood the correctness of his proof, then no other recognition was necessary. In 2010, the Clay Mathematical Institute announced that Dr. Grigoriy Perelman was the recipient of the First Millennium Prize for resolution of Poincaré conjecture. Perelman also turned down the prize saying that his contribution is no greater than that of Richard Hamilton who introduced the theory of Ricci flow for solving the conjecture.

The conjecture was formulated by a French mathematician Henri Poincaré in 1904 at the end of a sixty-five-page research paper [18]. Poincaré is regarded as one of the most creative mathematicians of the nineteenth century, and he was the founder of the subject topology. The version stated by Poincaré is equivalent to the following.

3.1 Poincaré Conjecture

Any closed simply connected 3*-manifold is diffeomorphic to the standard* 3*-dimensional sphere* S^3.

Note that a manifold is closed if it is compact and without boundary, and that a manifold is simply connected if every simple closed curve in the manifold can be deformed continuously to a point without leaving the manifold, that is, if the fundamental group of the manifold is trivial.

The confirmation of the conjecture has important implications, not only in mathematics but also in astrophysics, for example, in the formation of black hole (see Witten [23]). However, proving it mathematically was far from easy. To the astonishment of most mathematicians, it turned out that the manifolds of the fourth, fifth, and higher dimensions were more tractable than those of the third. By 1980s Poincaré conjecture had been proved in all dimensions except the third by Smale [20] and Freedman [5] (also Donaldson [4]).

The problem may be reduced to a more tractable form. Recall that a Riemannian metric g on a manifold M is a smooth assignment to each point $x \in M$ a positive definite inner product $g_x : T_x(M) \times T_x(M) \longrightarrow \mathbb{R}$ on the tangent space $T_x(M)$. The metric tensor g induces a distance function on M making it a metric space, where the distance between a pair of points is the infimum of the length of rectifiable curves joining them. The sectional curvature of M assigns to each 2-dimensional plane $P \subset T_x(M)$ a real number $K(P)$. It is a smooth real-valued function on the Grassmann bundle of 2-planes over M. A model space M^n_K is a complete simply connected Riemannian n-manifold of constant sectional curvature K. Following special cases are important: M^n_0 is the Euclidean n-space $\mathbb{R}^n$, M^n_1 is the standard n-sphere S^n and M^n_{-1} is the hyperbolic n-space $\mathbb{H}^n$. Each model spaces M^n_K, for $K \neq 0$, can be obtained from one of the special cases S^n, $\mathbb{H}^n$ by scaling the metric. Explicitly, if $K > 0$, M^n_K is obtained from S^n by multiplying the distance function by $1/\sqrt{K}$, and if $K < 0$, M^n_K is obtained from $\mathbb{H}^n$ by multiplying the distance function by $1/\sqrt{-K}$. A classical result says that if (M, g) is a complete connected Riemannian n-manifold of constant sectional curvature K, and M is endowed with the distance function induced by g, then M is isometric to the quotient space $\Gamma\backslash M^n_K$, where Γ is a subgroup of the group of isometries of M^n_K, naturally isomorphic to the fundamental group of M. In fact, the universal covering $\widetilde{M}$ is isometric to M^n_K. Therefore the Poincaré conjecture is a consequence of the following conjecture.

3.2 Positive Scalar Curvature Conjecture

Any closed simply connected 3*-manifold admits a Riemannian metric of strictly positive constant sectional curvature.*

Again, the last conjecture is a particular case of the geometrization conjecture of William Thurston. Note that a Riemannian manifold is homogeneous if the group of isometries acts transitively on the manifold. A complete Riemannian manifold M is a geometric structure modelled on a given homogeneous manifold if every point of M has a neighbourhood isometric to an open set of the homogeneous manifold. Then M may be described as the quotient $\Gamma\backslash G/H$, where G is the isometry group of the universal covering $\widetilde{M}$, and Γ and H are discrete and compact subgroups of the Lie group G, respectively. In 1982, Thurston established that there are only eight such simply connected geometries G/H in dimension 3 which admit compact quotients. Five of these geometries are $\mathbb{R}^3$, S^3, $\mathbb{H}^3$, $S^2 \times \mathbb{R}$, and $\mathbb{H}^2 \times \mathbb{R}$. The remaining three are (1) nontrivial S^1 bundle over the torus T^2 (Nil geometry), (2) nontrivial T^2 bundle over S^1 (Sol geometry), and (3) nontrivial S^1 bundle over a surface of genus > 1 ($\widetilde{SL_2\mathbb{R}}$ geometry) (good references are [19] and [22]). Thurston's geometrization conjecture is a far-reaching generalization of the Poincaré conjecture. A simple minded version of the conjecture says that the above eight geometries are the building blocks of any 3-manifold.

3.3 Geometrization Conjecture

Any closed orientable 3*-manifold can be canonically cut along embedded* 2*-spheres and* 2*-tori so as to decompose into the above eight geometric pieces.*

If this is confirmed, the Poincaré conjecture would be too. Thurston proved his conjecture in many important cases, the most celebrated one being for a Haken manifold, i.e. for a manifold which contains an incompressible surface of genus ≥ 1 [21].

In 1982, Hamilton wrote a paper [6], in which he crafted a well-developed program in the hope of resolving the Thurston conjecture. His strategy was to take an arbitrary 3-manifold M with a Riemannian metric g_0 and to deform g_0 in the space of Riemannian metrics on M by some regularizing process to a uniform metric with strictly positive sectional curvature. To guide the deformation, Hamilton [6] introduced a geometric evolution equation for metrics designed after the heat equation, which is the mathematical model for heat flow, and named it Ricci flow equation, after an early geometer Gregorio Ricci-Curbastro (1853–1925). Hamilton was motivated by the harmonic heat flow introduced by Eells and Sampson in 1964.

A one-parameter family $g(t)$ of Riemannian metrics on M is called a Ricci-flow if

$$\frac{\partial g}{\partial t} = -2\mathrm{Ric}_{g(t)}.$$

Here $\mathrm{Ric}_{g(t)}$ is the Ricci curvature of the metric $g(t)$, which is, like the metric itself, a symmetric bilinear form on each tangent space $T_x(M)$. Roughly speaking, the Ricci curvature is a measure of volume distortion, that is, it measures the degree to which 3-dimensional volumes in coordinate neighbourhoods of M differ from the volumes of corresponding neighbourhoods in $\mathbb{R}^3$. In dimension 3, the Ricci curvature completely determines the local geometry of the metric g on M. Under the Ricci flow, the Ricci curvature spreads around M. As heat gradually flows from hotter to cooler parts of a metallic body until the body reaches an equilibrium constant temperature, it was the intuition of Hamilton that in Ricci flow, regions of higher curvature will tend to diffuse into regions of lower curvature to produce an equilibrium geometry for M for which Ricci curvature is constant. Note that if the Ricci curvature is constant, i.e. $\mathrm{Ric}_g = l \cdot g$ for some $l \in \mathbb{R}$ (and in this case the metric g is called Einstein metric), then the sectional curvature is also constant. However, there is one significant difference between the two situations. The Ricci flow equation is nonlinear, unlike the heat equation which is linear. The equation involves a conflict between the diffusion (or linear) term, which tends to disperse the concentration of curvature uniformly on M, and the reaction (or nonlinear) term, which tends to build up the concentration of curvature. If the nonlinearity dominates, the curvature may explode to $+\infty$ developing singularity in finite time. The fact is that for any smooth initial metric on a compact M there is a maximal time interval $[0, T)$ on which a unique smooth solution $(M, g(t))$ of the Ricci flow exists, where either $T = +\infty$ or else the sectional curvature is unbounded as $t \to T$. In this case, the solution is said to become singular at time T (and T is called a singular time), because the solution cannot be extended further beyond T.

The most striking result of Hamilton [6] is the following theorem, which is the first step on way to the Poincaré conjecture, via the positive sectional curvature conjecture.

3.4 Positive Ricci Curvature Theorem

If a connected closed 3*-manifold M admits a Riemannian metric of strictly positive Ricci curvature, then M also admits a Riemannian metric of constant strictly positive sectional curvature.*

To get the proof of the theorem, one runs Ricci flow, starting with an initial metric $g(0)$ with positive Ricci curvature. This will cause the manifold shrink to a point at a singular time T, getting rounder and rounder as it shrinks. As t approaches T, if one continually rescales the flow so as to have constant volume, then the rescaled sectional curvatures become closer and closer to being a constant. Finally, one obtains in the limit a Riemannian metric on M with constant positive sectional curvature.

In general, the Ricci flow may behave wildly outside the class of positive Ricci curvatures. Also, the flow can go singular before it can shrink to a point. An example of this type singularity is the standard neckpinch, in which a cross-section $\{0\} \times S^2$ in a topological neck $(-1, 1) \times S^2 \subset M$ shrinks to a point in a finite time. In the case of neckpinch singularity, Hamilton's idea was to perform surgery on the neckpinch. The operation of surgery consists of removing a neighbourhood $(-\varepsilon, \varepsilon) \times S^2$ of the shrinking 2-sphere and glueing 3-dimensional balls onto the resulting boundary 2-spheres $\{-\varepsilon\} \times S^2$ and $\{-\varepsilon\} \times S^2$. The surgery operation changes the topology and the geometry of the manifold, but they are controlled, because the original manifold can be recovered from the ensuing manifolds by connected sums (recall that the connected sum of two manifolds is obtained by removing a small 3-ball from each summand and then glueing the boundaries together). One then lets the ensuing manifolds evolve by the Ricci flow again. If one encounters another neckpinch, then one performs a new surgery, lets the new manifolds evolve, and so it goes.

It follows that if all the singularities were actually caused by neckpinches, then the Thurston conjecture would have been proved following the above method of Hamilton. Thus the major steps in the proof of the Thurston conjecture involved examining what sort of singularities develop when a manifold undergoes the Ricci flow, determining whether this surgery process can be completed, and wondering whether the surgery might be needed to be repeated infinitely many times. These problems were completely solved by Perelman, as we shall glimpse in a moment little later.

Over the years, Hamilton and others obtained profound results about Ricci flow. Some of these results are the Hamilton-DeTurek work on the short-time existence and uniqueness of Ricci flow solution [3, 6], Hamilton's maximal principle for Ricci flow solution [7], the Hamilton-Chow work on Ricci flow on surfaces [2, 8], Hamilton's work on Harnack inequality for Ricci flow solutions with nonnegative curvature operator [9], Hamilton's compactness theorem for Ricci flow solutions [10] and the Hamilton-Ivey curvature pinching estimate for three-dimensional Ricci flow solutions [11, 12]. Another landmark result of Hamilton is the following theorem.

3.5 Normalized Flow Theorem

If a volume normalized Ricci flow (which is a variant of the Ricci flow in which the volume remains constant) on a closed connected orientable 3*-manifold M has a smooth solution that exists for all positive time, then the geometrization conjecture holds for M.*

In spite of these great deal of pioneering research on Ricci flow, Hamilton could not tame the singularities. For example, he encountered a troublesome possibility that might occur in a blowup limit as an ancient solution $\mathbb{R} \times$ (cigar soliton). Here an ancient solution is a solution that exists on a maximal time interval $(-\infty, T)$, $T < \infty$ and a cigar (the terminology is due to Hamilton) is a complete Riemannian surface $(\mathbb{R}^2, g_\Sigma)$, where g_Σ is a Kähler metric $g_\Sigma = dz \cdot d\bar{z}/(1 + |z|^2)$ on $\mathbb{C} \approx \mathbb{R}^2$. This surface is also known in the physics literature as Witten's black hole [23]. As $r = |z| \to \infty$, the cigar metric g_Σ becomes asymptotic to a cylinder of radius 1. This particular solution is undesirable, because there might not be 2-spheres along which to do surgery, and it would be impossible to achieve uniform geometry. Hamilton conjectured in [11, §26] that this type of solution could be avoided by means of a suitable generalization of his "Little Loop Lemma" [11, §15].

In his paper [15], Perelman realized that in addition to the properties that Hamilton had established there was one more crucial integral quantity that Hamilton had not considered. This quantity is a functional $\mathscr{W} : \mathscr{M} \longrightarrow \mathbb{R}$, $\mathscr{M}$ being the space of Riemannian metrics on M, and Perelman obtained this by enhancing the Einstein-Hilbert functional $\int_M R(g)$, where $R(g)$ is the scalar curvature of g, the trace of the Ricci curvature Ric_g, so that the gradient of $\mathscr{W}$ is the Ricci flow. By analogy with statistical mechanics (the mathematical model for the laws of thermodynamics), Perelman called the functional $\mathscr{W}$ "entropy". The entropy always increases along the Ricci flow. Using this concept, and another monotonic quantity he called "reduced volume", which is closely related to an eigenvalue of a certain "elliptic equation", Perelman proved his first ground breaking result, the No Local Collapsing Theorem [15].

3.6 No Local Collapsing Theorem

Let M be a closed n-manifold. If $(M, g(\cdot))$ is a given Ricci flow solution that exists on a time interval $[0, T)$, $T < \infty$, then for any $\rho > 0$, there is a number $\kappa > 0$ with the following property. Suppose that $r \in (0, \rho)$ and let $B_t(x, r)$ be a metric r-ball in a time-t slice. If the sectional curvatures on $B_t(x, r)$ are bounded in absolute value by $1/r^2$, then the volume of $B_t(x, r)$ is bounded below by κr^n.

This theorem rules out the possibility of $\mathbb{R} \times$ (cigar soliton) solution that had worried Hamilton.

The main achievement of the first paper [15] of Perelman is the following. For the classification of the singularities, Hamilton's technique was to rescale the flow, obtain Ricci flows with bounded curvatures, and then try to take limit of these rescaled flows. Then one would classify the limits and obtain a description of $g(t)$ near points of high curvature. An important problem in this strategy is the existence of these limits, and this involved controlling the injective radius of the metric. Perelman used his No Local Theorem to give a complete classification of these limits, called κ-solutions. Finally he used his Canonical Neighbourhood Theorem (Theorem 12.1) to show that the points with large curvature have neighbourhoods closed to κ-solutions, thus with canonical geometry. Perelman concluded his first paper by proving the following theorem.

3.7 Smooth Flow Theorem

If the Ricci flow on a closed orientable 3-manifold has a smooth solution that exists for all positive time, then M satisfies the geometrization conjecture.

In the second paper [16], Perelman constructed a surgery algorithm to deal with Ricci flow with singularities. Consider a 3-dimensional Ricci flow $(M, g(t))$, $0 \leq t < T$, going singular at time $T < \infty$, Perelman extended the flow past time T by constructing a Ricci flow with surgery. He deduced that at the singular time T there is a limiting metric (possibly not complete) defined on an open submanifold $\Omega \subset M$. The ends of Ω are diffeomorphic to $S^2 \times [0, 1)$ with metrics at any point as the product metric of a rescaled version of a round metric on S^2 with the Euclidean metric on the interval $[0, 1]$ and with curvature tending to $+\infty$ as one approaches the ends of these tubes. The surgery consists of cutting off the ends of these tubes along one of the 2-spheres in the product structure where the curvature explodes and sewing in 3-balls to construct a new compact Riemannian 3-manifold $(M', g(T))$. This is called time slice of the Ricci flow with surgery at time T. One then restarts the Ricci flow at time T with $(M', g(T))$ as the initial metric. This flow will go singular at some time $T' > T$. The point is that as we cross a singular time both the topology and the geometry of the time-slice change, but in a controlled way.

Perelman showed in [16] that starting with any compact Riemannian 3-manifold, this process can be repeated forever to construct a Ricci flow with surgery defined for all positive times. Moreover, the singular times are discrete, and the topology of the manifold before surgery can be deduced from the topology of the manifold after surgery. In particular, it can be shown from the description of the topological change as one crosses a singular time, that if the manifold after a surgery satisfies the geometrization conjecture, then the manifold just before surgery also satisfies the geometrization conjecture. Then arguing by induction, one sees that if any time-slice of the Ricci flow with surgery satisfies the geometrization conjecture, then so does the initial manifold.

The proof of the Poincaré conjecture is now straightforward. Start with a closed simply connected 3-manifold M with a Riemannian metric $g(0)$. Construct a Ricci flow with surgery defined for all time with $(M, g(0))$ as the 0 time-slice. As noted above, if any time-slice of this Ricci flow satisfies the geometrization conjecture, then so does M. So the proof of the Poincaré conjecture may be completed by showing that the time-slices of this Ricci flow with surgery at all sufficiently large times are empty, that is, the Ricci flow with surgery becomes extinct at some finite time. This was done in [17].

References

1. H.-D. Cao, X.-P. Zhu, A complete proof of the Poincaré and geometrization conjectures—application of the Hamilton-Perelman theory of the Ricci flow. Asian J. Math. **10**, 165–492 (2006)
2. B. Chow, The Ricci flow on the 2-sphere. J. Diff. Geom. **33**, 325–334 (1991)
3. D. DeTurck, Deforming metrics in the direction of their Ricci tensors. J. Diff. Geom. **18**, 157–162 (1983)
4. S. Donaldson, An application of gauge theory to four-dimensional topology. J. Diff. Geom. **18**, 279–315 (1983)
5. M. Freedman, The topology of four-dimensional manifolds. J. Diff. Geom. **17**, 357–453 (1982)
6. R. Hamilton, Three-manifolds with positive Ricci curvature. J. Diff. Geom. **17**, 255–306 (1982)
7. R. Hamilton, Four-manifolds with positive curvature operator. J. Diff. Geom. **24**, 153–179 (1986)
8. R. Hamilton, The Ricci, flow on surfaces, in *Mathematica and general relativity (Santa Cruz, CA)*, Contemp. Math. Amer. Math. Soc. Providence, RI, 1988. **71**, 237–262 (1986)
9. R. Hamilton, The Harnack estimate for the Ricci flow. J. Diff. Geom. **37**, 225–243 (1993)
10. R. Hamilton, A compactness property for solutions of the Ricci flow. Amer. J. Math. **117**, 545–572 (1995)
11. R. Hamilton, *The formation of singularities in the Ricci flow, surveys in Diff*, vol. II, Geom (International Press, Cambridge, MA, 1995), pp. 7–136
12. T. Ivey, Ricci solitons on compact three-manifolds. Diff. Geom. Appl. **3**, 301–307 (1993)
13. B. Kleiner, J. Lott, Notes on Perelman's papers. http://www.arxiv.org/abs/math.DG/0605667 (2006)
14. J. Morgan, G. Tian, Ricci flow and the Poincaré conjecture. http://www.arxiv.org/abs/math.DG/0607607 (2006)
15. G. Perelman, The entropy formula for the Ricci flow and its geometric applications. arXiv:math.DG/0211159 (2002)
16. G. Perelman, Ricci flow with surgery on three-manifolds. arXiv:math.DG/0303109 (2003)
17. G. Perelman, Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. arXiv:math.DG/0307245 (2003)

18. H. Poincaré, *Cinquième complèment à l'analysis situs* (Œuvres Tome VI, Gauthier-Villars, Paris, 1953)
19. P. Scott, The geometries of 3-manifolds. Bull. London Math. Soc. **15**, 401–487 (1983)
20. S. Smale, Generalized Poincaré's conjecture in dimensions greater than four. Ann. Math. **74**(2), 391–406 (1961)
21. W. Thurston, Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. Bull. Amer. Math. Soc. (N.S.) **6**, 357–381 (1982)
22. W. Thurston, Three-dimensional geometry and topology, vol. 1. Princeton Mathematical Series, 35 (Princeton University Press, New Jersey, 1997)
23. E. Witten, String theory and black holes. Phys. Rev. D **3**, 44 (1991). **2**, 314–324

Chapter 4
An Introduction to Data Assimilation

Amit Apte

Abstract This talk will introduce the audience to the main features of the problem of data assimilation, give some of the mathematical formulations of this problem and present a specific example of application of these ideas in the context of Burgers' equation.

Keywords Data Assimilation · Burgers' equation · Kalman filter

The availability of ever increasing amounts of observational data in most fields of sciences, in particular in earth sciences, and the exponentially increasing computing resources have together lead to completely new approaches to resolving many of the questions in these sciences, and indeed to formulation of new questions that could not be asked or answered without the use of these data or the computations. In the context of earth sciences, the temporal as well as spatial variability is an important and essential feature of data about the oceans and the atmosphere, capturing the inherent dynamical, multiscale, chaotic nature of the systems being observed. This has led to development of mathematical methods that blend such data with computational models of the atmospheric and oceanic dynamics—in a process called data assimilation—with the aim of providing accurate state estimates and uncertainties associated with these estimates.

This expository talk (and this short article) aims to introduce the audience (and the reader) to the main ideas behind the problem of data assimilation, specifically in the context of earth sciences. I will begin by giving a brief, but not a complete or exhaustive, historical overview of the problem of numerical weather prediction, mainly to emphasise the necessity for data assimilation. This discussion will lead to a definition of this problem. In Sect. 4.2, I will introduce the "least squares" or variational approach, relating it to the Kalman filter and the full-fledged Bayesian approach. In the final section, I will introduce the Burgers' equation for which I will present some results and ongoing research on variational and Bayesian approaches to data assimilation.

A. Apte (✉)
International Centre for Theoretical Sciences (ICTS-TIFR), Bangalore, India
e-mail: apte@icts.res.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_4

4.1 Data Assimilation in the Context of Earth Sciences

In order to highlight the need for incorporating observational data into numerical models of the atmosphere and the oceans, I will give a brief historical account of numerical weather prediction, using it to drive towards a definition of data assimilation that I find most convenient to keep in mind. Extensive descriptions of this history are available from various sources such as [12, 14–16] and the account below is far from complete, serving only to set the context for data assimilation.

It was Vilhelm Bjerknes who was the first one to develop, in 1904, the idea of predicting weather using the hydrodynamic and thermodynamic equations for the atmosphere. Note that this was around 90 years after Laplace discussed [13], the concept of what is now commonly known as *Laplace's deamon*: "Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes". This is one of the first but still most complete statements of determinism, ironically in an essay on probabilities! Bjerknes' program was also formulated (i) around 80–60 years after the formulation of Navier–Stokes equations for viscous flow and the equations of thermodynamics, (ii) around the same time as Poincaré's explorations of chaotic motion in celestial dynamics, but (iii) at least 60 years before the implications of chaos came to be appreciated widely by the scientific community.

The first actual attempt at executing Bjerknes' idea was made by Lewis Fry Richardson about a decade later. He basically attempted to solve the partial differential equations for the atmosphere by dividing Europe into 200 Km by 200 km blocks. The actual calculation for a six-hour forecast took him weeks of calculation by hand! And at the end of it all, his prediction for the change in pressure in six-hour period was larger by around 100 times than the actually observed change in pressure. Some of Richardson's comments that are of relevance to us were as follows [17]:

(1) "It is claimed that the above [predictions] form a fairly correct deduction from a somewhat unnatural initial distribution"—I would interpret this as pointing out the need for accurate and smooth initial condition for a good weather forecast.

(2) "... 64,000 computers would be needed to race the weather for the whole globe. ... Imagine a large hall like a theatre ... The walls are painted to form a map of the globe. ... A myriad computers are at work upon the weather of the part of the map where each sits ... The work is coordinated by an official of higher rank ..."—essentially he was foreseeing the use of supercomputers and parallelization of computational tasks in attempting to solve this problem.

There is also another reason for the failure of Richardson's attempt: the sparsity and non-smoothness of the observational data available to him. When his experiment was reproduced [15] with smoothed version of the same data, the prediction was quite accurate. In the context of data assimilation, the so-called "background" or the prior that we will discuss in detail later in Sect. 4.2 provides a systematic way of achieving this "smoothing".

Over decades, several new advances have been made in addressing the problem of weather prediction. The most notable of these are: (i) the realisation by Charney and others of the importance of quasi-geostrophy and the subsequent development of models based on these ideas; (ii) the use of electronic computers, pioneered by von Neumann in the 1950s, leading to ever increasing resolution for these models; (iii) improved observations, including explosion in the number of observations from satellites; and last but the most relevant to us (iv) improved mathematical methods for incorporating these observations into the models—this is the domain of data assimilation.

One of the fundamental characteristics of the atmospheric and oceanic dynamics is its chaotic nature, which is manifested in the sensitivity to the initial conditions and puts severe restrictions on the predictability of these systems. Chaotic systems have been studies extensively for a little over a hundred years, beginning with Poincaré's work on Hamiltonian chaos, and continuing with the work of Cartwright and Littlewood, Ed Lorenz's famous study of the "Lorenz system" and many, many others. One important implication of the presence of chaotic solutions is that for predicting the state of systems such as the atmosphere, it is necessary to continually use observational data in order to "correct" the model state and bring it closer to reality.

4.1.1 Incomplete Models; Inaccurate, Insufficient Data

Models of the atmosphere and the oceans, or I would argue, of *any* physical system in general, are necessarily incomplete, in the sense that they do not represent all the physical processes in the system, even though the models are based on sound and well-tested physical theories of nature. For example, it is well-known that processes associated with convection and cloud formation are not captured well by models, and even with the highest foreseeable resolution, there will be features of these processes that cannot be captured. Arguably, such inadequacies also apply to models of even "simple, low-dimensional" chaotic systems such as nonlinear electronic circuits, but I will not delve into this point further.

Additionally, even if we have an atmospheric model which is "perfect" (without precisely defining what I mean by "perfect"), if we do not use observational data, we will only be able to make qualitative and quantitative statements about "generic" states on the chaotic attractor of the model, or statistical statements about averages of quantities. Without observational data, we will be unable to make predictions of states of the atmosphere, or compare the time series of a specific variable such as temperature with observational time series. In this sense, data provide a crucial link to reality. More specifically, we hope that the data will help us make statements about one specific trajectory of the atmosphere; out of the infinitely many trajectories that comprise the chaotic attractor of the system. It is important to keep in mind that it is entirely unclear whether a single trajectory, or even a large ensemble of trajectories, of an incomplete and imperfect model of a chaotic system can provide useful predictive information about the real state of such a system, but again I will not dwell on this point further.

The problem of how the data can be used quantitatively is far from trivial since the data themselves are (i) inaccurate, in the sense that they contain inherent measurement errors and noise, and (ii) insufficient, in the sense that they do not completely specify the state of the system by themselves. For example, most atmospheric and oceanic data are not uniformly distributed in space or time and are quite sparse compared to the resolution of the numerical models being used. Additionally, there are inaccuracies in the relation of these observational data with the model of the system, which I will refer to as imperfections in the observational model. Thus the problem of determining the state of the atmosphere from the data alone is an under-determined, ill-posed inverse problem.

4.1.2 A Definition of Data Assimilation

The above discussion has hopefully made it clear that models and data must be brought together in order to attack the problems such as weather prediction. This is precisely the main idea behind data assimilation. Thus we could define data assimilation to be the art of

(1) *optimally incorporating*
(2) *partial and noisy observational data* of a
(3) *chaotic and thus necessarily nonlinear (and most often, complex, multiscale) dynamical system,* with an
(4) *imperfect model of the system dynamics and of the observations,* in order to obtain
(5) *predictions of and the associated uncertainty for the state of the system,* repetitively in time.

A caricature of this definition is captured in a schematic shown in Fig. 4.1. To recap, a few of the characteristics of data assimilation problem are as follows:

(1) We have good physical theories to describe the systems but the models based on these theories are incomplete and imperfect.
(2) The systems are described by nonlinear, chaotic, but usually deterministic dynamical models.
(3) The dynamics contain interactions across multiple temporal and spatial scales and occur on very high or infinite dimensional state space.
(4) The observational data of the system are

   (a) very high dimensional,
   (b) noisy and inaccurate,

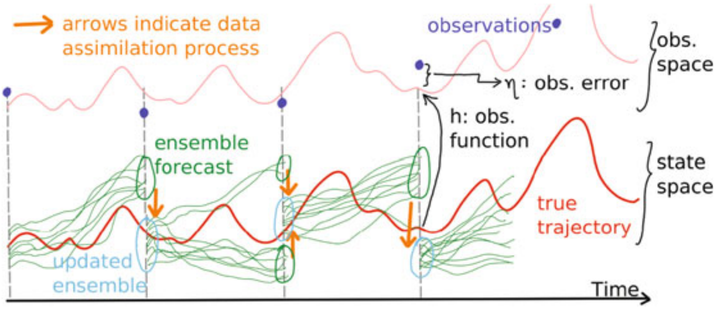


Fig. 4.1 A schematic representation of data assimilation process

(c) partial and usually sparse, and
(d) discrete in time.

I will now discuss some of the mathematical formulations that are used frequently to resolve this problem. It will become clear that each of these formulations takes into account only some, but not all, of the above characteristics. Thus formulating an overarching mathematical framework for data assimilation that addresses all the features above is a formidable and interesting modelling problem.

4.2 Least Squares Approach, Kalman Filter, and the Bayesian Formulation

Consider the problem of estimating n quantities $x = (x_1 \ldots, x_n) \in \mathbb{R}^n$ using m noisy observations $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$, which are related to the unknowns x through *observation function* (or matrix, in case of linear relation): h (or H): $\mathbb{R}^n \to \mathbb{R}^m$. That is, in the absence of noise, the observations $\hat{y}$ of unknowns x would be $y = h(x)$ (or $= Hx$). We assume that $m < n$ (and typically $m \ll n$). Thus even in the linear case, H will not be an invertible matrix.

For example, x could be the velocity field on a model grid whereas y could be the velocity measurement at locations which may or may not be on the same grid. In this case, H will be the matrix for projection or interpolation between model grid and measurement grid.

The "least squares" formulation of this problem is to find the minimum x^a of the following *cost function*:

$$\tilde{J}(x) = \frac{1}{2}\|y - Hx\|_R^2, \tag{4.1}$$

where the norm $\|z\|_R^2 = z^T R^{-1} z$ simply gives different weights to different observations. Of course, since H is not invertible, we immediately see that

$\nabla \tilde{J} = -H^T R^{-1}(y - Hx^a) = 0$ cannot be solved for x^a, we need to “regularise” this problem. This can be achieved by, e.g. modifying the cost function to

$$J(x) = \tilde{J}(x) + \frac{1}{2}\|x - x^b\|^2_{P^b}, \tag{4.2}$$

where x^b is some “background” (or a priori) information about x, and the norm $\|\cdot\|_{P^b}$ again gives different weights to different components of this background state x^b. This cost function is minimised by solving $\nabla J = (H^T R^{-1} H + (P^b)^{-1})x^a - (H^T R^{-1} y + (P^b)^{-1} x^b) = 0$ and the solution is given by

$$x^a = x^b + [H^T R^{-1} H + (P^b)^{-1}]^{-1} H^T R^{-1}(y - Hx^b). \tag{4.3}$$

Here $I := (y - Hx^b)$ is called the *innovation* vector and the prefactor of innovation

$$K = [H^T R^{-1} H + (P^b)^{-1}]^{-1} H^T R^{-1} = P^b H^T (H P^b H^T + R)^{-1} \tag{4.4}$$

is called the “Kalman gain matrix”. Equation (4.3) may also be rewritten as $x^a = x^b + K(y - Hx^b)$ to make it clear the *analysis* x^a is a linear combination of the *background* x_b and the innovation I. (Note that the final equality in Eq. (4.4) is obtained using the Sherman–Morrison–Woodbury identity.)

This calculation can be extended to the case when the unknown x is the initial condition of a dynamical system, whereas the observations are spread over time. We will now see that such a setup leads to the Kalman filter.

4.2.1 Observations over Time—Kalman Filter

Consider a *linear* dynamical model, $x_{i+1} = m(x_i) = Mx_i$ on $\mathbb{R}^n$. The initial condition x_0^+ is unknown and we wish to determine it based on some observations $y_i = h_i(x_i^+)$ (plus noise) for $i = 1, \ldots, N$. Here x_i^+ is the trajectory starting with the initial condition x_0^+, i.e. $x_i^+ = M^i x_0^+$. In parallel with our approach above, we can now consider a cost function

$$J(x_0) = \sum_{i=1}^{N} \frac{1}{2}\|y_i - H_i M^i x_0\|^2_{R_i} + \frac{1}{2}\|x_0 - x^b\|^2_{P^b}, \tag{4.5}$$

where the last term is the “background” term that regularises the minimisation problem. This is a quadratic function of x_0 and thus has a unique minimum which can be obtained by setting $\nabla_{x_0} J = 0$. But we will now consider the following interactive way of calculating this minimum, defined by a two stage process:

(1) The “forecast” step gives the dynamical evolution of the state from step $k - 1$ to next step k:

$$x_k^f = Mx_{k-1}^a \quad \text{and} \quad P_k^f = MP_{k-1}^a M^T. \tag{4.6}$$

(2) The "update" step gives the "analysis" x_k^a as a linear combination of the observation y_k at step k with the forecast x_k^f:

$$x_k^a = x_k^f + K_k I_k \quad \text{and} \quad P_k^a = (I - K_k H_k)P_k^f, \tag{4.7}$$

where $I_k = y_k - H_k x_k^f$ is the innovation and $K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}$ is the Kalman gain. This iteration is begun with $x_0^a = x^b$ and $P_0^a = P^b$ at $k = 1$ and continues until $k = N$. At the end of these N steps, we get the final "analysis" given by $(x_N^a; P_N^a)$. This two-step process is known as the Kalman filter.

The main relation of the result of this two-step process to the least squares problem of minimization of the cost function is as follows: suppose x_0^m is the minimum of the cost function $J(x_0)$ from Eq. (4.5), and P_0^m be its Hessian. Then the dynamical evolution of this minimum and the Hessian from step 0 to step N given by the following equations,

$$x_N^m = M^N x_0^m \text{ and } P_N^m = M^N P_0^m (M^N)T \tag{4.8}$$

is exactly the same as the analysis of the Kalman filter: $x_N^m = x_N^a$ and $P_N^m = P_N^a$. Thus, Kalman filter provides a way of solving the minimization problem. It is important to note that this equivalence of Kalman filter and the variational approach holds only in the case when the dynamical model is linear and the observations depend linearly on the state.

In practical problems in earth sciences, there are two main reasons why the Kalman filter is not usable. First, the size of the system n is quite large, usually $n = 10^6$ or more. In these cases, it is impossible to solve the above equations for Kalman filter, since they involve the $n \times n$ covariance matrices P_k^f which are impossible to store and manipulate for such large system sizes. The second equally serious reason is that these systems are nonlinear and chaotic as we saw in the previous section. Thus the above equations need to be modified appropriately.

I will only mention the two main modifications that address these two issues separately. The extended Kalman filter (EKF) is designed to work with nonlinear systems which are close to being linear. The ensemble Kalman filter (EnKF) is a set of methods designed to work with an ensemble of states, but without the explicit construction of the covariance matrices P_k^f. The EnKF and the variational methods are two of the most commonly used methods in the earth sciences, but there are several theoretical and practical problems that are still being investigated.

Before moving on to describing the variational and Bayesian approaches to data assimilation in the context of Burgers' equation, I will very briefly introduce the Bayesian framework in the next paragraph.

4.2.2 Bayesian Formulation

Let us consider a deterministic, discrete time dynamical model for $x \in \mathbb{R}^n$:

$$x_{n+1} = m(x_n) \text{ or equivalently } x_n = \Phi(x_0; n). \tag{4.9}$$

The initial condition x_0 is the unknown, but we will assume that we know a prior distribution for it, given by a density $\zeta(x_0)$. We will consider the case when noisy observations y_k at time k depend on the state x_k at that time and contain additive noise η_k:

$$y_k = h(x_k) + \eta_k = h(\Phi(x_0, k)) + \eta_k, \; k = 1, \dots, N. \tag{4.10}$$

I will only talk about the so-called *smoothing problem* which is to assimilate all these observations to get an estimate of the initial condition x_0. For this purpose, the observations are concatenated:

$$y = \{y_k\}_{k=1}^{y} = H(x_0) + \eta, \tag{4.11}$$

where

$$H(x_0) = \{h(\Phi(x_0, k))\}_{k=1}^{N} \text{ and } \eta = \{\eta_k\}_{k=1}^{N}. \tag{4.12}$$

We will assume that the noise η has a density. Then this density indeed gives the conditional probability of the observation y given the initial condition x_0, i.e. $p(y|x_0)$. For example, if $y \in R^m$ and $\eta \sim \mathcal{N}(0, \Sigma)$ (Gaussian observational errors), then

$$p(y|x_0) \propto \exp\left(-\hat{J}(x_0, y)\right), \quad \text{where } \hat{J}(x_0, y) = \frac{1}{2}\|y - H(x_0)\|_{\Sigma}^2. \tag{4.13}$$

But we are really interested in the probability density for the initial condition x_0. This will be obtained as *the posterior probability density* as given by the Bayes' rule:

$$p(x_0|y) = \frac{p(y|x_0)p(x_0)}{p(y)}, \quad \text{where } \; p(y) = \int p(y|x_0)p(x_0)\mathrm{d}x_0 \; \text{ is a constant.} \tag{4.14}$$

In the context of data assimilation,

$$p(x_0|y) \propto \zeta(x_0)\exp\left(-\hat{J}(x_0, y)\right). \tag{4.15}$$

For uncorrelated erros η_k (Σ is block diagonal with blocks R_k), this becomes

$$p(x_0|y) \propto \zeta(x_0)\prod_{k=1}^{K}\exp\left(-\frac{1}{2}\|y_k - h(\Phi(x_0, k))\|_{R_k}^2\right). \tag{4.16}$$

Now, we see that if the prior is Gaussian, e.g. $\zeta(x_0) \propto \exp\left(-\|x_0 - x_0^b\|^2_{Pb}/2\right)$, then the cost function introduced in Eq. (4.5) is exactly the logarithm of this density, if the dynamics and the observation function are both linear: $m(x) = Mx$ and $h(x) = Hx$. In this case, the minimum of the cost function of Eq. (4.5) is the *maximum a posteriori estimate.* Thus we see a close relation between the Bayesian framework, the least squares or the variational approach, and the Kalman filter.

All of the material in this section is discussed extensively in many existing reviews on various aspects of these problems. A fairly incomplete list of references is [3, 5, 6, 8, 11, 12] and references therein. In particular, the last two reference contain an excellent introduction to the relation of estimation theory to data assimilation, explaining in detail the various relations between the mean of the posterior distribution (conditional mean), the minimum variance estimator which in the linear case leads to the best unbiased linear estimator (BLUE) and the Kalman filter. There are several other topics that have a direct bearing on the data assimilation problem: stability and convergence of nonlinear filters and particle filtering, observability of nonlinear dynamical systems, probability measures and Bayes' theorem on infinite dimensional spaces that arise from partial differential equation models, etc. But a short introduction such as this certainly fails to provide a reasonable glimpse to these important relations, many of which are currently active topics of research.

Having introduced some of the main approaches to the data assimilation problem and their interrelations, I will now talk about a specific application in the case of Burgers' equation.

4.3 Data Assimilation for Burgers' Equation

We will work with the viscous Burgers' equation

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial z} = \nu\frac{\partial^2 v}{\partial z^2} \quad \text{with } v(t=0,z) = u(z) \text{ and } v(t,0) = 0 = v(t,1) \qquad (4.17)$$

on the domain $(z,t) \in \Omega \times (0,T)$ with $\Omega = (0,1)$. This is a nonlinear evolution equation which has unique solution: for $u \in H^1_0(\Omega)$, there exist a unique $v \in L^2(0,T;H^1_0(\Omega)) \cap C(0,T;H^1_0(\Omega))$ and I will indicate this map by $v(t) = \Phi(u,t)$. If fact, using Cole-Hopf transform, the exact solution can be written down, though I will not explicitly use this fact in this lecture.

There are several motivations for studying data assimilation problems for the Burgers' equation model. The dynamical models in earth sciences are based on the partial differential equations (PDE) of fluid dynamics of the air and water. These can be considered as infinite dimensional dynamical systems, of which the Burgers' equation is an example. It is also a nonlinear PDE whose solutions can be written analytically. Even though Burgers' equation does not exhibit the dynamical complexity of the atmospheric or oceanic flows, it acts as a toy model that is amenable to mathematical analysis whose qualitative features are still relevant to the more complicated scenarios in realistic applications.

For data assimilation problem in the context of Burgers' equation, I will now describe some of the known results as well as some of the ongoing research for three types of noisy observations. In all these cases, the problem will be determining the initial condition u given some observations, written as $y = H(u) + \eta$, with H being the observation operator. These three types of problems are as follows.

4.3.1 Observations Continuous in Space at a Specific Time

In this case, we will observe v for all $z \in \Omega$ at time T: thus $H(u) = \Phi(u; T)$ and η is a Gaussian measure on the Hilbert space $L^2(\Omega)$ supported on some appropriate subspace, e.g. $H_0^1(\Omega)$. Some of the questions of interest in this case are as follows. Suppose we consider a sequence of problems with observational noise $\eta_n = (1/n)\eta_0$— we have more accurate measurements as n increases. In the limit of $n \to \infty$, this reduces to a classical inverse problem of determining initial condition u from the solution at time T. The study of qualitative and quantitative behaviour of the posterior distribution for the initial condition u which is conditioned on the observations y, and in particular the limit $n \to \infty$, are some of the open problems, known as the problem of "posterior consistency" in the statistical literature. Some related results in context of linear models such as the heat equations and other related results are contained in, for example, [1, 2, 9, 10, 19]. A general overview of Bayesian approach to inverse problems, including extensive discussion of data assimilation is contained in [18].

4.3.2 Observations Continuous in Time

Here, we observe some function of v at all times $t \in (0, T)$. Thus at any given time, we assume a continuous map $C : H_0^1(\Omega) \to \mathbb{Z}$ into some Hilbert space $\mathbb{Z}$ (the observation space). Thus, if $Z^d(t)$ are the actual observations, then in terms of this map, the cost function for a variational formulation can be written as

$$J(u) = \frac{1}{2}\int_0^T \|C(\Phi(u,t)) - Z^d(t)\|_{\mathbb{Z}}^2 \mathrm{d}t + \frac{1}{2}\|u - u^b\|_{P^b}^2. \tag{4.18}$$

The question of uniqueness of minimum of this cost function has been studied for the cases of small enough observational time horizon T and for large time horizon T in [7, 20], respectively. The probabilistic formulation of this problem and the study of the corresponding posterior distribution is a possible direction for future research.

4.3.3 Observations Which Are Discrete in Space and Time

For this case, we take the observations of v at discrete space locations $\{z_i\}$ for $i = 1, \dots, K$ at discrete times $\{t_j\}$ for $j = 1, \dots, N$: Here, $H(u) = \{\Phi(u, t_j)(z_i)\}_{i=1,j=1}^{i=K,j=N}$ and η is a probability distribution on $\mathbb{R}^{NK}$. We do not yet have any theoretical results either about the minimum of a cost function in the variational formulation or about the behaviour of the posterior density in the Bayesian framework, e.g. in the case when the observational noise decreases. Some of the numerical results, possibly indicating presence of multiple minima, are presented in [4].

4.4 Concluding Remarks

The main aim of this lecture was to provide a short introduction to the topic of data assimilation, beginning with attempts to predict weather phenomena using numerical solutions of relevant partial differential equations. I discussed some of the most common approaches including variational methods, the variants of Kalman filter and the Bayesian framework providing interrelations between these. I also discussed some of these in the context of Burgers' equation which provides an example of a nonlinear partial differential equations where these techniques can be studies in great detail. All along, I tried to point out some of the open questions, emphasising the interdisciplinary nature of data assimilation research.

Acknowledgments I would like to thank the referee for comments which improved the exposition. A part of the research in the "data assimilation and dynamical systems" group of ICTS is supported by the Airbus International Teaching and Research Chair on "Mathematics of Complex Systems."

References

1. S. Agapiou, Aspects of bayesian inverse problems, Ph.D. thesis, University of Warwick, 2013
2. S. Agapiou, S. Larsson, A.M. Stuart, *Posterior Contraction Rates for the Bayesian Approach to Linear Ill-posed Inverse Problems* (2012). http://arxiv.org/abs/1203.5753
3. A. Apte, C.K.R.T. Jones, A.M. Stuart, J. Voss, Data assimilation: mathematical and statistical perspectives. Int. J. Numer. Meth. Fluids **56**, 1033–1046 (2008)
4. A. Apte, D. Auroux, M. Ramaswamy, Variational data assimilation for discrete burgers equation. Electron. J. Diff. Eq. Conf. **19**, 15–30 (2010)
5. F. Bouttier, P. Courtier, *Data Assimilation Concepts and Methods*, ECMWF Meteorological Training Course Lecture Series, March 1999
6. E.S. Cohn, An introduction to estimation theory. J. Met. Soc. Japan **75**, 257–288 (1997)
7. G. Cox, *Large-time Uniqueness in a Data Assimilation Problem for Burgers' Equation* (2012). http://arxiv.org/abs/1207.4782
8. G. Evensen, *Data Assimilation: The Ensemble Kalman Filter* (Springer, New York, 2007)
9. J.N. Franklin, Well-posed stochastic extensions of ill-posed linear problems. J. Math. Anal. **31**, 682–716 (1970)

10. A. Hofinger, H.K. Pikkarainen, Convergence rates for linear inverse problems in the presence of an additive normal noise. Stoch. Anal. Appl. **27**, 240–257 (2009)
11. A.H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic Press, 1970)
12. E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, 2003)
13. P.S. Laplace, *A Philosophical Essay on Probabilities* (Wiley, 1902), Translated from sixth French edition by F.W. Truscott, F.L. Emory. https://archive.org/details/philosophicaless00lapliala
14. E.N. Lorenz, Reflections on the conception, birth, and childhood of numerical weather prediction. Ann. Rev. Earth Planet. Sci. **34** (2006), 37–45. doi:10.1146/annurev.earth.34.083105.102317
15. P. Lynch, *The Emergence of Numerical Weather Prediction: Richardson's Dream* (Cambridge University Press, Cambridge, 2011)
16. A. Persson, F. Grazzini, User guide to ECMWF forecast products, Technical Report, ECMWF, 2005
17. L.F. Richardson, *Weather Prediction by Numerical Process* (Cambridge University Press, Cambridge, 1922)
18. A.M. Stuart, Inverse problems: a Bayesian perspective. Acta Numer. **19**, 451559 (2010)
19. S.J. Vollmer, *Posterior Consistency for Bayesian (elliptic) Inverse Problems Through Stability and Regression Results* (2013). http://arxiv.org/abs/1302.4101
20. W.L. White, A study of uniqueness for the initialization problem for Burgers' equation. J. Math. Anal. Appl. **172**, 412 (1993)

Chapter 5
Bose Einstein Condensation, Geometry of Local Scale Invariance, and Turbulence

Siddhartha Sen

Abstract Turbulent excitations have been observed in superfluid liquid helium, a Bose Einstein system, that obeys the Kolmogoroc $\frac{5}{3}$ scaling law for its energy spectrum. In recent joint work with Kouskik Ray of IACS we show how by regarding superfluid helium as a Bose Einstein quantum field theory with local 3 dimensional scale invariance leads to the Kolmogorov scaling law observed. In order to get local 3 dimensional scale invariance geometrical methods are needed, while in order to derive the observed turbulence scaling law the Bose Einstein condensation description of superfluid helium and Zakharov's weak wave turbulence method are used.

Keywords Bose Einstein condensate · Quasi-particle spectrum · Turbulence · Kolmogorov's scaling law

5.1 Introduction

I am happy to be participate in this special conference organised to celebrate the centenary of the Calcutta University, Department of Applied Mathematics and the 150th birth anniversary of the great patron of higher education and mathematics, Ashutosh Mukherjee. It is also an occasion where the great scientists S.N. Bose, M.N. Saha and N.R. Sen associated with the department are remembered. The work that I will describe today [1] uses the idea of Bose Einstein condensation and differential geometry, both topics of interest to Professor S.N. Bose. It also uses ideas of hydrodynamics, a topic of interest to Professor N.R. Sen.

Almost immediately after receiving Bose's paper on a new derivation of Planck's law of radiation where the idea of Bose statistics was first introduced Einstein applied these ideas to explain the then puzzling temperature dependence of the specific heat of molecules. Soon afterwards Einstein went on to suggest that the new statistics of

S. Sen (✉)
CRANN, Trinity College Dublin, Dublin 2, Ireland
e-mail: sen1941@gmail.com

S. Sen
R.K. Mission Vivekananda University, Belur 711202, West Bengal, India

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_5

Bose could also explain the low temperature superfluidity of helium He^4. Einstein's idea was that Bose statistics allowed a large number of helium molecules to condense to their ground state which would lead to the novel properties of liquid helium at low temperature that were observed. This idea that a large assembly of molecules can all settling down to the ground state of the system is called Bose Einstein condensation. It holds for integer spin molecules such as He^4 for which Bose statistics hold. Spin half objects, such as electrons, do not have this property however even for electrons the idea Of Bose Einstein condensation has been used to explain low temperature superconductivity by an ingenious method of creating zero spin objects from a pair of spin half electrons. The electrons being the carrier of current. In a series of experimental works [6] it was shown that if superfluid helium is heated filament excitations appear. The energy spectrum $E(k)$, where k is the wavenumber, of these filaments obey the scaling law $E(k) \approx k^{\frac{5}{3}}$ first predicted for ordinary turbulent liquid flows by Kolmogorov. An intuitive understanding of this law can be given as follows. For turbulent flows energy is pumped into the system at some large wavelength scale in the form of a circulating wave or large eddies. According to ideas of Richardson these large eddies break up into smaller eddies and this process continues. The dissipation of energy is supposed to happen for very small eddies. According to this picture there is thus a rather large region of wavelength where there is no dissipation but simply a progression of large eddies breaking up into smaller eddies. This is the Richardson cascade picture for the structure of turbulent flow in liquids. Let us show how this picture leads to Kolmogorov's law.

Suppose we have an eddy of size l_n, and circulating speed of v_n. Then there is a natural time scale of circulation $\tau_n \approx \frac{l_n}{v_n}$. The energy of the eddy is $\varepsilon_n \approx (v_n)^2$. A natural rate of transfer of energy from one scale to another can be taken to be r_n is $\approx \frac{(v_n)^2}{\tau_n}$. The intuitive idea used by Kolmogorov was that $r_n = r_0$ is independent of the length scale index n. Then

$$r_0 \approx \frac{(v_n)^2}{\tau_n} \approx \frac{(v_n)^3}{l_n}$$

Thus $(v_n)^2 \approx (r_0)^{\frac{2}{3}}(l_n)^{\frac{2}{3}}$. Changing to the wave number description by taking Fourier Transforms gives the Kolmogorov law. The key physically motivated input was the scale invariance of the energy transfer function r_n. A rigorous mathematical proof establishing this simple result starting from the basic equations of hydrodynamics remains an unsolved challenging problem.

Let us now turn to our system of interest which is, superfluid He^4, regarded as a Bose Einstein condensate and explain how such a system, described by a Schroedinger quantum field theory, can lead to Kolmogorv's scaling law. A quantum field theory is a standard way used to describe an assembly of helium molecules which also allows us to incorporate the idea of Bose Einstein condensation. As Kolmogorv scaling observed is a three dimensional property we ask the question: What structure is needed in order to make a free Schrodinger quantum field theory into one that has local 3 dimensional scale invariance? We find that we need to use the geometrical idea of gauge invariance to construct such a theory and that these geometrical ideas lead to the introduction of new fields and they completely fix the nature of the interactions between the helium molecules. It is these interactions that are responsible, as we show, for the emergence of scale invariance for the excitation energy spectrum. They lead to Kolmogorov's scaling law.

Thus we proceed to a construct an action for a 1 + 3-dimensional Schroedinger field theory which is invariant under local scaling in the three spatial dimensions. This is effected by introducing, following ideas of Weyl [2], a gauge field and a spatial metric. The gauge fields and spatial metric have kinetic energy terms that are also fixed by the requirement of three dimensional scale invariance. The invariance allows a Chern Simons term in the action but forbids a Maxwell term. The locally scale invariant action is unique in the sense that it contains all possible terms having polynomial interaction among the Schroedinger field, the gauge field and the metric. Moreover, gauge invariance for this system is rather novel due to the presence of the metric.

Historically, local scale invariance, was introduced by [2] in an attempt to unify the theories of gravitation and electromagnetism. Stipulating local scale invariance of the theory of General Relativity in four dimensions led to the extremely novel idea of introducing a gauge field with an additional term in the action resembling Maxwell's theory. Identifying this term as electromagnetism a unification of the theories of gravitation and electromagnetism was deemed to be have been achieved through purely geometric means. This approach was criticised, however, as being incompatible with the observed discrete spectra of atoms [3]. The idea was thus given up as a means to producing a unified field theory only to be revived later on with the local scaling of lengths replaced by a local change of phase of a quantum wave function [4]. This construction is now known as "gauge theory" although it is no longer related to length scales. What is retained, however, is the idea of introducing a gauge field in order to make a system invariant under a local symmetry.

In this article we construct a spatially scale invariant generalization of the Schroedinger field theory action following Weyl. Unlike the original approach, however, gauge variations, that is local changes of scale, that we consider are compensated for by a Ricci term rather than by a Maxwell term.

Since turbulence in superfluid liquid Helium [5] exhibits Kolmogorov scaling [6] and does not have any associated discrete spectrum it avoids Einstein's criticism of the idea and is thus could be system for testing Weyl's idea. The usual theoretical approaches to describe superfluid turbulence uses the non-linear Gross-Pitaevski (GP) equation [7, 8] where the nonlinearity present in the theory reflects the interaction between helium atoms in the field theory description of the system taking a superfluid condensate into account.

When energy is injected into the system, say by heating, excitations in the form of filaments appear. In this approach the observed filament excitations are understood with their location given by the zeros of the GP wave function. The filament excitations can also be modelled more directly with their dynamics described in analogy with interaction of wires carrying currents obeying the Biot-Savart law [9].

Distribution functions in superfluid turbulence are different from those arising for classical fluids. For instance, the velocity distribution function is not Gaussian but has a power law tail [10]. We find that the unique locally scale invariant theory constructed here contains the appropriate degrees of freedom for describing superfluid turbulence namely, a condensate and the filament excitations. But our main focus here is to show how the theory gives the Kolmogorov 5/3 law observed in a certain range of momenta of the quasi-particle excitations of the theory. While numerical studies of the GP equation yield similar results [11], the present analysis is completely analytic.

Let us point out that the present approach examines the consequences of the idea of local scale invariance but does not seek to furnish a physical picture for the emergence of Kolmogorov scaling microscopically, for instance, by vortex tangles and Kelvin-wave turbulence caused by Kelvin waves on a single vortex. It provides an effective theory for a locally scale invariant system, superfluidity being an interesting example.

5.2 Scale Invariant Action

In this section we construct an action invariant under spatial scaling starting from the action for a free Schroedinger field. The gauge group is $\mathbf{R}^{\star}$, the group of non-zero reals, which is non-compact. The free Schroedinger equation, in operator form, allows Bose-Einstein condensation and is thus an appropriate starting point for a theory of quantum turbulence. First, the free system is made invariant under global scaling by introducing a time-independent metric for the three spatial directions. It is then made invariant under local scaling by introducing a gauge field.

The action of the Schroedinger field ψ in $\mathbf{R}^1 \times \mathbf{R}^3$, with the first factor designating time, t, and the second one corresponding to the spatial coordinates $\mathbf{x} = (x^1, x^2, x^3) = (x, y, z)$ is

$$\mathscr{S}(\psi, g) = i \int \psi^{\star} \partial_t \psi \sqrt{g}\, \mathrm{d}t\, \mathrm{d}^3 x - \frac{1}{2m} \int g^{ij} \partial_i \psi^{\star} \partial_j \psi \sqrt{g}\, \mathrm{d}t\, \mathrm{d}^3 x, \tag{5.1}$$

where we have introduced a metric g on $\mathbf{R}^3$ and ∂_i denotes the derivative with respect to x^i and an asterisk designates complex conjugation. The second term of the action is invariant under the global scaling transformation of the field ψ and the metric

$$\begin{aligned} \psi &\longmapsto e^{-\Lambda/4} \psi, \\ g_{ij} &\longmapsto e^{\Lambda} g_{ij}, \end{aligned} \tag{5.2}$$

where Λ is a constant. Let us note that the scale invariance could not be effected without the metric. Moreover, as mentioned before, we do not impose scale invariance on the first term involving temporal derivative of the Schroedinger field. We now promote this global scaling symmetry to a local symmetry by allowing spatial dependence of Λ [12] and introducing a gauge field A_i and define covariant derivatives of the field ψ and the metric g as [13]

$$\begin{aligned} D_i \psi &= \partial_i \psi - \alpha A_i \psi, \\ D_i g_{km} &= \partial_i g_{km} + 4\alpha A_i g_{km}, \end{aligned} \tag{5.3}$$

where α is a real parameter. It appears from (5.3) that the parameter α may be dispensed with by a redefinition of the gauge field. However, the sign of α is of import in obtaining field configurations and will be fixed later. Then under the gauge transformation

$$\begin{aligned} \psi &\longmapsto e^{-\Lambda(\mathbf{x})/4} \psi, \\ g_{ij} &\longmapsto e^{\Lambda(\mathbf{x})} g_{ij}, \\ A_i &\longmapsto A_i - \frac{1}{4\alpha} \partial_i \Lambda(\mathbf{x}), \end{aligned} \tag{5.4}$$

with space-dependent Λ, the covariant derivatives of the scalar field ψ and the metric transform as

$$\begin{aligned} D_i \psi &\longmapsto e^{-\Lambda(\mathbf{x})/4} D_i \psi, \\ D_i g_{jk} &\longmapsto e^{\Lambda(\mathbf{x})} D_i g_{jk}. \end{aligned} \tag{5.5}$$

Hence replacing the derivatives with respect to the spatial coordinates in the second term of (5.1) by covariant derivatives we obtain the action

$$\mathscr{S}(\psi, A, g) = i \int \psi^\star \partial_t \psi \sqrt{g}\, \mathrm{d}t\, \mathrm{d}^3 x - \frac{1}{2m} \int g^{ij} D_i \psi^\star D_j \psi \sqrt{g}\, \mathrm{d}t\, \mathrm{d}^3 x, \tag{5.6}$$

which is invariant under the gauge transformations (5.4). One can add one more gauge-invariant term to the above action involving the curvature and the gauge field [14]. To this end let us define Christoffel symbols [13] as

$$\tilde{\Gamma}^i_{jk} = \frac{1}{2} g^{im} (D_j g_{mk} + D_k g_{mj} - D_m g_{jk}). \tag{5.7}$$

By (5.5), the Christoffel symbol is invariant under the local scaling transformations (5.4). Then the Ricci tensor ensuing from this Christoffel symbol defined as

$$\tilde{R}^i_{jkl} = \partial_l \tilde{\Gamma}^i_{jk} - \partial_k \tilde{\Gamma}^i_{jl} + \tilde{\Gamma}^i_{ml} \tilde{\Gamma}^m_{jk} - \tilde{\Gamma}^i_{mk} \tilde{\Gamma}^m_{jl} \tag{5.8}$$

is also invariant under the gauge transformation (5.4). The resulting scalar curvature defined as

$$\tilde{R} = g^{jl} \tilde{R}^i_{jil} \tag{5.9}$$

then transforms as $\tilde{R} \longmapsto e^{-\Lambda}\tilde{R}$ under (5.4). Hence,

$$\int |\psi|^2 \tilde{R}\sqrt{g}\, \mathrm{d}t\, \mathrm{d}^3 x \tag{5.10}$$

is invariant under the gauge transformation. It can be checked that no other term involving curvature tensors or derivatives of A or their combinations can be made gauge invariant in this fashion to yield a local polynomial action. In particular, the term F_{ij}^2 constructed from the gauge field A_i is not scale invariant in three dimensions, nor can it be made gauge invariant in a polynomial action.

The Ricci scalar $\tilde{R}$ defined above can be related to the Ricci scalar corresponding to the metric g by expanding $\tilde{\Gamma}^I_{jk}$ using (5.3) [13], resulting into

$$\begin{aligned}\tilde{R} &= R + 8\alpha\nabla_i A^i + 8\alpha^2 A^2,\\ &= R + \frac{8\alpha}{\sqrt{g}}\partial_i(\sqrt{g}A^i) + 8\alpha^2 A^2,\end{aligned} \tag{5.11}$$

where we used $A^2 = g^{ij}A_i A_j = A^i A_i$, ∇_i and R denote, respectively, the covariant derivative with respect to x^i and the scalar curvature corresponding to the metric g.

Putting (5.11) in (5.10) and adding to (5.6) along with a Chern Simons term for the gauge field we obtain the unique spatially scale-invariant action in $1+3$ dimensions given by

$$\begin{aligned}\mathscr{S}(\psi, A, g) = &\int \sqrt{g}\, \mathrm{d}t\, \mathrm{d}^3 x \left(i\psi^\star \partial_t \psi - \frac{1}{2m} g^{ij}(\partial_i \psi^\star \partial_j \psi - \alpha A_i \partial_j |\psi|^2 + \alpha^2 A_i A_j |\psi|^2)\right)\\ &+ \beta \int \mathrm{d}t\, \mathrm{d}^3 x |\psi|^2 \left(\sqrt{g}R + 8\alpha\partial_i(\sqrt{g}A^i) + 8\alpha^2\sqrt{g}A^2\right)\\ &+ \gamma \int \mathrm{d}t\, \mathrm{d}^3 x \varepsilon^{ijk} A_i \partial_j A_k,\end{aligned} \tag{5.12}$$

where ε^{ijk} denotes the rank three antisymmetric tensor, β and γ are real parameters. The Chern-Simons term being independent of the metric is locally scale invariant. From the four-dimensional perspective, this term is to be thought of as the unique potential term for the gauge field which has local three-dimensional scale invariance. Furthermore, while it does not contribute to the equations of motion, this term plays a crucial role, as we shall see below, in determining the interaction between filaments modelled by the gauge field. We now proceed to study the properties of this unique locally scale invariant three-dimensional system.

In view of this in the next section we proceed to construct an effective action for the system in terms of the wave function by integrating out the gauge field and in terms of the gauge field by integrating out the wave function.

5.3 The Condensate and Quasi-particle Spectrum

First let us integrate out the gauge field from the action (5.12) with a flat metric to obtain the effective action for the condensate ψ defined by the path integral

$$e^{iS_{\text{eff}}(\psi)} = \frac{1}{\sqrt{\pi}} \int \mathscr{D}A e^{iS(\psi,A,\eta)}. \tag{5.13}$$

Setting $g_{ij} = \eta_{ij}$ in (5.12), we obtain, up to boundary terms

$$\begin{aligned} S(\psi, A, \eta) = i \int \psi^\star \partial_t \psi - \frac{1}{2m} \int \partial_i \psi^\star \partial_i \psi - \frac{\widehat{g}}{4} \int \left(\partial_i \log |\psi|^2\right)^2 |\psi|^2 \\ + \widehat{g} \int \left(A_i - \frac{1}{2}\partial_i \log |\psi|^2\right)^2 |\psi|^2 + \gamma \int \varepsilon^{ijk} A_i \partial_j A_k. \end{aligned} \tag{5.14}$$

where we defined $\widehat{g} = 8\alpha\beta - \frac{1}{2m}$ and suppressed the measure $\mathrm{d}t\mathrm{d}^3x$ in the integrals. We now redefine the gauge field with a shift, namely,

$$\tilde{A}_i = A_i - \frac{1}{2}\partial_i \log |\psi|^2. \tag{5.15}$$

Then in the Chern-Simons term

$$\int \varepsilon^{ijk} A_i \partial_j A_k = \int \varepsilon^{ijk} \tilde{A}_i \partial_j \tilde{A}_k, \tag{5.16}$$

up to boundary terms. Integrating out with respect to the new filed $\tilde{A}$ we obtain the effective action

$$S_{\text{eff}}(\psi) = i \int \psi^\star \partial_t \psi - \frac{1}{2m} \int \partial_i \psi^\star \partial_i \psi - \frac{\widehat{g}}{4} \int \left(\partial_i \log |\psi|^2\right)^2 |\psi|^2 + \gamma \int \varepsilon^{ijk} \tilde{A}_i \partial_j \tilde{A}_k + \Gamma. \tag{5.17}$$

where the effective potential

$$\Gamma = -\frac{1}{2} \int_0^L \frac{d\xi}{\xi} \int \mathrm{d}^3x e^{-\xi\ \widehat{g}|\psi|^2}. \tag{5.18}$$

Expanding the exponential and performing the integration with respect to ξ, the effective potential becomes

$$\Gamma = -\frac{1}{2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{(\widehat{g}L)^n}{n} \int |\psi|^{2n}, \tag{5.19}$$

where we have neglected an infinite constant term ensuing from the unit term in the exponential.

Let us now consider the quasi-particle spectrum of this theory [15, 16]. Considering stationary configurations we expand ψ in Fourier modes

$$\psi(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} a_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \psi^\star(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} a^\dagger_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{5.20}$$

where $a^\dagger_{\mathbf{k}}$ and $a_{\mathbf{k}}$ are, respectively, creation and annihilation operators for the bosonic modes satisfying the commutation relation

$$[a^\dagger_{\mathbf{k}}, a_{\mathbf{k}'}] = \delta_{\mathbf{k}\mathbf{k}'}. \tag{5.21}$$

The sum is over all momentum modes. For each momentum mode we define a number operator $\widehat{n}(k) = a^\dagger_{\mathbf{k}} a_{\mathbf{k}}$, depending only on the magnitude of the momentum, thanks to the rotational symmetry. The states diagonalizing these number operators satisfy

$$\begin{aligned} \widehat{n}(k)|n(\mathbf{k})\rangle &= n(k)|n(\mathbf{k})\rangle, \\ a_{\mathbf{k}}|n(\mathbf{k})\rangle &= \sqrt{n(k)}|n(\mathbf{k}) - 1\rangle, \\ a^\dagger_{\mathbf{k}}|n(\mathbf{k})\rangle &= \sqrt{n(k)+1}|n(\mathbf{k}) + 1\rangle. \end{aligned} \tag{5.22}$$

For the zero momentum mode we also assume the existence of a state $|\psi_0\rangle = |n(0)\rangle$ with

$$a_0|\psi_0\rangle = a^\dagger_0|\psi_0\rangle = \sqrt{N}|\psi_0\rangle, \tag{5.23}$$

where we denoted $n(0) = N$ and assumed N to be sufficiently large so that $\sqrt{N} \sim \sqrt{N+1}$. This state corresponds to the condensate over which the non-zero modes are taken to be fluctuations. Substituting (5.20) in (5.19) we obtain

$$\Gamma = -\frac{1}{2} \sum_{\substack{\mathbf{k}'_1,\mathbf{k}'_2 \ldots,\mathbf{k}'_n \\ \mathbf{k}_1,\mathbf{k}_2 \ldots,\mathbf{k}_n}} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{g^n}{n} a^\dagger_{\mathbf{k}'_1} a^\dagger_{\mathbf{k}'_2} \ldots a^\dagger_{\mathbf{k}'_n} a_{\mathbf{k}_1} a_{\mathbf{k}_2} \ldots a_{\mathbf{k}_n} \\ \delta(\mathbf{k}'_1 + \mathbf{k}'_2 + \cdots + \mathbf{k}'_n - \mathbf{k}_1 - \mathbf{k}_2 - \cdots - \mathbf{k}_n), \tag{5.24}$$

where we denoted $g = \widehat{g}L/V$. So far we have not fixed the parameters. We now assume that $g = 1/N$. Then in Γ the quadratic terms $a^\dagger_{\mathbf{k}} a_{\mathbf{k}}$, $a_{-\mathbf{k}} a_{\mathbf{k}}$ and $a^\dagger_{-\mathbf{k}} a^\dagger_{\mathbf{k}}$ arise with N^{n-1} in the n-th term, while all other terms are lower order in N. The effective potential becomes

$$\Gamma = -\frac{1}{2} \sum_{\mathbf{k}} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{g}{n} \left(n^2 a^\dagger_{\mathbf{k}} a_{\mathbf{k}} + \binom{n}{2} a_{-\mathbf{k}} a_{\mathbf{k}} + \binom{n}{2} a^\dagger_{-\mathbf{k}} a^\dagger_{\mathbf{k}} \right) + \mathscr{O}(1/N) \\ \times \text{ quartic terms.} \tag{5.25}$$

The coefficients of the quadratic terms are determined by the number of ways of satisfying the momentum conservation constraint

$$\mathbf{k}'_1 + \mathbf{k}'_2 + \cdots + \mathbf{k}'_n = \mathbf{k}_1 + \mathbf{k}_2 + \cdots + \mathbf{k}_n. \tag{5.26}$$

For example, the term $a^\dagger_{\mathbf{k}} a_{\mathbf{k}}$ is obtained as one chooses any one $\mathbf{k}'$ as well as any single $\mathbf{k}$ to be non-zero, which can be chosen in $n \times n$ ways. The term $a_{-\mathbf{k}} a_{\mathbf{k}}$ is obtained by choosing all $\mathbf{k}'$ to be zero and two of the n $\mathbf{k}$'s to be non-zero. The third term is obtained similarly.

Now, along with the kinetic term, the Hamiltonian reads, upon performing the sum over n,

$$H = \sum_{\mathbf{k}\neq 0} \left(2\ell_1 a^\dagger_{\mathbf{k}} a_{\mathbf{k}} - \ell_2 (a_{-\mathbf{k}} a_{\mathbf{k}} + a^\dagger_{-\mathbf{k}} a^\dagger_{\mathbf{k}}) \right), \tag{5.27}$$

where we defined

$$\ell_1 = \frac{1}{2}\left(\frac{k^2}{2m} + \frac{g}{2e}\right), \quad \ell_2 = \frac{g}{2e}(\frac{e}{2} - 1). \tag{5.28}$$

Let us note that the third term involving the derivative of $\ln|\psi|^2$ comes with the coupling constant $\widehat{g} = gV/L$, which can be ignored compared with g. In order to obtain the quasi-particle spectrum we need to diagonalize the Hamiltonian. To this end we change basis as

$$a_{\mathbf{k}} = u\alpha_{\mathbf{k}} + v\alpha^\dagger_{-\mathbf{k}} \tag{5.29}$$

$$a^\dagger_{\mathbf{k}} = u\alpha^\dagger_{\mathbf{k}} + v\alpha_{-\mathbf{k}}, \tag{5.30}$$

where u and v are taken to be real parameters. Requiring the commutations relations

$$[\alpha^\dagger_{\mathbf{k}}, \alpha_{\mathbf{k}'}] = \delta_{\mathbf{k}\mathbf{k}'}, \tag{5.31}$$

in addition to (5.21) for any momentum $\mathbf{k}$, we obtain the constraint $u^2 - v^2 = 1$, so that the two bases are related by a Bogoliubov transformation

$$a_{\mathbf{k}} = \alpha_{\mathbf{k}} \cosh\theta + \alpha^\dagger_{-\mathbf{k}} \sinh\theta \tag{5.32}$$

$$a^\dagger_{\mathbf{k}} = \alpha^\dagger_{\mathbf{k}} \cosh\theta + \alpha_{-\mathbf{k}} \sinh\theta, \tag{5.33}$$

where θ is a real parameter. Then expressing the Hamiltonian in terms of the new oscillators α and demanding that the off-diagonal terms vanish, we obtain a relation among θ, ℓ_1 and ℓ_2, namely

$$\ell_1 \cosh 2\theta - \ell_2 \sinh 2\theta = \frac{1}{2}\varepsilon(k) \tag{5.34}$$

$$\ell_1 \sinh 2\theta - \ell_2 \cosh 2\theta = 0, \tag{5.35}$$

where $\varepsilon(k)$ is the dispersion depending on the magnitude k of the quasi-particle momentum $\mathbf{k}$ due to the rotational symmetry. Solving for the hyperbolic functions in terms of ℓ_1, ℓ_2 and $\varepsilon(\mathbf{k})$, and using the identity $\cosh^2 2\theta - \sinh^2 2\theta = 1$, yields an expression of $\varepsilon(k)$ in terms of ℓ_1 and ℓ_2, which in turn relates it to the g,

$$\begin{aligned}\varepsilon(k) &= 2(\ell_1^2 - \ell_2^2)^{1/2} \\ &= \left(\left(\frac{k^2}{2m} + \frac{g}{2e}\right)^2 - \left(\frac{g}{e}\right)^2 \left(\frac{e}{2} - 1\right)^2\right)^{1/2},\end{aligned} \tag{5.36}$$

and the Hamiltonian is

$$H = \sum_{\mathbf{k} \neq 0} \varepsilon(k)\, n(k), \tag{5.37}$$

where $n(k) = \alpha_{\mathbf{k}}^{\dagger} \alpha_{\mathbf{k}}$ is the occupation number of the quasi-particle state with energy ε and momentum $\mathbf{k}$, depending on the magnitude of $\mathbf{k}$ again thanks to the rotational symmetry.

5.3.1 Kolmogorov Scaling

From our locally scale invariant model we have seen that in order to describe the superfluid state a quasiparticle with energy that scales linearly with momentum in a certain range of momentum values emerges. Unlike the standard Bogoliubov quasiparticle result the second sound value is not fixed by the theory but has to be taken from experiment but the linear relationship between energy and momentum is present in both approaches in the small momentum region. This result was obtained, as in the Bogoliubov approach, by putting in details of the superfluid helium state in terms of a condensate. For this calculation to be valid the interaction between quasiparticles must be small. This is essential for the idea of quasiparticle to be useful. These two features, both present in our effective model, allow us to use the method of weak wave turbulence to determine the turbulent properties of superfluid helium. Essentially this means determining the way energy for momentum k scale with momentum and to check if the scaling exponent of energy calculated agrees with the observed Kolmogorov exponent.

We have already found the scaling exponent for quasiparticles. We now need to find if the occupation number for momentum k scales with k and to determine this exponent. The key calculation to do this, in weak turbulence, involves setting up a Boltzmann type of equation for the occupation number of quasiparticles with a given momentum and checking to see if it has a time independent scaling solution i.e. a solution where the occupation number scales with the momentum exists. The procedure outlined is a standard step of weak wave turbulence. The calculation for a quartic interaction term has been done and we can simply use the known results to write down scaling exponent for occupation number for our case. Once this exponent is determined the Kolmogorov exponent is fixed.

For our quasiparticle system with a weak quartic interaction term the Boltzmann time evolution equation for the quasiparticle excitation number of momentum $\mathbf{k}$ is obtained from the system Hamiltonian. Time independent solutions to this evolution equation with a scaling law behaviour have energy [16–19]

$$\begin{aligned} E(k) &= n(k)\varepsilon(k) \\ &\sim k^{-\gamma/3}, \end{aligned} \tag{5.38}$$

where the exponent γ is expressed as $\gamma = \frac{3d+2\beta}{\alpha} - 4$ in terms of the spatial dimension d and the exponents of scaling of the coefficient of the quartic term and the energy dispersion, namely

$$T(k) \sim k^{\beta} \tag{5.39}$$

$$\varepsilon(k) \sim k^{\alpha}. \tag{5.40}$$

We have so far discussed the terms quadratic in the raising and lowering operators in the Hamiltonian. The coefficient of the quartic term, which goes as $1/N$ in the large N limit that we are considering, is independent of $\mathbf{k}$, leading to $\beta = 0$. As can be seen from (5.36) if the momenta are in the range

$$\begin{aligned} \frac{k^2}{2m} &< \frac{g}{e} \\ \frac{k^2}{2m}\left(\frac{k^2}{2m} + \frac{g}{e}\right) &> \frac{g}{e}\left(e - \frac{e^2}{4} - \frac{3}{4}\right) \\ &= 0.12\left(\frac{g}{e}\right)^2, \end{aligned} \tag{5.41}$$

then the dispersion is linear in momentum and thus gives $\alpha = 1$. Hence, $\gamma = 5$, leading, according to (5.38), to the Kolmogorov scaling law, $E(k) \sim k^{-5/3}$, within this range of momentum. Thus, in an appropriate range of momentum we obtain linear dispersion relation and thus weak turbulence and Kolmogorov scaling law from the four wave resonance.

We thus conclude that an effective theory based on Weyl's original idea of gauge invariance as local scale invariance is compatible with the existing descriptions used to understand superfluid turbulence. Local scale invariance leads to correctly identifying the degrees of freedom and leads to the dynamics of the excitations in the superfluid turbulent phase. It is satisfying that the effective theory links filament locations, postulated to be filament currents which couple to scale gauge fields, with the zeros of the GP-like equation. The approach described to construct locally scale invariant systems is also of theoretical interest as it is a very general method for constructing locally scale invariant effective theories.

Acknowledgments SS acknowledges the hospitality of the Department of Theoretical Physics, IACS, during the period this work and Koushik Ray for an enjoyable collaboration.

References

1. K. Ray, S. Sen, Scale invariance and superfluid turbulence. Nucl. Phys. B [FS], 637 (2013)
2. H. Weyl, Gravitation and electricity. Sitzungsber. Preuss. Akad. Berlin 465 (1918) (Collected in [3])
3. L. O'Raifeartaigh, *The Dawning of Gauge Theory*, Princeton Series in Physics (1997)
4. F. London, Quantum-mechanical interpretation of Weyl's theory. Zeit. F. Phys. **42**, 375 (1927)
5. M. Tsubota, Quantum turbulence. J. Phys. Soc. Jpn. **77**, 111006 (2008) arXiv:0806.2737(cond-mat)
6. M. Paoletti et al., Velocity statistics distinguish quantum turbulence from classical turbulence. Phys. Rev. Lett. **101**, 154501 (2008)
7. E. Gross, Structure of a quantized vortex in boson systems, Il. Nuovo Cimento **20**, 454 (1961)
8. L. Pitaevskii, *Vortex Lines in an Imperfect Bose Gas, Soviet Physics JETP-USSR*, vol. 13, pp. 451 (American Institute of Physics, New York, 1961)
9. K. Schwarz, Three-dimensional vortex dynamics in superfluid ^{4}He: line-line and line-boundary interactions. Phys. Rev. B **31**, 5782 (1985)
10. D. Kivotides et al., Velocity spectra of superfluid turbulence. Europhys. Lett. **57**, 845 (2002)
11. M. Kobayashi, M. Tsubota, Kolmogorov spectrum of superfluid turbulence: numerical analysis of the Gross-Pitaevskii equation with a small-scale dissipation. Phys. Rev. Lett. **94**, 065302 (2005)
12. C. Connaughton, S. Sen, Local scale invariance and weak wave Turbulenc (Unpublished)
13. T. Padmanabhan, Conformal invariance, gravity and massive gauge theories. Class. Quantum Grav. **2**, L105 (1985)
14. A. Iorio, L. O'Raifeartaigh, I. Sachs et al., Weyl gauging and conformal invariance. Nucl. Phys. B **495**, 433 (1997) [hep-th/9607110]
15. S. Musher, A. Rubenchik, V. Zakharov, Weak Langmuir turbulence. Phys. Rep. **252**, 177 (1995)
16. M. Rakowski, S. Sen, Quantum kinetic equation in weak turbulence. Phys. Rev. E **53**, 586590 (1996). arXiv:cond-mat/9510107
17. V. Zakharov, V. L'vov, G. Falkovich, *Kolmogorov Spectra of Turbulence* (Springer, New York, 1992)
18. A.M. Balk, On the Kolmogorov Zakharov spectra of weak turbulence. Phys. D **139**, 137157 (2000)
19. D. Sanyal, S. Sen, Quantum weak turbulence. Ann. Phys. **321** 1327 (2006). arXiv:cond-mat/0402395

Chapter 6
A Brief Introduction to Quantum Phase Transitions

K. Sengupta

Abstract In this article, we are going to present a pedagogical review of basic properties of Ising and Heisenberg models. Using these properties, we shall study basic properties of the quantum phase transition in 1D Ising model and follow it with an analogous study of the Bose-Hubbard model which is relevant to the current experimental systems involving bosons in optical lattices.

Keywords Quantum phase transition · Ultracold atoms · Bose–Hubbard Model

6.1 Introduction

The study of quantum phase transition has gained tremendous impetus in recent advancement in the field of ultracold atoms. The theoretical development of this subject started a long time before these experiments. In the early days, specific spin models such as the Ising and the Heisenberg models served as theoretical test beds for studying properties of these transition. In this article, we shall therefore first review, in this section, the basic properties of several spin models. This will be followed by discussions on quantum phase transition and ultracold atoms in subsequent sections.

A study of spin models is probably one of the oldest topics in condensed matter physics since they turn out to be low energy-effective models describing many strongly correlated condensed matter systems. Typically, these models are aimed at describing a set of localized spins with a given symmetry in d dimensions interacting with themselves, possibly in the presence of magnetic field. Here we shall revisit the physics of the simplest of these models, namely, models which has nearest neighbor interactions and is subjected to a magnetic field. The simplest and probably the oldest of these models is the Ising model in a transverse field whose Hamiltonian is given by

K. Sengupta (✉)
Theoretical Physics Department, Indian Association for the Cultivation of Science,
Kolkata 700032, India
e-mail: ksengupta1@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_6

$$H_{\text{Ising}} = J \sum_{\langle ij \rangle} S_i^z S_j^z - h \sum_i S_i^x. \tag{6.1}$$

Here h denotes the transverse magnetic field (or μ_B times the magnetic field to be more accurate) and $\langle ij \rangle$ denote sum over nearest neighbors. Note that in the absence of the magnetic field, the model has a Z_2 symmetry, i.e., the Hamiltonian remains invariant under a global spin flip $S_i^z \to -S_i^z$. This is the simplest example of the class of spin models with discrete symmetries.

The phases of the Ising model in a hypercubic lattice in d dimension is quite straightforward to obtain. In the limit of infinite transverse field, the ground state involves all the spins pointing along x. Following standard notation in the literature, we shall call this phase "paramagnet." In the other limit, when $J \gg h$, the nature of the phase depends on the sign of J. With our sign convention in Eq. 6.1, for $J < 0$, the system gets into a ferromagnetic phase while for $J > 0$, the ground state is antiferromagnetic. Note that each of these ground states breaks Z_2 symmetry. This point is illustrated in Fig. 6.1.

Starting from a large $x = J/h$, if we adiabatically decrease this ratio by increasing the transverse field, the system undergoes a phase transition at some critical value of $x = x_c$. For the antiferromagnetic case ($J > 0$), the transition is first order. The simplest way to see this is to note that the net magnetization $m = \sum_{i,a=x,z} S_i^a$ undergoes a discontinuous change at the transition. On the other hand, for $J < 0$, the transition is continuous. We shall discuss this case in details during our study of quantum phase transitions.

Before going to discussion of other spin models, I would like to mention that the simplicity of the phases obtained for hypercubic lattices is more a property of the lattices than the model. Ising model in triangular or other non-bipartite lattices can have quite complicated phases due to a phenomenon called frustration. To see this, consider the Ising Hamiltonian (Eq. 6.1) on a 2D triangular lattice with $h = 0$ and $J > 0$ (antiferromagnetic interaction). Now consider a triangle in the lattice. The two vertices of the triangle can be occupied by spins pointing in opposite directions so as to minimize interaction between them as shown in Fig. 6.2. But there is no way to fill the third vertex which minimizes interaction with both the spins; the spin is

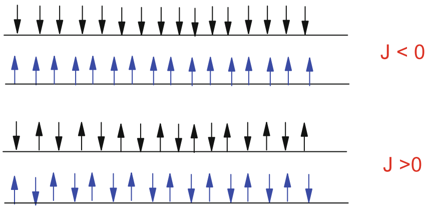


Fig. 6.1 Ground states of the Ising model for $J < 0$ and $J > 0$ for $|J| \gg h$. Note that the ground state spontaneously breaks Z_2 symmetry. The two states shown in each case are degenerate and can be reached from another by a global Z_2 transformation, i.e., a simultaneous flip of all spins. Since all spins need to be flipped simultaneously, one cannot connect these states via local perturbation

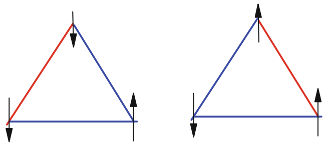


Fig. 6.2 Frustration for antiferromagnetic Ising model in a 2D triangular lattice. It is possible to satisfy only two of the three bonds in a triangle. The satisfied bonds are shown in *blue* while the unsatisfied ones are shown in *red*. The two states shown are degenerate and such a degeneracy grows exponentially with system size

therefore "frustrated." This leads to two possible ways to fill up the third site, and consequently to two degenerate ground states. It is easy to see that this degeneracy grows exponentially with the system size N and it turns out that for the Ising model in 2D triangular lattice, the number of degenerate ground state is $\exp(0.323N^{2=3})$. Such a huge degeneracy clearly complicates the problem of finding the true zero temperature ground state of the system. In fact, we shall see an example where this degeneracy has profound influence on phase transitions in the model.

The next class of models which will be of interest to us are the XXZ models which has the Hamiltonian

$$H_{\mathrm{XXZ}} = J_{\perp} \sum_{\langle ij \rangle} \left(S_i^x S_j^x + S_i^y S_j^y \right) + J_z \sum_{\langle ij \rangle} S_i^z S_j^z. \tag{6.2}$$

Note that here, a global rotation of the spins about the z axis leaves the system invariant, but a general rotation in 3D spin-space does not, since $J_{\perp} \neq J_z$. Consequently, the model has $U(1)$ symmetry. In contrast to Ising model discussed previously, this is an example of spin Hamiltonian with continuous symmetry. The special point $J_{\mathrm{perp}} = Jz = J$ is called the Heisenberg point of the model for which the H_{XXZ} reduces to the well-known Heisenberg model

$$H_{\mathrm{Heisenberg}} = J \sum_{\langle ij \rangle} \mathbf{S}_i \mathbf{S}_j \tag{6.3}$$

which has SU(2) symmetry since the hamiltonian $H_{\mathrm{Heisenberg}}$ remains invariant under global rotation in spin-space. The reader is urged to verify this point.

For the rest of this section, we shall discuss the methods of obtaining the ground state and the excitation spectrum of the XXZ and the Heisenberg model. First we shall take the ferromagnetic case, for which the ground state do not break translational symmetry. Our main tool for doing this will be a mapping of these spin models to a boson model using Holstein–Primakoff (HP) transformation. The HP transformation, which is a mapping between spins and bosons can be understood as follows. We know that quantum spins must satisfy the commutation relations $\left[S_p^i, S_q^j \right] = i\hbar\varepsilon_{ijk} S_p^k \delta_{pq}$, where ε_{ijk} is the antisymmetric tensor and δ denote Kronecker delta function. Now if we look at boson operators, they also satisfy commutation relations $\left[b_i^\dagger, b_j\right] = -i\delta_{ij}$. From this observation, one is led to the question as to whether it is possible to express the spins in terms of bosons and vice versa. The answer is of course yes, as figured out by Holstein and Primakoff in 1940. The transformation, for spin S, is given by

$$S_i^+ = \sqrt{2S}\left(1 - \frac{b_i^\dagger b_i}{2S}\right)^{1/2} b_i, \quad S_i^- = \left(S_i^+\right)^\dagger, \quad S_i^z = S - b_i^\dagger b_i, \tag{6.4}$$

where $S_i^+ = S_i^x \pm i S_i^y$ are the spin raising and lowering operators. Note that the factors correctly $\left(1 - b_i^\dagger b_i/2S\right)^{1/2}$ implements a finite Hilbert space for the spins, although that for the bosons is invite. The reader is urged to check the communication relation for the spins from Eq. 6.4.

Now let us consider the Heisenberg model on a hypercubic lattice in d dimensions with ferromagnetic interaction ($J < 0$). The ground state corresponds to all spins pointing along the z axis and hence corresponds to $S_z = S$ at every site of the lattice. In the boson language, this means that the ground state is a vacuum for bosons since $\langle b_i^\dagger b_i\rangle_{\text{gnd}} = 0$. We now wish to study low-lying excitations over the FM ground state. To do this, we reexpress the Heisenberg Hamiltonian (Eq. 6.3) in terms of the bosons using Eq. 6.4. This yields

$$H_{\text{excitation}} = JS\left[\sum_{\langle ij\rangle}\left(b_i^\dagger b_j + \text{h.c}\right) - 2\sum_i b_i^\dagger b_i\right] + \text{O}\left(b^4\right). \tag{6.5}$$

where we have neglected all quartic terms for the bosons. The last approximation amount to neglecting scattering among bosons. Since the ground state here corresponds to boson vacuum, for low-lying excitations, such scattering events are rare and can be neglected. The next task is to diagonalize $H_{\text{excitation}}$ by going to the Fourier space which yields, in d dimensions,

$$H_{\text{excitation}} = JS\sum_k E_k b_k^\dagger b_k, \quad E_k = 2|J|S\left(z - \gamma_k\right), \quad \gamma_k = \sum_{\mathbf{n}} e^{i\mathbf{k}\cdot\mathbf{n}}. \tag{6.6}$$

where $z = 2d$ is the coordination number for the d-dimensional hypercubic lattice and the sum over $\mathbf{n}$ denotes sum over nearest neighbors. $_{E}k$ here denotes energy corresponding to the low-lying excitations of the spin systems which are called spin-waves. The name originates from the fact that a finite density of the bosons at a small wave-vector k physically represents canting of spins by π from their ground state orientation over a length scale $2\pi/k$. At low k, one gets $E_k = 2JSa^2k^2$, where a is the lattice spacing. Thus the spin-waves here have quadratic dispersion which indicates vanishing group velocity at low momentum. For low $k \sim 1/L$, the spin-wave energy vanishes and hence a very gradual canting of spins become the lowest lying excitations over the FM ground state.

Finally, we come to the case of antiferromagnets for which $J > 0$. Here the additional complication that arises is that the expected ground state corresponds to spins pointing in opposite direction at neighboring lattice sites so that $\langle S_z \rangle_{\text{ground}} = \pm S$ on two neighboring site. This observation leads us to the fact that if we want to describe the low energy excitation over this ground state, one set of boson operators is not enough. To get around this obstacle, we divide the hypercubic lattice into two sublattices A and B such that the ground state corresponds to spins on $B(A)$ pointing up (down). Then the HP transformation for all spins on the B sublattice is given by Eq. 6.4 whereas those for spins on A sublattice is given by

$$S_i^+ = \sqrt{2S}\, a_i^\dagger \left(1 - \frac{a_i^\dagger a_i}{2S}\right) \quad S_i^- = \left(S_i^+\right), \quad S_i^z = -S + a_i^\dagger a_i. \tag{6.7}$$

Comparing Eqs. 6.4 and 6.7, we find that S_i^+ and S_i^- must switch roles to ensure a negative sign of S_z. Next, we express the Heisenberg Hamiltonian in terms of the bosonic operators a and b. Neglecting interaction between bosons, we find

$$H^{AF}_{\text{excitations}} = 2JS \sum_k \left[\gamma_k \left(a_k^\dagger b_k^\dagger + \text{h.c}\right) + z\left(a_k^\dagger a_k + b_k^\dagger b_k\right)\right]. \tag{6.8}$$

Note that $H^{AF}_{\text{excitations}}$, though quadratic, is not quite diagonal in Fourier space due to the presence of the off-diagonal terms. To diagonalize this, we use a Bogoliubov transformation which amounts to first writing

$$\alpha_k = u_k b_k - v_k a_k^\dagger, \quad \beta_k = u_k a_k - v_k b_k^\dagger, \quad u_k^2 - v_k^2 = 1, \tag{6.9}$$

and then finding u_k and v_k for which $H^{AF}_{\text{excitations}}$ becomes diagonal in terms of α_k and β_k. It turns out that one can write

$$H^{AF}_{\text{excitations}} = \sum_k E_k \left(a_k^\dagger \alpha_k + \beta_k^\dagger \beta_k + 1\right), \quad E_k = 2JzS\sqrt{\left(1 - \gamma_k^2/z^2\right)}. \tag{6.10}$$

The reader is urged to find the values of u_k and v_k which does the trick.

From Eq. 6.10, we note that the for low momenta the spin-waves (for hypercubic lattice) have linear dispersion: $E_k = 2Jzsak$ which implies a finite velocity of the spin-waves at low momentum. More interestingly, we find that our starting ground state ansatz is not the correct one. To see this, let us compute $\langle S_z^B \rangle_{\text{gnd}} = NS - \langle b_k^\dagger b_k \rangle$. In terms of the α_k and β_k operators, this is given by (Eq. 6.9)

$$\langle S_z^B \rangle_{\text{gnd}} = NS - \left\langle \sum_k u_k^2 \alpha_k^\dagger \alpha_k + v_k^2 \beta_k \beta_k^\dagger + \text{off} - \text{diagonal terms} \right\rangle = NS - \sum_k v_k^2. \tag{6.11}$$

Thus the sublattice magnetization deviates from its classical value due to quantum fluctuations, a feature that is hallmark of quantum antiferromagnets, but not of ferromagnets.

6.2 Quantum Phase Transitions

The subject of phase transition, i.e., transition between two states or phases of matter due to change of external parameters such as pressure, temperature, etc., is again an important topic in condensed matter physics. Here we shall explore only a few aspect of this important topic; for detailed study, one can refer to standard literature such as Refs. [2–4].

Standard finite temperature phase transitions occur as a result of competition of internal energy (U) and entropy (S) in the free energy of a system. Such a competition can arise in various contexts. A simple example to understand this is to consider the Ising model with $J < 0$ and $h = 0$, so that at low temperature the ground state can be assumed to be ferromagnetic. Note that this is an assumption and need not be correct in all dimensions. Now let us increase the temperature so that the spins can flip. Typically, this will lead to the formation of domain walls. In $d = 1$, the domain wall corresponds to a series of flipped spins along the chain, as demonstrated in the upper panel of Fig. 6.3. It is easy to see that such a domain wall has an energy cost of $2J$, whereas the entropy corresponding to such a configuration is $\simeq \ln N$ for large N. Thus the free energy of the system is $F \simeq 2J - k_B T \ln N < 0$ for all T in the thermodynamic limit. This situation changes in higher dimensions as $d = 2$ as shown in the lower panel of Fig. 6.3. Here the energy cost of forming the domain wall is $4NJ$, while the entropy gain is $\ln\left(N2^{2N}\right)$ so that the free energy becomes negative above at a critical temperature $T_c \simeq 2J/k_B \ln 2$. Thus dimensionality plays a crucial role in the phase transition.

Typically, finite temperature transitions are driven by thermal fluctuation and thus occur at a critical temperature. However, phase transitions can also occur at $T = 0$ (by which we mean situations where temperature is lower than all other energy

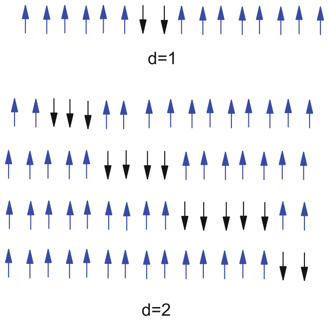


Fig. 6.3 Domain walls for Ising model in $d = 1$ and $d = 2$. Note that the energy of domain wall formation depends on system size in $d \geq 2$, but not at $d = 1$

scales in the problem, and not necessarily absolute zero) where quantum fluctuations leads to a change of phase. In this case, the phase transition occurs due to competition between different terms in its Hamiltonian (such as $-J\sum_{ij} S_i^z S_j^z$ and $h\sum_i S_i^x$ terms in the Ising Hamiltonian defined earlier) and entropy do not play any role. One of key ingredients in understanding the behavior of a system near such a transition is the Landau–Ginzburg–Wilson paradigm. According to this paradigm, the Lagrangian density of the system in the ordered phase near the critical point for a order–disorder transition can be written in terms of the order parameter Δ of the ordered phase as

$$f_{\mathrm{LGW}} = \Delta^* \left(\omega^2 - k^{2z} + r\right) \Delta + u|\Delta|^4 + \cdots \qquad (6.12)$$

where ellipsis denotes all higher order terms, $r = 0$ at the second-order transition point and z is called the dynamical critical exponent which determines relative scaling between space and time ($z = 1$ implies relativistic invariance). Here we have assumed that the terms which are odd in Δ are zero. This assumption, of course, need not be true in the general case. The basic point that one needs to take care of in constructing such a free energy is that it is consistent with all the basic symmetries of the microscopic Hamiltonian. Actually, in principle, such a free energy can be systematically derived from the microscopic Hamiltonian describing the system. However, except for very simple cases, this is in general impossible in practise. The key point to be emphasized here is that our failure to derive such a free energy from the microscopic Hamiltonian does not mean that we will not be able to guess its form. It only means that we will not be able to determine the precise values coefficients r, u, etc. However, the different possibilities of the physics near the phase transition can be captured without them which makes this method very powerful. For a detailed account, see Ref. [5]. For the rest of this lecture, we are going to consider a subclass of such phase transition, namely, second-order transitions.

Next, we shall introduce the concept of universality class. As is well-known [2, 4, 5], all physically important quantities (such as equal time correlation functions of the order parameter, or energy gap) of a system exhibits power law behavior close to a second-order phase transition. This is a manifestation of the fact that phase transitions are usually accompanied by divergent length and time scales. The divergence of the time scale comes from the vanishing of the energy gap of the system $\delta E \sim |J - J_c|^{z\nu}$. The divergent length ζ comes, for example for Ising model, since the characteristics decay length of the spin–spin correlator diverges as $\zeta^{-1} \sim \Lambda\, |J - J_c|^{\nu}$, where Λ is some unimportant cutoff scale. Such a power law behavior means that we can specify the behavior of the system near a phase transition by a set of exponents $\nu, z \ldots$ The physics near the transition is completely determined by these exponents; two transitions with same set of exponents will therefore have exactly same physics near the transition. The set of these exponents therefore determines the universality class of a transition. Thus the chief assertion of the universality is that the physics is independent of microscopic parameter values of the Hamiltonian which can be seen as a consequence of the presence of diverging length and time scales. In most phase transitions, the universality class of the transition can be guessed from the symmetries of the underlying Hamiltonian. However, more recent theories, have found exceptions to this rule.

We are going to see an example of such an exception in this lecture. We are now going to consider two specific examples of phase transitions. The main intention would be bringing out few key general points. First, let us consider the Ising model in $d = 1$ at $T = 0$ and for $J < 0$. As discussed before, there are two distinct phases of this system. The first corresponds to $h \gg |J|$ which is a quantum paramagnet with all spins pointing along x. The second is $|J| \gg h$ for which the ground state is a ferromagnet with all spins pointing wither along z or $-z$. Now as we change h/J, the system must go from one phase to another. The first question to ask is whether the change will occur as a transition or a smooth crossover. The answer to this question in the present model is easy to see from symmetry. We know that the system breaks Z_2 (discrete) symmetry in the ferromagnetic phase while there is no such broken symmetry in the paramagnet. This allows us to conclude that these two ground states cannot be smoothly connected—to go from one to the another one needs to have a transition.

To find out at what value of h/J this transition occurs, we consider the following. Imagine that the system is in the paramagnetic phase with the ground state corresponding to all the spins pointing along x. The basic excitation above this ground state corresponds to flipping a spin on-site i leading to an excited state $|i\rangle$. Such a process costs an on-site energy of $2h$. But now the flipped spin can move around between different sites. It is easy to see that all such states $|j\rangle$ have same energy. To compute the energy gain from such a move, consider the matrix element between states $|i\rangle$ and $|j\rangle$ are given $\langle j|H_{\text{Ising}}|i\rangle = J\delta_{i,j\pm 1}$. Thus, one is faced with a degenerate perturbation theory problem which is trivial to solve in momentum space, leading to the excitation energy $E(k)$

$$E(k) = 2h - 2J\cos(k), \quad E_{\min} = E(k=0) = 2(h - J), \tag{6.13}$$

where $E_{\min}$ is the minimum energy of the excited state. Note that this energy touches 0 (ground state energy) for $h/J = 1$. At this point, it becomes energetically favorable to flip spins spontaneously and the ground state is destabilized. One can carry out a similar exercise staring from the ferromagnetic side and arrive at an identical answer for the critical h and $E(k)$. It is also possible to obtain an exact result for the single particle excitations of this model for all J and h as shown in Ref. [4]:

$$E_{\text{exact}}(k) = 2\sqrt{J^2 + h^2 - 2hJ\cos(ka)} \tag{6.14}$$

which conforms to the perturbative results. All exponents of this transition can be found; the transition belongs to Ising universality class.

Finally, we shall consider the anisotropic Ising model with $J_{i,i+x}$, $J_{i,i+y} = J > 0$ and $J_{i,i+z} = J' < 0$ in 3D and finite temperature but in the absence of a transverse field and in a slightly different geometry. Our aim is to show that the effect of the frustration can change the universality class of a transition. The geometry we want to consider is that of a stacked 2D triangular lattices. At high temperature, such a model must show a paramagnetic or disordered phase. At $T = 0$, the state should clearly order. This can be seen by trying to create domain wall over an ordered state. It can be easily seen that such a domain wall creation is energetically costly. Hence we expect the ordered state to hold till some finite temperature T_c. The question is what is the nature of this ordered phase.

The answer to this question, for the considered geometry, is quite subtle. Since we are dealing with antiferromagnetic Ising model in a triangular lattice, the system is frustrated. It can be shown that the degeneracy corresponding to possible ground states grows exponentially with system size. Thus, although we are sure that there will be some ordered phase, it is not easy to guess this ordering. The aim of the rest of the section is to show that in the present case, the possible ordering comes out naturally from a proper theory of the phase transition.

To see how phase transition takes place in this model, let us rewrite the Ising model in a stacked triangular lattice in momentum space

$$\begin{aligned} H_{\text{Ising}}^{\text{3D}} &= \sum_k J(\mathbf{k}) S_k^z S_{-k}^z, \\ J(\mathbf{k}) &= K\left(\cos(k_x a) + 2\cos(k_x a/2)\cos\left(\sqrt{3}k_y a/2\right)\right) - J'\cos(k_z a). \end{aligned} \tag{6.15}$$

When $J > 0$, the minima of the dispersion occurs at $Q_\pm = (\pm 4\pi/3, 0, 0)$. Since the fluctuations about these minima are most important for destabilizing the ordered phase, we find that we need two fluctuating fields for describing this phase transition: $\psi_\pm = S(\mathbf{Q}_\pm + \mathbf{q}) = m\exp(\pm i\phi)$, where $|\mathbf{q}|/|\mathbf{Q}_\pm| \ll 1$. It can be shown that the presence of two low-energy fluctuating fields (instead of one as will be the case in absence of frustration when $J < 0$) changes the universality class of the transition from Ising to XY type. The next task is to write down the effective Landau–Ginzburg free energy functional in terms of the low-energy fluctuating fields that we have identified. The thing to keep in mind while doing this is that such a functional has to be invariant under all symmetries operations of the underlying stacked triangular lattice. Following Ref. [6], one can find that such a functional can be written as

$$\mathscr{F} = \sum_q \left(r + \mathbf{q}^2\right) m(q)m(-q) + u_4 \sum_4 m^4 + u_6 m^6 + v \sum_6 m^6 \cos(6\phi) \tag{6.16}$$

where $\sum_n \equiv \sum_{q_1,q_2,\ldots,q_n} \delta(q + 1 + q_2 + \cdots + q_n)$. The parameters r, u_4, u_6, and v can be computed from microscopic theory, but their precise form will not interest us for the moment. The transition to the ordered phase takes place when $r = 0$. We note that at the transition point, v, is zero since it is irrelevant in the RG sense. Therefore, the relative phase ϕ between the fields $\psi_\pm$ is not fixed at the transition. It turns out that as we go inside the ordered phase, the magnitude of v grows and pins the relative phase to 0 (if $v < 0$) or $\pi/6$ (if $v > 0$). These lead to two possible ordered phases as shown in Fig. 6.4 [6]. Thus the fate of the ordered phase is determined by a variable which is (dangerously) irrelevant. The quantum counterpart of such phase

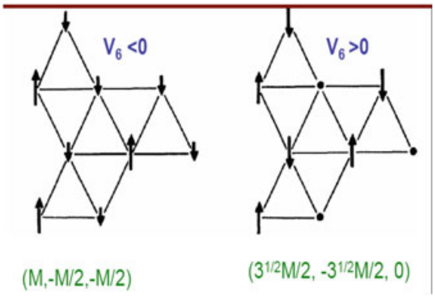


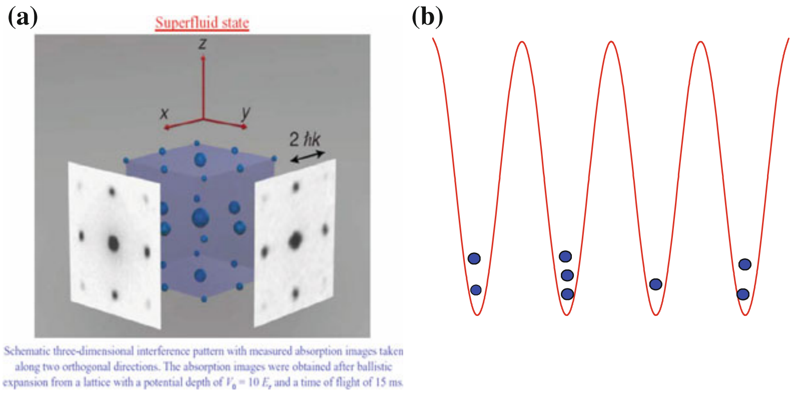
Fig. 6.4 Two possible three by three ordering for the stacked triangular lattice for $v > 0$ and $v < 0$. M_1; M_2; M_3 represents the value of the magnetization on the three sublattices

transition has recently been identified in extended Bose–Hubbard models and spin systems [7, 8].

6.3 Bose–Hubbard Model

The physics of bosons has the fascinating theoretical aspect called Bose–Einstein condensation (BEC), i.e., occupation of a single quantum state by macroscopic number of bosons at low enough temperature leading to fascinating phenomenon such as superfluidity. Moreover, there has been renewed interest in physics of these systems due to their recent experimental realizations in trapped atoms [9]. Such experiments can manipulate BECs with incredible precision. In particular, it has been possible to form an optical lattice in a system of these trapped bosons, which, when deep enough, may result in Mott localization of Bosons leading to destruction of the BEC state. Such a destruction is a result of a phase transition in the bosonic system. The physical temperatures relevant in these experiments are of the order of tens of nanokelvins (which makes these systems the coldest known place in the universe) and is at least 2–3 order of magnitudes lower than all other energy scales. Thus such a transition is an example of a quantum phase transition. In this section, we shall give a brief account of the physics associated with such a transition, by considering the simplest possible BECs, i.e., BECs formed from spin-less bosonic atoms such as ^{87}Rb.

The optical lattice is formed by applying six counter-propagating laser beams of fixed wavelengths to the condensed Bose atoms in a trap (which can be magnetic or optical). These lasers have a electric field **E** and form standing waves of light in all three directions. The atoms have a polarizibility $\alpha(\omega;\omega_0)$, where ω is the applied lasers frequency and ω_0 is some characteristic frequency of the atoms. As a result, the atoms feel a potential $V = -\alpha(\omega;\omega_0)|E|^2$. By tuning the frequency of the applied laser, one can now make α positive, so that the atoms have a tendency to sit at the bottom of the potential which acts as lattice sites as shown in right panel of Fig. 6.5. Once they do that, the kinetic energy of the atoms makes them hop from one site to the next. As the lattice becomes deeper, this process is exponentially suppressed since it can be shown that the hopping amplitude $t \sim \exp\left(-\sqrt{V/E_R}\right)$,



Schematic three-dimensional interference pattern with measured absorption images taken along two orthogonal directions. The absorption images were obtained after ballistic expansion from a lattice with a potential depth of $V_0 = 10\,E_r$ and a time of flight of 15 ms.

Fig. 6.5 **a** Sketch of absorption imaging of bosons from a free flight. **b** Schematic representation of bosons in a one dimensional optical lattice. The present figure is obtained from Ref. [9]

where $E_R = \hbar^2/2m\lambda^2$, called the recoil energy, is the basic energy scale created out of the mass (m) of the atoms and the wavelength (λ) of the laser. The bosons which form the condensate is neutral and so the interaction between them is shortrange Van der Walls type. In the presence of a lattice, the interaction between the boson is most significant when they are on the same lattice site which we shall call U. Interaction between the atoms in the neighboring site can be neglected as a first approximation. The key point to recognize is that this interactions, unlike the hopping strength t, do not depend exponentially on the strength of the lattice potential V.

Now consider an optical lattice with one boson per site. If the kinetic energy is large compared to the on-site interaction $(t \gg U)$, the bosons are free to hop around and therefore the ground state of the system is clearly the one in which a major number of bosons sit in the $k = 0$ state. Thus the bosons form a BEC. However, if we now increase the depth of the lattice t/U becomes small, and hence a stage comes when the bosons do not find it convenient to hop around since they have to pay too much interaction energy cost to do so. In this limit all the bosons become localized. Since this localization is induced by interaction, its called a Mott insulating state.

How do we see this transition experimentally? It turns out the easiest way to look at this bosons is to switch at the lattice and the trap at the same time and let the bosons fly out. After some time of such free flight, the position distribution of these bosons can be measured by absorption imaging of the bosons in a free flight as shown in left panel of Fig. 6.5. Since the position of the bosons after a time t of such a flight depends on their starting velocity or equivalently momentum, *the position distribution of these bosons actually reflect their momentum distribution inside the trap*. Now if there were no lattice, all the bosons would be in the $k = 0$ state (the condensate) and hence their momentum distribution will be localized around $k = 0$. On the other hand, if the bosons were in the Mott state, they are localized in real space which means their momentum can take all possible values. Thus the Mott state momentum distribution should reflect a featureless blur. As the strength of the optical lattice is increased, it is therefore expected that the momentum distribution of the bosons will crossover from a central peaked to a featureless blurred one. This is precisely what is seen in experiments as shown in Fig. 6.6. The phase transition occurs somewhere around $V_0 \simeq 14E_R$ where the central bright spot disappears.

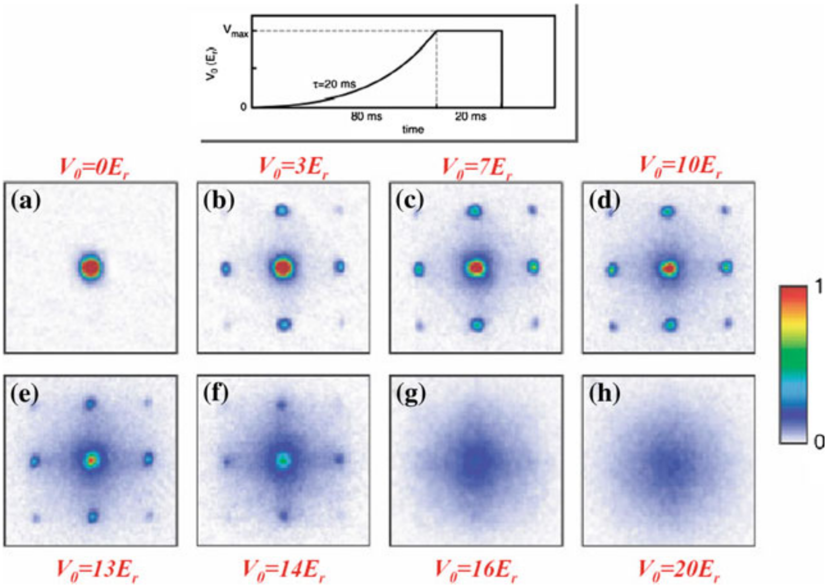


Fig. 6.6 Measurement of momentum distribution of the bosons. The lattice potential is ramped over a time period of 80 ms to its maximum value V_0 as shown in the *top* panel. The system is allowed to equilibrate for 20 ms and after that both the lattice and the trap potential is switched off. The position distribution of the bosons is measured after 10 ms of free flight. Note that the central peak which is the signature of superfluidity disappears at $V_0 \simeq 14E_R$ signifying the *onset* of the Mott state. The present figure is obtained from Ref. [9]

How do we develop a theory for this transition? Well, we could try doing what we did for the Ising model. Let us first look at the Mott state when $t \ll U$ and there is an integer number of bosons per site. Neglecting the effect of hopping of bosons here, we can see that the Hamiltonian is

$$H_{\text{Mott}} = U \sum_i \frac{1}{2} n_i \left(n_i - 1\right) - \frac{\mu}{U} n_i. \tag{6.17}$$

Since the Hamiltonian is on-site, one could easily find out the ground state wave function and energy. This state is given by

$$\Psi_{\text{ground}} = \prod_i |n_i = n_0\rangle \frac{E\left[n_0\right]}{U} = \frac{1}{2} n_0 \left(n_0 - 1\right) - \frac{\mu}{U} n_0, \tag{6.18}$$

where $n_0 \equiv (\mu/U)$ is the integer which minimizes $E\left[n_0\right]$. One can easily check that

$$\begin{aligned} n_0 &= 0 \text{ for } \mu \leq 0 \\ &= 1 \text{ for } 0 \leq \mu \leq 0 \\ &= 2 \text{ for } U \leq \mu \leq 2U \ldots . \end{aligned} \tag{6.19}$$

The Mott state is the stable ground state when $t/U \ll 1$. The next question to ask is what happens when we increase t. From the experiments, we already know the answer; the ground state becomes unstable when a critical t is reached. Now to find when the ground state is destabilized, we need to find out what are the possible excited states of the system over the ground state and when can they destabilize the ground state. Note that this line of thinking operates on the same basic principle as in study of quantum phase transition in Ising model, i.e., to find out the lowest lying excitations over the ground state of the ordered phase and check when their energy touches the ground state energy.

At finite t, let us now consider the excited state which corresponds to addition of an extra particle/hole over the Mott ground state with n_0 particles per site. The minimum excitation energy of such states are

$$\delta E_p = -\mu + U n_0 - zt\,(n_0 + 1) \quad \delta E_n = \mu - U\,(n_0 - 1) - ztn_0 \tag{6.20}$$

which destabilizes the Mott ground state at

$$t_c^p = \frac{-\mu + U n_0}{z\,(n_0 + 1)}\; t_c^h = \frac{\mu - U\,(n_0 - 1)}{zn_0} \tag{6.21}$$

leading to a critical hopping of $t_c = \text{Min}\left[t_c^p, t_c^h\right]$. The plot of the superfluid insulator boundary using this simple theory captures some essential features of the transition. First, we note that at the boundary between the Mott phases with n_0 and $n_0 + (-)1$ particles, $\mu = U n_0(n_0 - 1)$ so that t_c vanishes. At these points, the excited state energies $\delta E_p(\delta E_h)$ vanishes for $t_c^p\ \left(t_c^h\right) \simeq 0$ and there is no Mott state. Second at

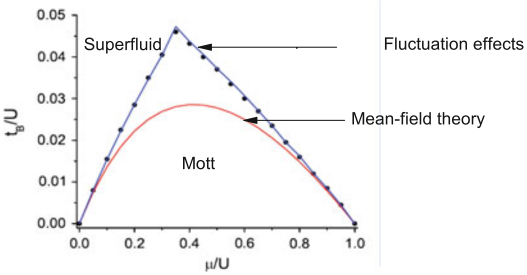


Fig. 6.7 Mott-Superfluid phase boundary for $d = 3$ and $n_0 = 1$. The *red curve* shows the mean-field phase boundary while the *blue curve* and the *black dots* denotes the phase boundary where fluctuation effects are taken into account. The Mott phase is in the shape of a lobe and has particle-hole symmetry at its tip

the tip of the Mott lobe (see Fig. 6.7), where $t_c^p = t_c^h = (2n_0 + 1)/z$ and $\mu = 2m_0(n_0 + 1)/z(2n_0 + 1)$, it becomes equally costly to add a particle or a hole to the system. In other words, the system possess particle-hole symmetry at this special point. This property has profound consequence on the universality class of this phase transition which we shall not dig into in details in the present article. More refined calculations such as a mean-field analysis and even those which keep track of higher order fluctuations can be done and the corresponding phase diagram is shown in Fig. 6.7. The qualitative symmetry issues that we have discussed above, however, do not change.

In conclusion, we have presented a brief pedagogical introduction to the subject of quantum phase transition in the context of spin and boson models. Such transition of course occurs in many other different systems and a detailed discussion of them is beyond the scope of the current article. However, it turns out that in many cases the insights gathered from simple models described above, provides us with powerful tools for understanding the properties of such transitions in more complicated settings. It is therefore expected that this pedagogical review is going to provide a basic introduction to the subject.

References

1. C. Kittle, *Quantum Theory of Solids* (Wiley, New York, 1987)
2. S.-K. Ma, *Modern Theory of Critical Phenomenon* (Addision-Wesley, New York, 1992)
3. R. Shankar, Rev. Mod. Phys. **66**, 129 (1994)
4. S. Sachdev, *Quantum Phase Transitions* (Cambridge University Press, Cambridge, 1999)
5. S.L. Sondhi et al., Rev. Mod. Phys. **69**, 315 (1997)
6. D. Blankschtein et al., Phys. Rev. B **29**, 5250 (1984)
7. T. Senthil et al., Science **303**, 1490 (2004)
8. L. Balents et al., Phys. Rev. B **71**, 144509 (2005)
9. M. Greiner et al., Nature (London) **415**, 39 (2002)

Chapter 7
Statistical Mechanics of Human Resource Allocation: A Mathematical Modeling of Job-Matching in Labor Markets

Jun-ichi Inoue and He Chen

Abstract We provide a mathematical model to investigate the human resource allocation problem for agents, for example, university graduates who are looking for their positions in labor markets. The basic model is described by the so-called Potts spin glass which is well known in the research field of statistical physics. In the model, each Potts spin (a tiny magnet in atomic scale length) represents the action of each student, and it takes a discrete variable corresponding to the company he/she applies for. We construct the energy to include three distinct effects on the students' behavior, namely, collective effect, market history, and international ranking of companies. In this model system, the correlations (the adjacent matrix) between students are taken into account through the pairwise spin–spin interactions. We carry out computer simulations to examine the efficiency of the model. We also show that some chiral representation of the Potts spin enables us to obtain some analytical insights into our labor markets.

Keywords Human resource allocation · Bose–Einstein condensation · Potts spin glass model

7.1 Introduction

our society. This is because they can produce not only various products and services in the society, but also they contribute to the society by paying their taxes. For this reason, in each scale of society, for example, from nation to companies or much smaller communities such as laboratory (or research group) of university, allocation

We are deeply shocked to hear the news of untimely death of Dr. Jun-ichi Inoue at a very young age, on May 17, 2015 after submitting his paper—Editors

H. Chen (✉) · J. Inoue
Graduate School of Information Science and Technology, Hokkaido University,
N14-W-9, Kita-ku, Sapporo 060-0814, Japan
e-mail: chen@complex.ist.hokudai.ac.jp

J. Inoue
e-mail: jinoue@cb4.so-net.ne.jp; j_inoue@complex.ist.hokudai.ac.jp

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_7

of human resources is one of the essential problems. Needless to say, such appropriate allocation of human resource is regarded as a "matching problem" between individuals and some "groups" such as companies, and the difference among individuals in their abilities or preference makes the problem difficult.

A typical example of the human resource allocation is found in simultaneous recruiting of new graduates in Japan. Students who are looking for their jobs might research several candidates of companies to enter and send the application letter through the web site (what we call "entry sheet"). However, the students incline to apply to well-established companies, whereas they do not like to get a job in relatively small companies. This fact enhances the so-called "mismatch" between labors (students) and companies. We can easily see the situation of job-searching process in Japan. At the job fair, we find that some booths could collect a lot of students (they are all wearing a dark suit even in midsummer!). On the other hand, some other booths could not attract the students' attentions. Therefore, the job-matching itself is apparently governed by some "collective behavior" of students. Namely, each student seems to behave by looking at their "neighbors" and adapting to the "mood" in their community, or they sometimes can share the useful information (of course, such information is sometimes extremely "biased") about the market via Internet or social networking service.

In macroeconomics, there already exist a lot of effective attempts to discuss the macroscopic properties [1–5] including so-called search theory [6–9]. However, apparently, the macroscopic approaches lack of their microscopic view points, namely, in their arguments, the behavior of microscopic heterogeneous agents such as labors or companies are neglected.

To investigate the collective effects on the job-matching process from the microscopic view point, we have proposed several models and carried out computer simulations [10–12] by considering some "aggregate data set" for the labor market.

In our previous successive studies [10–12], we succeeded in evaluating the macroscopic quantities such as unemployment rate U and labor shortage ratio Ω from the microscopic view point. However, our studies depend on numerical (computer) simulations for relatively small system size to calculate these quantities, and we definitely need some mathematically rigorous approaches to find the universal fact underlying in the job-matching process of labor markets. It is also important issue to be considered that we should take into account correlation between agents (students) when we consider the job-matching process in realistic labor markets. However in our previous studies [10–12], we have neglected the correlation in our modeling.

Motivated by the above background and requirement, here we propose a mathematical toy model to investigate the job-matching process in Japanese labor markets for university graduates and investigate the behavior analytically. Here we show our preliminary limited results for the typical behavior of the market.

This paper is organized as follows. In Sect. 7.2, we briefly review our previous study on the urn model with disorder [13] and several remarkable properties of the model such as Bose–Einstein condensation. We also mention that the urn model cannot take into account the interactions between agents. In Sect. 7.3, we introduce our toy model, the so-called Potts model, and explain several macroscopic quantities.

Our preliminary results for several job-searching and selection scenarios by students and companies are shown in Sect. 7.4. The last Sect. 7.5 is summary and discussion.

7.2 Urn Models and Bose Condensation: A Short Review

As a candidate of describing the resource allocation problem, we might use the urn models. In this model, one can show that a sort of Bose condensation takes place. Hence, here we introduce the urn model with a disorder and explain several macroscopic properties according to the Ref. [13].

We first introduce the Boltzmann weight for the system as

$$p(\varepsilon_i, n_i) = \begin{cases} \frac{\exp[-\beta E(\varepsilon_i, n_i)]}{n_i!} & \text{(Each ball is distinguishable)} \\ \exp[-\beta E(\varepsilon_i, n_i)] & \text{(Each ball is NOT distinguishable)} \end{cases} \tag{7.1}$$

where β stands for the inverse temperature. The former is called *Ehrenfest class*, whereas the latter is referred to as *Monkey class*.

$E(\varepsilon_i, n_i)$ denotes the energy function for the urn i possessing a disorder ε_i and n_i balls. Obviously, in the system with $E(\varepsilon_i, n_i) \propto n_i (>0)$, each urn (agent) is affected by attractive forces and they attempt to gather the balls (resources), whereas in the system of $E(\varepsilon_i, n_i) \propto -n_i$, each urn is affected by repulsive force and they refuse to collect the balls. The job-matching process in labor market is well-described by the former case. On the other hand, the problem of spent-nuclear-fuel reprocessing plant in Japan is a good example to consider using the latter case, namely, balls are "wastes" and urns are "prefectures."

In the thermodynamic limit: $N, M \to \infty, M/N = \rho = \mathscr{O}(1)$, the averaged occupation probability $P(k)$, which is a probability that an arbitrary urn possesses k balls is given by

$$\rho = \left\langle \frac{\sum_{n=0}^{\infty} n\, \phi_{E,\mu,\beta}(\varepsilon, n)}{\sum_{n=0}^{\infty} \phi_{E,\mu,\beta}(\varepsilon, n)} \right\rangle, \quad P(k) = \left\langle \frac{\phi_{E,\mu,\beta}(\varepsilon, k)}{\sum_{n=0}^{\infty} \phi_{E,\mu,\beta}} \right\rangle, \quad z_s = \exp(\beta\mu)$$

where z_s is a solution of the saddle point equation (S.P.E.) and we defined

$$\phi_{E,\mu,\beta}(\varepsilon, n) = \begin{cases} \frac{\exp[-\beta(E(\varepsilon,n) - n\mu)]}{n!} & \text{(Ehrenfest class)} \\ \exp[-\beta(E(\varepsilon, n) - n\mu)] & \text{(Monkey class)} \end{cases} \tag{7.2}$$

In following, we consider the case of Monkey class with the cost function:

$$E(\varepsilon, n) = \varepsilon n, \tag{7.3}$$

which leads to the Boltzmann weight:

$$\phi_{E,\mu,\beta}(\varepsilon, n) = \exp[-\beta n(\varepsilon - \mu)]. \tag{7.4}$$

Table 7.1 The possible scenario of Bose condensation controlled by the density ρ

Density of balls	Solution of S.P.E.	# of condensation/# of non-condensation
$\rho < \rho_c$	$z_s < 1$	$0/N\rho$
$\rho = \rho_c$	$z_s = 1$	$0/N\rho_c$
$\rho > \rho_c$	$z_s = 1$	$N(\rho - \rho_c)/N\rho_c$

We choose the distribution of disorder: $D(\varepsilon) = \varepsilon_0 \varepsilon^\alpha$. Then, the saddle point equation is given by

$$\rho = \int_0^\infty \frac{\varepsilon_0 \varepsilon^\alpha d\varepsilon}{z_s^{-1} \exp(\beta\varepsilon) - 1} + \rho_{\varepsilon=0} \tag{7.5}$$

where we should notice that $\rho_{\varepsilon=0}$ is negligibly small before condensation. We increase the density ρ keeping the temperature β^{-1} constant. Then, the possible scenario is shown in Table 7.1. It should be noted that we defined the critical density as

$$\rho_c = \int_0^\infty \frac{\varepsilon_0 \varepsilon^\alpha d\varepsilon}{\exp(\beta\varepsilon) - 1} \tag{7.6}$$

After simple algebra, we have

$$P(k) = \frac{z_s^k \varepsilon_0 \Gamma(3/2)}{\beta^{3/2}} k^{-3/2} - \frac{z_s^{k+1} \varepsilon_0 \Gamma(3/2)}{\beta^{3/2}} (k+1)^{-3/2} \tag{7.7}$$

for $\alpha = 1/2$. We show the $P(k)$ for several values of z_s in Fig. 7.1. From this figure, we find that before condensation, namely, for $z_s < 1$, $\rho < \rho_c$, the occupation probability is given by

$$P(k) = \frac{(1 - z_s)\varepsilon_0}{\beta^{3/2}} k^{-3/2} \mathrm{e}^{-k \log(1/z_s)} \tag{7.8}$$

On the other hand, after condensation, that is, for $z_s = 1$, $\rho \geq \rho_c$, we have

$$P(k) = \frac{3\varepsilon_0 \Gamma(3/2)}{2\beta^{3/2}} k^{-5/2} + \frac{1}{N} \delta(k - k_*) \tag{7.9}$$

The important remarks here are the fact that the condensation is specified by the power-law behavior of the occupation probability and for the case of without disorder, namely, for $D(\varepsilon) = \delta(\varepsilon - \varepsilon_0)$, the power-law behavior disappears.

As we saw, the urn model with disorder exhibits a rich physical phenomena such as condensation; however, there is no explicit interaction between agents (balls and urns). Actually, when we consider the job-matching process, it is impossible to accept the assumption that there is no correlation between urns (companies), balls

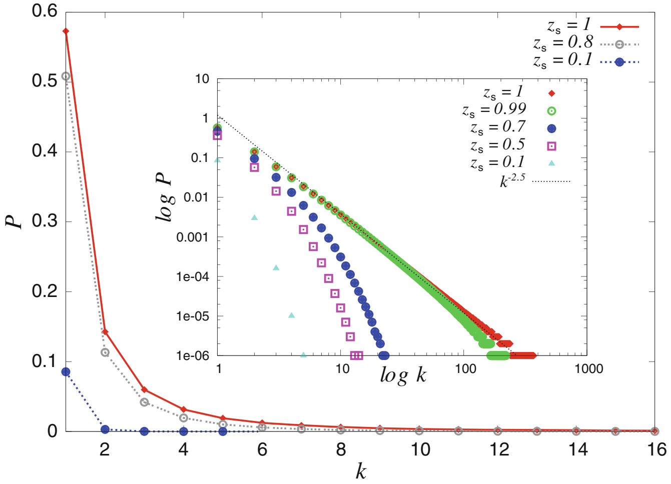


Fig. 7.1 The occupation probability of the Monkey class urn model with disorder. This figure was taken from our previous paper [13]

(students), or between urns and balls. Hence, we should use a different description of the system. In the next section, we use the so-called *Potts model* to describe the problem of human resource allocation.

7.3 Correlations: The Potts Model Descriptions

The basic model proposed here for this purpose is described by the so-called Potts spin glass model which is well-known in the research field of statistical physics. In the model, each Potts spin represents the action of each student, and it takes a discrete value (integer) corresponding to the company he/she applies for. The pairwise interaction term in the energy function describes cross-correlations between students, and it makes our previous model [10–12] more realistic. Obviously, labor science deals with empirical evidence in labor markets and it is important for us to look for the so-called "stylized facts" which have been discussed mainly in financial markets [14, 15]. We also should reproduce the findings from data-driven models to forecast the market's behavior.

In following, we show the limited results. Here we consider the system of labor market having N students and K companies. To make the problem mathematically tractable, we construct the energy (Hamiltonian) to include three distinct effects on the students' behavior:

$$H(\boldsymbol{\sigma}_t) = -\frac{J}{N}\sum_{ij} c_{ij}\,\delta_{\sigma_i^{(t)},\sigma_j^{(t)}} - \gamma\sum_{i=1}^{N}\sum_{k=0}^{K-1}\varepsilon_k\,\delta_{k,\sigma_i^{(t)}} + \sum_{i=1}^{N}\sum_{k=0}^{K-1}\beta_k\left|v_k^* - v_k(t-1)\right|\delta_{k,\sigma_i^{(t)}}, \tag{7.10}$$

where $\delta_{a,b}$ denotes a Kronecker's delta and a Potts spin $\sigma_i^{(t)}$ stands for the company which student i post his application letter to at stage (or time) t, namely,

$$\sigma_i^{(t)} \in \{0, \ldots, K-1\}, \quad i = 1, \ldots, N. \tag{7.11}$$

Therefore, the first term in the above Eq. (7.10) denotes a collective effect, the second corresponds to the ranking of companies and the third term is a market history. In order to include the cross-correlations between students, we describe the system by using the Potts spin glass (see the "quenched" random variables c_{ij} in (7.10)) as a generalization of the Sherrington-Kirkpatrick model, which is well-known as an exactly solvable model for spin glass so far. The overall energy function of probabilistic labor market is written explicitly by (7.10). c_{ij} is an adjacency matrix standing for the "interpersonal relationship" of students, and one can choose an arbitrary form, say

$$c_{ij} = \begin{cases} c & \text{(students } i, j \text{ are 'friendly')} \\ 0 & \text{(students } i, j \text{ are 'independent')} \\ -c & \text{(students } i, j \text{ are 'anti-friendly')} \end{cases} \tag{7.12}$$

for $c > 0$ and the ranking of the company k is defined by ε_k (see e.g. [11] for the detail).

Before investigating some specific cases below, we shall first provide a general setup. Let us introduce a microscopic variable, which represents the decision making of companies for a student as

$$\xi_i^{(t)} = \begin{cases} 1 \text{ (student } i \text{ receives an acceptance at stage } t) \\ 0 \text{ (student } i \text{ is rejected at stage } t) \end{cases} \tag{7.13}$$

Then, the conditional probability is given by

$$P\left(\xi_i^{(t)} | \sigma_i^{(t)}\right) = 1 - A\left(\sigma_i^{(t)}\right) - \left(1 - 2A\left(\sigma_i^{(t)}\right)\right)\xi_i^{(t)} \tag{7.14}$$

with the acceptance ratio

$$A\left(\sigma_i^{(t)}\right) \equiv \sum_{k=0}^{K-1} \delta_{k,\sigma_i^{(t)}} \Theta\left(v_k^* - v_k(t)\right) + \sum_{k=0}^{K-1} \delta_{k,\sigma_i^{(t)}} \frac{v_k^*}{v_k(t)} \Theta\left(v_k(t) - v_k^*\right), \tag{7.15}$$

where $v_k^*(=1/K,$ for simplicity in this paper) and $v_k(t)$ denote the quota and actual number of applicants to the company k per student at stage t, respectively. $\Theta\ (\cdots)$ is a conventional step function. Hence, when we assume that selecting procedure by companies is independent of students, we immediately have

$$
\begin{aligned}
P\left(\boldsymbol{\xi}_t|\boldsymbol{\sigma}_t\right) &= \prod_{i=1}^{N} P\left(\xi_1^{(t)}|\sigma_1^{(t)}\right)\cdots P\left(\xi_N^{(t)}|\sigma_N^{(t)}\right) \\
&= \exp\left[\sum_{i=1}^{N} \log\left\{1 - A\left(\sigma_i^{(t)}\right) - \left(1 - 2A\left(\sigma_i^{(t)}\right)\right)\xi_i\right\}\right]. \quad (7.16)
\end{aligned}
$$

Thus, we calculate the joint probability $P\left(\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t\right)$ by means of $P\left(\boldsymbol{\xi}_t|\boldsymbol{\sigma}_t\right) P\left(\boldsymbol{\sigma}_t\right)$ as

$$
\begin{aligned}
P\left(\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t\right) &= P\left(\boldsymbol{\xi}_t|\boldsymbol{\sigma}_t\right) P\left(\boldsymbol{\sigma}_t\right) \\
&= \frac{\exp\left[\sum_{i=1}^{N} \log\left\{1 - A\left(\sigma_i^{(t)}\right) - \left(1 - 2A\left(\sigma_i^{(t)}\right)\right)\xi_i^{(t)}\right\} - H\left(\boldsymbol{\sigma}_t\right)\right]}{\sum_{\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t} \exp\left[\sum_{i=1}^{N} \log\left\{1 - A\left(s_i^{(t)}\right) - \left(1 - 2A\left(s_i^{(t)}\right)\right)\xi_i^{(t)}\right\} - H(\boldsymbol{\sigma}_t)\right]}
\end{aligned}
\quad (7.17)
$$

where we assumed that the $P(\boldsymbol{\sigma}_t)$ obeys a Gibbs-Boltzmann distribution for the energy function (7.10) as $\sim \mathrm{e}^{-H(\boldsymbol{\sigma}_t)}$.

Therefore, the employment rate as a macroscopic quantity:

$$
1 - U(t) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^{(t)} \quad (7.18)
$$

is evaluated as an average over the joint probability $P\left(\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t\right)$, and in the thermodynamic limit $N \to \infty$, it leads to

$$
\begin{aligned}
1 - U(t) &= \frac{\sum_{\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t} \xi_i \exp\left[\sum_{i=1}^{N} \log\left\{1 - A\left(\sigma_i^{(t)}\right) - \left(1 - 2A\left(\sigma_i^{(t)}\right)\right)\xi_i^{(t)}\right\} - H\left(\boldsymbol{\sigma}_t\right)\right]}{\sum_{\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t} \exp\left[\sum_{i=1}^{N} \log\left\{1 - A\left(\sigma_i^{(t)}\right) - \left(1 - 2A\left(\sigma_i^{(t)}\right)\right)\xi_i^{(t)}\right\} - H\left(\boldsymbol{\sigma}_t\right)\right]} \\
&= \frac{\sum_{\boldsymbol{\sigma}_t} A\left(\sigma_i^{(t)}\right) \exp\left[-H\left(\boldsymbol{\sigma}_t\right)\right]}{\sum_{\boldsymbol{\sigma}_t} \exp\left[-H\left(\boldsymbol{\sigma}_t\right)\right]} \equiv \left\langle A\left(\sigma_i^{(t)}\right)\right\rangle, \quad (7.19)
\end{aligned}
$$

where we defined the bracket:

$$
\langle \cdots \rangle \equiv \frac{\sum_{\boldsymbol{\sigma}_t} (\cdots) \exp[-H(\boldsymbol{\sigma}_t)]}{\sum_{\boldsymbol{\sigma}_t} \exp[-H(\boldsymbol{\sigma}_t)]}. \quad (7.20)
$$

From the resulting expression (7.20), we are confirmed that the employment rate $1 - U(t)$ is given by an average of the acceptance ratio (7.15) over the Gibbs–Boltzmann distribution for the energy function (7.10). Using the above general formula, we shall calculate the employment rate exactly for several limited cases.

7.4 The Results

In following, we show our several limited contributions. Before we show our main result, we shall give a relationship between the Potts modeling and our previous studies [10–12] which are obtained by simply setting $J = 0$ in (7.10).

7.4.1 For the Case of $J = 0$

We first consider the case of $J = 0$. For this case, the energy function (7.10) is completely "decoupled" as follows.

$$H(\boldsymbol{\sigma}_t) = \sum_i H_i, \tag{7.21}$$

$$H_i = -\sum_{k=0}^{K-1} \left\{ \gamma \varepsilon_k - \beta |v_k^* - v_k(t-1)| \right\} \delta_{\sigma_i^{(t)},k} \tag{7.22}$$

where we set $\beta_k = \beta \, (\forall_k)$ for simplicity. Hence, the $v_k(t)$ is evaluated in terms of the definition (7.20) as

$$v_k(t) \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{\sigma_i^{(t)},k} = \left\langle \delta_{\sigma_i^{(t)},k} \right\rangle = \frac{\exp[-\gamma\varepsilon_k + \beta|v_k^* - v_k(t-1)|]}{\sum_{k=0}^{K-1} \exp[-\gamma\varepsilon_k + \beta|v_k^* - v_k(t-1)|]} \tag{7.23}$$

and from the expression of employment rate (7.19), we have

$$1 - U(t) = \frac{\sum_{k=0}^{K-1} \left\{ \frac{v_k^*}{v_k(t)} + \left(1 - \frac{v_k^*}{v_k(t)}\right) \Theta(v_k^* - v_k(t)) \right\} \exp[-\gamma\varepsilon_k + \beta|v_k^* - v_k(t-1)|]}{\sum_{k=0}^{K-1} \exp[-\gamma\varepsilon_k + \beta|v_k^* - v_k(t-1)|]}. \tag{7.24}$$

By solving the nonlinear equation (7.23) recursively and substituting the solution $v_k(t)$ into (7.24), we obtain the time-dependence of the employment rate $1 - U(t)$. In Fig. 7.2, we plot the time-dependence of the employment rate for the case of $K = 3$ (left) and the γ-dependence of the employment rate at the steady state at $t = 10$ for $K = 3$ and $K = 50$ (right). We set the job offer ratio defined in [10–12] as $\alpha = 1$. The ranking factor is also selected by

$$\varepsilon_k = 1 + \frac{k}{K}. \tag{7.25}$$

We here assumed that each agent posts only a single application letter to the market, namely, $a = 1$ in the definition of the previous studies [10–12]. It should be important for us to remind that the above Eq. (7.23) is exactly the same as the update rule for the

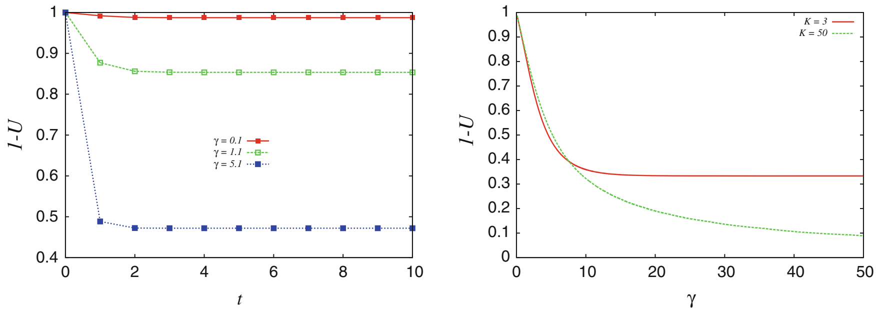


Fig. 7.2 The time-dependence of the employment rate for the case of $K = 3$ (*left*) and the γ-dependence of the employment rate at the steady state at $t = 10$ for $K = 3$ and $K = 50$ (*right*). We set the job offer ratio defined in [10–12] as $\alpha = 1$ and assume that each agent posts only a single application letter to the market, namely, $a = 1$ in the definition of the previous studies [10–12]

aggregation probability $P_k(t)$ in the Ref. [10]. However, when we restrict ourselves to the case of $\alpha = a = 1$, one can obtain the time-dependence of the employment rate exactly by (7.24). This is an advantage of this approach. It also should be noted that from the relationship:

$$U = \alpha\Omega + 1 - \alpha \tag{7.26}$$

(see [10] for the derivation), we have $U = \Omega$, namely, the unemployment rate is exactly the same as the labor shortage ratio for $\alpha = 1$.

It is important for us to notice that the aggregation probability of the system $P(\boldsymbol{\sigma}_t)$ is rewritten in terms of $P_k(t)$ in the Refs. [10–12] as

$$P(\boldsymbol{\sigma}_t) = \left\{ \prod_{k=1}^{K} P_k(t) \right\}^N \tag{7.27}$$

with $P_k(t) = \nu_k(t)$ (see (7.23)) even for $\alpha \neq 1$. For this case, the system parameters are only γ and β, and these unknown parameters are easily calibrated from the empirical data [12]. As the result, we obtained U-Ω curve using (7.26) for the past 17 years in Japanese labor market for university graduates. We plot the resulting and U-Ω curve in Fig. 7.3. The gap between the theoretical and empirical curves comes from the uncertainties in the calibration of average number of application letters a. In this figure, we simply chose the value as $a = 10$ in our calculations.

7.4.2 The Case of $J \neq 0$

We next consider the case of $J \neq 0$. Then, we should note that some "chiral representation" of the energy function (7.10) by means of the chiral Potts spin [16, 17] (Note: "i" appearing in "$2\pi i$" below is an imaginary unit):

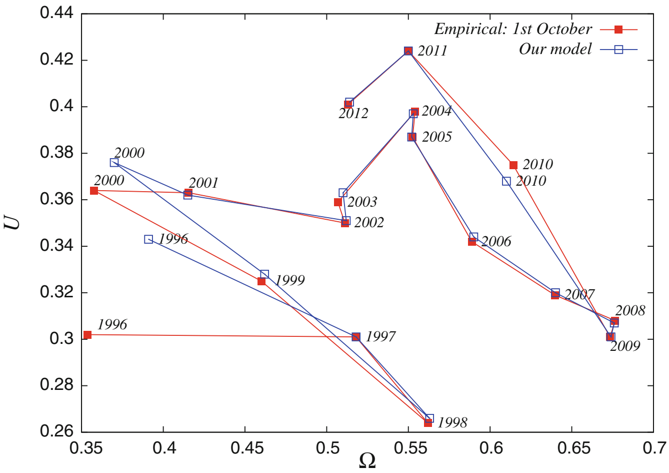


Fig. 7.3 The empirical and theoretical U-Ω curves. We clearly find that the large γ apparently pushes the U-Ω curve toward the *upper right* direction where the global mismatch between the students and companies is large. The picture was taken from our previous study [12]

$$\lambda_i = \exp\left(\frac{2\pi i}{K}\sigma_i^{(t)}\right),\ \sigma_i^{(t)} = 0, \ldots, K-1 \tag{7.28}$$

enables us to obtain some analytical insights into our labor markets.

7.4.2.1 The Case of $\gamma = \beta = 0$: Without Ranking and Market History

As a preliminary, we show the employment rate $1-U$ as a function of $J(>0)$ for the simplest case $\gamma = \beta = 0$ and $c_{ij} = 1\ (\forall_{ij})$ in Fig. 7.4 (right), and the actual number of applicants the company k obtains in Fig. 7.4 (left). We should keep in mind that for this simplest case with local energy

$$H_{ij} \equiv -J\delta_{\sigma_i,\sigma_j} = -\frac{J}{K}\sum_{r=0}^{K-1}\lambda_i^r\lambda_j^{K-r} = -\frac{J}{K}\left\{1+\sum_{r=1}^{K-1}\lambda_i^r\lambda_j^{K-r}\right\} \tag{7.29}$$

under the transformation (7.28) leading to the total energy $H(\boldsymbol{\sigma}) \equiv \sum_{ij} H_{ij}$, by evaluating the partition function:

$$Z = \sum_{\boldsymbol{\sigma}} \exp\left[\frac{J}{NK}\sum_{r=1}^{K-1}\sum_{ij}\cos\frac{2\pi r(\sigma_i-\sigma_j)}{K}\right] \tag{7.30}$$

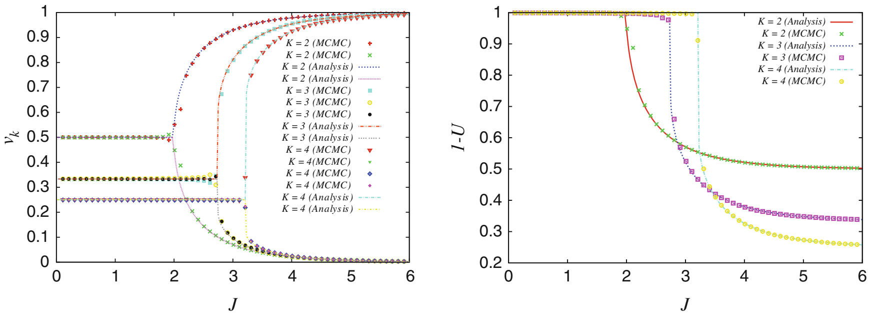


Fig. 7.4 The actual number of applicants v_k (*left*) and employment rate $1 - U$ (*right*) as a function of the strength of cooperation J. We find that the system undergoes a phase transition at the critical point. The transition is the second order for $K = 2$, whereas it is the first order for $K \geq 3$. These critical points are given by $J_c = 2$ for $K = 2$, $J_c = 2.73$ for $K = 3$ and $J_c = 3.21$ for $K = 4$. We should mention that the analytic results (*lines*) and the corresponding Monte Carlo simulations (MCMC) with the finite number of students $N = 1000$ (*dots*) are in an excellent agreement. We should notice that perfect employment phase is a "disordered phase," whereas the poor employment phase corresponds to an "ordered phase" in the literature of order-disorder phase transition. For large strength of cooperation J, as a company occupies all applications up to the quota, $\lim_{J \to \infty}(1 - U) = v_k^* = 1/K$ (the quota per student) is satisfied

in the limit of $N \to \infty$, one can obtain the employment rate $1 - U = \langle A(\boldsymbol{\sigma}) \rangle$ (see also Eq. (7.19)) exactly as

$$
\begin{aligned}
1 - U &= \frac{\sum_{\boldsymbol{\sigma}} A(\boldsymbol{\sigma}) \exp[-H(\boldsymbol{\sigma})]}{\sum_{\boldsymbol{\sigma}} \exp[-H(\boldsymbol{\sigma})]} \\
&= \frac{\left\{ \frac{v_0^*}{v_0} + \left(1 - \frac{v_0^*}{v_0}\right) \Theta(v_0^* - v_0) \right\}}{1 + (K-1)\mathrm{e}^{-\frac{Jx}{K-1}}} + \frac{(K-1)\left\{ \frac{v_k^*}{v_k} + \left(1 - \frac{v_k^*}{v_k}\right) \Theta(v_k^* - v_k) \right\} \mathrm{e}^{-\frac{Jx}{K-1}}}{1 + (K-1)\mathrm{e}^{-\frac{Jx}{K-1}}}
\end{aligned}
\tag{7.31}
$$

with

$$
v_k \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{\sigma_i,k} = \left\langle \delta_{\sigma,k} \right\rangle = \frac{\delta_{0,k} + \sum_{\sigma=1}^{K-1} \delta_{\sigma,k}\, \mathrm{e}^{-\frac{Jx}{K-1}}}{1 + (K-1)\,\mathrm{e}^{-\frac{Jx}{K-1}}}, \quad k = 0, \ldots, K-1,
\tag{7.32}
$$

where an order parameter x is determined as a solution of the following nonlinear equation:

$$
x = (K-1)\left(\frac{1 - \mathrm{e}^{-\frac{J}{K-1}x}}{1 + (K-1)\mathrm{e}^{-\frac{J}{K-1}x}} \right).
\tag{7.33}
$$

It should be noted that the above x is given by the extremum of the free energy density:

$$f = -\frac{Jx^2}{K(K-1)} + \log \sum_{\sigma=0}^{K-1} \exp\left[\frac{Jx}{K(K-1)} \sum_{r=1}^{K-1} \cos\left(\frac{2\pi r}{K}\sigma\right)\right]. \qquad (7.34)$$

The acceptance ratio $A(\boldsymbol{\sigma})$ is now given by

$$A(\boldsymbol{\sigma}) \equiv \sum_{i=1}^{N} A(\sigma_i) = \sum_{i=1}^{N} \sum_{k=0}^{K-1} \delta_{\sigma_i,k} \left\{\frac{v_k^*}{v_k} + \left(1 - \frac{v_k^*}{v_k}\right) \Theta(v_k^* - v_k)\right\}, \qquad (7.35)$$

and we omitted the time t-dependence in the above expressions because the system is no longer dependent on the market history, namely $v_k(t-1)$, for the choice of $\beta = \gamma = 0$ in the energy function (7.10).

In Fig. 7.4, we easily find that phase transitions take place when the strength of "cooperation" J increases beyond the critical point $J_{\rm c}$. Namely, for weak J regime, "random search" by students is a good strategy to realize the perfect employment state $(1 - U = 1)$, however, once J increases beyond the critical point, the perfect state is no longer stable and system suddenly goes into the extremely worse employment phase for $K \geq 3$ (first order phase transition). The critical point of the second-order phase transition for $K = 2$ is easily obtained by expanding (7.33) around $x = 0$ as

$$x = \frac{1 - {\rm e}^{-Jx}}{1 + {\rm e}^{-Jx}} \simeq Jx/2 \qquad (7.36)$$

and this reads $J_{\rm c} = 2$. For the first order phase transition, we numerically obtain the critical values, for instance, we have $J_{\rm c} = 2.73$ for $K = 3$ and $J_{\rm c} = 3.21$ for $K = 4$. As the number K is quite large far beyond $K = 3$ in real labor markets, hence the above finding for the discontinuous transition might be useful for discussing a mismatch between students and companies, which is a serious issue in recent Japanese labor markets (see the Ref. [12]).

We also carried out computer simulations to examine the efficiency of the model. We should mention that the analytic results (lines) and the corresponding Monte Carlo simulations (dots) with finite system size $N = 1000$ are in an excellent agreement in the figures. This preliminary result is a justification for us to conform that one can make a mathematically rigorous platform to investigate the labor market along this direction.

We next consider the case of $\beta, \gamma \neq 0$.

7.4.2.2 Ranking Effects

For the case of $\gamma \neq 0, \beta_k = 0\,(\forall_k)$, the saddle point equation is given by the following two-dimensional vector form:

$$(x_r, y_r) = \langle \boldsymbol{u}_r(s) \rangle_* = \left(\left\langle \cos\frac{2\pi r}{K}s\right\rangle_*, \left\langle \sin\frac{2\pi r}{K}s\right\rangle_*\right), \quad r = 0, \ldots, K-1 \qquad (7.37)$$

where we defined the bracket $\langle \cdots \rangle_*$ as

$$\langle \cdots \rangle_* \equiv \frac{\sum_{s=0}^{K-1} (\cdots) \exp[\psi_r(s : \{x_r\}, \{y_r\})]}{\sum_{s=0}^{K-1} \exp[\psi_r(s : \{x_r\}, \{y_r\})]}, \tag{7.38}$$

$$\psi_r(s : \{x_r\}, \{y_r\}) \equiv \sum_{r=0}^{K-1} \boldsymbol{X}_r \cdot \boldsymbol{u}_r(s) \tag{7.39}$$

with the following two vectors:

$$\boldsymbol{X}_r = \left(\frac{J}{K} x_r + \frac{\gamma}{K} \sum_{k=0}^{K-1} \varepsilon_k \cos \frac{2\pi r}{K} k, \ \frac{J}{K} y_r + \frac{\gamma}{K} \sum_{k=0}^{K-1} \varepsilon_k \sin \frac{2\pi r}{K} k \right) \tag{7.40}$$

$$\boldsymbol{u}_r(s) = \left(\cos \frac{2\pi r}{K} s, \sin \frac{2\pi r}{K} s \right). \tag{7.41}$$

From the energy function (7.10) and the above formula, we should notice that the ranking factor ε_k is regarded as a "state-dependent field" affecting each spin and the symmetry in the "perfect employment phase" for small J (see Fig. 7.4 (right)) might be broken by these unbiased effects. We also should keep in mind that for the case of $\gamma = 0$ or $\varepsilon_k = \varepsilon$ $(\forall_k)$, we find that the Eq. (7.37) possesses the solution of the type: $x_0, \ldots, x_{K-1} \neq 0, \ y_0 = \cdots = y_{K-1} = 0$. It should be also bear in mind that $K = 2$ is rather a special case and the solution of the above type is obtained simply as

$$x_0 = 1, \ x \equiv x_1 = \frac{1 - \mathrm{e}^{-Jx + \gamma(\varepsilon_1 - \varepsilon_0)}}{1 + \mathrm{e}^{-Jx + \gamma(\varepsilon_1 - \varepsilon_0)}}, \ y_0 = y_1 = 0. \tag{7.42}$$

However, for general case, we must deal with two-dimensional vectors $(x_r, y_y), \ r = 0, \ldots, K-1$ with each non-zero component $x_r, y_r \neq 0$ to specify the equilibrium properties of the system.

For the solution $(x_r, y_r), \ r = 0, \ldots, K-1$, we obtain the order parameters and employment rate as

$$v_r = \langle \delta_{r,s} \rangle_* \tag{7.43}$$

$$1 - U = \langle A(s) \rangle_*, \quad r = 0, \ldots, K-1. \tag{7.44}$$

In Fig. 7.5, we plot the J-dependence of the employment rate for $K = 2$ (left) and $K = 3$ (right). From this figure, we find that the employment rate decreases monotonically, however, within intermediate range of J, the $1 - U$ behaves discontinuously. We should notice that in this regime, the "ergodicity" of the system might be broken because the realized value of $1 - U$ by Monte Carlo simulation is strongly dependent on the choice of initial configuration (pattern) of Potts spins.

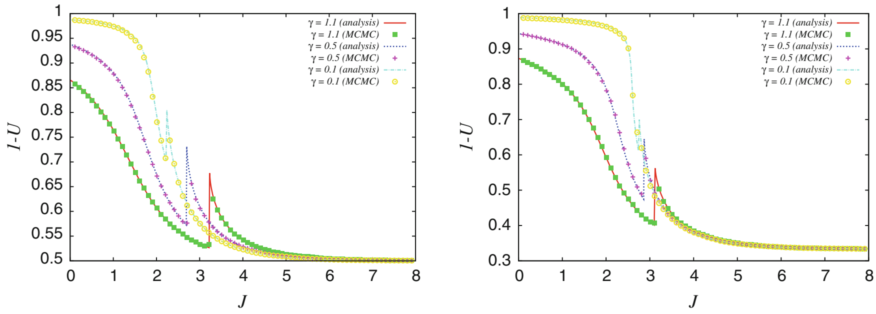


Fig. 7.5 The strength of cooperation J-dependence of the employment rate for the case of $\gamma \neq 0$, $\beta_k = 0$ $(\forall_k)$. We plot the case of $K = 2$ (*left*) and $K = 3$ (*right*). We find that the phase transition as shown in Fig. 7.4 disappears; however, the ergodicity breaking phase appears within intermediate range of J. We are conformed that $\lim_{J \to \infty}(1 - U) = 1/K$ is satisfied even for this case. The simulations (MCMC) are carried out for the system of size $N = 1000$

To see the result more explicitly, we should draw our attention to the initial condition dependence of the J-$(1 - U)$ curve. Actually, here we carry out Monte Carlo simulation to examine the initial configuration dependence of the $1 - U$ numerically and show the results in Fig. 7.6. From this figure, we confirm that the value of the $1 - U$ depends on the initial configuration of the Potts spins although the $1 - U$ is independent of the initial condition for $J < 3$ and $J \gg 1$. In this plot, we chose two distinct initial conditions so as to make the gap of order parameters $\mathscr{O}(1)$ object, that is,

$$\Delta x_r (\equiv x_r^{(\mathrm{a})} - x_r^{(\mathrm{b})}),\ \Delta y_r (\equiv y_r^{(\mathrm{a})} - y_r^{(\mathrm{b})}) \sim \mathscr{O}(1) \tag{7.45}$$

for $r = 0, \ldots, K - 1$.

It might be important for us to investigate the basin of attraction for the matching dynamics analytically as in the Ref. [18]; however, it is far beyond the scope of the current paper and it should be addressed our future study.

7.4.2.3 Market History Effects

We next consider the case of $\beta_k \neq 0$ $(\forall_k)$. For this case, we should replace the $\boldsymbol{X}_r$ in the saddle point equation (7.37) by

$$\begin{aligned}\boldsymbol{X}_r = \Bigg(& \frac{J}{K} x_r + \frac{1}{K} \sum_{k=0}^{K-1} (\gamma \varepsilon_k - \beta_k |v_k^* - v_k(t-1)|) \cos \frac{2\pi r}{K} k, \\ & \frac{J}{K} y_r + \frac{1}{K} \sum_{k=0}^{K-1} (\gamma \varepsilon_k - \beta_k |v_k^* - v_k(t-1)|) \sin \frac{2\pi r}{K} k \Bigg).\end{aligned} \tag{7.46}$$

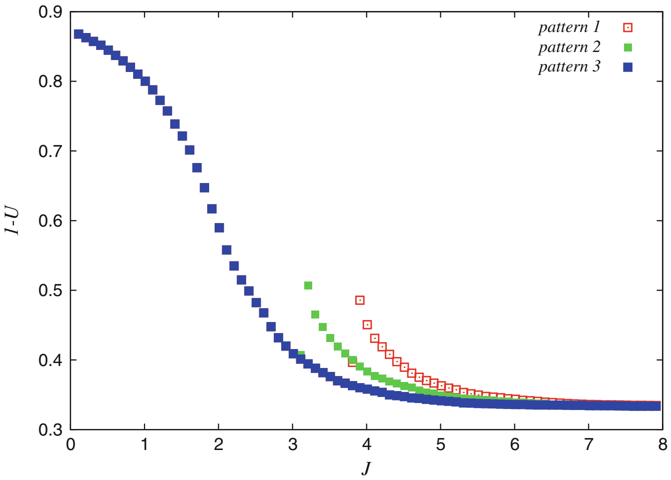


Fig. 7.6 The initial configuration dependence of the $1 - U$. We set $K = 3, \gamma = 1.1$ and choose three distinct initial configurations for Monte Carlo simulations. We find that $1 - U$ is strongly dependent on the initial condition (‘`pattern 1` $\sim$ 3’) within intermediate range of J. In this plot, we chose the two distinct initial conditions so as to make the gap of order parameters $\mathscr{O}(1)$ object, that is, $\Delta x_r (\equiv x_r^{(a)} - x_r^{(b)}), \Delta y_r (\equiv y_r^{(a)} - y_r^{(b)}) \sim \mathscr{O}(1)$ for $r = 0, \ldots, K-1$ (a, b = {`pattern1`, `pattern2`, `pattern3`})

It should be noticed that the v_k at the previous stage $t-1$ is regarded as an “external field” which affects the spin system at the current stage t. Hence, by substituting $v_k(0)$ as an initial state into the Eq. (7.37) with (7.46), we can solve the equation with respect to $v_k(1)$. By repeating the procedures, we obtain the “time series” as $v_k(0) \to v_k(1) \to \cdots v_k(t) \to$ for all k and $1 - U(t)$ as a function of t. In Fig. 7.7, we plot the time (stage) dependence of the employment rate $1 - U$ for the case of $K = 2, J = 1, \gamma = 0.1$ and $(\beta_1, \beta_0) = (1, 4)$ (left) and $(\beta_1, \beta_0) = (4, 1)$ (right).

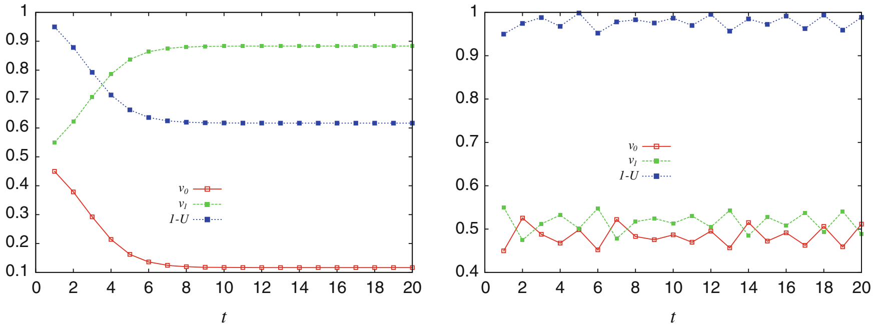


Fig. 7.7 The time (stage) dependence of the employment rate $1 - U$ for the case of $K = 2, J = 1, \gamma = 0.1$ and $(\beta_1, \beta_0) = (1, 4)$ (*left*) and $(\beta_1, \beta_0) = (4, 1)$ (*right*). The “zigzag behavior” in $v_k(t)$ is observed for $(\beta_1, \beta_0) = (4, 1)$

From this figure, we find that the larger weight of the market history effect for the highest ranking company β_1 in comparison with β_0 induces the periodical change of the order for v_1, v_0 due to the negative feedback (a sort of "minority game" [19] for the students). Namely, from the ranking gap $\varepsilon_1 - \varepsilon_0 = 1/2$ for $K = 2$, the company "1" attracts a lot of applications at time t even for a relatively small strength of the preference $\gamma = 0.1$. However, at the next stage, the ability of the aggregation for the company "1" remarkably decreases due to the large β_1. As the result, the inequality $v_1 > v_0$ is reversed as $v_0 > v_1$, and the company "0" obtains much more applications than the company "1" at this stage. After several time steps, the amount of $\beta_1|v_1^* - v_1(t-1)|$ becomes small enough to turn on the switch of the preference for the high ranking company "1," and eventually the inequality $v_1 > v_0$ should be recovered again. The "zigzag behavior" due to the above feedback mechanism in $v_k(t)$ is actually observed in Fig. 7.7 (right). On the other hand, when the strength of the history effect β_0 for the lower ranking company is larger than that of the higher ranking company β_1, the zigzag behavior disappears and v_0, v_1 converge monotonically to the steady states reflecting the ranking $\varepsilon_0 < \varepsilon_1$.

7.5 Summary and Discussion

In this paper, we proposed a mathematical toy model, the so-called chiral Potts model to investigate the job-matching process in Japanese labor markets for university graduates and investigated the behavior analytically. We found several characteristic properties in the system. Let us summarize them below. For the case without ranking effect and market history, we observed that the system undergoes fist-order phase transition for $K \geq 3$ by changing the strength of cooperation $J(>0)$. When we take into account the ranking effect without market history, the ergodicity breaking region in J appears. The market history affects on the dynamics of actual number of applicants to each company $v_k(t)$ to exhibit "zig-zag" behavior.

We would like to stress that the situation and our modeling are applicable to the other type of resource allocation (utilization) such as the so-called *Kolkata Paise Restaurant (KPR) problem* [20].

7.5.1 Inverse Problem of the Potts Model

However, from the view point of empirical science, in this model system, the cross-correlations (the adjacent matrix) between students and companies are unknown and not yet specified. Hence, we should estimate these elements by using appropriate empirical data sets. For instance, if we obtain the "empirical correlation" $\langle \delta_{\sigma_i,\sigma_j} \rangle_{\mathrm{empirical}}$ from the data, we can determine c_{ij} so as to satisfy the following relationship:

$$
\begin{aligned}
\langle \delta_{\sigma_i,\sigma_j} \rangle &= \frac{\partial}{\partial c_{ij}} \log \sum_{\boldsymbol{\sigma}} \exp[-H(\boldsymbol{\sigma} : \{c_{ij}\})] \\
&= \frac{\sum_{\boldsymbol{\sigma}} \delta_{\sigma_i,\sigma_j} \exp[-H(\boldsymbol{\sigma} : \{c_{ij}\})]}{\sum_{\boldsymbol{\sigma}} \exp[-H(\boldsymbol{\sigma} : \{c_{ij}\})]} = \langle \delta_{\sigma_i,\sigma_j} \rangle_{\mathrm{empirical}} \qquad (7.47)
\end{aligned}
$$

where $\langle \delta_{\sigma_i,\sigma_j} \rangle_{\mathrm{empirical}}$ might be evaluated empirically as a time-average by

$$\langle \delta_{\sigma_i,\sigma_j} \rangle_{\mathrm{empirical}} = (1/\tau) \sum_{t=t_0}^{\tau+t_0} \delta_{\sigma_i^{(t)},\sigma_j^{(t)}}. \qquad (7.48)$$

We might also use the EM (Expectation and Maximization)-type algorithm [21] to infer the interactions. Those extensive studies in this directions (the "inverse Potts problem") including collecting the empirical data are now working in progress.

7.5.2 Learning of Valuation Basis of Companies

In this paper, we did not take into account the details of valuation process by companies so far. In our modeling, we assumed that they randomly select suitable students from the candidates up to their quota. This is because the valuation basis is unfortunately not opened for the public and it is somewhat "black box" for students. However, recently, several web sites [22, 23] for supporting job hunting might collect a huge number of information about students as their "scores" of aptitude test.

Hence, we might have a N-dimensional vector, each of whose component represents a score for a given question, for each student $l = 1, \ldots, L$ as

$$\boldsymbol{x}^{(l)} = (x_1^{(l)}, x_2^{(l)}, \ldots, x_N^{(l)}) \qquad (7.49)$$

Then, we assume that each company $\mu = 1, \ldots, K$ possesses their own valuation basis (weight) as a N-dimensional vector $\boldsymbol{a}_\mu = (a_{\mu 1}, \ldots, a_{\mu N})$ and the score of student l evaluated by the company $\mu = 1, \ldots, K$ is given by

$$y_\mu^{(l)} = a_{\mu 1} x_1^{(l)} + a_{\mu 2} x_2^{(l)} + \cdots + a_{\mu N} x_N^{(l)}, \ \ \mu = 1, \ldots, K. \qquad (7.50)$$

It is naturally accepted that the company μ selects the students who are the v_μ^*-top score candidates. Therefore, For a given threshold θ_μ, the decision by companies is given by

$$\hat{y}_\mu^{(l)} = \Theta(y_\mu^{(l)} - \theta_\mu) = \begin{cases} 1 \ (\text{accept}) \\ 0 \ (\text{reject}) \end{cases} \qquad (7.51)$$

where $\Theta(\cdots)$ is a unit step function.

Thus, for L students and K companies, the situation is determined by the following linear equation:

$$\begin{pmatrix} y_1^{(l)} \\ \cdot \\ \cdot \\ \cdot \\ y_M^{(l)} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots\cdots\cdots & a_{1N} \\ \cdots & \cdots\cdots\cdots & \cdots \\ \cdots & \cdots\cdots\cdots & \cdots \\ \cdots & \cdots\cdots\cdots & \cdots \\ a_{M1} & \cdots\cdots\cdots & a_{MN} \end{pmatrix} \begin{pmatrix} x_1^{(l)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_N^{(l)} \end{pmatrix}, \quad l = 1, \ldots, L \tag{7.52}$$

namely,

$$\boldsymbol{y}^{(l)} = \boldsymbol{A}\boldsymbol{x}^{(l)}, \; l = 1, \ldots, L. \tag{7.53}$$

When we have enough number of data sets $(\boldsymbol{y}^{(l)}, \boldsymbol{x}^{(l)}), l = 1, \ldots, L$, one might estimate the valuation base $\boldsymbol{A}$ using suitable learning algorithm. When we notice that the above problem is described by "learning of a linear perception," one might introduce the following cost function:

$$E = \frac{1}{2LM} \sum_{l=1}^{L} \sum_{\mu=1}^{M} \delta_{s_\mu^{(l)},1} \left\{ y_\mu^{(l)} - \sum_{i=1}^{N} a_{\mu i} x_i^{(l)} \right\}^2 \tag{7.54}$$

where we defined $\delta_{a,b}$ as Kroneker's delta and

$$s_\mu^{(l)} = \begin{cases} 1 \text{ (student } l \text{ sends an application letter to company } \mu) \\ 0 \text{ (otherwise)} \end{cases} \tag{7.55}$$

Then, we construct the learning equation as

$$\frac{da_{\mu k}}{dt} = -\eta \frac{\partial E}{\partial a_{\mu k}} = \frac{\eta}{LM} \sum_{l=1}^{L} \delta_{s_\mu^{(l)},1} \left\{ y_\mu^{(l)} - \sum_{i=1}^{N} a_{\mu i} x_i^{(l)} \right\} x_k^{(l)} \tag{7.56}$$

for $\mu = 1, \ldots, M, k = 1, \ldots, N$.

We show an example of the learning dynamics through the error:

$$\varepsilon(t) = \frac{1}{NM} \sum_{\mu=1}^{M} \sum_{k=1}^{N} (a_{\mu k}^* - a_{\mu k}(t))^2, \tag{7.57}$$

where $a_{\mu k}^*$ denotes a "true weight," for artificial data sets in Fig. 7.8.

Here we showed just an example of learning from artificial data sets for demonstration; however, it should be addressed as our future work to apply the learning algorithm to realistic situation using empirical data set collected from [22, 23] or large-scale survey.

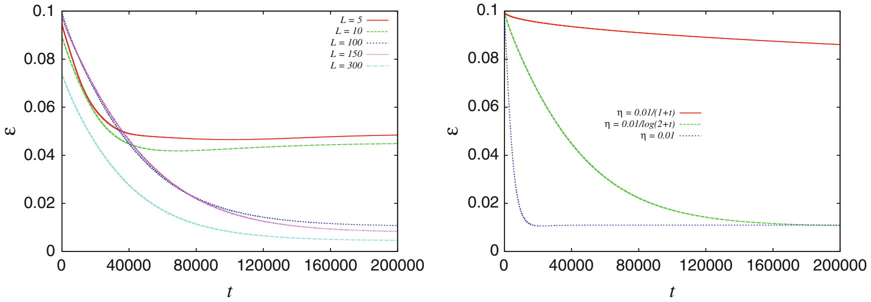


Fig. 7.8 Time-dependence of error $\varepsilon(t) = (1/NM)\sum_{\mu=1}^{M}\sum_{k=1}^{N}(a^{*}_{\mu k} - a_{\mu k}(t))^2$ for the learning Eq. (7.56) using artificial data sets. We choose the learning rate as $\eta = 0.01/\log(2+t)$. $N = M = 10$

Finally, it would be important for us to mention that it could be treated as "dictionary learning" [24] when the vector $\boldsymbol{x}^{(l)}, l = 1, \ldots, L$ is "sparse" in the context of *compressive sensing* [25–27].

Acknowledgments This work was financially supported by Grant-in-Aid for Scientific Research (C) of Japan Society for the Promotion of Science (JSPS) No. 2533027803 and Grant-in-Aid for Scientific Research on Innovative Area No. 2512001313. One of the authors (JI) thanks Hideaki Aoyama, Bikas K. Chakrabarti, Asim Ghosh, Siew Ann Cheong, Yoshi Fujiwara, Shigehiro Kato, Matteo Marsili, and Subinay Dasgupta for fruitful discussion and useful comments. One of the authors (HC) was financially supported by Grant-in-Aid for the JSPS Research Fellowship for Young Scientists. We acknowledge organizers of *International Conference on Emerging Trends in Applied Mathematics*, in particular, Susmita Sarkar for warm hospitality during our stay in Kolkata and editing the conference proceedings.

References

1. M. Aokiand, H. Yoshikawa, *Reconstructing Macroeconomics: A Perspective from Statistical Physics and Combinatorial Stochastic Processes* (Cambridge University Press, Cambridge, 2006)
2. T. Boeri, J. Van Ours, *The Economics of Imperfect Labor Markets* (Princeton University Press, Princeton, 2008)
3. G. Fagiolo, G. Dosi, R. Gabriele, Matching, bargaining, and wage setting in an evolutionary model of labor market and output dynamics. Adv. Complex Syst. **7**(2), 157–186 (2004)
4. R. Gabriele, Labor market dynamics and institution: an evolutionary approach. Working paper in laboratory of economics and management sant'anna school of advances studies, Pisa, Italy (2002)
5. M. Neugart, Complicated dynamics in a flow model of the labor market. J. Econ. Behav. Optim. **53**, 193–213 (2004)
6. S. Lippman, J.J. McCall, The economics of job search: a survey. Econ. Inq. **14**, 155–188 (1976)
7. P.A. Diamond, Aggregate demand management in search equilibrium. J. Polit. Econ. **90**, 881–894 (1982)
8. C.A. Pissarides, Short-run equilibrium dynamics of unemployment vacancie. Am. Econ. Rev. **75**, 676–690 (1985)

9. C.A. Pissarides, *Equilibrium Unemployment Theory* (MIT Press, Cambridge, 2000)
10. H. Chen, J. Inoue, *Dynamics of Probabilistic Labor Markets: Statistical Physics Perspective*. Lecture Notes in Economics and Mathematical Systems, vol. 662, pp. 53–64. Managing Market Complexity (Springer, New York, 2012)
11. H. Chen, J. Inoue, Statistical Mechanics of Labor Markets, *Econophysics of systemic risk and network dynamics, New Economic Windows*, vol. 2013, pp.157–171 (Springer, Milan-Italy, 2012)
12. H. Chen, J. Inoue, Learning curve for collective behavior of zero-intelligence agents in successive job-hunting processes with a diversity of Jaynes-Shannon's MaxEnt principle. Evol. Inst. Econ. Rev. **10**(1), 55–80 (2013)
13. J. Inoue, J. Ohkubo, Power-law behavior and condensation phenomena in disordered urn models. J. Phys. A: Math. Theor. **41**, 324020 (14 pp) (2008)
14. R. Cont, Empirical properties of asset returns: stylized facts and statistical issues. Quant. Financ. **1**, 223 (2001)
15. A. Chakraborti, Y. Fujiwara, A. Ghosh, J. Inoue, S. Sinha, Econophysics: physicists approaches to a few economic problems. J. Econ. Inter. Coord. (2013)
16. H. Nishimori, M.J. Stephen, Gauge-invariant frustrated Potts spin-glass. Phys. Rev. B **27**, 5644 (1983)
17. D.M. Carlucci, J. Inoue, Image restoration using the chiral Potts spin glass. Phys. Rev. E **60**, 2547 (1999)
18. J. Inoue, D.M. Carlucci, Image restoration using the Q-Ising spin glass. Phys. Rev. E **64**, 036121 (2001)
19. D. Challet, M. Marsili, Y.-C. Zhang, *Minority Games* (Oxford University Press, Oxford, 2005)
20. A.S. Chakrabarti, B.K. Chakrabarti, A. Chatterjee, M. Mitra, The Kolkata paise restaurant problem and resource utilization. Phys. A **388**, 2420 (2009)
21. J. Inoue, K. Tanaka, Dynamics of maximum marginal likelihood hyperparameter estimation in image restoration: Gradient descent versus expectation and maximization algorithm. Phys. Rev. E **65**, 016125 (9pp) (2002)
22. http://www.rikunabi.com
23. http://job.mynavi.jp
24. A. Sakata, Y. Kabashima, Statistical mechanics of dictionary learning. Europhys. Lett. **103**, 28008 (2013)
25. E.J. Candes, T. Tao, Decoding by linear programming. IEEE Trans. Info. Theory **51**, 4203 (2005)
26. Y. Kabashima, T. Wadayama, T. Tanaka, A typical reconstruction limit for compressive sensing on L_p-norm minimization. J. Stat. Mech. L09003 (2009)
27. S. Ganguli, H. Sompolinsky, Statistical mechanics of compressive sensing. Phys. Rev. Lett. **104**, 188701 (2010)

Chapter 8
What a Student Can Learn from the Saha Equation

Jayant V. Narlikar

Abstract This article describes the wide applicability of Saha's ionization equation which really launched astrophysics as an important branch of physics and astronomy.

Keywords Saha's ionization equation · Nucleosynthesis

I dedicate this presentation to the memory of Meghnad Saha whose association with Calcutta University is being honoured through this symposium. Although the main part of this work is related to Saha's ionization equation I would like to describe an event that shows how dedicated an academic Saha was. This event is in the form of a quote from Kameshwar Wali's book [1] on Chandra, that is Prof. S. Chandrasekhar who was awarded the Nobel Prize for his work in astrophysics in 1983.

A few months later, 2–8 January 1930, Chandra attended the Indian Science Congress Association meeting held in Allahabad.

The host and the president of the physics section of the Congress was Meghnad Saha, the eminent Indian astrophysicist, whose theory of ionization a decade earlier had unlocked the door to the interpretation of stellar spectra in terms of laboratory spectra of atoms of terrestrial elements, providing information about the state of stellar atmospheres, their chemical composition, the density distribution of various elements, and then about the most important physical parameter, the temperature.

Chandra had learned all of this from Eddington's book "The Internal Constitution of the Stars" and was aware of the high esteem Eddington had accorded to Saha and of Saha's election to the Royal Society in 1927. But Chandra was not aware that Saha was acquainted with his own work; so when he met Saha at the Congress and introduced himself, he was pleasantly surprised by Saha's compliment on his paper in the Proceedings of the Royal Society. Saha said that it was very suggestive and that one of his students was working on extending Chandra's ideas.

He introduced Chandra to this student, who also seemed to know about his work, and he invited Chandra to his home for lunch with a small group of research workers all older than Chandra. The small lunch turned later into a dinner invitation with such

J.V. Narlikar (✉)
IUCAA, Pune, India
e-mail: jayant@iucaa.ernet.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_8

distinguished senior Indian scientists as J.C. Ghosh, D.M. Bose and J.N. Mukherjee. Saha persuaded Chandra to extend his stay in Allahabad so that he and his students could discuss more with him. Chandra, so young, did not expect to be treated almost as an equal by an internationally renowned scientist of Saha's stature.

The above quote gives us a glimpse of the academic environment at Allahabad when Saha was a professor there. I now turn to the main part of this presentation.

8.1 The Ionization Equation

The seminal contribution of Meghnad Saha was his ionization equation, now well known as Saha's ionization equation. To say that it was an important work in astrophysics would be understating it: for, the subject of astrophysics really got going only as a result of the Saha equation. Let us begin with an examination of this assertion. For, to a student of physics the equation provides a menu of delicious results to be enjoyed and appreciated.

Till the second decade of the last century the main observational handle on studies of stars had been their luminosities and spectra. While the luminosity could give a crude estimate of the star's distance using the inverse square law of illumination, the spectrum contained a lot more information.

For example, the continuum spectrum did, in the first approximation resemble the black body spectrum which was well known in those days. If the star was generating energy inside it and radiating it away, then it was in a state of equilibrium and provided the amount radiated was negligible compared to the total store of radiation being scattered within the star one expected the equilibrium state to resemble the black body state. This enabled the astronomer to estimate the star's surface temperature.

With surface temperatures of the order of 3000 K and above, it became clear that the matter at the surface was not likely to be in a state of neutral gas. With large thermal motions and the resulting frequent collisions, it would be impossible for the typical atoms to retain all of their orbital electrons and so they would be ionized and the matter would be in a state of plasma. How would the state of equilibrium be in such circumstances? Naturally, we expect some of the atoms of the matter to remain unchanged, while some would exist as ions and free electrons. But what would be the proportions of these three ingredients?

The Saha equation answered this important question by giving the following relation:

$$\frac{N_i N_e}{N} = \left(\frac{m_e \kappa T}{2\pi h^2}\right)^{3/2} \exp\left(-\frac{B}{\kappa T}\right) \tag{8.1}$$

Here the number N_e, N_i, N denote the number densities of free electrons, ions and neutral atoms at temperature T, B being the binding energy of the atom. The ratio of the binding energy to temperature appears in the exponential form in this equation, thus underscoring its critical effect on the equilibrium abundances of these three component species. Let us try to understand it qualitatively.

What does the binding energy indicate? Recall that an atom contains a positively charged nucleus surrounded by negatively charged electrons. The latter are held close to the nucleus by the force of electrostatic attraction. It is this force that provides the *binding* and its energy denotes what work must be done to tear the electrons apart from the nucleus. Thus the larger the binding energy B the more likely that the atom would stay intact despite any attempt at disruption.

The disruption comes from collisions of the atom by other particles. The larger the velocity $\langle v \rangle$ of a colliding particle the greater the chance of a break-up of the atom. As statistical mechanics tells us, the measure of speed is through the temperature T of the system. The larger the temperature the greater the average velocity per particle. In fact, we know from this subject that the average kinetic energy per particle is proportional to T.

In the above equation we thus see that the larger the value of B the smaller the value of the ratio on the left-hand side. That is, we will expect a smaller proportion of ions and free electrons. However, as T the temperature is raised, the right-hand side increases and we get higher proportions of free ions and electrons. *In short, with rising temperature the matter moves towards the plasma state.*

In Saha's equation we therefore see the broad link between atomic physics, thermodynamics and observational astronomy. The appearance of the Boltzmann's constant κ in (8.1), the atomic binding energy and the temperature indicates this tripartite relationship. This was the beginning for astrophysics: it was here that a clue was made available to interpret the spectrum of a star including the strengths of the emission and absorption lines in it in terms of the ambient state of ionization of the stellar envelope and its temperature.

The surface conditions of the star serve as valuable boundary conditions for stellar models which seek to give details of the unseen stellar interior. In the model proposed by Eddington, the star is a spherical object, made of plasma held together under its own gravitation. In fact, the inward force of gravitation can be so large that unless the star has significant pressure gradients within, it cannot resist gravity.

The classic equations of Eddington [2] are differential equations which give the march of physical quantities like density, temperature, pressure, luminous flux, etc., from the centre outwards. To solve them completely the boundary conditions at the surface are required. This explains why the Saha Equation was such an important stimulus for the early astrophysics. Saha's paper [3] appeared in around 1920 and in the next 4–5 years Eddington's stellar models could be set up. That was the beginning of astronomers using the methods of laboratory physics and applying basic theories of physics to understand the large-scale behaviour of stars, galaxies and the whole universe.

The purpose here is to emphasize the wide applicability of the Saha Equation to astrophysics: for the general impression is that the equation has relevance to stellar scenarios only. I will select two scenarios to illustrate my point, both of them far removed from stellar astrophysics. Both are of critical importance to cosmology, the subject dealing with the large-scale structure of the universe. The first relates to the popular theory of the origin of the microwave background radiation in the universe and the second to the theory of the origin of light nuclei in the early universe.

8.2 The Microwave Background

The presently popular big bang framework of cosmology envisages the following sequence of events since the origin of the universe in a big explosion. In the early stages the universe was very hot, with typical particles of matter moving relativistically, i.e. as photons. Such a phase was said to be radiation dominated. Even electrons and protons moved with speeds close to that of light provided the universe had a temperature of about ten thousand billion. As the universe expanded, it cooled and the speeds of particles began to drop. A crude but very useful estimate tells us that the average energy per particle is comparable to kT. Thus the speeds of the more massive particles will be lower. As the speed falls significantly below the light speed c, the particle becomes 'non-relativistic'. So in our case, first the protons become non-relativistic and then the electrons.

Standard texts in cosmology, e.g. Ref. [4], give the relevant relations describing when this happens. The later cooler epochs have the universe 'dust dominated'. That is, the universe is mainly made of matter that has negligible random motions with respect to the cosmological rest frame. Denoting the scale factor of the expanding universe by S, the simple rule is that the temperature of the radiation drops in inverse proportion to S.

As we shall see in the next section, during the period 1–200 s after the big bang, the temperature of the universe dropped from about 10 billion degrees to a few hundred million degrees. This was when the synthesis of nuclei took place.

The presence of nuclei, free protons and electrons did not, however, have much effect on the dynamics of the universe, which was still radiation dominated. But, these particles especially the lightest of them, the electrons, because of their electric charge acted as scattering centres of the ambient radiation and kept it thermalized. The universe was therefore quite opaque to start with. For, with its frequent scattering light could not travel coherently in a straight line very far.

However, as the universe cooled, the Coulomb electrical attraction between the electron and the proton began to assert itself. In detailed calculations performed by P.J.E. Peebles, the mixture of electrons and protons and of hydrogen atoms was studied at varying temperatures. Because of Coulomb attraction between the electron and the proton, the hydrogen atom has a certain binding energy B. The problem of determining the relative number densities of free electrons, free protons (that is, ions), and neutral H-atoms in thermal equilibrium is therefore analogous to that we considered earlier for stars. The only difference is that the setting is cosmological rather than stellar. Following (8.1) we arrive at the formula relating the number densities of electrons (N_e), protons ($N_p = N_e$), and H-atoms (N_H) at a given temperature T:

$$\frac{N_e^2}{N_H} = \left(\frac{m_e \kappa T}{2\pi h^2}\right)^{3/2} \exp\left(-\frac{B}{\kappa T}\right), \tag{8.2}$$

where m_e = electron mass. This equation is a particular case of *Saha's ionization equation*.

Writing N_B for the total baryon number density, we may express the fraction of ionization by the ratio

$$x = \frac{N_e}{N_B}$$

Then, since $N_H = N_B - N_e$, we get from (8.2)

$$\frac{x^2}{1-x} = \frac{1}{N_B}\left(\frac{m_e \kappa T}{2\pi h^2}\right)^{3/2} \exp\left(-\frac{B}{\kappa T}\right) \tag{8.3}$$

For the H-atom, $B = 13.59$ eV. Substituting for various quantities on the right-hand side of (8.3), we can solve for x as a function of T. The results show that x drops sharply from 1 to near zero in the temperature range of ~5000–2500 K, depending on the value of N_B. For example, $x = 0.003$ at $T = 3000$ K for the case where the baryon density at present is about 2×10^{-30}g cm^{-3}.

Thus by this stage most of the free electrons were removed from the cosmological brew, and as a result the main agent responsible for the scattering of radiation disappeared from the scene. The universe thus became effectively transparent to radiantion. This is called the ‘epoch of last scattering’.

Thus the Saha Equation essentially fixes the epoch when radiation decoupled from matter. Subsequent to this epoch, the radiation cooled more or less undisturbed by whatever process went on with the formation of large-scale structures of matter. Since it had acquired a black body spectrum prior to the epoch of last scattering, it retained that but with a steadily diminishing temperature. In fact the formula $T \propto (1/S)$ continued to hold even when radiation decoupled from matter. The microwave background we observe today is believed to be that relic radiation and it should therefore carry signatures of the last scattering epoch intact. This conclusion has remarkable observational consequences since it enables us to probe the early universe by looking at the microwave background very minutely.

8.3 Primordial Nucleosynthesis

Let us now go further back in time to the 1–200 s epoch when the universe was hot enough for nucleosynthesis to have taken place. Here we encounter the Saha Equation in a different setting, with atomic binding replaced by nuclear binding. Free protons and neutrons could combine to form bigger and bigger nuclei provided their random speeds were slow enough for them to be trapped in the nuclear potential wells. The calculation which was first attempted by George Gamow in the late 1940s is described briefly as follows.

A typical nucleus Q is described by two quantities A = atomic mass and Z = atomic number, and is written[1]

$$^{A}_{Z}Q.$$

This nucleus has Z protons and $(A - Z)$ neutrons. If m_Q is the mass of the nucleus, its binding energy is given by

$$B_Q = [Zm_p + (A - Z)m_n - m_Q]c^2. \tag{8.4}$$

Let us now consider a unit volume of cosmological medium containing N_N nucleons, bound or free. Since the masses of protons and neutrons are nearly equal, we may denote the typical nucleon mass by m. Thus $m_n \approx m_P = m$. If there are N_n free neutrons and N_p free protons in the mixture

$$X_n = \frac{N_n}{N_N}, \quad X_p = \frac{N_p}{N_N} \tag{8.5}$$

will denote the fractions by weight of free neutrons and free protons. If a typical bound nucleus Q has atomic mass A and there are N_Q of them in our unit volume, we may denote the weight fraction of Q by

$$X_Q = \frac{N_Q A}{N_N} \tag{8.6}$$

Now at very high temperatures ($T \gg 10^{10}$ K), the nuclei are expected to be in thermal equilibrium. However, at the temperatures around 10^{10} K the usual formulae for non-relativistic thermodynamics will apply. Further, since we are now concerned with relative number densities, we need to consider the chemical potentials. Thus

$$N_Q = g_Q \left(\frac{m_Q \kappa T}{2\pi h^2}\right)^{3/2} \exp\left(\frac{\mu_Q - m_Q c^2}{\kappa T}\right) \tag{8.7}$$

where we have explicitly used the chemical potentials μ_Q. Since chemical potentials are conserved in nuclear reactions,

$$\mu_Q = Z\mu_p + (A - Z)\mu_n \tag{8.8}$$

assuming that the nuclei were built out of neutrons and protons by nuclear reactions.

The unknown chemical potentials can be eliminated between (8.7) and similar relations for N_P and N_n, with the result expressed in this form:

[1] Sometimes the suffix Z is suppressed.

$$X_Q = \frac{1}{2} g_Q A^{5/2} X_p^Z X_n^{A-Z} \times \xi^{A-1} \exp\left(\frac{B_Q}{\kappa T}\right) \tag{8.9}$$

where

$$\xi = \frac{1}{2} N_N \left(\frac{m\kappa T}{2\pi h^2}\right)^{-3/2} \tag{8.10}$$

Note that Eq. (8.9) is a reincarnation of the Saha equation (8.1) with nuclear binding replacing the atomic Coulomb binding!

For an appreciable build-up of complex nuclei, T must drop to a low enough value to make $\exp(B_Q/\kappa T)$ large enough to compensate for the smallness of ξ^{A-1}. This happens for nucleus Q when T has dropped down to

$$T_Q \sim \frac{B_Q}{\kappa(A-1) \mid ln\xi \mid}. \tag{8.11}$$

Let us consider what happens when we apply the above formula to the nucleus of ^4He. The binding energy of this nucleus is $\cong 4.3 \times 10^{-5}$ erg. If we substitute this value in (8.11) and estimate N_N from the presently observed value of nucleon density of around 10^{-6} cm^{-3}, we find that T_Q is as low as $\sim 3 \times 10^9$ K. However, at this low temperature the number densities of participating nucleons are so low that four-body encounters leading to the formation of ^4He are extremely rare. Thus the underlying assumption of thermodynamic equilibrium (which requires frequent collisions) leading to (8.11) becomes invalid. We therefore need to proceed in a less ambitious fashion in order to describe the buildup of complex nuclei.

Hence, we try using two-body collisions (which are not so rare) to describe the build-up of heavier nuclei. Thus deuterium (d), tritium (^3H) and helium (^3He, ^4He) are formed via reactions like

$$\begin{aligned} p + n &\leftrightarrow d + \gamma \\ d + d &\leftrightarrow {}^3He + n \leftrightarrow {}^3H + p \\ {}^3H + d &\leftrightarrow {}^4He + n. \end{aligned} \tag{8.12}$$

Since formation of deuterium involves only two-body collisions, it quickly reaches its equilibrium abundance as given by

$$X_d = \frac{3}{\sqrt{2}} X_p X_n \xi \exp\left(\frac{B_d}{\kappa T}\right). \tag{8.13}$$

However, the binding energy B_d of deuterium is low so that unless T drops to less than 10^9 K, X_d is not high enough to start further reactions leading to ^3He, ^3H and ^4He. In fact the reactions given in (8.12) with the exception of the first one do not proceed fast enough until the temperature has dropped $\sim 8 \times 10^8$ K.

Although at such temperatures nucleosynthesis does proceed rapidly enough, it cannot go beyond ${}^4He$ and bigger nuclei cannot be made. This is because there are no stable nuclei with $A = 5$ or 8, and so nuclei heavier than ${}^4He$ cannot be made. So the process terminates there. Detailed calculations by several authors have now established this result quite firmly. For making heavier nuclei like C, O, Ne, etc., one has to study a similar process taking place in stars.

There is a fairly good agreement between these calculations and observational estimates of the light nuclei, agreement at least good enough to generate confidence in the big bang picture of an early hot universe. One could equally well argue that the success of the calculation generates confidence in the thermodynamic equilibrium picture conceived by Meghnad Saha.

8.4 Conclusion

I have illustrated these two examples from cosmology to emphasize the wide applicability of Saha's work today. The equation comes up again when we consider the synthesis of nuclei in stars also. Indeed, Saha himself may not have imagined that this result would have important implications for cosmology. However, work of a fundamental nature in physics inevitably finds unexpected applications. The Saha equation is a striking example.

References

1. K. Wali, *Chandra* (Chicago University Press, Chicago 1991)
2. A.S. Eddington, *The Internal Constitution of the Stars* (Cambridge University Press, Cambridge, 1926)
3. M.N. Saha, On a Physical Theory of Stellar Spectra. Proc. R. Soc. A **99**, 135 (1921)
4. J.V. Narlikar, *An Introduction to Cosmology* (Cambridge University Press, Cambridge, 2002)

Chapter 9
Computational Fluid Dynamics with Application to Aerospace Problems

S.K. Chakrabartty

Abstract The subject Computational Fluid Dynamics will be introduced with its applications to compute transonic inviscid and viscous flow past two- and three-dimensional bodies of practical interest in aerospace design and development. Development and implementation of vertex-based finite-volume methods for Euler and Navier–Stokes equations with an algebraic turbulence model will be discussed with some typical examples of aerospace applications, such as (i) Analysis of transonic flow past aerofoils including study of shock-induced separation at the foot of a strong shock, (ii) Analysis and design of aerofoils with flap configurations (GA(W)-2 and HANSA) including Gurney flaps, (iii) Internal flows through nozzles and cascades, (iv) Navier–Stokes analysis of round leading edge delta wing at high angles of attack and (v) Design and analysis of complete SARAS aircraft with sideslip will be presented.

Keywords Computational fluid dynamics · Transonic flow · Turbulence modelling

9.1 Introduction

Computational Fluid Dynamics (CFD), a mature discipline now, can contribute considerably to the design, analysis and development of engineering systems involving fluid flows. The main advantage lies in its ability to cut down developmental costs by minimizing scaled model testing leading to reduction in the design cycle time and design cost. Under the assumption of continuous media, Reynolds Averaged Navier–Stokes (RANS) equations with a suitable turbulence model can be accepted as valid equations governing the fluid flows in general. The basic features of the fluid flows of aerodynamic interest can be simulated by the Euler equations, obtained from the RANS equations under the assumption of inviscid flows. Euler equations can handle the rotational flows although there is no mechanism to generate vorticity (see Fig. 9.1). Computation of transonic flow field, where both subsonic and supersonic regions are present and significant in the determination of overall character of

S.K. Chakrabartty (✉)
National Aerospace Laboratories, Bangalore, India
e-mail: drskc@hotmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_9

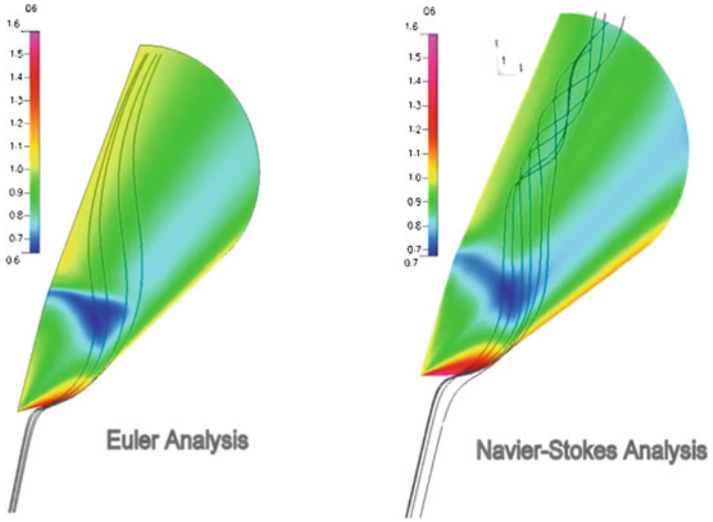


Fig. 9.1 Surface pressure distribution and streamlines, $M_\infty = 0.95$, $\alpha = 22°$, *Grid*: $201 \times 41 \times 75$

the flow, was initiated in India during early seventies of last century. Mathematical complexities of this flow field were governed by the non-linear mixed-type partial differential equations, characterized by the fact that subsonic and supersonic regions exist adjacent to each other, separated by sonic line, and not known a priori and the usual occurrence of shock waves terminating the supersonic region drew attention of the mathematical community throughout the world, particularly after the famous discovery of transonic controversy, Bers [1]. Early successful computations of transonic small perturbation equation representing two-dimensional inviscid flow past aerofoils were done by Magnus and Yoshihara [2] and Murman and Cole [3] using finite-difference methods and by Chakrabartty [4, 5] using integral equation method. A rapid progress was observed in next two decades to develop computational methods for transonic Full Potential (FP) equation, Euler equations and Reynolds Averaged Navier–Stokes (RANS) equations in two and three dimensions of Niyogi et al. [6] and Chakrabartty [7]. Added gradually are different geometrical configurations and physical complexities like viscosity, flow separation, turbulence, etc. Visualization of flow field, surface load distribution and various aerodynamic forces and moments are the criteria for basic design of aerospace configurations. CFD complements experimental and theoretical fluid dynamics by providing an alternative and cost effective means to simulate real flow phenomena. One advantage of CFD is that the entire flow field solution can be stored and the required analysis can be done later. In experiments, analysis of flow field is difficult; each experiment serves a specific purpose which makes the process expensive. To understand structure of the flow field inside the boundary layer and at the foot of the shock, spiralling vortex structure, vortex breakdown, three-dimensional boundary layer separation, etc. are some of the phenomena which need CFD analysis.

Considering the potential of the growing subject and visualizing its utility in national projects of civil and military aircraft, National Aerospace Laboratories (NAL) at Bangalore decided to initiate the development of a comprehensive code for the analysis of three-dimensional viscous compressible flows past complete aircraft and its components capable of capturing the typical flow features and predicting the aerodynamic coefficients for design purpose with adequate accuracy. So the project was started in 1991 continuing till today passing successfully all the critical stages of development and delivering results of high accuracy for many applications of aerospace interest. Along with the main solver codes, the corresponding pre- and post-processing codes were also developed to complete CFD package consisting: (i) pre-processing and grid generation, (ii) solution of the governing equations and (iii) post-processing of the solution.

9.2 Transonic Flows and Mathematical Complexities

Distinguishing features of a transonic flow field are as follows:

(i) Both subsonic and supersonic regions are present and significant in determining the overall character of the flow.
(ii) These are separated by a surface called sonic surface, where local Mach no. $M = 1$, which is a part of the solution.
(iii) Weak shocks are present.
(iv) Local Mach number, $M \sim 1$ throughout.

Such flow fields occur in a wide range of aerodynamic problems like flow through nozzles, around blunt bodies moving supersonically, near airplane wings flying close to Mach number unity and around propellers and turbine blades. Guiding differential equation for potential flow is non-linear even under the assumption of small perturbation and is of mixed elliptic–hyperbolic type. Euler and Navier–Stokes (N–S) equations are non-linear and coupled sets of equations: For three-dimensional flows, steady N–S equations are elliptic–hyperbolic, whereas corresponding unsteady equations are parabolic–hyperbolic. Unsteady Euler equations are hyperbolic with respect to time. With respect to space, for both steady/unsteady supersonic flows, the characteristic matrix has five real eigen values with respect to all spatial directions within the Mach cone with axis along the velocity and apex angle $\cos^{-1}(1/M)$, so they are of hyperbolic type; whereas for subsonic flow, the characteristic matrix has three equal real and two complex eigen values with respect to every spatial direction, so they are of Hybrid type [6]. As viscosity → zero, N–S equations lose its parabolic and elliptic nature which implies singular behaviour in limiting sense.

Apart from these, there are physical complexities like Compressibility, Viscosity, Shocks, Transition, Turbulence, Wake, Separation, Reattachment, Shock-Induced Separation, Vorticity, etc., and geometrical complexities like Multielement aerofoils, Satellite Launch Vehicles, Aircraft, Helicopter, Turbo machinery, etc.

9.3 Governing Equations

Under the assumption of continuous media, Navier–Stokes equations are valid governing equations for fluid flows. Direct Numerical Simulation (DNS) or Large Eddy Simulation (LES) requires enormous resources, not yet reached the maturity to be used for practical problems. Reynolds Averaged Navier–Stokes (RANS) equations with some turbulence model have been accepted to govern the fluid motion.

In solving a differential equation, a serious difficulty occurs when the highest space derivative terms of this equation are multiplied by a small parameter, such as with the Navier–Stokes and related equations for high Reynolds number flow ($1/Re \to 0$, or viscosity, $\mu \to 0$). We are considering a physical theory, viz. that of inviscid flow governed by the Euler equations, as an approximation to another physical theory, that of viscous flow governed by the Navier–Stokes equations. The solutions of the Euler equations admit discontinuities of different kinds, e.g. discontinuity in velocity from one streamline to an adjacent streamline, although the pressure is continuous. Such discontinuities, known as **contact discontinuities**, are consistent with the theory of characteristics and with the integral conservation laws whose differential forms are the Euler equations. Another form of discontinuity, known as a **shockwave**, may exists in which the normal velocity, pressure, density, entropy and absolute temperature jump across a surface. Since such discontinuities may not occur in the Navier–Stokes equations, we expect the perturbation problem to be singular. This implies that the Euler limit of a Navier–Stokes solution may not be uniformly valid. The solution of the Euler equations (weak solution) for a given set of boundary conditions is not unique. Often, it is the inherent numerical dissipation in the solution scheme that makes the solution unique. We shall call a solution of the Euler equation, which is the Euler limit of the Navier–Stokes solution for the same conditions, a relevant Euler solution [6]. We emphasize again the fundamental difference, in principle, between the cases of small viscosity and large viscosity. The physical theory expressed by the Euler equations assumes that the viscosity is zero. An asymptotic expansion for small μ must tend to the Euler solution in the limit $\mu \to 0$. There is no physical theory for compressible flow in which the viscosity is infinite. The Stokes flow is valid for incompressible flow. The governing equations for large values of coefficient of viscosity are still the Navier–Stokes equations.

9.4 Eddy Viscosity Approach for Turbulence Modelling

- Eddy viscosity μ_t, related to mean strain using Prandtl's mixing length approach.
- No additional differential equations are to be solved.
- Need minimum computer time and storage.
- Easy to implement.
- Give reasonably accurate results for attached and separated flows past simple geometries.

For two- and three-dimensional Reynolds Averaged Navier–Stokes (RANS) equations, solver codes JUMBO2D and JUMBO3D, respectively, with algebraic Baldwin Lomax type of turbulence models along with corresponding Euler versions JUEL2D and JUEL3D have been developed. Salient features of these codes are multi-block structured, capability of handling H-, C- or O-type grids with vertex-based finite-volume space discretization, five-stage Runge–Kutta time stepping and second- and fourth-order artificial dissipation with convergence acceleration schemes like grid sequencing, local time stepping, implicit residual smoothing and enthalpy damping. Grid generation code JUMGRID was also developed in NAL for two- and three-dimensional multi-block grid for complex external and internal flow configurations. The code can generate a multi-block grid for any arbitrary, geometrically complex bodies, for both external and internal flows. The grid is structured in any particular block but the blocks may be unstructured. The geometry data for a complex configuration is generally available for each component separately, sometimes in its own coordinate system. The code JUMGRID follows the following steps to achieve the desired multi-block grid: (1) Read geometrical data for each component. (2) Redefine complete geometry in global coordinate system. (3) Form blocks suitable to the geometry and the topology intended. (4) Define all six faces in each block. (5) Redistribute points on each face as necessary. (6) Fill up initial internal points in each block. (7) Establish inter-block connectivity/boundary condition. (8) Smooth the grid by solving elliptic equations.

Details of these codes, grid generation algorithms, algebraic turbulence models and time-stepping schemes used in the present analysis, different types of boundary conditions, convergence acceleration schemes for Euler and Navier–Stokes equations and their implementation in the vertex-based finite-volume discretization are available in [8–11].

9.5 Results and Discussions

These codes have been applied for analysis and design of two- and three-dimensional aerodynamic problems like transonic flow past different aerofoils, wings, wing–body combinations, satellite launch vehicles and complete civilian and military aircraft. Details of these applications are available in [12–25]. Some typical cases are illustrated below for the sake of completeness. To illustrate the inability of Euler equations to generate vorticity, a flow past slender body of revolution at high angle of attack has been considered. Geometrical discontinuity and viscosity are the two major sources of vorticity generation in the flow field. If it is generated by the kink or sharp leading edge of the body, then Euler equation can take care of its motion, but it has no mechanism to generate it. So the smooth round leading edge body of revolution at angle of attack, vortex roll-up phenomenon, can be obtained only through the viscous flow solution. Surface pressure distribution and streamlines are shown in Fig. 9.1. Inviscid flow computed using Euler code JUEL3D shows no vortex roll-up, whereas

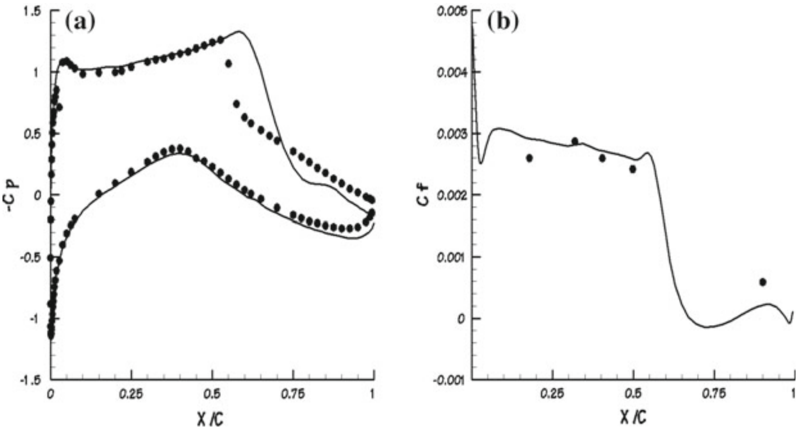


Fig. 9.2 **a** Comparison of surface pressure C_p and **b** skin friction distribution C_f with experiments

Navier–Stokes analysis at $Re_\infty = 6 \times 10^6$ using JUMBO3D code shows vortex roll-up.

Transonic flow past RAE28222 Aerofoil with shock-induced separation has been computed using two-dimensional Navier–Stokes solver JUMBO2D code at $M_\infty = 0.75, \alpha = 2.81°$, $Re_\infty = 6.2 \times 10^6$. A C-type algebraic grid (257 × 61) where minimum normal spacing of 10^{-5} with chord length as unity has been used such that the maximum law-of-the-wall coordinate y^+ value [6] at the first grid node is of the order of four. Distributions of pressure coefficient, CP and surface skin friction coefficient, C_f are shown in Fig. 9.2. Apart from the shock position, both C_P and C_f compare well with experiment [26] throughout the airfoil. Shock position can be improved using better turbulent model, but then it was observed that the discrepancies appear on the lower side of the body surface. Negative skin friction distribution shows the point of separation and reattachment after the shock. Mach contour distribution is shown in Fig. 9.3 showing the supersonic zone terminated with shock.

Next example shown here is flow through VKI-LS-59 Turbine Cascade for inlet flow angle, $\alpha = 30°$, inlet total pressure and total temperature corresponding to isentropic Mach number, $M1_{is} = 0.282$, inlet Reynolds number, $Re_\infty = 0.8 \times 10^6$ and exit static pressure corresponding to isentropic Mach numbers, $M2_{is} = 0.81$ and 1.12. Figure 9.4 shows the C-type periodic grid used. Proper resolution of the boundary layer has been achieved by the minimum cell height near the boundary as $5 \times 10^{4-}$ which gives an average y^+ of the order of four. Figures 9.5 and 9.6 give the Mach contour distributions and Figs. 9.7 and 9.8 show the comparison of isentropic Mach number distributions using various grid levels and with experiential results (NAL) for subsonic and supersonic exits, respectively [17]. In both the cases, grid convergence of the solutions was obtained and the fine grid results show excellent comparison with experiments. Figures 9.9 and 9.10 show the streamline distributions near the trailing edge. Here, the typical vortex structure of the flow has been clearly predicted.

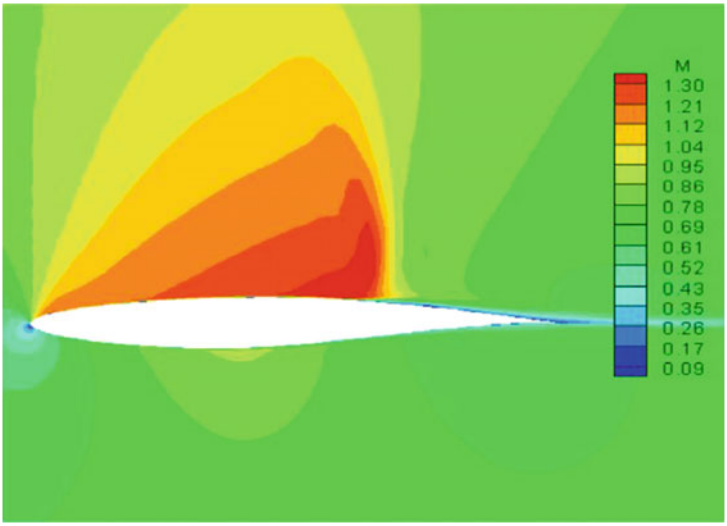
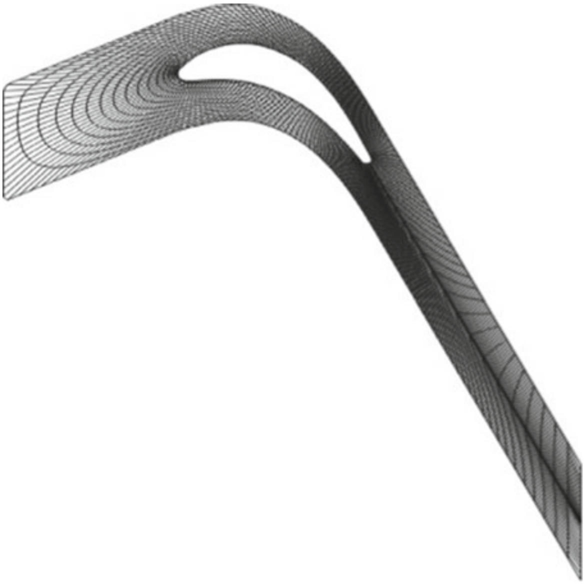


Fig. 9.3 Mach Contour Distribution on RAE28222 Aerofoil at $M_\infty = 0.75$, $\alpha = 2.81°$, $Re_\infty = 6.2 \times 10^6$

Fig. 9.4 C-type grid (595 × 51) for flow through Turbine Cascade



Effectiveness of the control surface is of critical importance for any aircraft. JUMBO2D code was used for the analysis and subsequent design of airfoil flap geometry for better performance with interest to in-house aircraft projects. It was observed that the existing flap on the HANSA aircraft lacks the required effectiveness

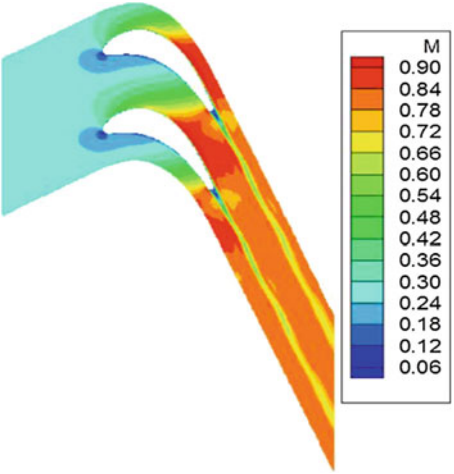


Fig. 9.5 Mach Contour for subsonic exit

at moderate angles of incidence. JUMBO2D code was used to analyse the flow and after several cycles of design and analysis showed a significant improvement in the flow behaviour [18]. Figure 9.11 compares the shape of the original aerofoil flap with that of the modified one. Figure 9.12 shows the improvement of the flow pattern

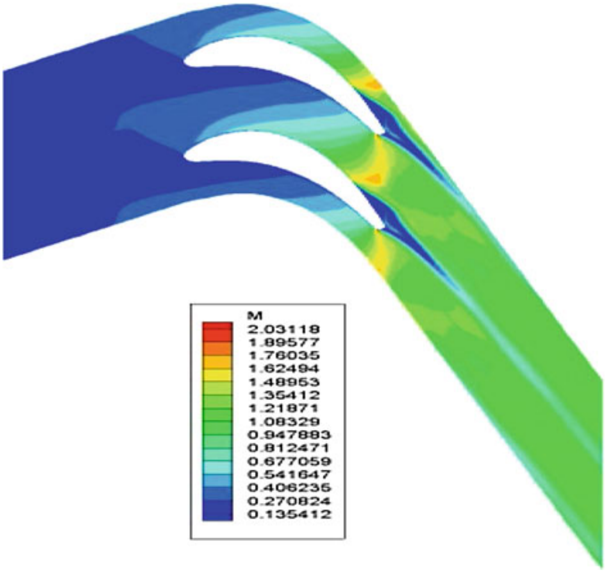


Fig. 9.6 Mach Contour for supersonic exit

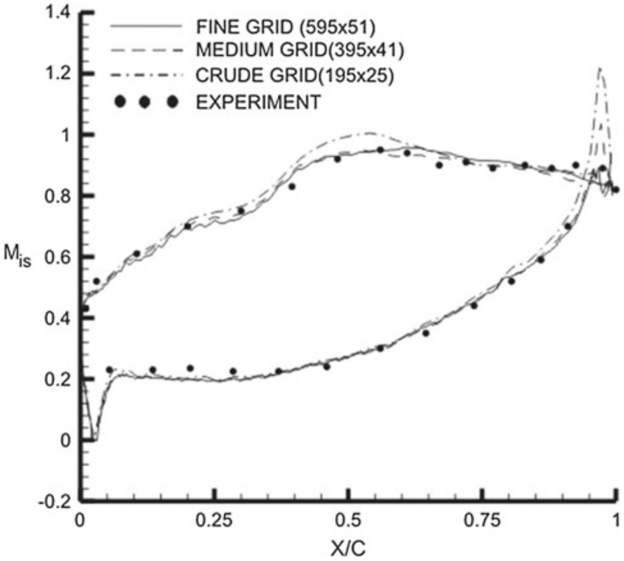


Fig. 9.7 M_{is} Distribution for subsonic exit

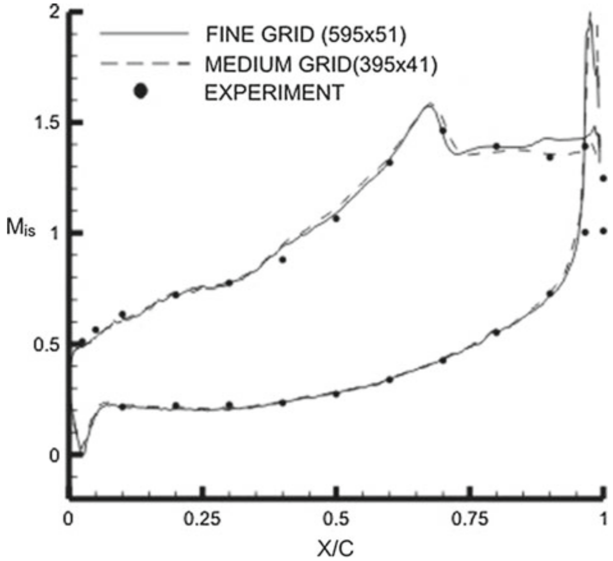


Fig. 9.8 M_{is} Distribution for supersonic exit

for the modified geometry at $M_\infty = 0.3$, angle of attack, $\alpha = 10°$, flap deflection angle, $\delta = 20°$ and $Re_\infty = 2.0 \times 10^6$. The separation free flow was achieved with smooth behaviour having attached flow throughout with a small cove vortex in the gap close to the trailing edge.

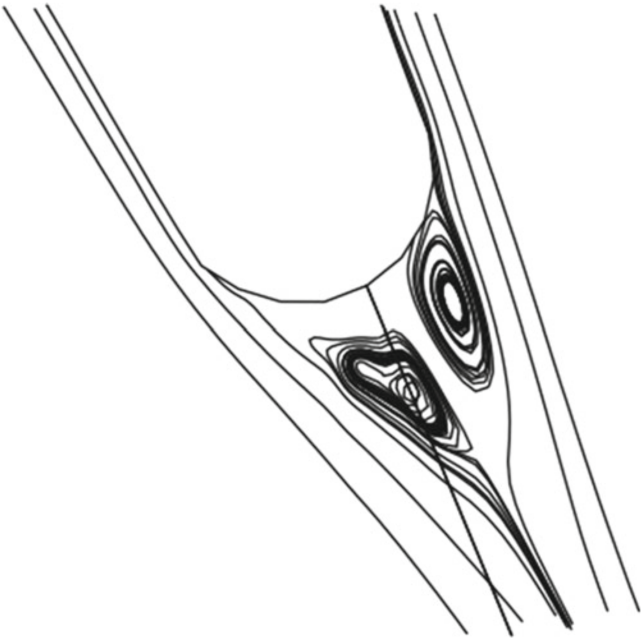


Fig. 9.9 Streamlines close to trailing edge subsonic exit

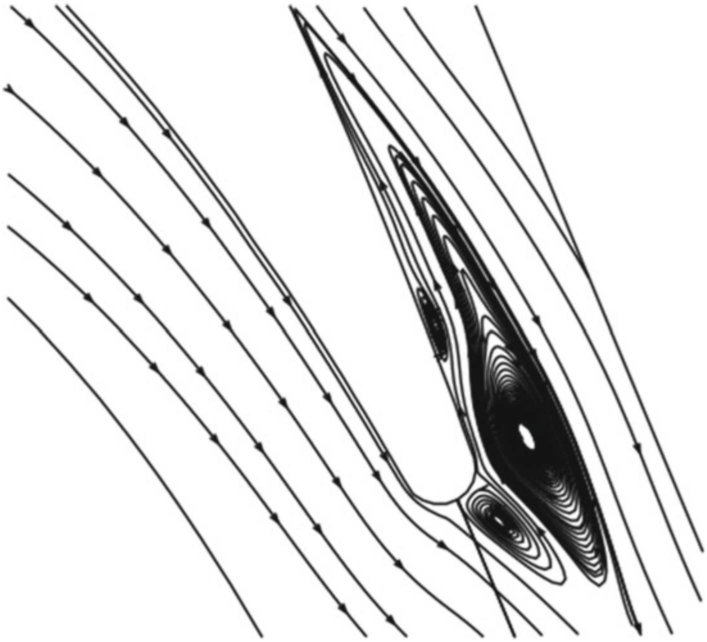


Fig. 9.10 Streamlines close to trailing edge supersonic exit

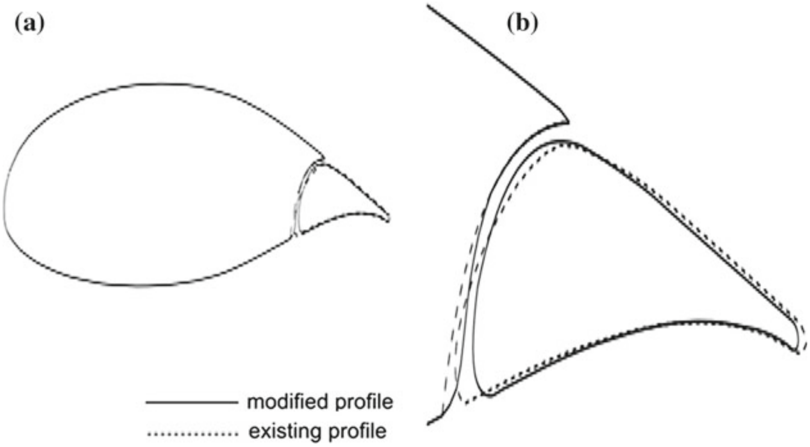


Fig. 9.11 Existing HANSA-3 and modified aerofoil-flap configurations: **a** Full configuration, **b** Enlarged view near the gap and flap

The complicated vortex flow over a 65° cropped delta wing with round leading edge at moderate to high angles of attack has been studied in detail using JUMBO3D code [15, 16]. A typical example has been shown here for $M_\infty = 0.85$, $\alpha = 10°$ and $Re_\infty = 2.38 \times 10^6$. The topology of three-dimensional separated flows can be studied using the surface skin friction lines. Skin friction lines on the upper surface of the wing have been shown in Fig. 9.13. Primary separation point starts inboard and close to the apex and gradually moves towards the leading edge until it

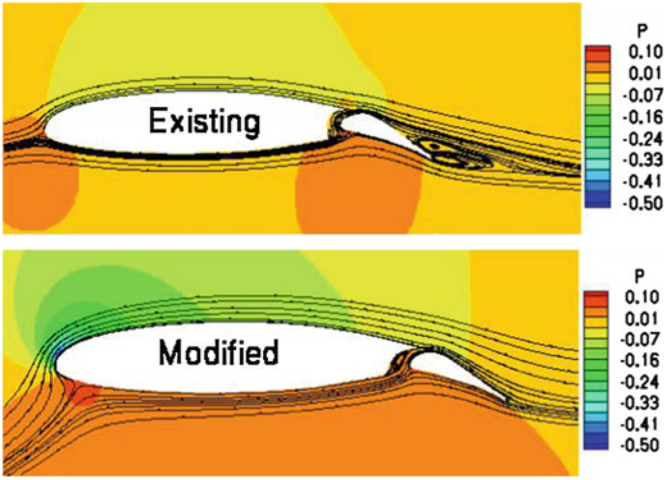


Fig. 9.12 Streamlines and pressure contours of existing and modified flap configurations at $M_\infty = 0.3$, $\alpha = 10°$, $\delta = 20°$ and $Re_\infty = 2.0 \times 10^6$

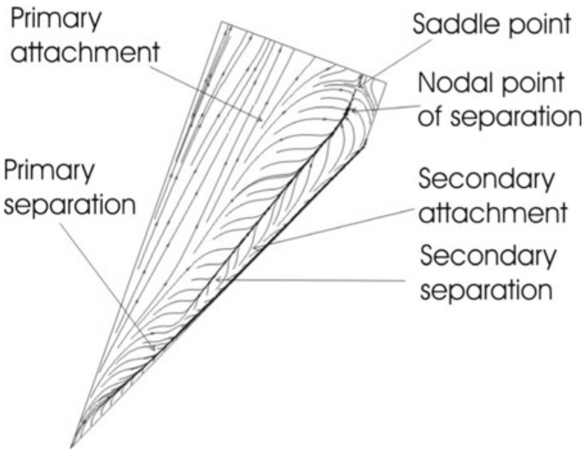


Fig. 9.13 Skin friction lines on the *upper* surface of the delta wing

meets the kink. The secondary separation starts at around 35 % of the root chord and terminates at a node at around 90 % of the root chord. At around 95 % of the root chord, a corresponding saddle point on the surface appears. Appearances of primary and secondary vortices are more clearly shown in Fig. 9.14, where the streamline patterns of the particles leaving the leading edge to form the roll-up vortex have been plotted. The leading edge connectivity of core vortex is lost due to the cropped tip, but the particle path remains smooth in the core and no abrupt lifting of vortex near this junction is observed. Appearance of secondary vortex with opposite direction can

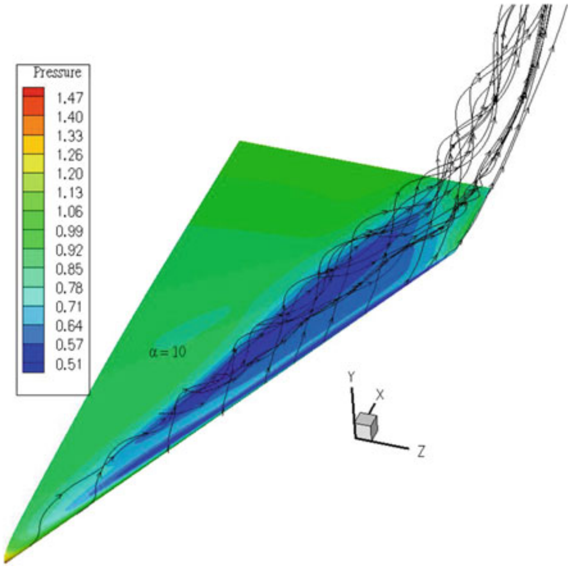


Fig. 9.14 Surface pressure contour and streamlines showing primary and secondary vortices

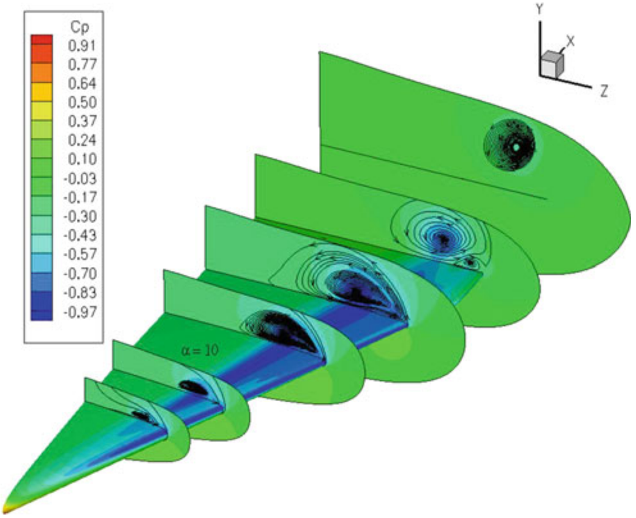


Fig. 9.15 C_p Contours and particle traces on the *upper* surface and at different cross-flow planes

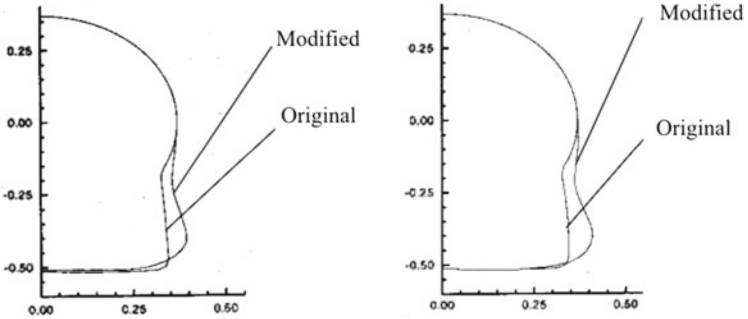


Fig. 9.16 Typical fuselage sections showing original and modified fuselage

also be seen here. Particle traces at different cross-flow planes are shown in Fig. 9.15. The trajectories of the particles in these planes roll-up around what appears to be the vortex core. Here, the primary and secondary vortices are present up to about 80 % of the chord. For complete analysis of this complex, vortex flows at different angles of attack are available in Ref. [15].

Design of wing–fuselage junction is very critical for an aircraft to minimize the mutual interference effect. In order to improve the aerodynamic flow pattern, the SARAS fuselage has been modified by repeated analysis using JUMBO3D code and modification of the wing–body configuration was obtained from the visualization of the surface flow [19, 20]. Figure 9.16 shows some typical cross sections of the fuselage with both original and modified contours. Figure 9.17 shows the streamlines on the original and modified configurations. The modified fuselage gives the smooth

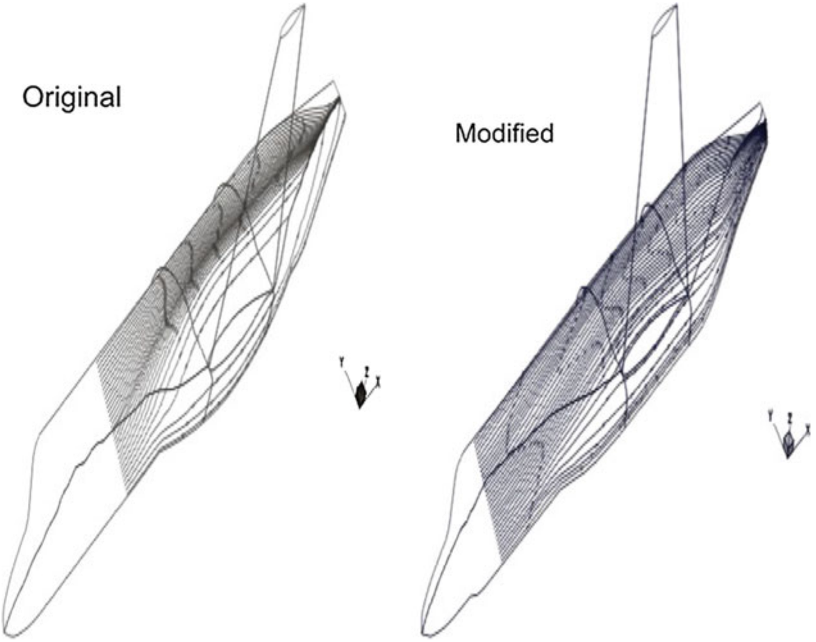


Fig. 9.17 Surface flow patterns on original and modified wing–fuselage fairings at $M_\infty = 0.5$, $\alpha = 0°$

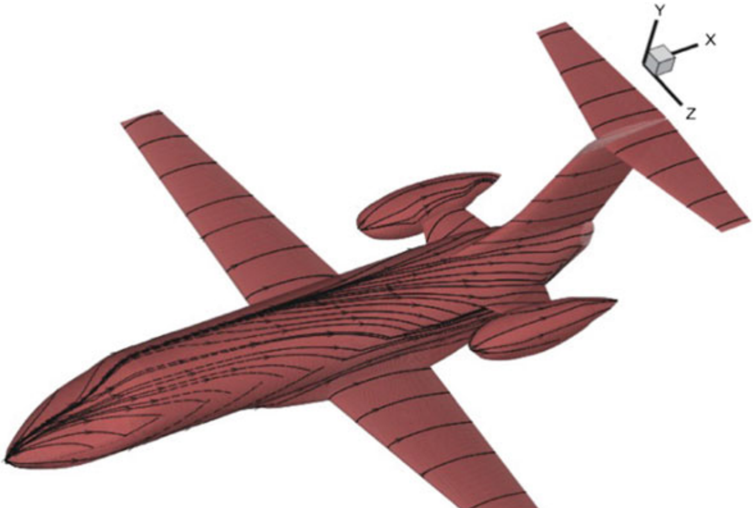


Fig. 9.18 Surface flow patterns on complete SARAS aircraft at $M_\infty = 0.5$, $\alpha = 5°$

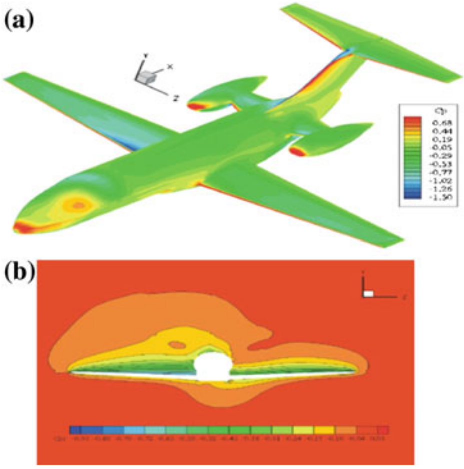


Fig. 9.19 C_p contours on **a** surface and **b** i-const. *cross section* across wing and Fuselage, at $M_\infty = 0.4$, $\alpha = 0°$ and sideslip $\beta = -15°$

streamlines and does not have the clustering of streamlines on the rear fuselage as seen in the original one. This smooth behaviour of the flow is maintained with complete aircraft and computed surface streamline patterns are shown in Fig. 9.18. Finally, the inviscid flow past complete SARAS aircraft with sideslip was analysed and Fig. 9.19 shows Cp contours on the surface and i-constant cross section across wing and fuselage, at $M_\infty = 0.4$, $\alpha = 0°$ and sideslip $\beta = -15°$.

9.6 Conclusion

Three-dimensional Euler and Navier–Stokes codes along with grid generation and post-processing packages reached its maturity to make the analysis of practical aerospace vehicles with reasonable accuracy and affordable computing cost. In the near future, CFD will make the design process faster and less expensive. As the confidence in the analysis codes becomes more and more reliable, aircraft designers will be more likely to use CFD-generated data, though it cannot completely replace the traditional design and experimental methods, but CFD techniques will certainly play a major and irreplaceable role in any future aerospace projects.

Still there is enough scope of improvements like add better turbulence models, reformulate the codes to solve equations in non-inertial frame of reference keeping in mind its application to turbo machinery and helicopter rotor blades, compute unsteady time-accurate flow, improve capability to use unstructured grids, add computer graphics and improve user-friendliness, etc.

Acknowledgments Let me take this opportunity to thank the Department of Applied Mathematics, University of Calcutta and the organizing committee of International conference on "Emerging Trends in Applied Mathematics dedicated to the Memory of Sir Asutosh Mukherjee and Con-

tributions of S.N. Bose, M.N. Saha and N.R. Sen" for their kind interest and for giving me this opportunity to share my experience here.
I also express my sincere thanks to Prof. P. Niyogi, under whose guidance the work on CFD was started in Jadavpur University, Kolkata as early as 1971 and whose continuous encouragement helped me a lot to continue the work till now.
I sincerely acknowledge the proper facilities and encouragements I received from my colleagues at CTFD Division and from the authority of the National Aerospace Laboratories, Bangalore. Computed results presented and discussed here are from the work done by the speaker jointly with his colleagues Mrs. K. Dhanalakshmi, Dr. J.S. Mathur, Dr. V. Ramesh and Mr. Manish Singh at the CTFD Division, National Aerospace Laboratories, Bangalore. I express my sincere thanks to all of them.

References

1. L. Bers, *Mathematical Aspects of Subsonic and Transonic Gasdynamics* (Wiley, New York, 1958)
2. R. Magnus, H. Yoshihara, Inviscid transonic flow over airfoils. AIAA J. **8**, 2157–2162 (1970)
3. E.M. Murman, J.D. Cole, Calculation of plane steady transonic flows. AIAA J. **9**, 114–121 (1971)
4. S.K. Chakrabartty, An analytical and numerical study of plane transonic profile flow at zero and non-zero incidence, Ph.D. Thesis, Jadavpur University, Calcutta (1974)
5. S.K. Chakrabartty, Approximate shock-free transonic solution for lifting aerofoils. AIAA J. **13**(8), 1094–1097 (1975)
6. P. Niyogi, S.K. Chakrabartty, M.K. Laha, *Introduction to Computational Fluid Dynamics* (Pearson Education (Singapore) Pte. Ltd., New Delhi, 2005)
7. S.K. Chakrabartty, Computation of transonic potential flow past wing body configurations. Acta Mech. **81**(3–4), 201–209 (1990)
8. S.K. Chakrabartty, Numerical solution of Navier-Stokes equations for two dimensional viscous compressible flows. AIAA J. **27**(7), 843–844 (1989)
9. S.K. Chakrabartty, K. Dhanalakshmi, Computation of transonic flows with shock induced separation using algebraic turbulence models. AIAA J. **33**(10), 1979–1981 (1995)
10. S.K. Chakrabartty, A finite volume nodal-point scheme for solving two dimensional Navier-Stokes equations. Acta Mech. **84**(1–2), 139–153 (1990)
11. S.K. Chakrabartty, Vertex-based finite-volume solution of the two dimensional Navier-Stokes equations. AIAA J. **28**(10), 1829–1831 (1990)
12. S.K. Chakrabartty, K. Dhanalakshmi, J.S. Mathur, Computation of three-dimensional transonic flow using a cell vertex finite volume method for the Euler equations. Acta Mech. **115**(1–4), 161–177 (1996)
13. S.K. Chakrabartty, K. Dhanalakshmi, Navier-Stokes analysis of Korn aerofoil. Acta Mech. **118**(1–4), 235–239 (1996)
14. S.K. Chakrabartty, K. Dhanalakshmi, J.S. Mathur, Computation of three dimensional transonic viscous flow using the JUMBO3D code. Acta Mech. **119**(1–4), 181–197 (1996)
15. S.K. Chakrabartty, K. Dhanalakshmi, J.S. Mathur, Navier-Stokes analysis of vortex flow over a cropped delta wing. Acta Mech. **131**, 69–87 (1998)
16. S.K. Chakrabartty, K. Dhanalakshmi, J.S. Mathur, Computation of flow past aerospace vehicles. Curr. Sci. **77**(10), 1295–1302 (1999)
17. S.K. Chakrabartty, K. Dhanalakshmi, J.S. Mathur, Navier-Stokes analysis of flow through two-dimensional cascades. Comput. Fluid Dyn. J. **10**(2), 233–241 (2001)
18. S.K. Chakrabartty, K. Dhanalakshmi, V. Ramesh, Navier-Stokes analysis and design of aerofoil-flap configurations for low speed aircraft. Comput. Fluid Dyn. J. **11**(3), 285–289 (2002)

19. S.K. Chakrabartty, J.S. Mathur, K. Dhanalakshmi, Application of advanced CFD codes for aircraft design and development at NAL. J. Aerosp. Sci. Technol. (Formerly J. Aeronaut. Soc. India) **55**(1), 74–88 (2003)
20. J.S. Mathur, K. Dhanalakshmi, S.K. Chakrabartty, Application of advanced CFD codes for design and development SARAS aircraft. J. Aerosp. Sci. Technol. **55**(3), 174–185 (2003)
21. S.K. Chakrabartty, K. Dhanalakshmi, V. Ramesh, Navier-Stokes analysis of GA(W)-2 aerofoil with deflected flap and redesign of HANSA flap for better performance. Comput. Fluid Dyn. J. **12**(1), 89–97 (2003)
22. K. Dhanalakshmi, S.K. Chakrabartty, J.S. Mathur, V. Ramesh, Computation of inviscid flow past SARAS Aircraft with side-slip using a multi-block grid. J. Aerosp. Sci. Technol. **56**(1), 66–76 (2004)
23. S.K. Chakrabartty, Role of computational fluid dynamics in design of aerospace configurations. J. Aerosp. Sci. Technol. **57**(1), 24–32 (2005)
24. M.K. Singh, K. Dhanalakshmi, S.K. Chakrabartty, Navier-Stokes analysis of airfoils with Gurney flaps. AIAA J. Aircr. **44**(5), 1487–1493 (2007)
25. J.S. Mathur, K. Dhanalakshmi, V. Ramesh, S.K. Chakrabartty, Aerodynamic design and analysis of SARAS aircraft. Comput. Fluid Dyn. J. **16**(3), 320–334 (2008)
26. P.H. Cook, M.A. McDonald, M.C.P. Firmin, Aerofoil RAE2822 pressure distribution and boundary layer and wake measurements, AGARD-AR-138 (1979)

Chapter 10
Contact Problem in Elasticity

Arabinda Roy

Abstract We review the classical Hertz contact theory under normal load and formulate a new unified method valid for the Hertz contact theory and a variety of frictionless elliptic contact problem with an elliptic contact connection both for a rigid punch and a conical indenter. We also give a direct way to evaluate the stress and displacement field in the medium. As a limiting case, we derive the results for circular connection as well as line contact problems in the two-dimensional case. Relations of the contact stresses of the wheel of a locomotive rolling on the rails of straight and curved railway with failure of the rail are discussed. Use of such study in hardness testing is discussed.

Keywords Hertz contact theory · Frictionless elliptic contact problem · Rigid punch · Conical indenter · Stresses and deflections

10.1 Introduction

The subject of contact mechanics began with the publication of Hertz in 1882. The contact problem to be considered here goes by the name of Hertz. Hertz, basically an electric engineering while investigating the phenomenon of Newton's optical interference fringes in the gap between two glass lenses, became interested in the localized deformation and the distribution of pressure between the optical lenses. During this investigation, he wrote his paper "on the contact of elastic solids" in 1881 at the age 24. Even after 140 years, his interest in this topic has not waned to mechanical engineers because of its application.

Many authors (Boussinesq [1], Huber [2], Huber and Fuchs [3] and in recent years Sneddon [4–7] and Johnson [8]) have studied the same problem in detail.

The stresses and deflections arising from the contact between two elastic bodies have practical applications in hardness testing, wear and impact damage of engi-

A. Roy (✉)
Department of Applied Mathematics, University of Calcutta, Kolkata, India
e-mail: roy_arabinda1@yahoo.co.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_10

neering ceramics, the design of gear teeth, ball and roller bearings. Rolling contact under normal loads occurs in rail wheels and under normal and tangential load in locomotive driving wheels and braked rail wheels.

10.2 Basic Equations

Let two convex-shaped bodies be in contact initially at a point at the upper body S_1 and be pressed with a vertical pressure P. Referred to the tangential plane at the initial point, the profile of S_1 is an ellipse given by

$$z = \frac{1}{2R_1'}x_1^2 + \frac{1}{R''{}_1}y_1^2,$$

in terms of R_1' R_2'', the principal radii of curvature of the surface at the origin.

There is a similar set for S_2.

Due to contact pressure, the surfaces of two bodies are displaced relative to fixed points by an amount $u_z^1 + u_z^2 = \delta_1 + \delta_2 = \delta - Ax^2 - By^2$.

Outside the contact area, $u_z^1 + u_z^2 > \delta - Ax^2 - By^2$.

Guided by his observation of interference fringes between two cylindrical lenses, Hertz made the following assumptions:

(a) the contact area is in general elliptical.
(b) for local deformation, each body can be regarded as an elastic half space so that the displacement and stresses satisfy the differential equations of equilibrium for elastic bodies and the stresses are localized.
(c) the surfaces are assumed to be frictionless so that only normal stress is present.
(d) The normal pressure is zero outside and equal and opposite inside the contact region.

Our interest is to determine the stress and displacement in the medium, $z > 0$ associated with specified displacement w_1 on elliptic contact area S.

The contact problem can be described as a mixed boundary value problem of the half space, $z > 0$ with the following boundary condition:

$$\tau_{zx}(x, y, 0) = \tau_{zy}(x, y, 0) = 0, \ \forall\ (x, y).$$

$$w(x, y, 0) = w_1(x, y)\ (x, y) \in S. \tag{10.1}$$

$$\tau_{zz}(x, y, 0) = 0\ (x, y) \notin S.$$

The displacements satisfy the elastic equilibrium equation and the displacements **u** are given in terms of three harmonic potentials ϕ, ψ and χ as

$$\mathbf{u} = \nabla\phi - z\nabla\psi(3 - 4\nu)\mathbf{k}\psi + \nabla \times (\mathbf{k}\chi)\,, \tag{10.2}$$

where $\nabla^2(\phi, \psi, \chi) = 0$.

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

ν is the Poisson's ratio and **k** is an unit vector in the positive z direction

$$(\phi, \psi, \chi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [P(\xi, \eta),\ Q(\xi, \eta), R(\xi, \eta)]\ e^{i(\xi x + \eta y) - z\left(\xi^2 + \eta^2\right)^{1/2}}\, d\xi d\eta.$$

First of the boundary condition (10.4) yields $R(\xi, \eta) = 0$.

$$\left(\xi^2 + \eta^2\right)^{1/2} P(\xi, \eta) = (1 - 2\nu)\, Q(\xi, \eta). \tag{10.3}$$

We introduce a new unknown $B(\xi, \eta)$ related to $Q(\xi, \eta)$ as

$$B(\xi, \eta) = \frac{1}{1 - 2\nu} \left(\xi^2 + \eta^2\right) P(\xi, \eta) = \left(\xi^2 + \eta^2\right)^{1/2} Q(\xi, \eta).$$

We now take the indentation w on the elliptic contact area overlying on an elastic half space under as $w(x, y) = \delta - f(\rho)$, where $f(0) = 0$ and $\rho = \sqrt{\left(x^2/a^2\right) + \left(y^2/b^2\right)}$.

We make the following transformation in the integrand such that the ellipse is transformed to a circular region:

$$\begin{aligned} x' &= a\rho' \cos\phi & \qquad y' &= b\rho' \sin\phi \\ \xi a &= k \cos\chi & \qquad \eta b &= k \sin\chi. \end{aligned}$$

so that $k = \left(\xi^2 a^2 + \eta^2 b^2\right)^{1/2}$, $\rho = a\rho' = \left(x^2 + \dfrac{a^2}{b^2} y^2\right)^{1/2}$.

Also,

$$e^{\pm ik\rho'(\cos\chi - \phi)} = \sum_{n=0}^{\infty} \varepsilon_n (\pm i)^n J_n(k\rho') \cos n(\chi - \phi)$$

where $\varepsilon_n = 2 - \delta_{n0}$, δ_{n0} is the Kronecker's delta function and $J_n(z)$ is the Bessel function of order n.

We assume the Fourier expansions for the transformed quantities, namely

$$B(\xi, \eta) = B(k, \chi) = \sum_{0}^{\infty} B_n(k) \cos n\chi + \sum_{0}^{\infty} B_n^s(k) \sin n\chi.$$

On making the transformation, we obtain the surface displacement as

$$w_1(x,y) = \frac{\left(1-\nu^2\right)}{\pi E}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(\xi,\eta)\left(\xi^2+\eta\right)^{-1/2} e^{-i(\xi x+\eta y)} d\xi d\eta$$
$$= \frac{\left(1-\nu^2\right)}{\pi Eab}\left[I_{00}\int_0^{\infty} B_0(k)J_0(k\rho)dk + \sum_{n=1}^{\infty}\sum_{s=1}^{\infty} I_{ns}B_n(k)J_s(k\rho)\,dk\cos s\phi\right].$$

The set of integral equation on satisfying the contact conditions in (10.1) reduces to

$$\int_0^{\infty} I_{00}B_0(k)J_0(kr)\,dk = \frac{\pi Eab}{1-\nu^2}\left[\delta - f(\rho)\right],\ r<1.$$

$$\int_0^{\infty} B_0(k)kJ_0(kr)\,dk = 0,\ r>1 \tag{10.4}$$

and $B_n = 0$, $n \neq 0$, $r = \rho/a$, where

$$I_{00} = \int_0^{2\pi}\frac{d\chi}{\left(\frac{\cos^2\chi}{a^2}+\frac{\sin^2\chi}{b^2}\right)^{1/2}} = 4b\int_0^{\pi/2}\frac{dk}{\left(1-k_0^2\sin^2\chi\right)^{1/2}} = 4bK\left(k_0\right). \tag{10.5}$$

The solution of the integral equation (10.4) is taken as

$$B_0(k) = a_1\int_0^1 \phi(t)\cos kt\,dt, \tag{10.6}$$

where $a_1 = \dfrac{\pi Eab}{I_{00}\left(1-\nu^2\right)}$.

Substituting in the first equation, we get on using the integral

$$\int_0^{\infty}\cos\left(\xi t\right)J_0(\xi r)\,d\xi = \frac{H(r-t)}{\sqrt{r^2-t^2}},$$

$$\int_0^r \frac{\phi(t)\,dt}{\sqrt{r^2-t^2}} = \left[\delta - f(r)\right] \text{ with } r = \left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)^{1/2} < 1,$$

which is an Abel-type integral.

On inverting

$$\phi(t) = \frac{2}{\pi}\left[\delta = \frac{d}{dt}\int_0^t \frac{yf(y)\,dy}{\sqrt{t^2-y^2}}\right] = \frac{2}{\pi}\left[\delta - t\int_0^t \frac{f'(y)}{\sqrt{t^2-y^2}}dy\right]. \tag{10.7}$$

On integrating by parts and on evaluating at $t = 1$, the penetration δ is given by

$$\delta = \int_0^1 \frac{f'(y)\,dy}{\sqrt{1-x^2}} + \frac{\pi}{2}\phi(1). \tag{10.8}$$

The normal surface stress is on inverting the second contact condition

$$\tau_{zz}(x,y,0) = \int_0^\infty B(k)kJ_0(kr)\,dk = \frac{1}{r}\cdot\frac{d}{dr}r\int_0^\infty B(k)J_1(kr)\,dk.$$

$$\tau_{zz} = \frac{\pi Ea}{8\left(1-\nu^2\right)K(k_0)}\left[\frac{\phi(1)}{\sqrt{1-r^2}} - \int_\rho^1 \frac{\phi'(t)}{\sqrt{t^2-r^2}}\,dt\right]. \tag{10.9}$$

We list the surface stress for various indenters:
(a) elliptic punch $w = \delta =$ constant so that, $a_1\phi(t) = 2\delta/\pi$

$$B(k) = \frac{\sin k}{k} \qquad \tau_{zz}(x,y) = \frac{P}{\pi ab}\cdot\frac{1}{\sqrt{a^2-\rho^2}}\,H(a-\rho).$$

(b) Hertz contact problem

$$w = \delta - 2C\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)$$

$a_1\phi(t) = \delta - 2Cr^2$ and if $\phi(1) = 0$ so that

$$B(k) = \frac{\sin k}{k^2} - \frac{\cos k}{k} \qquad \tau_{zz}(x,y,0) = -\frac{3}{2}p_m\left(1-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right)^{1/2} H\left(1-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right)$$

in terms of the mean pressure $p_m = P/4ab$.
(c) For a conical punch, the indentation is given by

$$w_1 = \delta - f\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)$$

with $f(0,0) = 0$ and

$$f(x,y) = a\tan\beta\sqrt{\frac{x^2}{a^2}+\frac{y^2}{b^2}},$$

where $\pi - \beta/2$ is the semi-vertical angle of the cone.
Then

$$\phi(t) = \frac{2}{\pi}(\delta - ta\tan\beta).$$

$$\tau_{zz}(x, y, 0) = \frac{\pi E a}{8\left(1-\nu^2\right) K\left(k_0\right)}\left[\frac{\phi(1)}{\sqrt{a^2-\tau^2}} H(a-\rho)\right]$$

for non-adhesive case, we have in addition $\phi(1) = 0$, i.e. $\delta = a \tan \beta$.

Substituting the values of $B(k)$, we obtain the stress in the medium for Hertz problem as

$$\tau_{zz} = \frac{\mu a_1}{2\pi a b}\int_0^{2\pi}\left[I_1 + z\left(\frac{\cos^2\chi}{a^2} + \frac{\sin^2\chi}{b^2}\right)^{1/2} I_3\right] d\chi + \tau_{zz}^*,$$

where

$$I_1 + z\left(\frac{\cos^2\chi}{a^2} + \frac{\sin^2\chi}{b^2}\right)^{1/2}$$

and u is the positive root of

$$\frac{x^2}{a^2+u} + \frac{y^2}{b^2+u} + \frac{z^2}{u} - 1 = 0.$$

The corresponding tangential stresses are

$$\tau_{zx}(x, y, z) + i\tau_{zy}(x, y, z) = \frac{\mu}{2\pi a b}\left(\frac{\cos\phi}{a} + i\frac{\sin\phi}{b}\right)\int_0^{2\pi} I_6 d\chi + \tau_{zx}^*(x, y, z) + i\tau_{zy}^*(x, y, z)$$

with

$$\zeta = z\left(\frac{\cos^2\chi}{a^2} + \frac{\sin^2\chi}{b^2}\right)^{1/2}.$$

$$I_6 = \frac{\zeta a^2 z\sqrt{u}}{\left(u^2 + a^2\zeta^2\right)} \cdot \frac{\left(\cos^2\chi, \sin^2\chi\right)}{\left(\frac{\cos^2\chi}{a^2} + \frac{\sin^2\chi}{b^2}\right)^{1/2}}.$$

The starred quantities can be obtained (see [9]).

In obtaining the values of the stress and deformations for various indenters, we observe that one needs to compute integrals of the type

$$Z_n^m = C_n^m - i S_n^m,$$

where

$$\left(C_n^m,\ S_n^m\right) = \int_0^\infty k^{n-2}(\cos k, \sin k)\, e^{-k\zeta}\ J_m(k\rho)\, dk = (\mathrm{Re}, \mathrm{Im})\int_0^\infty k^{n-1}\, e^{-(\zeta+i)k}\ J_m(k\rho)\, dk,$$

where $\zeta = z\left(\frac{\cos^2\chi}{a^2} + \frac{\sin^2\chi}{b^2}\right)^{1/2}$, $\rho = \left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)^{1/2}$.

Using the recurrence relation for the Bessel function, any value of Z_n^m for $(m, n) > 2$ can be obtained in terms of the values for $(n.m) = (0, 1, 2)$. From the relation of the type

$$Z_{n-1}^m = \frac{\rho}{2m}\left[Z_n^{m-1} + Z_n^{m+1}\right].$$

An integration over is present for the elliptic case.

In particular,

$$C_1^0 - iS_1^0 = \int_0^\infty \frac{1}{k}e^{-k(\zeta_1+i)} J_0(k\rho)\,dk = -\log\left[i + \zeta + \sqrt{(\zeta+i)^2 + \rho^2}\right].$$

The integrals C_n^m, S_n^m were first evaluated by Elliot [10]. Sneddon gave the values of the integral which has been tabulated in convenient form in Appendix 1 by Murgis [11] in his book. Fabrikant [12, 13] gave the values in terms of elementary functions for circle. Roy and Basu [8] listed those values for elliptic contact problem. For Hertz problem, the values are

$$\begin{aligned}
I_1 &= S_0^0 - C_1^0 = \zeta\,\sin^{-1}\left(\frac{a}{l_2}\right) - \frac{\zeta}{\sqrt{u}}\\
I_2 &= -\int I_1 dz = S_{-1}^0 - C_0^0 = \frac{1}{4}\\
&\quad\left[-\left(l_2^2 - a^2\right)^{1/2} + 3\frac{\zeta^2}{\left(l_2^2 - a^2\right)^{1/2}} - \left(2\zeta^2 - \rho^2 + 2\right)\sin^{-1}\left(\frac{a}{l_2}\right)\right]\\
I_3 &= S_1^0 - C_2^0 = \sin^{-1}\left(\frac{a}{l_2}\right) - \frac{\sqrt{u}}{s}\\
I_4 &= S_0^1 - C_1^1 = -\frac{\rho}{2}\left(\frac{\sqrt{u}}{a^2+u} - \sin^{-1}\left(\frac{a}{l_2}\right)\right)\\
I_6 &= S_1^1 - C_2^1 = \frac{\zeta}{\sqrt{u}}\cdot\frac{\rho}{s\left(a^2+u\right)},
\end{aligned}$$

where $I_j(\chi)$ are given by

$$(l_1, l_2) = \frac{1}{2}\left[\left((\rho+a)^2 + \zeta^2\right)^{1/2} \mp \left((\rho-a)^2 + \zeta^2\right)^{1/2}\right].$$

$$\rho = \left(x^2 + \frac{a^2}{b^2}y^2\right)^{1/2} \qquad \zeta = a\left(\cos^2\chi + \frac{a^2}{b^2}\sin^2\chi\right)^{1/2}.$$

The corresponding quantities for elliptic punch under constant indentation are L_j where

$$\begin{array}{ll} L_1 = S_2^0 & L_2 = \sin^{-1}(l_2/a) \\ L_3 = S_3^0 & L_4 = S_1^1 \\ L_5 = S_0^1 & L_6 S_3^1. \end{array}$$

Since

$$Z_{n-1}^m = \frac{\rho}{2m}\left[Z_n^{m-1} + Z_n^{m+1}\right]$$

all the values of Z_n^m for all successive values of n and m can be generated from Z_1^0, Z_1^{-1}, Z_2^0, Z_2^2. The appropriate modifications in the elliptic contact problem with constant indentation have been considered by Roy and Basu [8]. The reader is referred to the paper for details.

We note that the starred quantities τ_{zz}^*, τ_{zx}^*, etc., are the effect of the contact area being elliptic and can be easily calculated [8]. In case if $a = b$, i.e. if the contact area is circular, all the starred quantities vanish. The integration over is now elementary and we get back the stresses for the indentation by a sphere (Johnson [14]).

For $b \to \infty$, our result agrees with that of an indentation by a line crack (see [15]).

We have considered only normal loading. The analysis can be easily extended to the case of tangential loading (see Roy and Basu [8]).

10.3 Discussion

We now mention some applications of the result discussed by various researchers.

The stresses beneath contact plane can be used to predict the region of failure of the body under indentation load. The stresses and deflections arising from the contact between two elastic bodies have practical applications in hardness testing, wear and impact damage of engineering ceramic.

The study of the contact stresses in line contact can be used to study the failure of the rail during the motion of a locomotive wheel on rails of straight or curved railway. In this case, besides the weight of the locomotive acting normal to the rail, the frictional force arising from the use of brakes applied on the wheel gives rise to tangential load. In the case of the rail on curved track, frictional forces arise because of wheel slippage, since the wheels are rigidly attached to the axle and the tangential force is the thrust on the outside rail of a curve due to centrifugal forces on the train and is larger than on straight rail. Thus the rail failures are more severe on curved rails

Failure from contact stresses starts as a localized inelastic deformation (yielding or distortion) and by fracture by progressive spreading of a crack.

Two failure criterions are commonly used in literature. One is the Tresca failure criterion, namely, the maximum value of

$$\tau_{\max} = \frac{1}{2}\left|\sigma_1 - \sigma_2\right|$$

σ_1, σ_2 being the maximum and minimum principal stresses.

The other is the Von Mises criterion which depends on the deviatoric stresses

$$\tau_{G\max} = \frac{1}{3}\sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_1 - \sigma_3)^2}.$$

Various authors have drawn contours of principal stresses for line, punch and spherical indentation to get information about the failure region.

Smith and Liu [16] drew the contours of principal stresses on $y = 0$. From a study of the plot of the principal stresses in the xz-plane, they observed that the maximum values of the principal stresses occur at the surface at $(0.3a, 0)$

$$\sigma_{1\max} = -1.39p_0, \quad \sigma_{2\max} = -0.72p_0, \quad \sigma_{3\max} = -0.53p_0$$

so that $\tau_{\max} = 0.43p_0$ and $\tau_{G\max} = 0.37p_0$.

When only normal force acts $f = 0$,

$$\tau_{\max} = 0.30p_0 \text{ and } \tau_{G\max} = 0.27p_0.$$

These values are attained on the z-axis at $z = 0.78a$ underneath.

Thus the location and magnitude changes due to the presence of tangential stress and moves to the surface. For $f > 1/3$, the maximum shear stress occurs at a point on the surface, while if $f < 1/3$, this stress is underneath the surface.

Hamilton and Goodman [17] plotted the lines of constant $J_2^{1/2}/p$ for various values of friction μ observed that for circular sliding contact that the region of maximum yield parameter moves towards the surface less rapidly. And the most likely region of failure is the front edge of the circle of contact.

The second type of failure is associated with repeated applications of the loads with a fracture (fatigue) that starts as a localized crack with very little visual evidence of inelastic evidence. The crack starts either at the surface or underneath the surface and grows progressively as the stress is repeated until some of the metal breaks out of the surface causing pitting, shelling or other damaging effects to the surface.

Type characterized by inelastic strain in the surface layer on top of the rail head and shelly failures of railway rails usually occur in the outside rail on the rail-head edge in contact with the wheel flange. After repeated movement, cracks in the rail head appear and progress either horizontally or transversely across the rail head. The above is also valid for shelly failures in the rims of the wheels on diesel locomotives due to the combined action of the normal weight and tangential forces due to the driving torque applied through these wheels.

Poritsky [15] used the stresses in line contact problem for normal as well as tangential loading and applied it to study contact between gear teeth and also rolling motion of rails.

He noted that for cylinder in rolling contact, slipping takes place in part of the contact region where the friction μ is critical so that $T = \mu P$, while for $T < \mu P$ the two surfaces are locked in $b < |x| < a$, where b is related to the tangential stress T', given by

$$T' = \mu P' \left(1 - \frac{b^2}{a^2}\right).$$

Poritsky describes the method of computing the creep in the locked region resulting from the difference in strain between two surfaces and gave the creep rate as a ratio of the rotation rate of the railway wheels, namely

$$1 + |\Delta e_{xx}| = \frac{2\pi R \times \text{number of wheel rotations}}{\text{length of track covered}}.$$

We note that for punch problem, the vertical stress tends to infinity as the rim of the circular contact is approached. For Hertzian case, the radial stress σ_r is compressive inside the contact circle. Outside the circle, it is tensile with a maximum value on the edge of the Hertz circle. For brittle material, it is responsible for the initiation of the Hertzian cone cracks penetrating below the surface.

Acknowledgments This paper is the abridged version of a paper to be communicated soon.

References

1. J. Boussinesq (Gauthier-Villars, Paris, 1885)
2. M.T. Huber, Ann. Phys. **43**, 153–163 (1904)
3. M.T. Huber, S. Fuchs, Phys. Z. 1282–1285 (1913)
4. I.N. Sneddon, Proc. Glasgow Math. Assoc. **4**, 108 (1969)
5. I.N. Sneddon, *The Use of Integral Transform* (McGraw Hill, New York, 1972)
6. I.N. Sneddon, Proc. Camb. Phil. Soc. **42**, 29 (1946)
7. I.N. Sneddon, Proc. R. Soc. A **287**, 229 (1946)
8. K.L. Johnson, J. Appl. Mech. **25**, 260 (1956)
9. A. Roy, U. Basu, Z.A.M.M, **91**, 544–564 (2011)
10. H.A. Elliot, N.F. Mott, Proc. Camb. Phil. Soc. **44**, 522 (1948)
11. D. Maugis, *Contact Adhesion and Rupture of Solids* (Springer, Berlin, 2000)
12. V.I. Fabricant, J. Appl. Mech. **55**, 604 (1988)
13. V.I. Fabricant, Adv. Appl. Math. **27**, 153 (1990)
14. K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, 1985)
15. H. Poritsky, J. Appl. Mech. **72**, 191 (1950)
16. J.C. Smith, C.K. Liu, J. Appl. Mech. **21**, 156 (1953)
17. D.M. Hamilton, L.E. Goodman, J. Appl. Mech. **33**, 371 (1966)
18. H. Hertz, J. Reine Angew. Math. **92**, 156 (1881)

Chapter 11
Stochastic Analysis and Bounds on Noise for a Holling Type-II Model

Gaurav Pachpute and Siddhartha P. Chakrabarty

Abstract A deterministic predator–prey model with Holling type II functional response is modified to a stochastic one by incorporating multiplicative Gaussian noise about the interior equilibrium point of the deterministic model. A Lyapunov function is constructed so as to analyze the stability of the stochastic model. Bounds on the intensities of environmental fluctuations are derived. Low fluctuations in one population allows, up to a certain limit, for a higher fluctuation in the other population. The bounds on fluctuation in predator density attain a maximum at an intermediate value in the allowable range for deterministic model parameters and disappear at the boundaries of this range, demonstrating a tradeoff between upward and downward fluctuations in predation. The results are illustrated through simulations.

Keywords Holling type II · Stochastic model · Environmental fluctuations · Bounds

Mathematics Subject Classification: 92D25

11.1 Introduction

The classical model of predator–prey interaction due to Lotka and Volterra [1] was greatly advanced by the pioneering work of Holling [2–4]. Holling's analysis of the rate of consumption of prey by the predator was based on experimental studies. Holling [3] proposed three different kinds of functional response [1, 5, 6] to encapsulate this rate of consumption by the predator. The type I functional response is applicable in cases where the consumption of the prey by the predator is a linear function of the prey up to a point where the consumption rate does not change irrespective of availability of more prey [1, 6]. The type II functional response comprises a hyperbolic function that saturates as a result of the time required by the predator to handle (capture and consume) the prey [1, 6]. Finally, type III functional

G. Pachpute · S.P. Chakrabarty (✉)
Department of Mathematics, Indian Institute of Technology Guwahati,
Guwahati 781039, Assam, India
e-mail: pratim@iitg.ernet.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_11

response is a sigmoidal curve and is typical of cases where the consumption rate is low, below a certain density threshold level of the prey population. However, this rate increases as the prey density goes above this threshold level, followed by an eventual saturation level being attained [1, 6]. Hassell et al. [7] suggested that the type III functional response may be more common in case of invertebrate predators as opposed to the widely accepted notion that type II response is more common. Crawley [1, 8] proposed a fourth kind of functional response.

The capture and consumption rate of the prey would typically be expected to increase as the prey density increases. This rate of increase, however, is likely to decrease with time and eventually reach a saturation level, due to several factors such as handling time of the prey [1]. This is manifested in terms of a Holling type II functional response of the form $f(N) = cN/(a+N)$, where N is the density of the prey population. Real [6] presents an interesting review of the ecological motivation of the three types of functional response due to Holling as well as their mathematical formulation.

Abrams [9] presented an adaptive variation into the disk equation. He argued about the likelihood of violation of the assumption in the disk equation which forms the basis of the mathematical formulation of the Holling type functional response, and presented the variations in the functional forms as a consequence of such violations. Oaten and Murdoch [10] examined the effects that predators can have on the stability of the prey population in an environment. Berryman [11] in his article discussed the origins and subsequent development of predator–prey theory, starting with the original theoretical analysis of population dynamics due to Malthaus and Verhulst. He dwelled upon the classical model of Lotka–Volterra, the Holling type models, and their variations as well as the ratio-dependent model (a modification of the Holling type II model) proposed by Arditi and Ginzburg [12]. Sugie et al. [13] derived the necessary and sufficient condition for uniqueness of limit cycle, for a predator–prey model with a general functional response of Holling type. Global stability analysis for predator–prey system with Holling type functional response was carried out by Hsu and Huang [14]. Ruan and Xiao [15] carried out the global analysis for a predator–prey system with the functional response being nonmonotonic in nature. Extensive analysis of models with Holling type functional response has been carried out by several researchers [16–18]. Zhang and Chen [16] analyzed a food chain model with a Holling type II functional response in the presence of impulsive perturbations. A Holling type model with delay is investigated in [17], whereas a reaction–diffusion predator–prey model with Holling type II functional response is examined in [18]. Srinivasu et al. [19] considered the standard predator–prey model with Holling type II functional response and incorporated the supply of additional food to the predator population. They analyzed the consequences of this addition and concluded that the handling time for the food is crucial in determining the eventual evolution of the system. Bandyopadhyay and Chakrabarti [20] presented a deterministic and stochastic analysis of a nonlinear predator–prey system, with the stochastic analysis being accomplished by the method of statistical linearization. The stochastic analysis by Maiti and Samanta [21] on a similar model involved the method of spectral density analysis.

The paper is organized as follows. In Sect. 11.2, we discuss the deterministic model. In Sect. 11.3, we present the stochastic model by incorporating multiplicative Gaussian noise. We analyze the model and obtain moving upper bounds on the noise in the context of the stability of the system. Finally, in Sect. 11.4, we discuss the implications on the evolution of the stochastic system.

11.2 Deterministic Model

We consider the following classical predator–prey model with a Holling type II functional response [1]:

$$\frac{dN}{dT} = rN\left(1 - \frac{N}{K}\right) - \frac{cNP}{a+N} \tag{11.1}$$

$$\frac{dP}{dT} = \frac{bNP}{a+N} - mP \tag{11.2}$$

Here, $N(T)$ and $P(T)$ denote the prey and predator densities, respectively, at time T. The intrinsic growth rate for the prey populations is r and the carrying capacity is K. The capture rate of predators and the conversion rate of prey into predator biomass are denoted by c and b, respectively, with a as the half-saturation constant. Finally, the predator population is assumed to have a natural death rate of m. After introducing the new variables [1, 19] $x = \frac{N}{a}$, $y = \frac{cP}{ar}$ and $t = rT$, the system (11.1) and (11.2) reduces to

$$\frac{dx}{dt} = x\left(1 - \frac{x}{\gamma}\right) - \frac{xy}{1+x} \tag{11.3}$$

$$\frac{dy}{dt} = \frac{\beta xy}{1+x} - \delta y, \tag{11.4}$$

where $\gamma = \frac{K}{a}$, $\beta = \frac{b}{r}$ and $\delta = \frac{m}{r}$. The reduced system (11.3) and (11.4) admits three equilibrium points where the prey and predator null-isoclines intersect. They are

- $E_0(0, 0)$ (trivial).
- $E_a(\gamma, 0)$ (axial).
- $E_i(x^*, y^*)$ (interior or co-existing), where $x^* = \dfrac{\delta}{\beta - \delta}$ and $y^* = \left(1 - \dfrac{\delta}{\gamma(\beta - \delta)}\right)\dfrac{\beta}{\beta - \delta}$.

The trivial equilibrium point E_0 is a saddle point. The axial equilibrium point E_a is stable if $\gamma < \dfrac{\delta}{\beta - \delta}$ and a saddle point otherwise. The interior equilibrium point E_i exists if $\beta > \delta$ and $\gamma > \dfrac{\delta}{\beta - \delta}$, and is stable if $\gamma < \dfrac{\beta + \delta}{\beta - \delta}$. Also, the system exhibits a

Hopf bifurcation at $\gamma = \dfrac{\delta}{\beta - \delta}$ (see [20, 21] for details). On the other hand, it is stable on the right of the peak of the prey null-isocline, i.e., $x^* > \dfrac{\gamma - 1}{2}$ or $\gamma < \dfrac{\beta + \delta}{\beta - \delta}$. The system has a limit cycle on the left of the peak.

11.3 Stochastic Model

The deterministic nature of the parameters and hence that of the model is inconsistent with most natural phenomena, which tend to fluctuate about some average value [1, 22–24]. Under realistic conditions, model parameters exhibit variations, which cannot be captured by a deterministic model alone. One of the ways to mathematically incorporate such variations in a system of interacting species is to vary the model parameters about some average value. More specifically, a parameter p in a deterministic model can be replaced by $p_0 + \zeta p_1(t)$, where p_0 represents the average value about which the parameter p fluctuates, $p_1(t)$ is the noise function, and ζ is the intensity of noise.

The equilibrium points of the system are the combinations of model parameters and consequently the equilibria of the system also fluctuate about some average value. In this paper, we incorporate the environmental fluctuations by varying the population density about an equilibrium as opposed to adding stochastic perturbations to the parameters. This is accomplished by adding multiplicative noise to each equation. It is important to note that, while deterministic equilibrium points and their respective stabilities are invariant in time, with added fluctuations this invariance is lost. For a stochastic system, one must study the probabilistic stability of these points through the equilibrium probability distribution. The deterministic part of the model is responsible for the restitution of the population densities to their average values from the diffusion caused by the random fluctuations. Systems exhibiting this characteristic are called stochastically stable within fluctuating environment [23].

The physical origin of stochasticity in biological and ecological models can be attributed to a wide variety of factors. Schaffer et al. [25] study the effects of random perturbations caused by weather fluctuations and outbreaks of epidemics. Beretta et al. [26] incorporate the environmental fluctuations caused by epidemic resulting from the spread of infectious diseases in a population model system. Epidemic induced by virulent phages on bacteria in a marine environment [27] and spontaneous regression and progression of a malignant tumor system [28] are modeled and analyzed using stochastic perturbations. We introduce stochastic perturbations about the interior equilibrium point (x^*, y^*) in the Holling type II model (11.3) and (11.4) to incorporate environmental fluctuations. The source of these fluctuations in our model could be the outbreak of an epidemic or a result of vagaries in weather. These perturbations are assumed to be Gaussian white noise and proportional to the distances of $x(t)$ and $y(t)$ from the equilibrium values (x^* and y^*) [23, 27, 28].

The deterministic model (11.3) and (11.4) is converted into a stochastic differential equations as given below:

$$dx = \left(x\left(1-\frac{x}{\gamma}\right) - \frac{xy}{1+x}\right)dt + \sigma_1(x-x^*)d\xi^1(t) \tag{11.5}$$

$$dy = \left(\frac{\beta xy}{1+x} - \delta y\right)dt + \sigma_2(y-y^*)d\xi^2(t), \tag{11.6}$$

where $\xi^i(t)$ for $i = 1, 2$ represent independent Wiener processes with $\sigma_1, \sigma_2 > 0$ being the stochastic intensity. The system (11.5) and (11.6) can also be represented in the following form:

$$dX(t) = F(X(t))dt + g(X(t))d\xi(t), \tag{11.7}$$

where

$$X(t) = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(X(t)) = \begin{pmatrix} x\left(1-\dfrac{x}{\gamma}\right) - \dfrac{xy}{1+x} \\ \dfrac{\beta xy}{1+x} - \delta y \end{pmatrix}, \quad \xi(t) = \begin{pmatrix} \xi^1(t) \\ \xi^2(t) \end{pmatrix} \quad \text{and}$$

$$g(X(t)) = \begin{pmatrix} \sigma_1(x-x^*) & 0 \\ 0 & \sigma_2(y-y^*) \end{pmatrix}. \tag{11.8}$$

The function $F(X(t))$, known as the drift coefficient, represents the continuous deterministic part of the system [23]. Also, $g(X(t))$ represents the random component of the system and is called the diffusion matrix. The vector $\xi(t)$ is the multi-dimensional Wiener process, the components ($\xi^i(t)$ for $i = 1, 2$) of which are independent of each other. Due to the dependence of the function g on the state variables $X(t)$ in the equation, the system (11.7) is said to have multiplicative noise. The interior equilibrium point E_i of the deterministic system (11.3) and (11.4) is also an equilibrium point for the stochastic model (11.5) and (11.6). We study the local characteristics by linearizing the deterministic part $F(X(t))$ around this point. Introducing new variables, $u_1 = x - x^*$ and $u_2 = y - y^*$, we rewrite Eq. (11.7),

$$dU(t) = f(U(t))dt + g(U(t))d\xi(t), \tag{11.9}$$

where

$$U(t) = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{11.10}$$

$$f(U(t)) = \begin{pmatrix} \left(1 - \dfrac{2x^*}{\gamma} - \dfrac{y^*}{1+x^*} + \dfrac{x^*y^*}{(1+x^*)^2}\right)u_1 - \dfrac{x^*}{1+x^*}u_2 \\ \left(\dfrac{\beta y^*}{1+x^*} - \dfrac{\beta x^* y^*}{(1+x^*)^2}\right)u_1 + \left(\dfrac{\beta x^*}{1+x^*} - \delta\right)u_2 \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{\delta}{\beta}\left(1-\dfrac{\beta+\delta}{\gamma(\beta-\delta)}\right)u_1-\dfrac{\delta}{\beta}u_2 \\ \\ \left(\beta-\delta-\dfrac{\delta}{\gamma}\right)u_1 \end{pmatrix} \tag{11.11}$$

$$\text{and} \quad g(U(t)) = \begin{pmatrix} \sigma_1 u_1 & 0 \\ 0 & \sigma_2 u_2 \end{pmatrix}. \tag{11.12}$$

The trivial equilibrium $U(t) = (0,0)^\top$ of the above system corresponds to the equilibrium E_i of the system (11.7). By defining the constants

$$P = \frac{\delta}{\beta}\left(1-\frac{\beta+\delta}{\gamma(\beta-\delta)}\right) < 0. \quad Q = \beta-\delta-\frac{\delta}{\gamma} > 0 \text{ and } R = \frac{\delta}{\beta} > 0,$$

we can write the expression for $f(U(t))$ in the following simplified form:

$$f(U(t)) = \begin{pmatrix} Pu_1 - Ru_2 \\ Qu_1 \end{pmatrix} \tag{11.13}$$

The following theorem from Afanas'ev et al. [29] (as in [27, 28]) enables us to study the mean square stability of the system given in Eq. (11.9).

Theorem 11.1 *Let* $\mathbf{D} = (t \geq t_0) \times \mathbb{R}^2, t_0 \in \mathbb{R}^+$ *and suppose* $V(t,U) \in C_2^0(\mathbf{D})$ *is a twice continuously differentiable function with respect to U and a continuous function of t, satisfying the inequalities*

$$K_1|U|^p \leq V(t,U) \leq K_2|U|^p, \tag{11.14}$$

$$LV(t,U) \leq -K_3|U|^P, \tag{11.15}$$

where $p > 0$ *and* $K_i > 0$ *for* $i = 1, 2, 3$. *Then the trivial equilibrium of the system (11.9) is exponentially p-stable over* $t \geq t_0$. *The special case,* $p = 2$, *refers to the exponential mean square stability of the system.*

We now define the Lyapunov function

$$V(t,U) = \frac{1}{2}(\omega_1 u_1^2 + 2\omega_2 u_1 u_2 + u_2^2), \tag{11.16}$$

where ω_1 and ω_2 are the positive constants to be defined later. The condition (11.14) for $p = 2$ in Theorem 11.1 is satisfied for this Lyapunov function, provided

$$\omega_2^2 - \omega_1 < 0. \tag{11.17}$$

The expression for $LV(t, U)$ is defined by

$$\begin{aligned} LV(t,U) &= \frac{\partial V(t,U)}{\partial t} + f^{\top}(U(t))\frac{\partial V(t,U)}{\partial U} \\ &\quad + \frac{1}{2}Tr\left(g^{\top}(U(t))\frac{\partial^2 V(t,U)}{\partial U^2}g(U(t))\right), \end{aligned} \tag{11.18}$$

where the partial derivatives of $V(t, U)$ with respect to U are defined as follows:

$$\frac{\partial V(t,U)}{\partial U} = \begin{pmatrix} \dfrac{\partial V}{\partial u_1} \\ \\ \dfrac{\partial V}{\partial u_2} \end{pmatrix} \quad \text{and} \quad \frac{\partial^2 V(t,U)}{\partial U^2} = \begin{pmatrix} \dfrac{\partial^2 V}{\partial u_1^2} & \dfrac{\partial^2 V}{\partial u_1 u_2} \\ \\ \dfrac{\partial^2 V}{\partial u_2 u_1} & \dfrac{\partial^2 V}{\partial u_2^2} \end{pmatrix} \tag{11.19}$$

Using Eq. (11.16), we obtain the following expressions:

$$\frac{\partial V(t,U)}{\partial U} = \begin{pmatrix} \omega_1 u_1 + \omega_2 u_2 \\ \omega_2 u_1 + u_2 \end{pmatrix} \quad \text{and} \quad \frac{\partial^2 V(t,U)}{\partial U^2} = \begin{pmatrix} \omega_1 & \omega_2 \\ \omega_2 & 1 \end{pmatrix}. \tag{11.20}$$

Substituting $f(U(t))$ and $g(U(t))$ in (11.18) and using (11.20), we get the following expression for $LV(t, U)$:

$$LV(t,U) = \left(\left(P + \frac{\sigma_1^2}{2}\right)\omega_1 + Q\omega_2\right)u_1^2 + (P\omega_2 - R\omega_1 + Q)u_1u_2 + \left(\frac{\sigma_2^2}{2} - R\omega_2\right)u_2^2. \tag{11.21}$$

We choose $\omega_2 = \frac{R\omega_1 - Q}{P}$, which is positive provided $\omega_1 < \frac{Q}{R}$. Let $\omega_1 = \frac{Q}{nR} > 0$, where $n > 1$. Under these conditions, we can rewrite $LV(t, U)$ as

$$\begin{aligned} LV(t,U) &= \left(\left(P + \frac{\sigma_1^2}{2}\right)\omega_1 + Q\omega_2\right)u_1^2 + \left(\frac{\sigma_2^2}{2} - R\omega_2\right)u_2^2. \\ &= -U^{\top}MU, \end{aligned} \tag{11.22}$$

where

$$M = \begin{pmatrix} -\left(P + \dfrac{\sigma_1^2}{2}\right)\omega_1 - Q\omega_2 & 0 \\ \\ 0 & -\dfrac{\sigma_2^2}{2} + R\omega_2 \end{pmatrix}. \tag{11.23}$$

In order for the second-order square matrix M to be positive definite, it must have positive real eigenvalues, i.e., the two eigenvalues $\lambda_1 = -\left(P + \frac{\sigma_1^2}{2}\right)\omega_1 - Q\omega_2$ and $\lambda_2 = -\frac{\sigma_2^2}{2} + R\omega_2$, both must be positive real numbers. Substituting the values of ω_1 and ω_2 and imposing positivity of σ_i^2 for $i = 1, 2$, we get the following condition on n:

$$1 < n < 1 + \frac{P^2}{QR}. \tag{11.24}$$

For this range of n and the chosen values of ω_1 and ω_2, the condition stated in (11.17) is satisfied. This gives us the bounds on the intensities of environmental fluctuations:

$$\begin{aligned} \sigma_1^2 &< 2\left[\frac{(n-1)QR}{P} - P\right] \ (> 0) \quad \text{and} \\ \sigma_2^2 &< -\frac{2(n-1)QR}{nP} \ \ (> 0). \end{aligned} \tag{11.25}$$

If $\lambda_m = \min\{\lambda_1, \lambda_2\}$, then for the real symmetric matrix M, we obtain the following result:

$$LV(t, U) \leq -\lambda_m |U|^2. \tag{11.26}$$

This completes the proof for the stability of the system described in Eq. (11.9).

Thus, (11.24) and (11.25), along with the necessary condition for the local stability of the deterministic system (11.3) and (11.4), i.e., $\frac{\delta}{\beta-\delta} < \gamma < \frac{\beta+\delta}{\beta-\delta}$, form the set of necessary conditions for the stochastic stability of the system. The bounds on environmental fluctuations given in Eq. (11.25) are functions of a variable n, with a particular domain (Eq. 11.24). If the bound on one of the fluctuations is known, then one can determine the stability of the system and the maximum bound on the other fluctuation, i.e., the maximum bounds on σ_1 and σ_2, say, $\sigma_{1,\max}$ and $\sigma_{2,\max}$ lie on the following curve in the first quadrant:

$$\sigma_{1,\max}^2 + \frac{2\sigma_{2,\max}^2 QR}{2QR + \sigma_{2,\max}^2 P} + 2P = 0. \tag{11.27}$$

11.4 Discussion

The analysis and biological implications of the deterministic model have been extensively studied in [1, 19]. The exposition of the corresponding stochastic model is presented in this section.

For the interior equilibrium point to be stable, the parameter γ in the Holling Type II model (11.3) and (11.4) must lie between $\delta/(\beta-\delta)$ and $(\beta+\delta)/(\beta-\delta)$. For notational convenience, we define $\gamma_{\text{min}} = \delta/(\beta-\delta)$ and $\gamma_{\text{max}} = (\beta+\delta)/(\beta-\delta)$. Both γ_{min} and γ_{max} are functions of the ratio β/δ. We can thus write γ_{max} in terms of γ_{min} as $\gamma_{\text{max}} = 2\gamma_{\text{min}} + 1$. Furthermore, there is an inverse relation between γ_{min} (equivalently, γ_{max}) and the ratio β/δ.

The first bound (11.24) derived in the previous section is on the values of n. The maximum possible value of n for stochastic stability, in terms of γ_{min} and γ_{max}, is

$$n_{\text{max}} = 1 + \frac{\gamma_{\text{min}}^2(\gamma_{\text{max}}-\gamma)^2}{\gamma\delta(\gamma-\gamma_{\text{min}})(1+\gamma_{\text{min}})}.$$

For a given γ and constants γ_{min}, γ_{max} (consequently constant β/δ), smaller values of δ (or, equivalently, smaller β) result in greater ranges for n. The maximum possible value for n is attained as $\gamma \to \gamma_{\text{min}}$ and symmetrically, the minimum is attained as $\gamma \to \gamma_{\text{max}}$.

The upper bound on σ_1^2 can be written as

$$2\left[\frac{\gamma_{\text{min}}(\gamma_{\text{max}}-\gamma)}{\gamma(1+\gamma_{\text{min}})} - \frac{\delta(\gamma-\gamma_{\text{min}})(n-1)}{\gamma_{\text{min}}(\gamma_{\text{max}}-\gamma)}\right].$$

This bound attains a maximum, for a particular value of n, as $\gamma \to \gamma_{\text{min}}$, with the corresponding maximum value of 2 (i.e., $\sigma_{1,\text{max}} = \sqrt{2}$). For a fixed n, the upper bound decreases as γ increases and reduces to 0 as $\gamma \to \gamma_{\text{max}}$. On the other hand, for a particular γ, $\sigma_{1,\text{max}}$ is inversely related to n. The maximum is realized as $n \to 1$ and $\sigma_{1,\text{max}}$ reduces to 0 as n reaches its maximum. The limiting nature of the maximum of $\sigma_{1,\text{max}}$ in γ (i.e., $\gamma \to \gamma_{\text{min}}$) is coincident with the extinction of the predators ($y^* = 0$). An increase in γ increases the equilibrium population density (y^*) of the predators. This, expectedly, reduces the sustainability of the prey population. Furthermore, the maximum with respect to n is inversely related to the ratio β/δ. This is consistent with the fact that lower predation and higher mortality in predators lead to a higher sustainability in preys.

Likewise, the maximum bound on σ_2^2 can be written as

$$\frac{2\delta(\gamma-\gamma_{\text{min}})(n-1)}{n\gamma_{\text{min}}(\gamma_{\text{max}}-\gamma)}.$$

Evidently, the maximum of this bound is correlated with the magnitude of n, i.e., a larger n leads to a larger maximum. The bound on σ_2^2 can now be restated as $\left(0, \frac{2\gamma_{\text{min}}(\gamma_{\text{max}}-\gamma)}{n_{\text{max}}\gamma(1+\gamma_{\text{min}})}\right)$, where n_{max} is the maximum value of n, as given earlier in this section. Interestingly, the maximum for the upper bound, $\sigma_{2,\text{max}}$, is not attained at either γ_{min} or γ_{max}, but at a point in between the two values. The maximum bound becomes 0 for $\gamma \to \gamma_{\text{min}}$ since this corresponds to extinction, whereas, when $\gamma \to \gamma_{\text{max}}$, the deterministic system loses its stability and so does the stochastic system. The value for σ_2^2 can be written as

$$\sigma_2^2 = \frac{2\gamma_{\min}(\gamma_{\max}-\gamma)(\gamma-\gamma_{\min})\delta}{\gamma\delta(\gamma-\gamma_{\min})(1+\gamma_{\min})+\gamma_{\min}^2(\gamma_{\max}-\gamma)^2}.$$

Differentiating this with respect to γ and setting it equal to zero, we get,

$$\begin{aligned}(\gamma_{\max}+\gamma_{\min}-2\gamma)&\left(\gamma\delta(\gamma-\gamma_{\min})(1+\gamma_{\min})+\gamma_{\min}^2(\gamma_{\max}-\gamma)^2\right)\\ =&\left((2\gamma-\gamma_{\min})\delta(1+\gamma_{\min})-2\gamma_{\min}^2(\gamma_{\max}-\gamma)\right)(\gamma_{\max}-\gamma)(\gamma-\gamma_{\min}).\end{aligned}$$

Simplifying the equation and using the identity $\gamma_{\max} = 2\gamma_{\min}+1$, we obtain

$$\begin{aligned}&\gamma_{\min}^2(1+\gamma_{\min})(\gamma_{\max}-\gamma)^2 = \gamma_{\max}\delta(1+\gamma_{\min})(\gamma-\gamma_{\min})^2\\ \implies\quad &\gamma_0 = \frac{\gamma_{\min}(\gamma_{\max}+\sqrt{\delta\gamma_{\max}})}{\gamma_{\min}+\sqrt{\delta\gamma_{\max}}}.\end{aligned}$$

Thus the maximum for the bound with respect to γ occurs at

$$\gamma = \gamma_0 = \frac{\gamma_{\min}(\gamma_{\max}+\sqrt{\delta\gamma_{\max}})}{\gamma_{\min}+\sqrt{\delta\gamma_{\max}}}$$

and for γ_0 at

$$n \to n_{\max} = 1 + \frac{\sqrt{\gamma_{\max}}}{\sqrt{\gamma_{\max}}+\sqrt{\delta}}$$

Thus, we have

$$\sigma_{2,\max}^2 = \frac{2\sqrt{\delta}}{2\sqrt{\gamma_{\max}}+\sqrt{\delta}}.$$

Thus, for a general value of n, the maximum bound $\sigma_{2,\max}$ increases as γ increases. On the contrary, $\sigma_{1,\max}$ decreases as γ increases. One expects the maximum bound for fluctuations in one population density to increase at the expense of the other bound. However, increasing γ leads to a smaller value of $n_{\max}$. As a consequence, the expected relative behavior between the two bounds ($\sigma_{1,\max}$ and $\sigma_{2,\max}$) is not always realized. This is explored further in the discussion that follows. The preceding discussion reflects that the bounds obtained, with the exception of $\sigma_{2,\max}$, are interlinked in a way such that the maximization of one bound nullifies at least one of the other bounds.

Two simulations for the model are presented in Figs. 11.1 and 11.2. While the former (Fig. 11.1) illustrates the trajectory of prey–predator population densities under stable conditions, the latter (Fig. 11.2) is under unstable conditions. The parameters

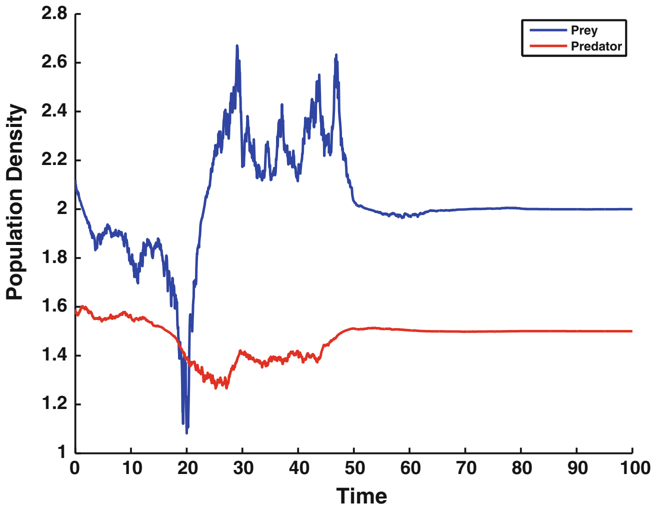


Fig. 11.1 Simulation for stable dynamics

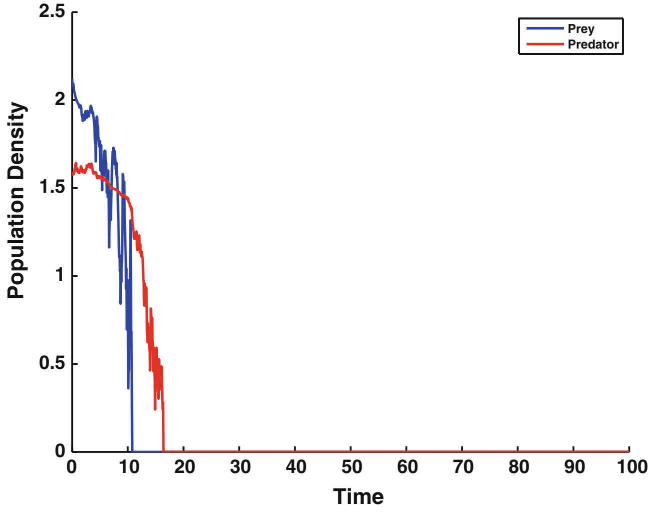


Fig. 11.2 Simulation for unstable dynamics

β and δ were chosen to be 0.3 and 0.2, respectively. Consequently, the stability is preserved for $\gamma \in (\delta/(\beta - \delta), (\beta + \delta)/(\beta - \delta)) = (2, 5)$, and $\gamma_0 = 4.0$. For the simulation, we chose $\gamma = \gamma_0$. The value of $n = 17/12$ was chosen to be the average of the allowed range $(1, 11/6)$. Based on these values, the maximum bounds on σ_1 and σ_2 are 0.4082 and 0.3430, respectively. Accordingly, for the simulation of stable dynamics, we chose $\sigma_1 = 0.4$ and $\sigma_2 = 0.2$, and for unstable dynamics simulation, these were $\sigma_1 = 0.8$ and $\sigma_2 = 0.4$. The simulations were run for a time window of 100. In Fig. 11.1, the densities of the two populations converge toward the interior equilibrium point $(x^*, y^*) = (2, 3/2)$. However, in case of unstable dynamics, population densities drop to zero (i.e., extinction) due to large fluctuations in the population densities, as seen in Fig. 11.2.

As noted before, the maximum bounds $\sigma_{1,\max}$ and $\sigma_{2,\max}$ lie on the curve (11.27). The preceding discussion suggests that as γ increases from $\gamma_{\min}$ to $\gamma_{\max}$, the maximum bound $\sigma_{1,\max}$ decreases from $\sqrt{2}$ to 0. The bound $\sigma_{2,\max}$, however, increases from 0 till γ_0 and decreases to 0 thereafter. Therefore, in the range $(\gamma_{\min}, \gamma_0]$, increasing one maximum bound leads to a decrease in the other. For the range $[\gamma_0, \gamma_{\max})$, however, an increase in one maximum bound is accompanied by an increase in the other and vice versa. This is illustrated in Figs. 11.3 and 11.4 where we present the bound curves $(\sigma_{1,\max}, \sigma_{2,\max})$ for different values of γ.

As γ increases, the equilibrium population density for predators increases. This results in higher predation and consequently lower permissible stochastic perturbations in prey density. The resulting diminished variation in stable prey density as well as the increased stable equilibrium density of the predators enables the predator density to withstand larger fluctuations. However, the behavior contrary to this (for $\gamma > \gamma_0$) is a consequence of a system with high predation. A high positive fluctuation in predators drives the preys toward extinction in a high-predation environment. Thus, the existence of such an extremum is a balance between less restricted downward fluctuations in preys and more restricted upward fluctuations in predators.

It can be shown that increasing β leads to an increase in $\sigma_{2,\max}$, followed by a maximum (corresponding to $\gamma = \gamma_0$) and a decrease thereafter. This is again

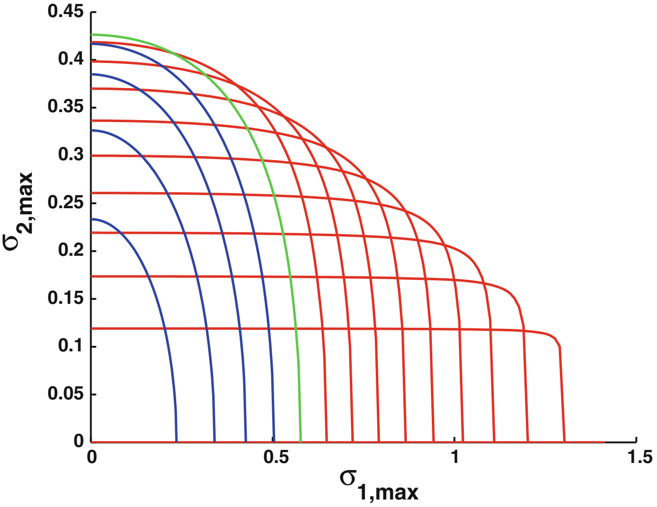


Fig. 11.3 Less dense contours of γ for (11.27). Here, *blue* represents $\gamma > \gamma_0$, *green* represents $\gamma = \gamma_0$, and *red* represents $\gamma < \gamma_0$

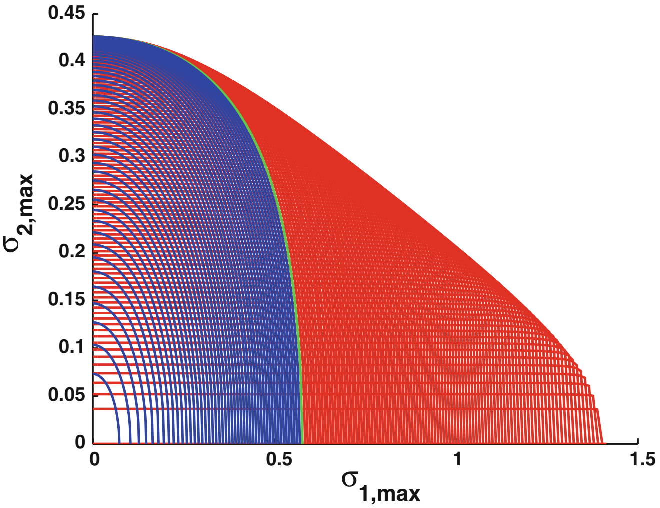


Fig. 11.4 More dense contours of γ for (11.27). Here, *blue* represents $\gamma > \gamma_0$, *green* represents $\gamma = \gamma_0$, and *red* represents $\gamma < \gamma_0$

a balance between below par predation, resulting in lower conversion to predator biomass, and above par predation, resulting in lower prey densities. Varying δ results in a behavior opposite to that for β. Similar observations were made by Srinivasu et al. [19] in the context of a model that incorporates additional food supply to the predators.

11.5 Conclusion

In this paper, we present the mathematical analysis for the stochastic stability of a Holling type II system. The choice of the Lyapunov function used is different from other previous approaches and the stability conditions for the stochastic system result in moving bounds on the environmental fluctuations that depend on an external variable. This results in the bounds lying on a curve in $\mathbb{R}^2$. Further scrutiny of the bounds shows that the increase in one bound is at the expense of the other, and ideally, they disappear on either end of the condition for deterministic stability, i.e., $\gamma_{\min}$ and $\gamma_{\max}$.

Bounds on variations in predator density disappear on both ends of the range for γ ($\gamma_{\min}$ and $\gamma_{\max}$) and have a maxima at an intermediate point γ_0, as defined in the previous section. The justification for this is a tradeoff between upward and downward fluctuations in the predator population density. An increase in γ increases equilibrium population density for predators allowing higher downward variation. This is undermined by the unaltered prey density, which cannot sustain high upward variation in predator population. The bounds at γ_0 provide the most sustainable environment for the predators.

We also discuss the dependence of the stochastic stability on the model parameters. For each of the parameters, the bounds on the sustainable environmental fluctuations in predator density have a peak. The bounds also substantially depend on the ratio β/δ and its inverse relation with the sustainable variations in prey density can be attributed to lower predation and higher mortality of predators.

References

1. M. Kot, *Elements of Mathematical Ecology* (Cambridge University Press, Cambridge, 2001)
2. C.S. Holling, The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. Can. Entomol. **91**(5), 293–320 (1959)
3. C.S. Holling, The functional response of predators to prey density and its role in mimicry and population regulation. Mem. Entomol. Soc. Can. **97**(S45), 5–60 (1965)
4. C.S. Holling, The functional response of invertebrate predators to prey density. Mem. Entomol. Soc. Can. **98**(S48), 5–86 (1966)
5. X. Liu, L. Chen, Complex dynamics of Holling type II Lotka-Volterra predator-prey system with impulsive perturbations on the predator. Chaos, Solitons Fractals **16**(2), 311–320 (2003)
6. L.A. Real, The kinetics of functional response. Am. Nat. **111**(978), 289–300 (1977)
7. M.P. Hassell, J.H. Lawton, J.R. Beddington, Sigmoid functional responses by invertebrate predators and parasitoids. J. Anim. Ecol. **46**(1), 249–262 (1977)
8. M.J. Crawley, *Natural Enemies: The Population Biology of Predators, Parasites, and Diseases* (Blackwell Scientific Publications, Oxford, 1992)
9. P.A. Abrams, The effects of adaptive behavior on the type-2 functional response. Ecology **71**(3), 877–885 (1990)
10. A. Oaten, W.W. Murdoch, Functional response and stability in predator-prey systems. Am. Nat. **109**(967), 289–298 (1975)
11. A.A. Berryman, The origins and evolution of predator-prey theory. Ecology **73**(5), 1530–1535 (1992)
12. R. Arditi, L.R. Ginzburg, Coupling in predator-prey dynamics: ratio-dependence. J. Theor. Biol. **139**(3), 311–326 (1989)
13. J. Sugie, R. Kohno, R. Miyazaki, On a predator-prey system of Holling type. Proc. Am. Math. Soc. **125**(7), 2041–2050 (1997)
14. S.-B. Hsu, T.-W. Huang, Global stability for a class of predator-prey systems. SIAM J. Appl. Math. **55**(3), 763–783 (1995)
15. S. Ruan, D. Xiao, Global analysis in a predator-prey system with nonmonotonic functional response. SIAM J. Appl. Math. **61**(4), 1445–1472 (2001)
16. S. Zhang, L. Chen, A Holling II functional response food chain model with impulsive perturbations. Chaos, Solitons Fractals **24**(5), 1269–1278 (2005)
17. R. Xu, M.A.J. Chaplain, F.A. Davidson, Periodic solutions for a predator-prey model with Holling-type functional response and time delays. Appl. Math. Comput. **161**(2), 637–654 (2005)
18. J. Zhou, C. Mu, Coexistence states of a Holling type-II predator-prey system. J. Math. Anal. Appl. **369**(2), 555–563 (2010)
19. P.D.N. Srinivasu, B.S.R.V. Prasad, M. Venkatesulu, Biological control through provision of additional food to predators: a theoretical study. Theor. Popul. Biol. **72**(1), 111–120 (2007)
20. M. Bandyopadhyay, C.G. Chakrabarti, Deterministic and stochastic analysis of a nonlinear predator-prey system. J. Biol. Syst. **11**(2), 161–172 (2003)

21. A. Maiti, G.P. Samanta, Deterministic and stochastic analysis of a prey-dependent predator prey system. Int. J. Math. Educ. Sci. Technol. **36**(1), 65–83 (2005)
22. M. Bandyopadhyay, J. Chattopadhyay, Ratio-dependent predator-prey model: effect of environmental fluctuation and stability. Nonlinearity **18**(2), 913–936 (2005)
23. T. Saha, C. Chakrabarti, Stochastic analysis of prey-predator model with stage structure for prey. J. Appl. Math. Comput. **35**(1–2), 195–209 (2011)
24. F.B. Hanson, D. Ryan, Optimal harvesting with both population and price dynamics. Math. Biosci. **148**(2), 129–146 (1998)
25. W.M. Schaffer, S. Ellner, M. Kot, Effects of noise on some dynamical models in ecology. J. Math. Biol. **24**(5), 479–523 (1986)
26. E. Beretta, V. Kolmanovskii, L. Shaikhet, Stability of epidemic model with time delays influenced by stochastic perturbations. Math. Comput. Simul. **45**(3–4), 269–277 (1998)
27. M. Carletti, On the stability properties of a stochastic model for phage-bacteria interaction in open marine environment. Math. Biosci. **175**(2), 117–131 (2002)
28. R.R. Sarkar, S. Banerjee, Cancer self remission and tumor stability—a stochastic approach. Math. Biosci. **196**(1), 65–81 (2005)
29. V.N. Afanas'ev, V.B. Kolmanowskii, V.R. Nosov, *Mathematical Theory of Control Systems Design* (Kluwer Academic, Dordrecht, 1996)

Chapter 12
Graph Theoretical Invariants of Chemical and Biological Systems: Development and Applications

Subhash C. Basak, Ramanathan Natarajan and Dilip K. Sinha

Abstract Chemical graph theory has been extensively applied in the characterization of structure in many areas of science, chemistry and biology in particular. Numerical graph invariants of molecules or topological indices have been used in the characterization of structure, discrimination of pathological structures like isospectral graphs, prediction of property/ bioactivity of molecules for new drug discovery and environment protection as well as quantification of intermolecular similarity. More recently, methods of discrete mathematics have found applications in the characterization of complex biological objects like DNA/ RNA/ protein sequences and proteomics maps. This chapter reviews the latest results in applications of discrete mathematics, graph theory in particular, to chemical and biological systems.

Keywords Topological indices · Pathological graphs · Molecular similarity · DNA sequence and proteomics maps · Isospectral graphs · Chirality · Vertices · Edges · Hydrogen-filled graph · Hydrogen-suppressed graph · Adjacency matrix · Distance

S.C. Basak (✉)
International Society of Mathematical Chemistry,
1802 Stanford Avenue, Duluth, MN 55811, USA
e-mail: sbasak@nrri.umn.edu

S.C. Basak
Natural Resources Research Institute, University of Minnesota Duluth,
5013 Miller Trunk Highway, Duluth, MN 55811, USA

R. Natarajan
V.K.A. Polymers Pvt. Ltd., 3-A Coimbatore Road,
Karur 639 002, Tamil Nadu, India
e-mail: rn@vkapolymers.com

D.K. Sinha
Eastern Indian Chapter, International Society of Mathematical Chemistry, Kolkata, India
e-mail: dilipkumarsinha@rediffmail.com

D.K. Sinha
Presidency University, Kolkata, India

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_12

matrix · Information content · Chemodescriptors · Quantitative structure property/activity relationship (QSPR/QSAR) · Biodescriptor

12.1 Introduction

The second half of the twentieth century witnessed a tremendous upsurge in research on applications of graph theory to various fields and this trend is continuing even today. To name just a few, graph and network theory was applied in formulating useful structural models in the physical sciences, social sciences, linguistics, biology, statistics, and operational research [1–4]. Graph theory serves as the mathematical model of representing structure in many fields. For example, the evolution of diverse and complex systems like the World Wide Web, business, and citation networks has been explored in terms of Bose–Einstein (BE) condensations of their respective network models [5].

In chemistry, invariants of molecular graphs, numerical graph invariants or topological indices (TIs) in particular, are used in the characterization of molecular structure. Recently, this approach has also been extended to biological systems like DNA/RNA sequence, proteins, and proteomics maps [6]. Such invariants encode information about various structural aspects, viz., size, shape, cyclicity, branching pattern, complexity, of the molecules, and biomolecules under investigation [3, 4]. TIs and related graph invariants have been used in the prediction of physicochemical, pharmacological, and toxicological properties of chemicals as well as quantification of proteomics maps and pathogenicity of organisms [3, 4, 6, 7].

12.2 Graph Theoretic Characterization of Structure

12.2.1 Topological Indices: Graph Theoretic Definitions and Calculation Methods

A graph, G, is defined as an ordered pair consisting of two sets V and $R, G = [V(G), R]$, where $V(G)$ represents a finite nonempty set of points and R is a binary relation defined on the set $V(G)$. The elements of $V(G)$ or V are called vertices and the elements of R, also symbolized by $E(G)$ or E, represent the edges. Such an abstract graph can be visualized by representing elements of $V(G)$ as points and by connecting each pair (u, v) of elements of $V(G)$ with a line if and only if $(u, v) \in R$. Two vertices in G are called adjacent if $(u, v) \in R$, i.e., they are connected by an edge. A walk of a graph is a sequence beginning and ending with vertices in which vertices and edges alternate and each edge is incident with vertices immediately preceding and following it. A walk of the form $v_0, e_1, v_1, e_2, \ldots, v_n$ joins vertices v_0 and v_n. The length of a walk is the number of edges in the walk. A walk is closed if $v_0 = v_n$,

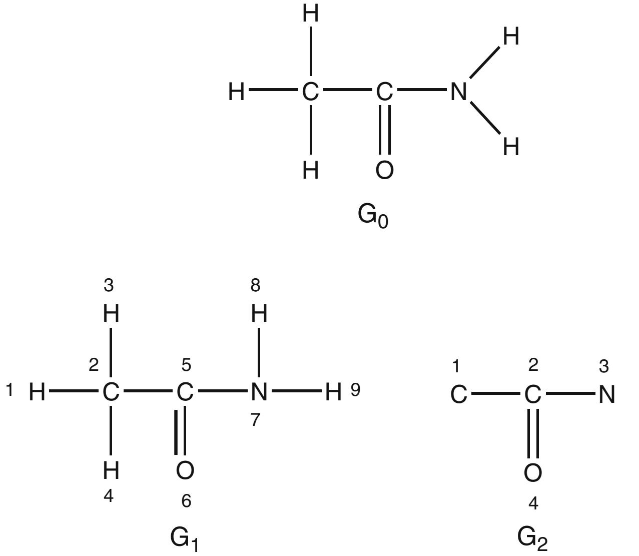


Fig. 12.1 Structural formula (G_0), labeled hydrogen-filled graph (G_1) and labeled hydrogen-suppressed graph (G_2) of acetamide. (Reprinted from Journal of Mathematical Chemistry, 4, 1990, S.C. Basak, G.J. Niemi, and G.D. Veith, Optimal characterization of structure for prediction of properties, 185–205, with kind permission of Springer Science and Business Media)

otherwise it is open. A closed walk with n points is a cycle if all its points are distinct and $n \geq 3$. A path is an open walk in which all vertices are distinct. A graph G is connected if each pair of its vertices is connected by a path. The distance $d(u, v)$ between vertices u and v in G is the length of the shortest path connecting u and v. The degree of vertex v, denoted by deg v, is equal to the number of edges incident with v. In molecular graphs, V represents the set of atoms or set of atomic nuclei and E represents the collection of covalent bonds in the molecule. The elements of E, however, may symbolize any type of bond, viz., covalent, ionic, or hydrogen bonds. It was emphasized by Basak et al. [8] that weighted pseudographs constitute a very versatile model for the representation of a wide range of chemical species.

In depicting a molecule by a connected graph $G = [V(G), E(G)]$, the set $V(G)$ may consist of either all atoms present in the molecule or only non-hydrogen (heavier) atoms. Hydrogen-filled graphs are preferable to hydrogen-suppressed graphs when hydrogen atoms are involved in chemically or physically important interactions. The structural formulas, labeled hydrogen-filled, and the labeled hydrogen-suppressed graphs for acetamide, are shown in Fig. 12.1.

Many topological indices can be conveniently derived from various matrices including the adjacency matrix $A(G)$ and the distance matrix $D(G)$ of a molecular graph G. These matrices are usually constructed from labeled graphs of hydrogen-suppressed molecular skeletons. For such a graph, $A(G)$ is defined to be the $n \times n$ matrix (a_{ij}), where a_{ij} may have only two different values as follows:

$$a_{ij} = \begin{cases} 1, & \text{if vertices } v_i \text{ and } v_j \text{ are adjacent in G} \\ 0, & \text{otherwise.} \end{cases}$$

The distance matrix $D(G)$ of a nondirected graph G with n vertices is a symmetric $n \times n$ matrix (d_{ij}), where d_{ij} is equal to the distance between vertices v_i and v_j in G. Each diagonal element d_{ii} of $D(G)$ is equal to zero.

The adjacency matrix $A(G_2)$ and the distance matrix $D(G_2)$ for the labeled graph G_2 in Fig. 12.1 are written as

$$A\left(G_2\right) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$D\left(G_2\right) = \begin{bmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{bmatrix}.$$

Wiener [9] was the first to put forward the idea of a structural index (topological index) for the estimation of properties of molecules from their structure. Hosoya [10] showed that this index, popularly known as the Wiener index, W, can be calculated from the distance matrix $D(G)$ of a hydrogen-suppressed graph G as the sum of entries in the upper triangular submatrix:

$$W = \sum_h h \cdot g_h = \frac{1}{2} \sum_{ij} d_{ij}. \tag{12.1}$$

From the adjacency matrix of a graph with n vertices, it is possible to calculate δ_i, the degree of the ith vertex, as the sum of all entries in the ith row:

$$\delta_i = \sum_{j=1}^{n} a_{ij}. \tag{12.2}$$

The zero-order connectivity index, ${}^0\chi$, is defined as [11]

$${}^0\chi = \sum_j \left(\delta_j\right)^{-1/2}. \tag{12.3}$$

The first-order connectivity index, also known as Randić's connectivity index [12], ${}^1\chi$, is defined as

$${}^1\chi = \sum_{\text{alledges}} \left(\delta_i \delta_j\right)^{-1/2}. \tag{12.4}$$

Kier and Hall [11, 13] extended the connectivity approach to calculate various types of connectivity indices and electrotopological invariants.

Information-theoretic topological indices are calculated by the application of information theory to chemical graphs. An appropriate set A of n elements is derived from a molecular graph G depending upon certain structural characteristics. On the basis of an equivalence relation defined on A, the set A is partitioned into disjoint subsets A_i of order

$$n_i \ (i = 1, 2, \ldots, h).$$

A probability distribution is then assigned to the set of equivalence classes:

$$\begin{array}{c} A_1, A_2, \ldots, A_h \\ p_1, p_2, \ldots, p_h, \end{array}$$

where $p_i = n_i/n$ is the probability that a randomly selected element of A will occur in the ith subset.

The mean information content of an element A is defined by Shannon's relation [14]:

$$IC = -\sum_{i=1}^{h} p_i \log_2 p_i. \tag{12.5}$$

12.2.2 Available Computer Software for the Calculation of Topological Indices

A large number of graph theoretic indices or topological indices and substructural descriptors can now be calculated using various computer programs including Dragon [15], Molconn-Z [16], POLLY [17], and APProbe [18]. They include simple connectivity indices, valence connectivity indices electrotopological state indices, Triplet indices, developed by Filip et al. [19], and neighborhood complexity indices [20]. When TIs are computed for small molecules, these indices were termed "chemodescriptors" by Basak [6].

12.3 Application of Graph Theoretic Indices

12.3.1 Comparing Pathological Graphs

One important use of graph invariants involves the characterization and discrimination of structures of closely related graphs for chemical documentation purposes. TIs and orthogonal parameters derived from them have been used in the discrimination of isospectral graphs which are well-known “pathological graphs” having the same value for many invariants [21].

12.3.2 *Molecular Similarity*

Molecular similarity is another area where TIs and subgraphs (substructures) have found wide application for practical purposes like new drug discovery and hazard assessment of chemicals. For a review of this topic, please see Basak et al. [22].

12.3.3 *Quantitative Structure Property/Activity Relationship (QSPR/QSAR)*

Quantitative structure property/activity relationships (QSPRs/QSARs) are mathematical models developed to predict property/bioactivity/toxicity of molecules. Initially, such techniques were formulated based on experimental data or properties derived from them. But for many practical situations, such properties are not available for the majority of chemicals under investigation. So now QSAR scientists routinely use computed properties which include numerical graph invariants calculated by the various programs mentioned above. For a review of the topic, please see references [6, 11, 13, 23].

12.3.4 *Biodescriptors for the Characterization of DNA Sequence and Proteomics Maps*

In the post-genomic era, catapulted by the Human Genome Project, a lot of interesting biological information is being generated by the “omics” (viz., genomics, proteomics, and metabolomics) technologies. Discrete mathematical methods, including those from graph theory, have been used for the characterization of DNA/RNA sequences and proteomics patterns relevant to human health and environmental protection [24–27].

12.3.5 *Characterization of Chirality (Handedness) Of Molecules*

Chemicals with one or more chiral centers can have multiple isomeric forms which may have different biological properties. Natarajan et al. [28] developed novel graph invariants of chiral molecules which can discriminate among the different chiral forms. Such invariants have found application in understanding the pharmacological and toxicological properties of molecules.

12.4 Conclusion

Graph theoretic methods have been used extensively in the representation and characterization of molecular structure as well as prediction of properties. Both weighted and unweighted molecular graphs of different types have been used to represent salient features of molecular structure. Such graphs may be looked upon model objects [29]. Invariants derived from the graphs are mathematical models useful for structure characterization. This article has given a short overview of the plethora of applications of graph theory to chemistry, biology, and the omics sciences.

Mathematicians and chemists continue to develop new ways of characterizing structure using graph theoretical methods. This trend of research is expected to significantly enrich the twenty-first century landscape of molecular descriptors and graph invariants. Hyle [30], an exclusive journal for the philosophy of chemistry, recently dedicated two of its issues to mathematical chemistry. The journal Current Computer-Aided Drug Design [31] also has published many papers on the development and use of graph invariants written by distinguished authors. These publications contain many interesting ideas, which will have important implications both for basic and applied researches in mathematical chemistry.

References

1. F. Harary, *Graph Theory and Theoretical Physics* (Academic Press, London, 1967)
2. M. Dehmer, S.C. Basak, *Statistical and Machine Learning Approaches for Network Analysis* (Wiley, Hoboken, 2012)
3. S.C. Basak, Philosophy of mathematical chemistry: A personal perspective. HYLE–Int. J. Philos. Chem. **19**, 3–17 (2013)
4. N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, 1983)
5. G. Bianconi, A. Barabasi, Bose-Einstein condensation in complex networks. Phys. Rev. Lett. **86**, 5632–5635 (2001)
6. S.C. Basak, Mathematical descriptors for the prediction of property, bioactivity, and toxicity of chemicals from their structure: a chemical-cum-biochemical approach. Curr. Comput. Aided Drug Des. **9**, 449–462 (2013)

7. A. Ghosh, A. Nandy, P. Nandy, B.D. Gute, S.C. Basak, Computational study of dispersion and extent of mutated and duplicated sequences of the H5N1 influenza neuraminidase over the period 1997–2008. J. Chem. Inf. Model. **49**, 2627–2638 (2009)
8. S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, Determining structural similarity of chemicals using graph theoretic indices. Discrete Appl. Math. **19**, 17–44 (1988)
9. H. Wiener, Structural determination of paraffin boiling points. J. Am. Chem. Soc. **69**, 17–20 (1947)
10. H. Hosoya, Topological Index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. Bull. Chem. Soc. Jpn. **44**, 2332–2339 (1971)
11. L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976)
12. M. Randić, Characterization of molecular branching. J. Am. Chem. Soc. **97**, 6609–6615 (1975)
13. L.B. Kier, L.H. Hall, *Molecular Structure Description: The Electrotopological State* (Academic Press, San Diego, 1999)
14. C.E. Shannon, A mathematical theory of communication. The Bell Syst. Tech. J. **27**, 379–423 (1948)
15. Dragon software, http://www.vcclab.org/lab/edragon/
16. Molconn-Z Version 3.5, Hall Associates Consulting. Quincy, MA (2000)
17. S.C. Basak, D.K. Harriss, V.R. Magnuson, POLLY v. 2.3: Copyright of the University of Minnesota (1988)
18. S.C. Basak, G.D. Grunwald, APProbe. Copyright of the University of Minnesota (1993)
19. P.A. Filip, T.S. Balaban, A.T. Balaban, A new approach for devising local graph invariants: derived topological indices with low degeneracy and good correlation ability. J. Math. Chem. **1**, 61–83 (1987)
20. S.C. Basak, in *Topological Indices and Related Descriptors in QSAR and QSPR*, ed. by J. Devillers, A.T. Balaban (Gordon and Breach Science Publishers, Netherlands, 1999), pp. 563-593
21. K. Balasubramanian, S.C. Basak, Characterization of isospectral graphs using graph invariants and derived orthogonal parameters. J. Chem. Inf. Comput. Sci. **38**, 367–373 (1998)
22. S.C. Basak, B.D. Gute, D. Mills, Similarity methods in analog selection, property estimation and clustering of diverse chemicals. ARKIVOC **9**, 157–210 (2006)
23. S.C. Basak, Role of mathematical chemodescriptors and proteomics-based biodescriptors in drug discovery. Drug Dev. Res. **72**, 1–9 (2010)
24. M. Randić, M. Vracko, A. Nandy, S.C. Basak, On 3-D graphical representation of DNA primary sequences and their numerical characterization. J. Chem. Inf. Comput. Sci. **40**, 1235–1244 (2000)
25. A. Nandy, M. Harle, S.C. Basak, Mathematical descriptors of DNA sequences: Development and applications. ARKIVOC **9**, 211–238 (2006)
26. M. Randić, J. Zupan, A.T. Balaban, D. Vikic-Topic, D. Plavsic, Graphical representation of proteins. Chem. Rev. **111**, 790–862 (2011)
27. S.C. Basak, B.D. Gute, Mathematical biodescriptors of proteomics maps: background and significance. Curr. Opin. Drug Disc. Dev. **11**, 320–326 (2008)
28. R. Natarajan, S.C. Basak, T.J. Neumann, A novel approach for the numerical characterization of molecular chirality. J. Chem. Inf. Model. **47**, 771–775 (2007)
29. M. Bunge, *Method, Model and Matter* (D. Reidel Publishing Co., Boston, 1973)
30. Hyle: Int. J. Philos. Chem. Available from: http://www.hyle.org/journal/issues/19-1/index.html
31. Curr. Comput. Aided Drug Des. Available from: http://www.benthamscience.com/ccadd/EBM.htm

Chapter 13
Mathematical Modeling of Breast Cancer Treatment

Suhrit K. Dey and S. Charlie Dey

Abstract Mathematical models to treat breast cancer both in situ, in which the cancer does not spread to other locations, and ones which have metastasized to a different location, have been developed and discussed in this article. The models consist of nonlinear-coupled ordinary differential equations for in situ cancer and partial differential equations for metastatic cancers. This model has been labeled as the Attacker–Defender model and was solved numerically using a predictor–corrector method. The results have been validated and the findings are very promising.

Keywords Beast cancer · Attacker–Defender model

13.1 Introduction

Cancer cells grow uncontrolled, violating the principles of homeostasis, the dynamic equilibrium, which all normal cells of the body must maintain. Cancer cells are abnormal, they are foreign antigens, and as a result, they trigger an immune response by the body's defenses. Because the p53 gene is missing from these cells, they are not prone to apoptosis, the cellular level preprogrammed death. Cancer cells continue growing as long as they have sufficient access to glucose and oxygen to feed on.

A simple uncontrolled growth model for any species u is represented by

$$\frac{du}{dt} = \lambda u,\ \lambda > 0. \tag{13.1}$$

$$u\left(t_0\right) = u_0$$

S.K. Dey (✉)
Professor Emeritus, Eastern Illinois University, Charleston, IL 61920, USA
e-mail: suhrit.day@gmail.com

S.C. Dey
Staff Researcher, Texas Advanced Computing Center, University of Texas,
Austin, TX 78758, USA

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_13

with a solution given by

$$u = u_0 e^{\lambda t}. \tag{13.2}$$

If $\lambda < 0$, this represents a model of decay as $t \to \infty$.

The common sense approach to prevent an uncontrolled growth is adding a term v to (13.1) such that $v = -\mu u$

$$\mu > 0, \tag{13.3}$$

where $\mu \geq \lambda$, then (13.1) becomes

$$\frac{du}{dt} = \lambda u - \mu u = -(\mu - \lambda)u \tag{13.4}$$

with a solution

$$u = u_0 e^{-(\mu-\lambda)t}. \tag{13.5}$$

If $\mu = \lambda, u = u_0$ will be a constant, so uncontrolled growth is prevented. But this will also cause u to not decay instead it will remain constant.

If $\mu > \lambda$, as $t \to \infty$, then $u \to 0$.

With this, the mathematical models have been designed in hopes of curing and at the very least, containing the cancerous cells. These models should prove to be a valuable asset for biochemists and other medical professionals to develop more effective tools and procedures to help fight and prevent cancer.

According to the Cancer Prevention Charity, World Cancer Research Fund International, the highest breast cancer rates were observed in 2012 in North America, 92 per 100,000 for the United States and 80 per 100,000 in Canada. It has also been observed that more women are beginning to be diagnosed with breast cancer at younger ages and that 75 % of breast cancer cases are hormone related.

Earlier, a model was introduced to analyze preventive techniques, which was very well received by oncologists, Dr. C. Wiseman [15], Dr. D. Characieju [3]; chemical biologist, Dr. H. Majumdar [11]; and mathematician, Dr. G. Webb [14].

This newer model has taken a more simplistic and application-oriented approach to the disease compared to the earlier models. One key difference is the parameter λ, previously λ was thought of being a constant, but upon further evaluation, it has been determined that λ can be time dependent. Some cancers have a very slow growth rate, keeping itself hidden among the fatty adipose tissue, which is abundant in breasts. When a weakness is detected, i.e., a drop in the immune system due to aging, medications, diet, environmental factors, or a combination of any of these, these cancerous cells begin growing and at a much more rapid pace.

The model for this in situ carcinoma is

$$\frac{du}{dt} = -(\mu(t) - \lambda(t))u, \tag{13.6}$$

where at any

$$t \geq t_0, \mu(t) > \lambda(t) \tag{13.7}$$

so the solution is then

$$u = u_0 e^{-\int_{t_0}^{t} (\mu(t) - \lambda(t))dt}. \tag{13.8}$$

With these as the guiding concepts, we can design a mathematical model for the treatment of breast cancer in situ, i.e., staying at the point of origin and not metastasizing through the lymphatic or cardiovascular systems including capillary blood vessels elsewhere.

13.2 Mathematical Model for in situ Breast Cancer

Ductal carcinoma in situ, or DCIS, begins in the milk ducts and is one of the earliest stages of breast cancer and can only be successfully treated with early detection. Unfortunately, due to its characteristics, there may not be any noticeable lump, and as such, is not totally detectable through self-examination. To make matters worse, mammograms are only 20 % effective in detecting these types of small-calcified dead cancer cells.

Another type of in situ cancer is lobular carcinoma in situ, or LCIS, which appears in the milk producing glands of the breasts. And once again, without proper detection, this type of cancer can have dangerous consequences. But, unlike other types of breast cancer, LCIS often occurs before menopause, and hence puts much larger percentage of the population at risk. Although its behavior is more toward a neoplasia, early tumor formation, than a true cancerous growth, it is still a foreign antigen and if left untreated can become an invasive carcinoma.

The mathematical model for in situ breast cancer is as follows:

Attacker:

$$\frac{du}{dt} = a_1 u - a_2 uv. \tag{13.9}$$

Defender:

$$\frac{dv}{dt} = b_1 u - b_2 uv, \tag{13.10}$$

where

u is the foreign antigen, the cancer;
v is a combination of the immune response and the medical response;
a_1 is the growth rate of each cancer cell per unit of time;

a_2 is the destruction rate of each cancer cell per unit of the defense per unit of time;
b_1 is the rate of gain of the defense for each unit of cancer cell; and
b_2 is the rate of loss for the defense per unit of cancer cells per unit of time.

The fight between the attacker and defender begins at the initial condition:

$$t = 0,\ u = u_0 \text{ and } v = v_0.$$

The goal is for v to overpower u. Mathematically, it means

$$\begin{aligned} a_1 &< b_1 \\ a_2 &> b_2 \end{aligned} \tag{13.11}$$

thus

$$\frac{a_1}{b_1} < 1 < \frac{a_2}{b_2}. \tag{13.12}$$

These are the conditions required to simulate a successful treatment.

13.2.1 Solution

From (13.9) and (13.10), we get

$$\frac{du}{dv} = \frac{a_1 - a_2 v}{b_1 - b_2 v} \tag{13.13}$$

when,

$$u = u_0,\ v = v_0$$

we get

$$u - u_0 = \alpha(v - v_0) + \beta \cdot \ln\left(\frac{b_1 - b_2 v}{b_1 - b_2 v_0}\right), \tag{13.14}$$

where

$$\begin{aligned} \alpha &= \frac{a_2}{b_2} \\ \beta &= \frac{(a_2 b_1 - a_1 b_2)}{b_2^2} = \frac{b_1}{b_2}\left(\frac{a_2}{b_2} - \frac{a_1}{b_1}\right). \end{aligned} \tag{13.15}$$

13.2.2 Critical Points

If

$$v = \frac{a_1}{a_2},$$

then from (13.9), we get

$$\frac{du}{dt} = 0 \text{ giving } u = u_0 \text{ for all } t.$$

If the values of the reaction coefficients a_1 and a_2 are chosen such that the drug promotes a stronger immune response, the cancer cells will not proliferate.

If

$$v = \frac{b_1}{b_2},$$

we get

$$\frac{dv}{dt} = 0 \text{ giving } v = v_0 \text{ for all } t$$

then

$$v_0 = \frac{b_1}{b_2} \text{ and } \frac{du}{dt} = -\lambda u, \tag{13.16}$$

where

$$\lambda = \left(\frac{1}{b_1}\right)\left(\frac{a_2}{b_2} - \frac{a_1}{b_1}\right) \tag{13.17}$$

since

$$\frac{a_2}{b_2} > \frac{a_1}{b_1}$$

the solution of (13.16) is

$$u = u_0 e^{-\lambda t} \tag{13.18}$$

thus, as $t \to \infty$, $u \to 0$.

Hence, for $v = \frac{b_1}{b_2}$, in time, the in situ tumor will shrink.

It should be noted that, whereas u is a free variable that can practically take any positive value, v does have certain restrictions. Regardless of what v is representing, if it is immunotherapy, chemotherapy, or radiation therapy, the patient can only tolerate a certain degree of dosage.

In this model, since

$$\frac{a_1}{b_1} < \frac{a_2}{b_2}$$

giving

$$\frac{a_1}{a_2} < \frac{b_1}{b_2}, \tag{13.19}$$

where $a_1 > 0, a_2 > 0, b_1 > 0, b_2 > 0$.

The dosage of v should be prescribed such that

$$\frac{a_1}{a_2} < v < \frac{b_1}{b_2}. \tag{13.20}$$

This should be the dosage of v that is capable of containing the cancer, u, and cause it to atrophy. The values of b_1 and b_2, as well as, a_1 and a_2, can be adjusted to reflect different types of cancers and different patients.

Also from (13.5) and (13.12), we find

$$\frac{du}{dv} = -\frac{a_2}{b_2} \cdot \frac{\left(v - \frac{a_1}{a_2}\right)}{\left(\frac{b_1}{b_2} - v\right)} < 0. \tag{13.21}$$

Thus, if the dosage of v as given by (13.21) is maintained, u will be a decreasing function of v.

This conclusion has been validated computationally.

13.2.3 Numerical Solution

Applying the conditions (13.20) and imposing no other restrictions, the nonlinear system (13.9) and (13.10) has been solved using Charlie's Method [4, 9, 10], an explicit finite difference Predictor–Corrector method where two parameters define the convex mappings.

Predictor:

$$\begin{aligned} \hat{U} &= U^n + \Delta t \cdot f\left(U^n, V^n\right) \\ \hat{V} &= V^n + \Delta t \cdot g\left(U^n, V^n\right). \end{aligned} \tag{13.22}$$

Corrector:

$$\begin{aligned} U^{n+1} &= (1-\gamma_1)\,\hat{U} + \gamma_1 \left\{U^n + \Delta t \cdot f\left(\hat{U}, \hat{V}\right)\right\} \\ V^{n+1} &= (1-\gamma_2)\,\hat{V} + \gamma_2 \left\{V^n + \Delta t \cdot g\left(\hat{U}, \hat{V}\right)\right\}, \end{aligned} \tag{13.23}$$

where

U and V are the net functions corresponding to u and v,
U^n corresponds to the value of U at time $t = t_n$,
V^n corresponds to the value of V at time $t = t_n$,
Δt is the time step,
$f\,(U^n, V^n)$ signifies the value of $(a_1 U - a_2 U V)$ at t_n,
$g\,(U^n, V^n)$ signifies the value of $(b_1 U - b_2 U V)$ at t_n,
$\hat{U}$ and $\hat{V}$ are the predicted values of U and V at t_{n+1},
U^{n+1} and V^{n+1} are the corrected values of U and V at t_{n+1}, and
γ_1 and γ_2 are the convex parameters such that

$$\begin{aligned} &0 < \gamma_1 < 1 \\ &0 < \gamma_2 < 1. \end{aligned} \tag{13.24}$$

If $\gamma_1 = \gamma_2 = \frac{1}{2}$, this is essential the same as the second Runge–Kutta method.

13.2.4 *Treatment of Invasive Breast Cancer, A One-Dimensional Model*

The cancer is considered invasive when it metastasizes and invades other organs, and can be mathematically modeled using Reaction–Diffusion equations. A simple one-dimensional model is

$$\frac{\partial u}{\partial t} = a_1 u - a_2 u v + \nu_1 \frac{\partial^2 u}{\partial x^2} \tag{13.25}$$

$$\frac{\partial v}{\partial t} = b_1 u - b_2 u v + \nu_2 \frac{\partial^2 v}{\partial x^2} \tag{13.26}$$

at $t = 0,\ u = u_0,\ v = v_0$.

ν_1 and ν_2 are the respective coefficients of dispersion.

As before, a technique similar to (13.22) will be used for the numerical experiment.

If $v = v_0 = \frac{a_1}{a_2}$, the tumor will begin shrinking, giving

$$\frac{\partial u}{\partial t} = \nu_1 \frac{\partial^2 u}{\partial x^2}. \tag{13.27}$$

The solution is of the following form, which will be further discussed later:

$$u = u_0 e^{-\lambda t} \cos\left(\sqrt{\lambda} \cdot x + h\right),\ \lambda > 0. \tag{13.28}$$

As $t \to \infty,\ u \to 0$, this property is natural and expected. If u cannot grow and must be dispersed, then it must be diffused.

If

$$v = \frac{b_1}{b_2}$$

then

$$\frac{\partial v}{\partial t} = \frac{\partial v}{\partial x} = 0,\ \text{(because } v \text{ is a constant)}$$
$$\frac{\partial u}{\partial t} = -\alpha u + \nu_1 \frac{\partial^2 u}{\partial x^2}, \tag{13.29}$$

$$\alpha = b_1 \left(\frac{a_2}{b_2} - \frac{a_1}{b_1} \right) > 0. \tag{13.30}$$

Let

$$u = X(x) \cdot T(t), \tag{13.31}$$

where $X(x)$ is a function of x and $T(t)$ is a function of t. Substituting (13.23) into (13.21), we get

$$\frac{1}{\nu_1} \left(\frac{T'}{T} + \alpha \right) = \frac{1}{X} \cdot X'' = \lambda. \tag{13.32}$$

With $\lambda > 0$, solution of $X'' + \lambda X = 0$ is $X = a \cdot \cosh\left(\sqrt{\lambda} \cdot x + b\right)$

$$\text{if } \lambda < 0,\ X = a \cdot \cos\left(\sqrt{\lambda} \cdot x + b\right),$$

where a and b are two constants.

For $\lambda > 0$, u will unbounded—but u *cannot* be unbounded since its growth has been suppressed by (13.20).

Thus, from (13.28),

$$u = u_0 e^{-(\alpha + \lambda \nu)t} a \cdot \cos\left(\sqrt{\lambda} \cdot x + b\right), \tag{13.33}$$

which shows that as $t \to \infty,\ u \to 0$, hence, the tumor must shrink.

This concept has been successfully extended to a three-dimensional mode:

$$\frac{\partial u}{\partial t} = a_1 u - a_2 uv + \nu_1 \nabla^2 u \tag{13.34}$$
$$\frac{\partial v}{\partial t} = b_1 u - b_2 uv + \nu_1 \nabla^2 u. \tag{13.35}$$

13.2.5 Three-Dimensional Numerical Model

Here Charlie’s method has been applied to the three-dimensional model.

Predictor:

$$\begin{aligned}\hat{U}_{ijk} &= U_{ijk}^{n} + \Delta t \cdot f_{ijk}\left(U^{n}, V^{n}\right)\\ \hat{V}_{ijk} &= V_{ijk}^{n} + \Delta t \cdot g_{ijk}\left(U^{n}, V^{n}\right).\end{aligned}$$

Corrector:

$$\begin{aligned}U_{ijk}^{n+1} &= (1-\gamma_1)\,\hat{U}_{ijk} + \gamma_1 \left\{U_{ijk}^{n} + \Delta t \cdot f_{ijk}\left(\hat{U}, \hat{V}\right)\right\}\\ V_{ijk}^{n+1} &= (1-\gamma_2)\,\hat{V}_{ijk} + \gamma_2 \left\{V_{ijk}^{n} + \Delta t \cdot g_{ijk}\left(\hat{U}, \hat{V}\right)\right\},\end{aligned}$$

where

$$\begin{aligned}f(U,V) =& a_1 U_{ijk} - a_2 (UV)_{ijk}\\ &+ \nu_1 \left\{\frac{U_{i-1jk} - 2U_{ijk} + U_{i+1jk}}{\Delta x^2} + \frac{U_{ij-1k} - 2U_{ijk} + U_{ij+1k}}{\Delta y^2}\right.\\ &\left. + \frac{U_{ijk-1} - 2U_{ijk} + U_{ijk+1}}{\Delta z^2}\right\}\\ g(U,V) =& b_1 U_{ijk} - b_2 (UV)_{ijk}\\ &+ \nu_2 \left\{\frac{V_{i-1jk} - 2V_{ijk} + V_{i+1jk}}{\Delta x^2} + \frac{V_{ij-1k} - 2V_{ijk} + V_{ij+1k}}{\Delta y^2}\right.\\ &\left. + \frac{V_{ijk-1} - 2V_{ijk} + V_{ijk+1}}{\Delta z^2}\right\}.\end{aligned}$$

$$\begin{aligned}U_{ijk}^{n} &= U\left(x_i, y_j, z_k, t_n\right)\\ V_{ijk}^{n} &= V\left(x_i, y_j, z_k, t_n\right)\end{aligned}$$

$\Delta t =$ time step

$\Delta x,\ \Delta y,\ \Delta z =$ mesh size

$$\begin{aligned}0 < \gamma_1 < 1\\ 0 < \gamma_2 < 1.\end{aligned}$$

$$\begin{aligned}i &= 1, 2, 3, \ldots, I\\ j &= 1, 2, 3, \ldots, J\\ k &= 1, 2, 3, \ldots, K.\end{aligned}$$

Field size = $I \cdot J \cdot K$.

13.2.6 Results

The computational results match all the mathematical predictions very precisely. Three different cases have been considered. Case one involves a ductal carcinoma in situ—where the cancer stayed local; case two looks at an invasive carcinoma—where the cancer is attempting to metastasize elsewhere; and case three looks at a metastatic carcinoma—a case in which the cancer has already metastasized to a different location. In each case, the inequality (13.20) plays a very critical role. When this inequality is satisfied, the treatment is shown to be successful. Even if the condition is violated, as long as it is not grossly violated, the computational solution still shows a positive result, i.e., a successful treatment and a good prognosis.

Figures 13.1 (in situ) and 13.2 (metastasizing) represent how the cancer is contained and/or the tumor beginning to atrophy. Even in Fig. 13.3—the worst case scenario, where the cancer has already metastasized, as long as (13.20) is satisfied, computationally, the cancer is seen as contained.

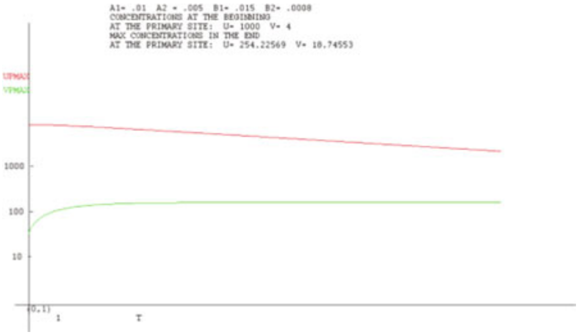


Fig. 13.1 Tumor in situ decreasing

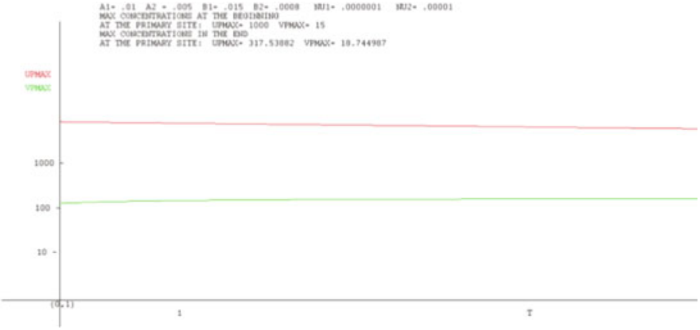


Fig. 13.2 Metastasizing tumor is contained

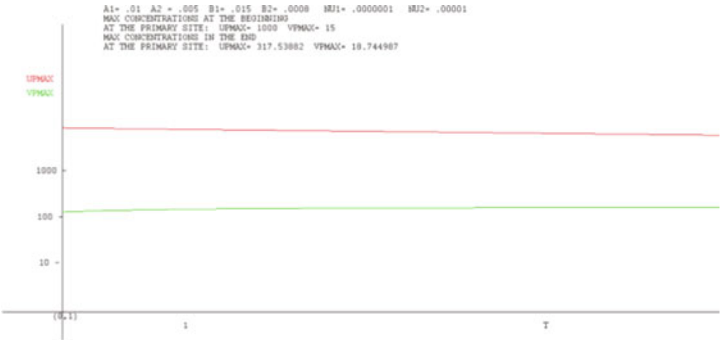


Fig. 13.3 Metastasized tumors are contained

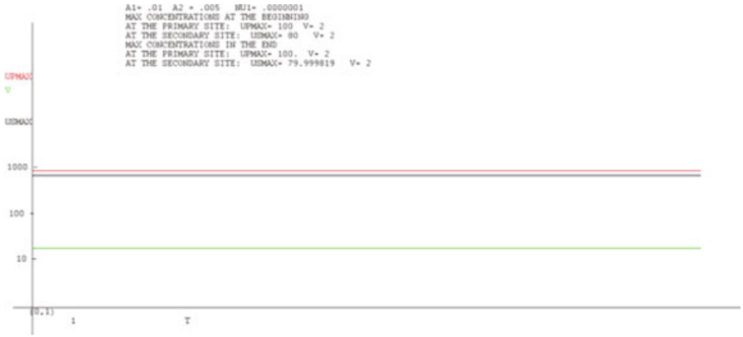


Fig. 13.4 Very slow shrinking of tumor in keeping with slow diffusion of immunotherapy

For the computational runs, the following conditions were used:

$$\begin{aligned} a_1 &< b_1 \\ a_2 &> b_2 \\ \nu_1 &< \nu_2, \end{aligned}$$

the cancer diffuses slower than that of the immunotherapy (Fig. 13.4).

In each case, the growth rate of the tumors was diminished. Only when these conditions were pungently violated, the tumors did begin growing.

The second phase of this study, with the introduction of a more stringent form of a medical response, is being looked into attempts to further contain or annihilate the cancerous cells. A new approach is also being considered where the user may interact with the model, changing the aspects of the cancer and the dynamics of the response and observe the results in real time. This level of user interaction is necessary to fully develop the computational model, as the properties of the nonlinear system cannot be fully comprehended by the mathematical model alone.

13.3 Conclusion

Cancer is a group of diseases that affect different patients in varying ways, and as such, the medical response treatment needs to be tailored for the individual patient and the individual cancer satisfying condition (13.19), else, according to this model [1–3, 5–8, 11–15], the treatment may not be effective, leading to a negative prognosis for the patient, whereas when the conditions are satisfied, the cancer is contained and, in some cases, totally annihilated. This model is also absolutely customizable for the patient, for the cancer, and for the individual immune and/or medical response.

Things being measured for future adaptations of this model include changing some assumptions. For the model discussed here, it was assumed that the variables

$$a_1, a_2, b_1, b_2$$

are constants meaning that the aspects of the cancer and patient will remain the same throughout the life of the cancer and the patient; however, they can be made to be time dependent, and hence have a constantly evolving attacker and a defender. The next phase of this model will be taking these ideas into consideration.

References

1. American Cancer Society, Breast Cancer Facts and Figures 2014. Atlanta, GA (2014)
2. American Joint Committee on Breast Cancer, *AJCC Cancer Stating Manual*, 7th edn. (Springer, New York, 2010)
3. D. Characiejus, *Personal Communication*, Department of Oncology, Vilnius Hospital, Vilnius, Lithuania
4. S.K. Dey, C. Dey, *An Explicit Predictor-Corrector Solver with Application to Burgers' Equation*, NASA Technical Memorandum. 84402 Sept 1983 (1983)
5. S.K. Dey, C. Dey, Accurate explicit finite difference solution of the shock tube problem, in *Numerical Methods and Approximation Theory III*, ed. by G.V. Milovanovic. University of Nis, 18–21 Aug 1987, pp. 191–197 (1987)
6. S.K. Dey, Computational modeling of breast cancer treatment by immunotherapy, radiation, estrogen inhibitors. Sci. Math. Jpn. **58**(2), 307–322 (2003)
7. H. Enderling, M. Chaplain, A. Anderson, J. Vaidya, A mathematical model of breast cancer development, local treatment and recurrance. J. Theor. Biol. **246**(2), 245–259 (2007)
8. F. Fuller Royal, G. Olson, Illness as a delusion, understanding the meaning of symptoms. Cent. Front. Sci. **7**(1), 138–144 (1998)
9. D. Kirschner, J. Panetta, Modeling immunotherapy of the tumor-immune interaction. J. Math. Biol. **37**, 235–252 (1998)
10. J.K. Koontz, S.C. Dey, S. Chatterjee, S.K. Dey, MPI implementation of PFI code for numerical modeling of the anatomy of breast cancer. Int. J. Comput. Math. **81**(8), 157–167 (2004)
11. H. Majumdar, *Personal Communication*, Institute of Chemical Biology, Kolkata, India
12. D. Mackenzie, Mathematical modeling and cancer. SIAM News. **37**, 1 (2004)
13. H. Osman, A.S. Wood, Stability of Charlie's method on linear heat conduction equation. Mathematika (2001)
14. G. Webb, *Personal Communication*. Mathematics Department. Vanderbilt University
15. C. Wiseman, *Personal Communication*, Department of Oncology, St. Vincent Hospital, Los Angeles, CA

Chapter 14
Effect of Dual Splitter Plate Attached with a Square Cylinder Immersed in a Uniform Flow

Bhanuman Barman and Somnath Bhattacharyya

Abstract In this paper we made a numerical study on control of vortex shedding and drag reduction of a cylinder by attaching splitter plates. The wake structure of the cylinder of square cross-section with attached splitter plate is analyzed for Reynolds number, based on the incident stream and height of the cylinder, up to 150. The length of the splitter plate, L is taken between the range $0 \leq L \leq 6$. The Navier-Stokes equations governing the flow is solved by the finite volume method over staggered grid arrangement. We have used the SIMPLE (Semi-Implicit Method for Pressure-Linked Equation) algorithm for computation. Our results show that the presence of a splitter plate upstream of the cylinder reduces the drag but it has a small impact on the vortex shedding frequency. It is found that an upstream splitter plate leads to a significant reduction is drag force when the length of the plate L is $0 \leq L \leq 3$. However, for $3.5 \leq L \leq 4.75$ the reduction in drag force is low and when $L \geq 5$, there is no effect of the splitter plate on the drag experienced by the cylinder. The presence of a downstream splitter plate damps the vortex shedding frequency. The entrainment of fluid into the inner side of the separated shear layers is obstructed by the downstream splitter plate. Our results suggest that by attaching splitter plates both upstream and downstream of the cylinder vortex shedding can be suppressed as well as a reduction in drag can be obtained. We made a parametric study to determine optimal length of the splitter plates attached upstream and downstream of the cylinder so as to achieve low drag and low vortex shedding frequency.

Keywords Square cylinder · Splitter plate · Vortex shedding · Drag reduction

B. Barman (✉)
University of Gour Banga, West Bengal, India
e-mail: bhanuman@maths.iitkgp.ernet.in

S. Bhattacharyya
IIT Kharagpur, Kharagpur, India
e-mail: somnath@maths.iitkgp.ernet.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_14

14.1 Introduction

The vortex shedding and formation of Karman Vortex Street behind a cylinder has been the object of numerous experimental and numerical studies because of the fundamental mechanism that this flow exhibits and its several practical relevance. The classical view of a vortex street in cross section consists of regions of concentrated vorticity shed into the downstream flow from alternate sides of the body (and with alternate sense of rotation), giving the appearance of an upper row of negative vortices and lower row of positive vortices. This alternate shedding of vortices in the near wake leads to large fluctuating pressure forces in a direction transverse to the flow and may cause structural vibrations, acoustic noise or resonance.

There have been numerous investigations in the past aiming to alter or suppress the pattern of vortex shedding. Reducing the drag is critically important in certain engineering applications, and both active and passive control techniques have been proposed to achieve that goal. In general, flow control techniques for reducing the aerodynamic drag exerted on a bluff body are classified into two types: active and passive control techniques. Active control methods control the flow by supplying external energy through several means such as rotational oscillatory motion of the bluff body or jet blowing. Passive control techniques control the vortex shedding by modifying the shape of the bluff body or by attaching additional devices in the flow stream. Therefore, the passive control technique is energy free and often easier to implement. Among the passive control techniques, the splitter plate has been considered as one of the most successful devices to control the vortex shedding behind a cylinder Ali et al. [1]. Another approach of controlling the flow behind a bluff body is to place a smaller bluff body in tandem with the main body. The reattachment of shear layers emerging from the front bluff body with the main body causes a substantial drag reduction and damping in oscillation.

Roshko [12] was among the first to study the control of vortex shedding behind a circular cylinder through attaching a splitter plate along the downstream. The study was expanded numerically by Kwon and Choi [7] for various plate lengths and at low Reynolds numbers (Re) i.e., Re between 80 and 160. Vortex shedding from bluff bodies with splitter plates was experimentally investigated by Nakamura [9]. The free shear layers emerging from either sides of a bluff body roll up to form vortices, and these vortices are shed alternately from each side of the body. Flow visualization study due to Kawai [6] suggests that a splitter plate reduces the three-dimensionality in the formation region by stabilizing the transverse flapping of the shear layers. Lin and Wu [8] found that a splitter plate having a length of $2D$ attached to the cylinder could suppress the vortex shedding. Tiwari et al. [14] carried out a numerical study to understand the effect on the flow characteristics of the splitter plate mounted on the back of the circular tube. When a geometric modification, either through attached/ detached splitter plate or placing a in-line bluff body, the flow approaches the bluff body is subjected to a momentum loss. This causes a significant reduction in pressure along the upstream face of the bluff body. Thus, the drag experienced by the body gets modified. Several studies on drag reduction and control of flow by placing another cylinder in tandem have been reported (e.g., [2, 5, 15]). Hwang and Yang [4] studied numerically the drag reduction by placing an attached splitter plate upstream of the cylinder.

Flows over square cylinders are important in many engineering applications, especially in flows around bridges, buildings, marine risers and in the context of augmentation of heat transfer from PCBs. The basic difference between the flow field past a square cylinder and a circular cylinder is that the points of separation for the former one are fixed at the upstream corners, whereas the points of separation for the latter case move back and forth depending on the oncoming fluid velocity. The fluid travels downstream at a large trajectory angle from each of these points and a comparatively larger recirculation zone is generated by Ali et al. [1]. Ozono [10] studied the effects on vortex shedding frequency by placing splitter plate along the centreline of the square cylinder. Ali et al. [1] investigated the effect of the length of a downstream attached splitter plate on the wake of a square cylinder. In this paper, the control of vortex shedding as well as drag reduction by attached splitter plates both upstream and downstream of a square cylinder is investigated. The splitter plates are placed along the horizontal centreline of the cylinder.

14.2 Problem Formulation

A long square cylinder of side D placed in a uniform flow (from left to right) with velocity U_∞ is considered. In the upstream of the cylinder, a splitter plate of fixed length L_1 is attached with the cylinder and in the downstream of the cylinder, another splitter plate of length L_2 attached with the cylinder. The length of the downstream splitter plate is varied. The thickness of the splitter plate is h. A schematic diagram of the flow configuration is presented in the Fig. 14.1. Here we have used U_∞ as velocity scale, D as scale for length, ρU_∞^2 as scale for pressure, where ρ is the density of the fluid and D/U_∞ is considered as time scale. With these scale, the non-dimensional Navier-Stoke's governing the fluid flow characteristics are given by

$$\nabla \cdot \mathbf{V} = 0 \tag{14.1}$$

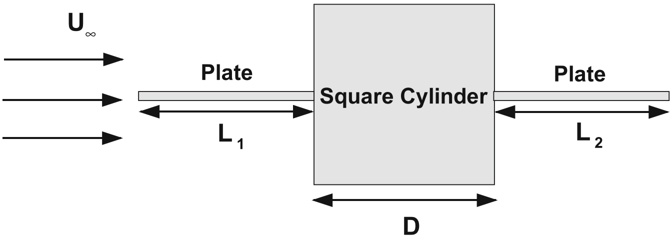


Fig. 14.1 Sketch of the test case and annotations

$$\frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla)\mathbf{V} = -\nabla p + \frac{1}{Re}\nabla^2 \mathbf{V} \tag{14.2}$$

where the non-dimensional parameter Reynolds number (Re) is defined as $Re = U_\infty D/\nu$, where ν is the kinematic viscosity.

We have considered U_∞ as far field (upstream) velocity. Symmetric boundary condition is considered along the downstream boundary. The drag and lift coefficients (C_D, C_L) are obtained by considering the viscous and pressure forces acting on the cylinder, are determined as

$$C_D = \frac{F_D}{0.5\rho U_\infty^2 D^2},\ C_L = \frac{F_L}{0.5\rho U_\infty^2 D^2} \tag{14.3}$$

where F_D and F_L are the integrated drag and lift forces experienced by the cylinder, respectively. The non-dimensional pressure coefficient (C_p) and base pressure coefficient (C_{pb}) are defined by

$$C_p = \frac{p^*}{0.5\rho U_\infty^2},\ C_{pb} = \frac{p^*}{0.5\rho U_\infty^2} \tag{14.4}$$

where p^* is the dimensional pressure on the surface of the cylinder.

14.3 Numerical Method and Validation

The governing Eqs. (14.1)–(14.2) are solved numerically by using a finite volume over a staggered grid system. In the staggered grid arrangement, the velocity components are stored at the midpoints of the cell sides to which they are normal. The scalar quantity pressure is stored at the center of the cell. The discretized forms of the governing equations are obtained by integrating over an elemental rectangular cells using finite volume method. A pressure correction-based iterative algorithm, SIMPLE [11] is used for solving the governing equation with boundary conditions specified previously. A first-order implicit scheme is used for discretizing the time derivatives. The pressure link between continuity and momentum is accomplished by transforming the continuity equation into Poisson equation for pressure. The Poisson equation implements a pressure correction for a divergent velocity field. At each time step the resulting tri-diagonal system of algebraic equations are solved through a block elimination method. In Fig. 14.2a, it is found that the time-averaged drag coefficient (C_{Dav}) changes 3 % for the grid sizes 420 × 310 compare to the grid size 471 × 352 due Ali et al. [1]. In Fig. 14.2b, the Strouhal number (St) are compared with Ali et al. [1] to validate our established code.

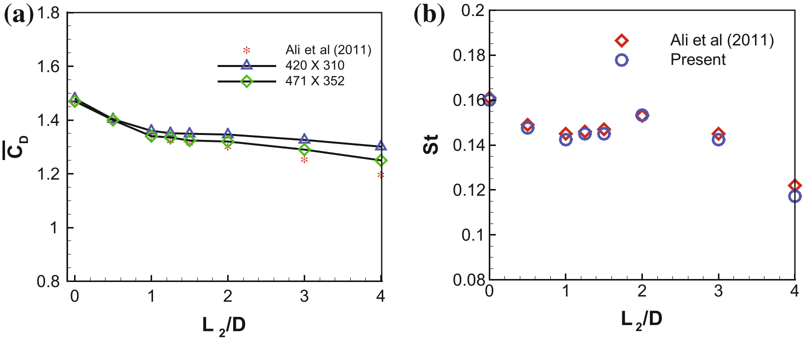


Fig. 14.2 For $Re = 150$ with Ali et al. [1], **a** time-averaged total drag coefficients (C_{Dav}) and **b** Strouhal number (St)

14.4 Results and Discussion

14.4.1 Upstream Splitter Plate

The effect due to the presence of an upstream splitter plate on wake and forces of a cylinder is discussed first. The plate is attached to the cylinder placed horizontally along the centreline. When the splitter plate is attached in front face of the cylinder, the vortex shedding is increasing whereas the drag coefficient is decreasing significantly. In Fig. 14.3a, b, two flow regime is found. Regime (1) $0 \leq L_1/D < 1.5$, the Strouhal number is increasing. At these moment the vortex shedding is too high and the drag force is decreasing sufficiently. Regime (2) $1.5 \leq L_1/D \leq 4$, the Strouhal number becoming steady and the drag force is decreasing rapidly. But when $L_1/D > 4$, there is no effect on the flow. The instantaneous vorticity lines for different length of splitter plate for $Re = 150$ is separated shear layers of the plate reattach on the downstream cylinder. The top and bottom shear layers of the plate have similar vorticity distribution and do not disturb the manner of the vortex shedding from the cylinder. The separated shear layer from the rod reattaches to the cylinder and rolls up in a quasi-steady manner. The pattern of the vortex shedding behind the cylinder remains unaltered due to the presence of the upstream plate. The upstream splitter plate decreases the momentum of the fluid impinging on the cylinder due to skin friction thereby decreasing the stagnation pressure on the surface of the cylinder facing the flow which results in reduction of drag. In Fig. 14.3c, it is found that the surface pressure on all faces of the cylinder is dropping down. Due to the thin plate in the front face of the cylinder, the shear layer of the flow get separated and diminished the pressure at the centre of the front face. As in Fig. 14.3c, it is shown that as the length of the thin plate increases the centre pressure in decreasing. The time average surface pressure distribution around the cylinder at Reynolds number, $Re = 150$ is presented in Fig. 14.3c for different values of the upstream splitter plate length i.e., $L_1/D = 0,\ 0.5,\ 1,\ 1.5,\ 2,\ 2.5$. The pressure is positive along the front

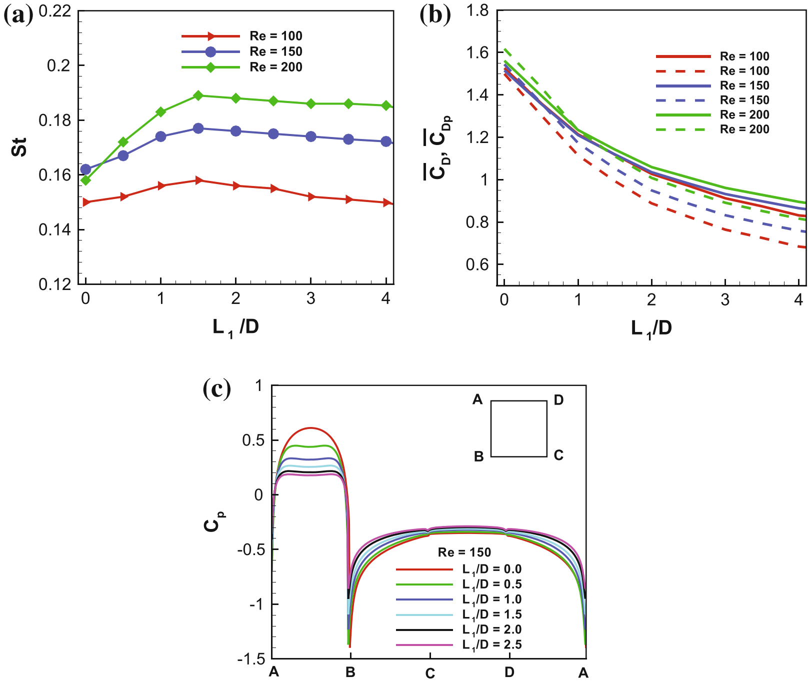


Fig. 14.3 **a** Strouhal number, **b** time-averaged drag coefficient and **c** time-averaged wall pressure effect for different length of upstream splitter plate

face AB of the cylinder with the stagnation pressure occurs at a point on AB. Fluid separates from the upper and lower corners A and B, where the pressure distribution attains minima. The pressure along other sides of the cylinder is negative as the flow separates from lower and upper corners are never reattaches with the cylinder. The pressure distribution on the front face of the cylinder changes due to the variation of the plate length is found. However, the surface pressure distribution along the faces other than the front face follow the similar trend as that the case of a cylinder without an attached plate ($L_1/D = 0$). Since the upstream plate decreases the momentum of the flow impinging on the cylinder the stagnation pressure is reduced.

The variation of Strouhal number (St) and average drag coefficient (C_D) due to the variation of Reynolds number at different values of the upstream splitter plate length is presented in Fig. 14.3a, b. At lower range of Re, it is found that the upstream splitter plate have a negligible effect on the vortex shedding frequency. At higher range of Re, the vortex shedding frequency is enhanced due to the increase of splitter plate length. The variation of the drag coefficient with Reynolds number at different upstream splitter plate length shows that the average drag reduces with the introduction of the upstream splitter plate. As the length of the plate increases the drag reduces at any given value of the Reynolds number. It is evident from the Fig. 14.3b that the variation of C_D with Re follow the same trend at different $L_1/D(=0,\ 0.5,\ 1.0,\ 2.0)$.

This suggests that the upstream splitter plate does not influence the vortex shedding mechanism downstream of the cylinder and the reduction in C_D is only due to the reduction of pressure at the front face of the cylinder. Hwang and Yang [4] observed the similar trend in C_D due to the presence of a detached splitter plate upstream of the cylinder.

14.4.2 Dual Splitter plate

The influence of a downstream splitter plate on the wake of a square cylinder in the laminar range of Reynolds number have been studied at length by Ali et al. [1]. There they found that in presence of the downstream attached plate, the free shear layers which are emerging from the opposite sides of the cylinder are convected further downstream before rolling-up. When plate length is bigger, a secondary vortex forms around the trailing edge of the plate. This secondary vortex influences the vortex shedding behind the cylinder. The flow in which dual splitter plates are attached with the cylinder is investigated. Figure 14.4 shows the influence of dual splitter plate on the flow around a square cylinder. We considered the flow at Reynolds number, $Re = 150$ and varied the length of the downstream plate ($0 \leq L_2/D \leq 4.0$) when the length of the upstream splitter plate is L_1/D is taken to be 1.0. The vortex shedding behind the cylinder occurs due to the interaction of shear layers of opposite sign emerging from either sides of the cylinder. The attached downstream plate interfere the fluid entrainment in the shear layers and consequently, the vortex shedding is influenced. Several authors have already reported on the influence of downstream splitter plate on vortex shedding. A review of which is already provided in the introduction section. It is found that with the introduction of the downstream plate the shear layer convects further downstream compared to an unbounded cylinder before being roll-up. Increase in the length of the plate causes the shear layers to extend further downstream before they entrain into each other shows that wake becomes steady at $Re = 150$ when L_2/D is bigger than 2. For lower range of the downstream plate length, a periodic vortex shedding occurs at Re = 150. The downstream splitter plate attached along the centreline produces symmetry in the wake structure. A tip vortex at the trailing edge of the plate is clearly visible at this Reynolds number. At steady state, the interaction between the top and bottom shear layers is suppressed. The shear layer emerging from top and bottom faces of the cylinder transport downstream almost horizontally without forming any vortex in the near wake of the cylinder. It is found that any given value of Reynolds number ($Re \leq 200$), there exists a critical value of the downstream splitter plate length for which the vortex shedding is suppressed. This critical value of L_2/D depends on the length of the upstream splitter plate (L_1/D). Later in this section, on suppression of vortex shedding by varying L_2/D through computation of the Strouhal number at different Reynolds number is discussed. The instantaneous vorticity contours for different values of the length of the downstream plate ($0 \leq L_2/D \leq 4.0$) at Reynolds number 150 and upstream plate length $L_1/D = 1$ is shown in Fig. 14.4. It is clear

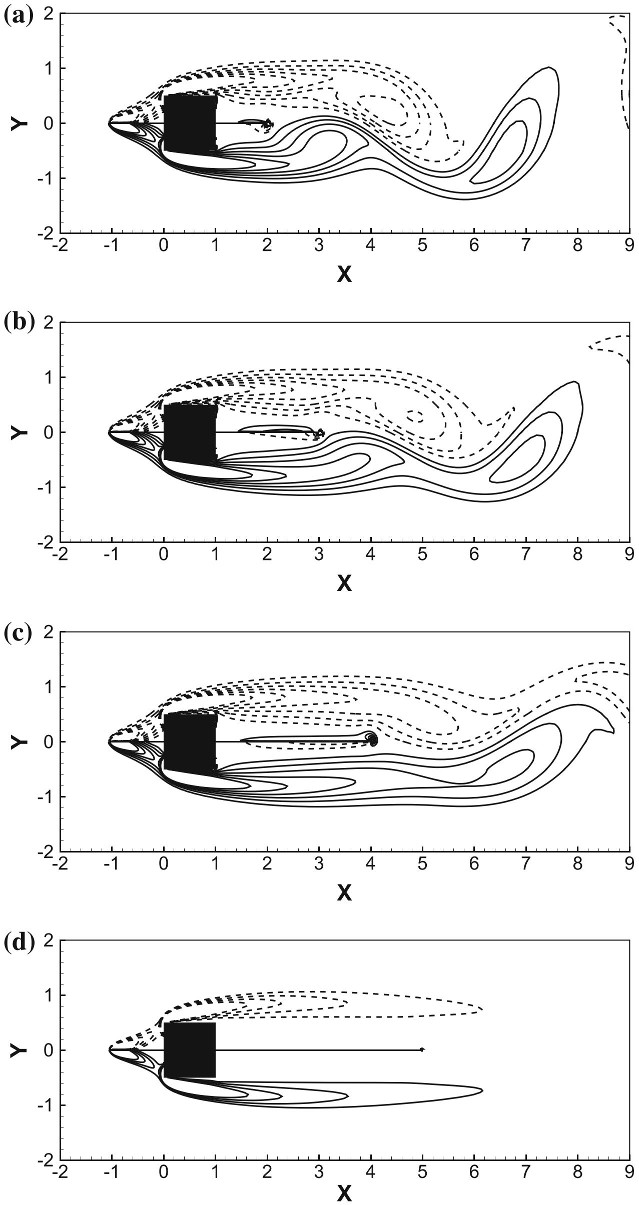


Fig. 14.4 Instantaneous vorticity contours for different downstream splitter plate at fixed upstream splitter plate of length $L_1/D = 1$ and $Re = 150$. **a** $L_2/D = 1.0$, **b** $L_2/D = 2.0$, **c** $L_2/D = 3.0$, **d** $L_2/D = 4.0$

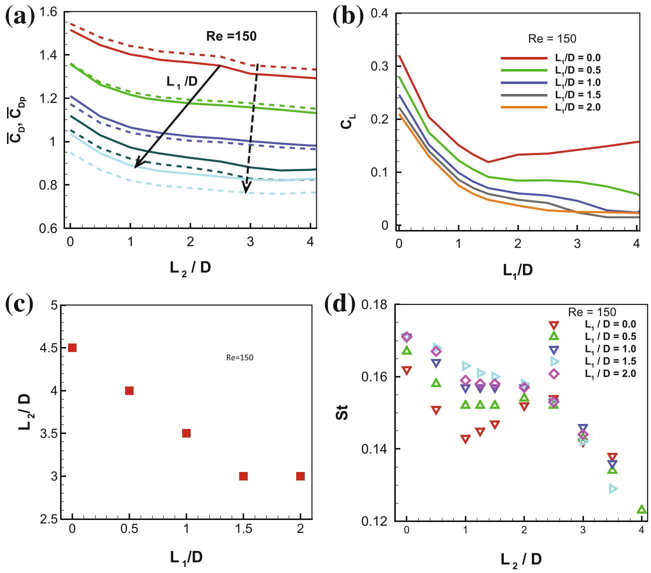


Fig. 14.5 **a** Time-averaged total drag coefficient $\overline{C}_D$ (*solid lines*) and time-averaged total pressure drag $\overline{C}_{Dp}$ (*dashed lines*), **b** time-averaged total lift coefficients ($\overline{C}_L$), **c** Critical length of splitter plate and **d** Strouhal number effect for double splitter plate at $Re = 150$

from Fig. 14.4 that the interaction of the shear layers of opposite sign occurs further downstream of the cylinder in comparison of a cylinder without any splitter plate.

The C_D and C_L is presented in Fig. 14.5a, b as a function of L_2/D for different choice of L_1/D when $Re = 150$. Drag is reduced markedly below the plain-cylinder value by a very short splitter plate. In Fig. 14.5c, the critical length of the splitter plates is shown and Fig. 14.5d is presented the Strouhal number effect for different length of the splitter plates for Reynolds number 150.

14.5 Conclusion

A parametric study has been carried out in order to assess the effectiveness of dual attached "thin" splitter plates on drag reduction of a square cylinder immersed in a free stream. The length of front splitter plate $L_1/D = 1.0$ is kept at constant and the length of the back splitter plate L_2/D is varying, and they are placed along the centerline, upstream and downstream of the cylinder, respectively. The numerical simulations have been carried out at a Reynolds number 150. The main objective of this study is to investigate the effect of splitter plate lengths on the flow structure about a square cylinder. The study has clearly shown that a splitter plate can fundamentally change the flow structure of the square cylinder wake via a variety of hydrodynamic interaction mechanisms. The structure of the flow is heavily influenced by the splitter plate length. When only the upstream splitter plate is attached to the cylinder, it has been observed that the drag forces reduces rapidly but the vortex shedding increases behind the cylinder, which may create a oscillation on the body. To reduce the vorticity, we had attached a splitter plate in the downstream of the cylinder. The effect of the back plate in this plate length range is to convect the free shear layers further downstream before roll up occurs. When the back plate length is further increased to $L_2/D = 1.5$, there is an decrease in the Strouhal number. This is an indication of the transition into a new flow regime. For the plate length range $1.5 \leq L_2/D \leq 2.0$, a secondary vortex about the splitter plate trailing edge is clearly observed. The Strouhal number in this flow regime decreases. A sudden drop in Strouhal number is noticed when the length of the plate is increased from $L_2/D = 2.0$ to $L_2/D = 3.0$. This is an indication of the transition into the third flow regime $L_2/D \geq 3.0$ at which the free layers reattach to the splitter plate. In this flow regime, the Strouhal number is almost unchanged with increasing the plate length. This study has also found the most suitable length scale for estimating a universal Strouhal number.

References

1. M. Ali, C. Doolan, V. Wheatley, Low Reynolds number flow over a square cylinder with a splitter plate. Phys. Fluids **23**, 033602 (2011)
2. S. Bhattacharyya, S. Dhinakaran, Vortex shedding in shear flow past tandem square cylinders in the vicinity of a plane wall. J. Fluids Struct. **24**(3), 400–417 (2008)
3. R. Franke, W. Rodi, B. Schönung, Numerical calculation of laminar vortex-shedding flow past cylinders. J. Wind Eng. Ind. Aerodyn. **35**, 237–257 (1990)
4. J. Hwang, K. Yang, Drag reduction on a circular cylinder using dual detached splitter plates. J. Wind Eng. Ind. Aerodyn. **95**(7), 551–564 (2007)
5. J. Hwang, K. Yang, S. Sun, Reduction of flow-induced forces on a circular cylinder using a detached splitter plate. Phys. Fluids **15**, 2433 (2003)
6. H. Kawai, Discrete vortex simulation for flow around a circular cylinder with a splitter plate. J. Wind Eng. Ind. Aerodyn. **33**(1), 153–160 (1990)
7. K. Kwon, H. Choi, Control of laminar vortex shedding behind a circular cylinder using splitter plates. Phys. Fluids **8**, 479 (1996)
8. S. Lin, T. Wu, Flow control simulations around a circular cylinder by a finite volume scheme. Numer. Heat Trans. A Appl. **26**(3), 301–319 (1994)
9. Y. Nakamura, Vortex shedding from bluff bodies with splitter plates. J. Fluids Struct. **10**(2), 147–158 (1996)
10. S. Ozono, Flow control of vortex shedding by a short splitter plate asymmetrically arranged downstream of a cylinder. Phys. Fluids **11**, 2928 (1999)
11. S. Patankar, *Numerical Heat Transfer and Fluid Flow* (Hemisphere Publication, Washington, 1980)
12. A. Roshko, On the wake and drag of bluff bodies. J. Aeronaut. Sci. **22**(2), 124–132 (l955)
13. A. Saha, G. Biswas, K. Muralidhar, Three-dimensional study of flow past a square cylinder at low reynolds numbers. Int. J. Heat Fluid Flow **24**(1), 54–66 (2003)
14. S. Tiwari, D. Chakraborty, G. Biswas, P. Panigrahi, Numerical prediction of flow and heat transfer in a channel in the presence of a built-in circular tube with and without an integral wake splitter. Int. J. Heat Mass Trans. **48**(2), 439–453 (2005)
15. L. Zhou, M. Cheng, K. Hung, Suppression of fluid force on a square cylinder by flow control. J. Fluids Struct. **21**(2), 151–167 (2005)

Chapter 15
The Possibility of the Existence of Superluminal Neutrinos: A Theoretical Framework

Indranath Bhattacharyya

Abstract The amazing result of the first OPERA experiment has explored a possibility for the existence of superluminal neutrinos. Although the successive experiments would have negated such possibility, but in the framework of extended standard model a mechanism is proposed to address that superluminal nature of neutrinos. The idea is based on the assumption that the neutrinos might behave as superluminal in a particular experiment, whereas in the most of the experiments that remain subluminal. In the proposed model, the mass matrix of Dirac as well as Majorana neutrino field is diagonalized to obtain two eigen values; one becomes imaginary and larger in magnitude, whereas the other one is real but smaller. The resulting fields are found to be the mixture of left- and right-handed fields, unlike the concept of seesaw mechanism. The mass generation of neutrinos in the left-right symmetric model is examined. It is also proposed here that the oscillation of neutrinos is the two-fold process consisting of mass eigenstates doublet having imaginary mass and that with real mass; the domination of one over another results the subluminal as well as superluminal channel of neutrino oscillation.

Keywords Superluminal neutrinos · Mass generation

The first result of the OPERA experiment [1] (OPERA+[1]) has brought the idea of superluminal nature of neutrinos leading to a threat of the feasibility of special theory of relativity (STR) [4]. In the $\nu_\mu \rightarrow \nu_\tau$ channel of neutrino oscillations OPERA+ has measured the neutrino velocity as $(v-c)/c = [2.48 \pm 0.28\,(\text{stat}) \pm 0.30\,(\text{sys})] \times 10^{-5}$, contradicting the stringent limit obtained by SN1987A data in the energy range of few MeV [5–7]. It indicates that the muon neutrinos propagate faster than light in vacuum. The subsequent experiment OPERA- has negated the result of OPERA+ and rescued the special relativity. But the idea of superluminal nature of neutrinos motivated to build a consistent model without violating the STR. Under the circumstances, one can

[1] Throughout the literature OPERA+ stands for the result interpreting superluminal, whereas OPERA- [2, 3] is that in which neutrinos found to be subluminal as usual case.

I. Bhattacharyya (✉)
Department of Mathematics, Acharya Prafulla Chandra Roy Government College, Himachal Vihar, Matigara, Siliguri 734 010, West Bengal, India
e-mail: i_bhattacharyya@hotmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_15

think about a model which can incorporate both of the possibilities—superluminal as well as subluminal neutrinos. In the present article, a theoretical model is proposed in the backdrop of that superluminal issue. The motivation comes from the OPERA+ result. According to the STR, any particle having its velocity greater than that of light (Tachyon) must have purely imaginary mass; and therefore, it is not possible to detect such particle. Therefore, if the neutrinos are superluminal then they must be undetectable and should have imaginary mass contradicting the smallness of the neutrino mass, evident from various experiments. Then one can think there must be an underlying mechanism to have the neutrino velocity more than that of light in the $\nu_\mu \to \nu_\tau$ channel of OPERA+ experiment. In this article, a theoretical model is proposed to explain such mechanism.

15.1 Mechanism of the Imaginary Mass Generation of Neutrino

According to the well-accepted conventional seesaw mechanism, the left-handed neutrino acquires a very little mass; whereas, the right-handed counterpart have an extremely heavy mass. To address the superluminal issue, such understanding may be reconsidered. As the neutrino mass includes the Dirac as well as Majorana mass terms, the Lagrangian of the neutrino field incorporates such option. The neutrino mass matrix can be diagonalized to obtain the expression

$$2\mathscr{L} = m_1\overline{\phi}_1\phi_1 + m_2\overline{\phi}_2\phi_2 \tag{15.1}$$

with introducing the angle θ so that

$$\tan 2\theta = \frac{2m_D}{m_R - m_L} \tag{15.2}$$

The corresponding mass eigenvalues are calculated as

$$m_{1,2} = \frac{\varepsilon_{1,2}}{2}\left[(m_L + m_R) \mp \sqrt{(m_L - m_R)^2 + 4m_D^2}\right] \tag{15.3}$$

with $|\varepsilon_{1,2}| = 1$.

The fields ϕ_1 and ϕ_2 take the form

$$\phi_1 = \cos\theta\left(\nu_L + \varepsilon_1\nu_R^c\right) - \sin\theta\left(\nu_L^c + \varepsilon_1\nu_R\right) \tag{15.4}$$

$$\phi_2 = \sin\theta\left(\nu_L + \varepsilon_2\nu_R^c\right) + \cos\theta\left(\nu_L^c + \varepsilon_2\nu_R\right) \tag{15.5}$$

The ν_L and ν_R^c are the spinors of the active neutrinos; whereas, ν_R and ν_L^c are those which cannot be observed in nature. In the seesaw mechanism $m_R \gg m_D$ and

$m_L = 0$; the mass eigenvalues become $m_1 \approx \frac{m_D^2}{m_R}$ and $m_2 \approx m_R$. Here ϕ_1 is evolved as left-handed neutrino and right-handed antineutrino field, whereas ϕ_2 becomes right-handed neutrino and left-handed antineutrino field, which is essentially hard to observe since the mass eigenvalue associated to it is much high. In spite of the success in every other aspect the conventional seesaw fails to explain the superluminal nature of neutrinos. Therefore, a new model being able to interpret the OPERA+ result is needed. The basic assumption of this model is that the field ν_L and ν_R are symmetric in every respect; which follows the mass $m_L = m_R = m_M$ (say). Under the circumstances equation (15.3) gives the mass eigenvalues as

$$m_{1,2} = \varepsilon_{1,2}(m_M \mp m_D) \tag{15.6}$$

Since the theoretical framework is guided by the result of OPERA+ experiment the possibility of the existence of imaginary mass of the neutrino, causing superluminal effect, is to be taken into account. It is to be remembered that the neutrino is not superluminal all the time, because OPERA+ is the sole experiment in which the superluminal behavior has been observed. In all other instants, including OPERA- [2, 3], the neutrinos are found to be subluminal with tiny mass. To fit the criteria that the neutrinos may be superluminal as well as subluminal it is assumed m_D and m_M are much high in magnitude, but very close to each other and therefore the absolute value of $\mid m_D - m_M \mid > 0$ becomes very small. It results

$$m_1 =\mid m_D - m_M \mid \rightarrow 0 \qquad\qquad m_2 = i(m_D + m_M) \tag{15.7}$$

where, $\varepsilon_2 = i$ and $\varepsilon_1 = \mp 1$ accordingly m_D is greater or less than m_M so that the sign of m_1 must be positive.

If it is further assumed $\theta = -\frac{\pi}{4}$. The ϕ_1 and ϕ_2 fields become

$$\phi_1 = \phi_{Re} = \frac{1}{\sqrt{2}}\left[\nu + \nu^c\right] \tag{15.8}$$

$$\phi_2 = \phi_{Im} = \frac{1}{\sqrt{2}}\left[\left(\nu_L^c - \nu_L\right) + i\left(\nu_R - \nu_R^c\right)\right] \tag{15.9}$$

ϕ_{Re} is the real neutrino field having tiny mass. This is not necessarily left handed, the right-hand field is also included here. Such field is therefore clearly subluminal. But, the field ϕ_{Im} bears imaginary mass and becomes superluminal. The neutrino mass evolved in that sense is also seesaw in nature, although the $|m_2|$ may not be high enough. No doubt this is a new kind of seesaw mechanism, which is essentially different from the earlier concept of seesaw mechanisms. The framing of the model can be generalized for three flavors with 3×3 mass matrices M_D, M_L and M_R. That results doublet of mass eigenstates Φ_{Re}^k (bradyonic) and Φ_{Im}^k (tachyonic).

15.2 Discussion

Introducing right-handed neutrino in the standard model, the global $B-L$ symmetry becomes gaugable and the $SU(2)_L \times SU(2)_R \times U(1)_{B-L}$ becomes the gauge group of the left-right symmetric model [8, 9] and the seesaw structure of neutrino mass [10] emerges in this left-right symmetric model. According to the model proposed in this article, the neutrino mass generation may occur in the framework of $SO(4)$ model [11], by which no Majorana mass is created by the spontaneous breakdown of $SU(2)_L$ symmetry. Such mass term, which is approximately equal to the Dirac mass term may be incorporated by introducing dimension five Operator [12] in the unbroken Lagrangian. According to the conventional seesaw model, the right-handed neutrino is difficult to find in nature due to its extremely high mass. But in the model proposed here, one wing of a neutrino flavor cannot be found, not because of its high mass, but as it has imaginary mass resulting superluminal in nature. The other wing exists with its extremely small mass subject to the experimental verification of its right-handedness. In the framework of this model, both of the mass eigenstates are mixtures of left-handed as well as right-handed states. Therefore, unlike the earlier concept the flavor eigenstates are the mixtures of all complex mass eigenstates. In the OPERA+ experiment, the neutrino oscillation in the $\nu_\mu \rightarrow \nu_\tau$ has been taken into account. Here the flavor eigenstates ν_μ and ν_τ are the mixtures of $\Phi^1 = \Phi^1_{Re} + i\Phi^1_{Im}$ and $\Phi^2 = \Phi^2_{Re} + i\Phi^2_{Im}$, respectively. In the low energy limit, only the real part of Φ^1 and Φ^2 are expected to take part in the oscillation process. But in the intermediate stage one cannot exclude the possibility of the participation of Φ^1_{Im} and Φ^2_{Im} assuming $\left(E^k\right)^2 = (p)^2 - |m^k_1|^2$ $(k = 1, 2)$, where $p \gg |m^k_1|$ but E^k remains in the same energy range as that of initial ν_μ. Thus the neutrino oscillation may occur in two different channels, one is conventional $\left(\Phi^1_{Re}, \Phi^2_{Re}\right)$ occurring within our light cone and the other is $\left(\Phi^1_{Im}, \Phi^2_{Im}\right)$ taking place outside of it, without changing the overall energy range. If the oscillation event in the superluminal region is dominated by that of subluminal region, then the overall neutrino speed due to $\nu_\mu \rightarrow \nu_\tau$ oscillation crosses the luminal barrier; in the reverse case, the neutrino speed remains subluminal. The OPERA+ experiment found the neutrino speed greater than that of light since here superluminal event would dominate the subluminal one. Measuring the neutrino speed less than that of light by OPERA—experiment cannot lead to negate the revolutionary result of OPERA+, rather one can say that most of the cases the subluminal event dominates over the superluminal event.

References

1. T. Adam et al., arXiv:1109.4897 (2011) [first version]
2. T. Adam et al., arXiv:1109.4897 (2012) [revised version]
3. M. Antolleno et al., arXiv:1203.3433 (2012)
4. A. Einstein, Ann. Phys. **17**, 891 (1905)

5. J.R. Ellis, N. Harries, A. Meregaglia, A. Rubbia, A. Sakharov, Phys. Rev. D **78**, 033013 (2008). arXiv: 0805.0253 [hep-ph]
6. M.J. Longo, Phys. Rev. D **36**, 3276 (1987)
7. D. Fargion and D. D'Armiento, arXiv:1109.5368
8. B. Brahmachari, E. Ma, U. Sarkar, Phys. Rev. Lett. **91**, 011801 (2003)
9. R.N. Mohapatra, R.E. Marshak, Phys. Lett. **91 B**, 222 (1980)
10. R.N. Mohapatra, G. Senjanovic, Phys. Rev. Lett. **44**, 912 (1980)
11. I. Bhattacharyya, Commun. Theor. Phys. **54**, 305 (2010)
12. S. Weinberg, Phys. Rev. Lett. **43**, 1566 (1979)

Chapter 16
Dependence of Brans–Dicke Parameter on Scalar Field

Surajit Chattopadhyay and Sudipto Roy

Abstract For the scale factor as $a(t) = At^n e^{bt}$ we have discussed the behavior of the Brans–Dicke parameter ω as a function of scalar field $\phi(a) = \phi_1 e^{\alpha a}$. We have examined viability of the scale factor and subsequently studied its impact on ϕ as well as ω.

Keywords Brans–Dicke parameter

16.1 Introduction

Riess et al. [1] in the High-redshift Supernova Search Team and Perlmutter et al. [2] in the Supernova Cosmology Project Team have independently reported that the present universe is expanding with acceleration. Cosmological observations on expansion history of the universe can be interpreted as evidence either for existence of some exotic matter components or for modification of the gravitational theory. In the first route of interpretation, one can take a mysterious cosmic fluid with sufficiently large and negative pressure, dubbed dark energy. In the second route, however, one attributes the accelerating expansion to a modification of general relativity. The representative models belonging to the second class are known as "modified gravity" models, which include $f(R)$ gravity, $f(T)$ gravity, $f(G)$ gravity, $f(R, T)$ gravity, etc.

Brans–Dicke (BD) theory is a special case of scalar–tensor theories, which is originally motivated by the search for a theory containing Machs principle. The Brans–Dicke cosmology has been well studied considering different models. Sheykhi et al. [3] considered the power-law entropy corrected version of BD theory defined

S. Chattopadhyay (✉)
Pailan College of Management and Technology, Bengal Pailan Park, Kolkata 700104, India
e-mail: surajcha@iucaa.ernet.in; surajitchatto@outlook.com

S. Roy
Department of Physics, St. Xaviers' College, Kolkata 700016, India
e-mail: roy.sudipto1@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_16

by a scalar field ϕ and a coupling function ω. As the simplest and best-studied generalizations of general relativity, we have the Holographic DE (HDE) and the new agegraphic DE (NADE) models in the framework of BD cosmology.

16.2 Choice of Scale Factor

The scale factor, which is a function of cosmic time t, represents the relative expansion of the universe. It relates the proper distance (which can change over time) between a pair of objects, e.g., two galaxy clusters, moving with the Hubble flow in an expanding or contracting FRW universe at any arbitrary time to their distance at some reference time. The formula for this is $d(t) = d_0 a(t)$, where $d(t)$ is the proper distance at epoch t, d_0 is the distance at the reference time t_0, and $a(t)$ is the scale factor. In the present work, we have proposed the scale factor or the expansion factor in the from

$$a(t) = At^n \exp[bt], \tag{16.1}$$

where $A,\ n$ and b are positive constants considering $a(t)$ and $\dot{a}(t)$ to be positive quantities. The deceleration parameter q is defined as $q = -a\ddot{a}\dot{a}^{-2}$. For the above choice of scale factor, the deceleration parameter comes out to be

$$q(t) = -1 + \frac{n}{(n+bt)^2}, \tag{16.2}$$

We need to examine the cosmological viability of the above choice of scale factor. If we assume that t_1 is the time point, when the universe transits from decelerated to accelerated phase of expansion, then from (16.1) it can be derived that

$$t_1 = -\frac{n-\sqrt{n}}{b}, \tag{16.3}$$

If $a(t = t_1) = a_1$, then we have

$$a(t) = a_1 \left(\frac{t}{t_1}\right)^n \exp\left[(\sqrt{n}-n)\left(\frac{t}{t_1}-1\right)\right], \tag{16.4}$$

and

$$q(t) = -1 + n\left[n + (\sqrt{n}-n)\frac{t}{t_1}\right]^{-2} \tag{16.5}$$

For the scale factor (16.1), the Hubble parameter is

$$H = \frac{\dot{a}}{a} = \frac{\sqrt{n}-n}{t_1} + \frac{n}{t} \tag{16.6}$$

For the present epoch, i.e., $a = 1$ let us take $q = q_0$ and $H = H_0$. Then using (16.1)–(16.6) it can be obtained that

$$\frac{H}{H_0} = \frac{1 + n\left\{\left(\frac{C+D\exp[-a/f(a_1)]+1}{n}\right)^{-\frac{1}{2}} - n\right\}^{-1}}{1 + n\left\{\left(\frac{C+D\exp[-1/f(a_1)]+1}{n}\right)^{-\frac{1}{2}} - n\right\}^{-1}} \tag{16.7}$$

Finally, we have

$$a(t) = \left(\frac{t}{t_0}\right)^n \exp\left[(H_0 t_0 - n)\left(\frac{t}{t_0} - 1\right)\right] \tag{16.8}$$

$$q(t) = -1 + n\left[n + (H_0 t_0 - n)\frac{t}{t_0}\right]^{-2} \tag{16.9}$$

$$a_1 = \left(\frac{\sqrt{n} - n}{H_0 t_0 - n}\right)^n \exp(\sqrt{n} - H_0 t_0) \tag{16.10}$$

In Fig. 16.1, we have plotted the deceleration parameter against a, and we observed that the signature flip is happening roughly in the range $0.5 < a < 1$. This is consistent with the present accelerated universe. For some values of n and for $a = 1$,

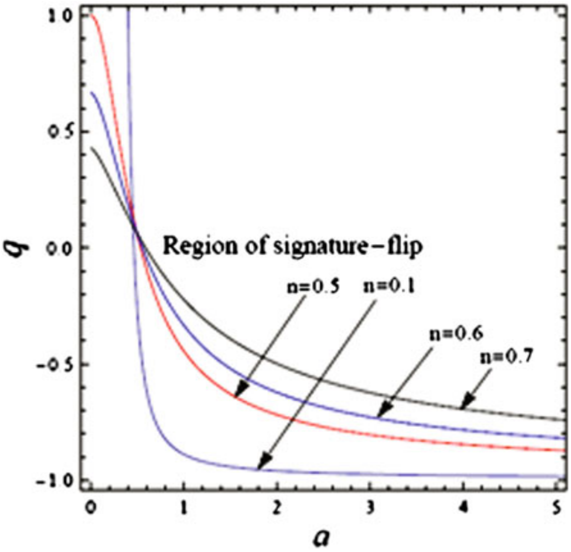


Fig. 16.1 The deceleration parameter shows transition from decelerated to accelerated phase of expansion for the scale factor (16.1)

we have $q = 0.2$. This result is consistent with the study of Giostri et al. [4], which states that combining BAO/CMB observations with SNIa data processed with the MLCS2k2 light-curve fitter gives $q_0 = -0.31 \pm 0.11$ at 68 % confidence level.

16.3 ω as function of ϕ

For flat FRW universe (which corresponds to a curvature parameter k equal to zero), the field equations in the generalized BD theory are given by

$$3H^2 = \frac{\rho}{\phi} + \frac{\omega(\phi)}{2}\left(\frac{\dot{\phi}}{\phi}\right)^2 - 3H\frac{\dot{\phi}}{\phi} \tag{16.11}$$

$$2\frac{\ddot{a}}{a} + H^2 = -\frac{\omega(\phi)}{2}\left(\frac{\dot{\phi}}{\phi}\right)^2 - 2H\frac{\dot{\phi}}{\phi} - \frac{\ddot{\phi}}{\phi} \tag{16.12}$$

where ρ represents the energy density of the matter distribution and an over-dot indicates a derivative with respect to the cosmic time t. The issue of considering ω as function of ϕ has been discussed in the work of Das and Banerjee [5] and has been further studied in [6].

We have considered the scalar field in the form

$$\phi(a) = \phi_1 \exp[\alpha a] \tag{16.13}$$

and matter density as

$$\rho = \rho_0 a^{-3} \tag{16.14}$$

In (16.13), there is no a priori physical motivation for this choice of ϕ. This is purely phenomenological, which leads to the desired behavior of the deceleration parameter q of attaining a negative value at the present epoch from a positive value during a recent past. Subsequently, using the solutions in the previous section and taking $f_0 = \rho_0 \phi_0^{-1}$, we can get the following quadratic equation

$$\alpha^2 + \left[6 - \frac{n}{H_0^2 t_0^2}\right]\alpha + \left[6 - \frac{2n}{H_0^2 t_0^2} - \frac{f_0 t_0^2}{H_0^2 t_0^2}\right] = 0; \quad \text{where } f_0 = \frac{\rho_0}{\phi_0} \tag{16.15}$$

The CMB measurements [1] put 1.05 as upper limit of the value of $H_0 t_0$. For this reason, we consider $H_0 t_0 = f_1 < 1$. It has been already established in the previous section that, in order to have a signature flip, we require $n < H_0^2 t_0^2 = f_1^2 < 1$. Hence, we can have $f_2 < 1$ such that $n = f_2 f_1^2$. We further define the parameter $f_3 = f_0 t_0^2$. Hence, Eq. (16.15) takes the form

$$\alpha^2 + (6 - f_2)\alpha + \left(6 - 2f_2 - \frac{f_3}{f_1^2}\right) = 0 \tag{16.16}$$

that leads to

$$\alpha_\pm = \frac{f_2 - 6 \pm \sqrt{f_2^2 - 4f_2 + 4f_3/f_1^2 + 12}}{2} \tag{16.17}$$

Let us denote:

$$\phi_+(a) = \phi_0 \exp[\alpha_+(a-1)] \tag{16.18}$$

$$\phi_-(a) = \phi_0 \exp[\alpha_-(a-1)] \tag{16.19}$$

Hence, from the second field equation we have the BD parameter as

$$\omega_\pm(\phi) = (4q-2)(\alpha_\pm a)^{-2} + (2q-4)(\alpha_\pm a)^{-1} - 2 \tag{16.20}$$

We shall now plot the BD parameter ω against the time t and $G = 1/\phi$ against the scale factor a. In the figures, we shall consider $H_0 t_0 = 0.95$. The blue and green lines would represent ϕ_+ and ϕ_-, respectively. The solid line corresponds to $n = 0.75$ and the dashed line corresponds to $n = 0.2$. We observe that, for the negative root, we have $\dot{G}/G > 0$ and for positive root $\dot{G}/G < 0$. However, $|\dot{G}/G|$ is increasing in both cases. The rate of increasing is sharper for $n = 0.60$ than for $n = 0.80$. For the current universe, for positive α we have $|\dot{G}/G| < 4\times 10^{-10}\, yr^{-1}$, which is well-documented upper limit of $|\dot{G}/G|$. However, for negative α, $|\dot{G}/G| > 4 \times 10^{-10}\, yr^{-1}$, and hence we discard the model with negative α (Figs. 16.2 and 16.3).

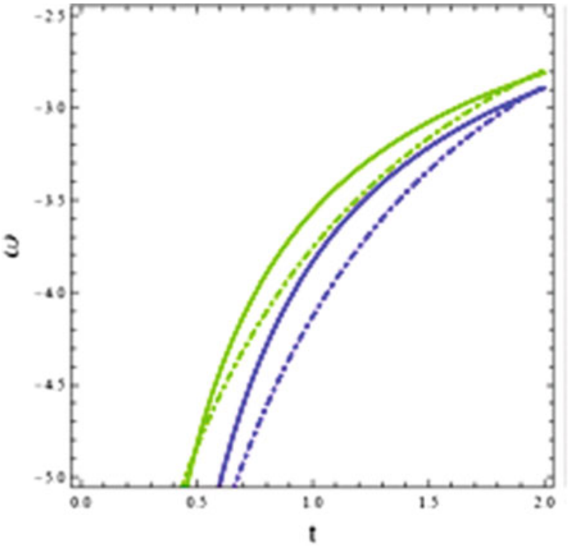
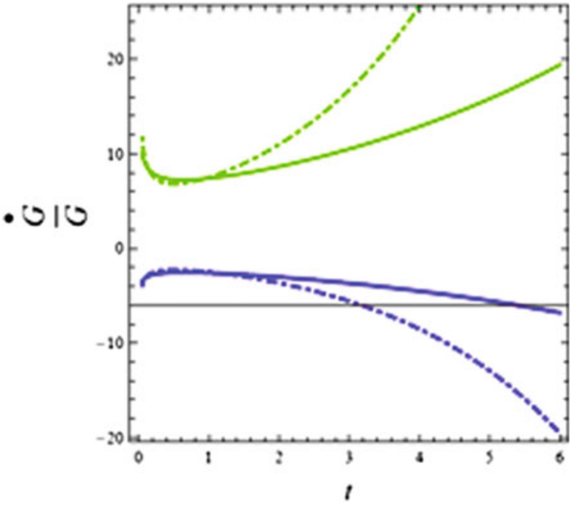


Fig. 16.2 Plot of the BD parameter against t

Fig. 16.3 $\dot{G}/G$ against t



16.4 Concluding Remarks

The present paper reports a study on the dimensionless parameter ω in Brans–Dicke theory. Based on a particular choice of scale factor a, we have investigated the signature flip of the deceleration parameter q to see whether the transition from decelerated to accelerated expansion of the universe is achievable under this choice of scale factor. Restrictions on the parameters obtained for this choice of scale factor have been subsequently used for discussing the Brans–Dicke parameter for scalar field $\phi(a) = \phi_1 \exp[\alpha a]$.

Acknowledgments Surajit Chattopadhyay acknowledges financial support from DST, Govt. of India under Fast Track Project Grant No. SR/FTP/PS-167/2011.

References

1. A.G. Riess et al., Astron. J. **116**, 1009 (1998)
2. S. Perlmutter et al., Astrophys. J. **517**, 565 (1999)
3. A. Sheykhi, K. Karami, M. Jamil, E. Kazemi, M. Haddad, Int. J. Theor. Phys. **51**, 1663 (2012)
4. R. Giostri, M.V. dos Santos, I. Waga, R.R.R. Reis, M.O. Calvao, B.L. Lago, JCAP **03** 027 (2012)
5. S. Das, N. Banerjee, Phys. Rev. D **78**, 043512 (2008)
6. S. Das, N. Banerjee, Gen. Relativ. Gravit. **38**, 785 (2006)

Chapter 17
Water-Wave Scattering by a Sphere in a Two-Layer Fluid with an Ice-Cover

Dilip Das and Nityananda Thakur

Abstract Using linear water–wave theory, wave scattering (both heave and sway) by a sphere submerged in a two-layer ocean consisting of a layer of fresh water of finite depth with an ice-cover and an infinite layer of salt water. The sphere is submerged in the lower layer of the two layers. Employing the method of multipoles, the problem is reduced to an infinite system of linear equations which are solved numerically by standard technique after truncation. The vertical and horizontal exciting forces on the sphere are obtained and depicted graphically against the wave number for various values of the submersion depths of the sphere in the lower layer in a number of figures to show the effect of the presence of ice-cover.

Keywords Water wave scattering · Two-layer fluid · Ice-cover

17.1 Introduction

Problems on water-wave radiation and scattering by spherical objects have been studied in the literature due to their importance in the construction of wave power devices. Linton and McIver [1] developed a general theory for two-dimensional wave propagation in such a two-layer fluid with a free surface and studied wave scattering by a long horizontal cylinder submerged in either of the two layers. Cadby and Linton [2] extended this work to three dimensions and investigated wave radiation and diffraction by a sphere submerged in either of the two layers. In two-layer fluid wherein the upper layer is of finite depth and bounded above by a thin but uniform layer of ice-cover modeled as a thin elastic sheet and the lower layer is infinitely

D. Das (✉)
Department of Mathematics, Shibpur Dhinbundhoo Institution (College),
412/1, G.T. Road, Howrah 711102, India
e-mail: dilipdas99@gmail.com

N. Thakur
Durga Charan Rakshit Banga Vidyalaya, Surpara, Bagbazar,
Chandannagar, Hooghly, India
e-mail: nitya.iitm@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_17

deep below the interface, time-harmonic waves with given frequency can propagate with two different wave numbers. Computations in [3] demonstrate this for the long horizontal circular cylinder submerged in either layer of a two-layer fluid with an ice-cover. Also Das and Mandal [4] recently studied the water-wave radiation by submerged sphere in either layers of two-layer fluid using the method of multipoles. The two-layer fluid water-wave problem arose from modeling an underwater pipe bridge across Norwegian fjords consisting of a layer of fresh water on the top of a deep layer of salt water. During winter, the fjords are covered by ice, and this has motivated us to extend the problem of [2] to an ice-covered two-layer fluid wherein the ice-cover is modeled as a thin elastic plate. Using linear water–wave theory, we consider wave scattering by a sphere submerged in the lower layer of the two-layer fluid. Employing the method of multipoles, the problem is reduced to an infinite system of linear equations which are solved numerically by standard technique after truncation. The vertical and horizontal forces on the sphere are obtained and depicted graphically against the wave number for various values of the submersion depths of the sphere in the lower layer in a number of figures to show the effect of the presence of ice-cover.

17.2 Mathematical Formulation

A Cartesian coordinate system is chosen in which the y-axis points vertically upwards with the plane $y = 0$ denoting the undisturbed interface of a two-layer ocean with an ice-cover. The plane $y = h$ denotes the mean position of the ice-cover, the lower fluid extends infinitely downwards. The fluid in each layer is assumed to be inviscid and incompressible. Under the usual assumptions of linear theory and irrotational motion, a velocity potential $\Phi(x, y, z, t) = Re\left\{\phi(x, y, z)\mathrm{e}^{-i\omega t}\right\}$ describing the fluid motion exists, where $\phi(x, y, z)$ is a complex valued function and ω is the angular frequency. The upper layer $0 < y < h$, is referred to as region I, and the lower layer $y < 0$ as region II. Let the potential in the upper layer of density ρ_I be ϕ_I^m and that in the lower layer of density ρ_{II} be ϕ_{II}^m ($m = 0, 1$, the potential functions for the heave and sway problems being denoted by ϕ^0 and ϕ^1 respectively). The potential functions ϕ_I^m, ϕ_{II}^m satisfy

$$\nabla^2\phi_I^m = 0, \quad \nabla^2\phi_{II}^m = 0, \quad \text{in appropriate layers.} \tag{17.2.1}$$

The ratio of the densities of the two fluids, ρ_I/ρ_{II} (<1), is denoted by ρ. The linearized boundary conditions at the interface are

$$\phi_{Iy}^m = \phi_{IIy}^m \quad \text{on} \quad y = 0, \tag{17.2.2}$$

$$\rho(\phi_{Iy}^m - K\phi_I^m) = \phi_{IIy}^m - K\phi_{II}^m, \quad \text{on} \quad y = 0. \tag{17.2.3}$$

The linearized ice-cover condition is

$$\left(D\nabla^4_{x,z} + 1 - \varepsilon K\right)\phi^m_{Iy} + K\phi^m_I = 0, \quad \text{on} \;\; \mathrm{y} = \mathrm{h}, \tag{17.2.4}$$

where

$$\nabla^4_{x,z} = \frac{\partial^4}{\partial x^4} + 2\frac{\partial^2}{\partial x^2}\frac{\partial^2}{\partial z^2} + \frac{\partial^4}{\partial z^4}, \quad K = \frac{\omega^2}{g}, \quad D = \frac{L}{\rho_I g},$$

where

$$L = \frac{Eh_0^3}{12\left(1-\nu^2\right)}$$

being the flexural rigidity of the elastic ice-cover, E being the Young's modulus and ν being the Poisson's ratio of the material of the ice-cover and $\varepsilon = \frac{\rho_0}{\rho_I}h_0$, ρ_0 being the density of the ice and h_0 being the very small thickness of the ice-cover. The boundary conditions (17.2.2) and (17.2.3) are obtained from the continuity of the normal velocity and pressure at the interface, respectively. Also the condition at large depth is

$$\nabla\phi^m_{II} \to 0 \quad \text{as} \;\; y \to -\infty. \tag{17.2.5}$$

Now the total scattering potential can be decomposed into two parts:

$$\phi = \phi_{\text{inc}} + \phi_s \tag{17.2.6}$$

where ϕ_{inc} is the potential representing the incident plane wave and ϕ_s must satisfy (17.2.1)–(17.2.5) and body boundary condition

$$\frac{\partial\phi_s}{\partial r} = -\frac{\partial\phi_{\text{inc}}}{\partial r} \qquad \text{on} \quad r = a, \tag{17.2.7}$$

and behave as an outgoing cylindrical wave far from the sphere. Without loss of generality, we can assume that the incident wave is from $x = -\infty$ so that $\alpha_{\text{inc}} = 0$.

17.3 Sphere in the Lower Layer

The center of the sphere is taken at $(0, f, 0)$ $(f < 0)$. Spherical polar coordinates (r, θ, α), are defined by $x = r\sin\theta\cos\alpha$, $y = f + r\cos\theta$, $z = r\sin\theta\sin\alpha$, where θ is the angle made with the upward vertical and α is the azimuthal angle. The multipole potentials ϕ^m_{In} and ϕ^m_{IIn} are constructed as (cf. [4])

$$\phi^m_{In} = \frac{a^{n+1}}{(n-m)!}\int_0^\infty k^n V(k) J_m(kR)\mathrm{d}k, \tag{17.3.1}$$

$$\phi^m_{IIn} = \left(\frac{a}{r}\right)^{n+1} P^m_n(\cos\theta) + \frac{a^{n+1}}{(n-m)!}\int_0^\infty k^n C(k) e^{ky} J_m(kR)\mathrm{d}k, \tag{17.3.2}$$

where $R = \left(x^2 + z^2\right)^{1/2}$, $V(k) = A(k)\mathrm{e}^{ky} + B(k)\mathrm{e}^{-yk}$ and

$$A(k) = K\left\{k\left(Dk^4 + 1 - \varepsilon K\right) + K\right\}\mathrm{e}^{k(f-h)}/H(k), \tag{17.3.3}$$

$$B(k) = K\left\{k\left(Dk^4 + 1 - \varepsilon K\right) - K\right\}\mathrm{e}^{k(f+h)}/H(k), \tag{17.3.4}$$

$$C(k) = (KsH_1(k) - ((1-s)k + K)H_2(k))\,\mathrm{e}^{kf}/H(k), \tag{17.3.5}$$

$$H(k) = KsH_1(k) - ((1-s)k - K)H_2(k) \tag{17.3.6}$$

with

$$H_1(k) = k\left(Dk^4 + 1 - \varepsilon K\right)\cosh kh - K\sinh Kh$$
$$H_2(k) = k\left(Dk^4 + 1 - \varepsilon K\right)\sinh kh - K\cosh Kh.$$

The path of integration in the integrals in (17.3.1) and (17.3.2) is indented below the poles at $k = \lambda_j$, $j = 1, 2$, where λ_1 and λ_2 ($\lambda_1 < \lambda_2$) are the only two real positive roots of the dispersion equation $H(k) = 0$.

The far-field form of the multipole, in the lower layer, is given by

$$\phi^m_{IIn} \sim \frac{(-i)^{m+1}a^{n+1}}{(n-m)!}\left(\frac{2\pi}{R}\right)^{\frac{1}{2}}\left(\lambda_1^{n-1/2}C_{\lambda_1}\mathrm{e}^{i\lambda_1 R + \lambda_1 y} + \lambda_2^{n-1/2}C_{\lambda_2}\mathrm{e}^{i\lambda_2 R + \lambda_2 y}\right)\mathrm{e}^{-i\pi/4} \tag{17.3.7}$$

as $R \to \infty$, where C_{λ_1} and C_{λ_2} are the residues of $C(k)$ at $k = \lambda_1$ and $k = \lambda_2$, respectively, which are given by

$$C_{\lambda_j} = \left(KsH_1(\lambda_j) - ((1-s)\lambda_j + K)H_2(\lambda_j)\right)\mathrm{e}^{\lambda_j f}/H'(\lambda_j), \quad j = 1, 2. \tag{17.3.8}$$

Using the result

$$\mathrm{e}^{\pm k(y-f)}J_m(kR) = (\pm 1)^m \sum_{q=m}^{\infty}\frac{(\pm kr)^q}{(q+m)!}P^m_q(\cos\theta) \tag{17.3.9}$$

where $P^m_q(\cos\theta)$ are associated Legandre functions, $J_m(z)$ is the Bessel function of first kind, (17.3.2) can be expanded in terms of polar coordinates as

$$\phi^m_{IIn} = \left(\frac{a}{r}\right)^{n+1} P^m_n(\cos\theta) + \sum_{q=m}^{\infty} A^m_{nq} r^q P^m_q(\cos\theta), \tag{17.3.10}$$

where

$$A_{nq}^{m} = \frac{1}{(n-m)!(q+m)!}\int_{0}^{\infty} k^{q+n} C(k) \mathrm{e}^{kf}\, \mathrm{d}k. \tag{17.3.11}$$

Incident wave train of wave number λ_1

We consider an incident plane wave of wave number λ_1 and amplitude A on the ice-covered surface $y = h$ whose potential can be expanded in spherical polar coordinates and get

$$\phi_{\text{inc}} = -\frac{igAK}{\omega\lambda_1}\mathrm{e}^{\lambda_1 y}\mathrm{e}^{i\lambda_1 R\cos\alpha} \tag{17.3.12}$$

$$= -\frac{igAK}{\omega\lambda_1}\mathrm{e}^{\lambda_1 f}\sum_{m=0}^{\infty}\varepsilon_m i^m \cos m\alpha \sum_{q=m}^{\infty}\frac{(\lambda_1 r)^q}{(q+m)!}P_q^m(\cos\theta), \tag{17.3.13}$$

where $\varepsilon_0 = 1, \varepsilon_m = 2$ for $m \geq 1$. For the scattering problems considered here, the dependence on the azimuthal angle α is unknown and so we must use a more general multipole expansion. We write

$$\phi_S = -\frac{igAK}{\omega\lambda_1}\sum_{m=0}^{\infty}\sum_{n=m_1}^{\infty} c_n^m\, \phi_n^m \cos m\alpha, \tag{17.3.14}$$

where $m_1 = max(m, 1)$ and ϕ_n^m is given (in the lower fluid layer) by (17.3.10). If we then apply the boundary condition (17.2.7) and use the orthogonality of the associated Legendre functions and the functions $\cos m\alpha$, we can derive an infinite system of equations for the sets of coefficients c_n^m, $n \geq m_1$ for each $m \geq 0$, which is

$$c_q^m - \frac{q}{q+1}\sum_{n=m_1}^{\infty} A_{ns}^m\, c_n^m = \frac{\varepsilon_m i^m qKa(\lambda_1 a)^{q-1}}{(q+1)(q+m)!}\mathrm{e}^{\lambda_1(f-h)}, \quad q \geq m_1. \tag{17.3.15}$$

These systems can be solved by truncation as before, but now there is an additional truncation parameter, namely the number of systems that are solved. In the computations presented below, two 4 × 4 systems were solved.

The hydrodynamic force on the body in the ith mode of motion can be written $F_i(t) = Re\left\{f_i \mathrm{e}^{-i\omega t}\right\}$, where f_i is found by integrating the dynamic pressure times the appropriate component of the normal over the body surface. In other words,

$$f_i = i\rho^{II}\omega\int_{S_B}\phi n_i \mathrm{d}s,$$

where S_B is the body boundary and n_i is the component of the inward normal to the body in the ith mode of motion. The vertical and horizontal exciting forces on the sphere, $\bar{f}_{\lambda_1}^0$ and $\bar{f}_{\lambda_1}^1$, can be obtained as

$$\bar{f}^0_{\lambda_1} = -\frac{4}{3}\Pi a^2 \rho_{II} g A(Kae^{\lambda_1(f-h)} + c^0_1 + \sum_{n=1}^{\infty} A^0_{n1}\, c^0_n), \tag{17.3.16}$$

and

$$\bar{f}^1_{\lambda_1} = -\frac{4}{3}\Pi a^2 \rho_{II} g A(iKae^{\lambda_1(f-h)} + c^1_1 + \sum_{n=1}^{\infty} A^1_{n1}\, c^1_n). \tag{17.3.17}$$

These can be simplified using (17.3.18) with $q = 1$ giving

$$f^0_{\lambda_1} = |\frac{\bar{f}^0_{\lambda_1}}{a^2 \rho_{II} g A}| = 4\pi |c^0_1|, \tag{17.3.18}$$

and

$$f^1_{\lambda_1} = |\frac{\bar{f}^1_{\lambda_1}}{a^2 \rho_{II} g A}| = 4\pi |c^1_1|. \tag{17.3.19}$$

The constants c^0_1 appearing in (17.3.18) and c^1_1 appearing in (17.3.19) can be obtained numerically by solving the linear system (17.3.15) after truncation. Here the linear system (17.3.15) is truncated up to four terms.

Incident wave train of wave number λ_2

Now, we consider the case of an incident plane wave of amplitude A on the interface $y = 0$ and the wave number λ_2 described by

$$\phi^I_{\text{inc}} = -\frac{igAK}{\omega\lambda_2}\, g_2(y)\mathrm{e}^{i\lambda_2 R\cos\alpha}, \tag{17.3.20}$$

$$\phi^{II}_{\text{inc}} = -\frac{igAK}{\omega\lambda_2}\, \mathrm{e}^{\lambda_2 y + i\lambda_2 R\cos\alpha}, \tag{17.3.21}$$

where

$$g_2(y) = \frac{\{\lambda_2(1-\rho) - K\}}{K\rho H_1(\lambda_2)}\left[M_1\mathrm{e}^{\lambda_2(y-h)} + M_2\mathrm{e}^{-\lambda_2(y-h)}\right]$$

where $M_1 = \lambda_2(D\lambda_2^4 + 1 - \varepsilon K) + K$ and $M_2 = \lambda_2(D\lambda_2^4 + 1 - \varepsilon K) - K$. The analysis is very similar to that given above for an incident wave of wave number λ_2 we use the same expansion for ϕ_s as before, Eq. (17.3.14), but denote the unknown coefficients by d^m_n, and we obtain the infinite system of equations

$$d^m_q - \frac{q}{q+1}\sum_{n=m_1}^{\infty} A^m_{ns}\, d^m_n = \frac{\varepsilon_m i^m qKa(\lambda_2 a)^{q-1}}{(q+1)(q+m)!}\mathrm{e}^{\lambda_2 f}, \qquad q \geq m_1, \tag{17.3.22}$$

for each $m \geq 0$.

The expressions for the vertical and horizontal exciting forces are

$$f^0_{\lambda_2} = |\frac{\bar{f}^0_{\lambda_2}}{a^2 \rho_{II} g A}| = 4\pi \, |d^0_1|, \tag{17.3.23}$$

and

$$f^1_{\lambda_2} = |\frac{\bar{f}^1_{\lambda_2}}{a^2 \rho_{II} g A}| = 4\pi \, |d^1_1|. \tag{17.3.24}$$

The constants d^0_1 appearing in (17.3.23) and d^1_1 appearing in (17.3.24) can be obtained numerically by solving the linear system (17.3.22) after truncation. Here the linear system (17.3.22) is truncated up to four terms. This provides an accuracy up to five decimal places, because if the system is truncated up to five or six terms, there is practically no change in the numerical results.

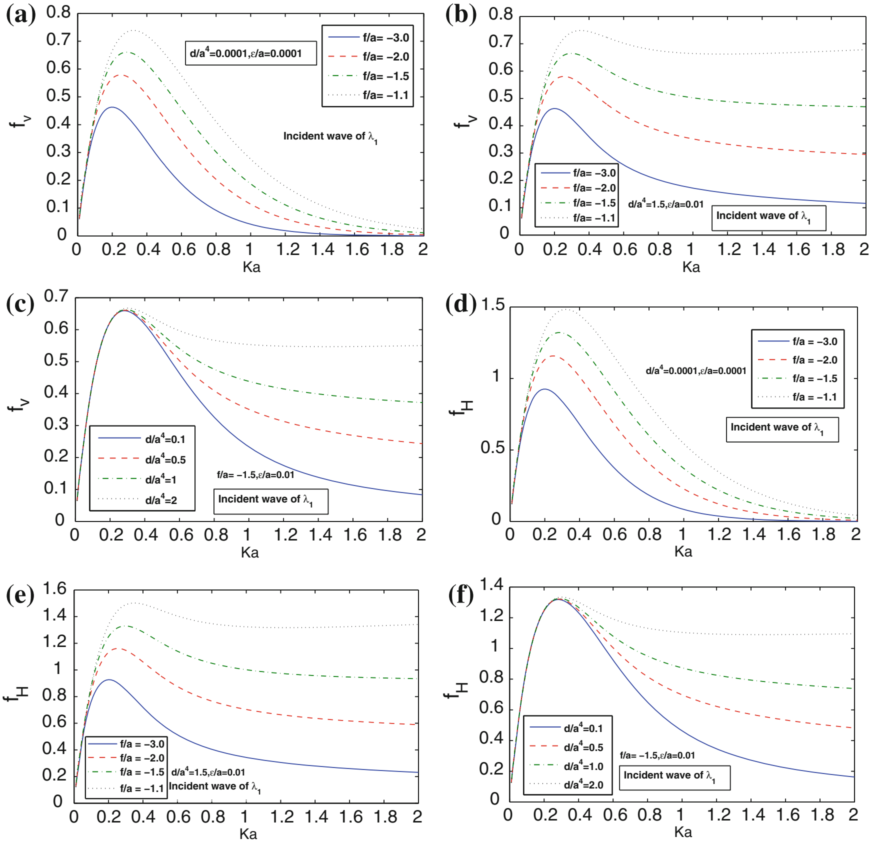


Fig. 17.1 **a**–**c** Vertical force against the wave number ka for a sphere in the lower layer. **d**–**f** Horizontal force against the wave number ka for a sphere in the lower layer

For the sphere in the upper layer, the expressions for vertical and horizontal forces have also been obtained but are not presented here.

17.4 Numerical Results

Figure 17.1a–f show, respectively, the nondimensional vertical and horizontal exciting forces on the sphere due to an incident wave of wave number λ_1 for $\rho = 0.95$, $h/a = 2$. Figure 17.1a, d show f_V and f_H, plotted against Ka for $D/a^4 = 0.0001$, $\varepsilon/a = 0.0001$ and different values of f/a and similarly Fig. 17.1b, d for $D/a^4 = 1.5$, $\varepsilon/a = 0.01$ and different values of f/a and Fig. 17.1c, f for $f/a = -1.5$, $\varepsilon/a = 0.01$ and different values of D/a^4. The curves of vertical and horizontal forces are very similar and show that, as one would expect, the forces increase the closer the sphere is to the interface. From figures we observed that as Ka increases, vertical and horizontal forces increase, attain a maximum value and then decrease as Ka increases. The vertical and horizontal forces are somewhat larger in comparison to that of [2]. This is due to the presence of ice-cover. When the flexural rigidity is taken to be very small, the numerical results for these quantities almost coincide with those for a two-layer ocean with a free surface (cf. [2]).

The vertical and horizontal exciting forces for the case of an incident wave of wave number λ_2 are not shown here. They can of course be determined. The forces are an order of magnitude smaller than the corresponding forces due to an incident wave of wave number λ_1 and display the same qualitative effects as the sphere approaches the interface.

17.5 Conclusion

In this paper, we have studied the problem of water-wave scattering by a sphere submerged in the lower layer of a two-layer ocean. The upper layer is of finite depth and is bounded above by an ice-cover which is modeled as a thin elastic plate and the lower layer extends infinitely downwards. The method of multipoles is employed to study this problem. The problem of water wave scattering by a body of an arbitrary shape is somewhat difficult to solve analytically. However, a nearly spherical body can be treated by an approximate method.

Acknowledgments The authors thank the Reviewer for his comments and suggestions to revise the paper.

References

1. C.M. Linton, P. McIver, The interaction of waves with horizontal cylinder in two-layer fluids. J. Fluid Mach. **304**, 213–229 (1995)
2. J.R. Cadby, C.M. Linton, Three-dimensional water wave scattering in two-layer fluids. J. Fluid Mech. **423**, 155–173 (2000)
3. D. Das, B.N. Mandal, Wave scattering by a horizontal circular cylinder in a two-layer fluid with an ice-cover. Int. J. Eng. Sci. **45**, 842–872 (2007)
4. D. Das, B.N. Mandal, Wave radiation by a sphere submerged in a two-layer ocean with an ice-cover. Appl. Ocean Res. **32**, 358–366 (2010)

Chapter 18
An Improved Adomian Decomposition Method for Nonlinear ODEs

Prakash Kumar Das and M.M. Panja

Abstract This work deals with getting approximate solution of boundary value problem consists of nonlinear ordinary differential equations in a series of exponential instead of power of independent variable in traditional Adomian decomposition method (TADM). As a consequence: (i) in contrast to TADM the vanishing boundary condition for localized solution can be implemented in a straightforward way, (ii) the convergence of the series obtained through the modification proposed here found to be faster than the same obtained by employing TADM, and (iii) for most of the problems, the sum of the series converges to the exact analytic solution to the equation involved. The efficiency of the modification of TADM has been illustrated for physical problems with varied nonlinearities.

Keywords Nonlinear ordinary differential equation · Boundary value problem · Improved Adomian decomposition method

18.1 Introduction

In many branches of applied mathematics, physical, biological, and engineering sciences, evolution of physical processes are found to be described by nonlinear ordinary or partial differential equations (ODEs/PDEs). The solution of such equations helps one to understand the nature of evolution of the process. But in most of the cases, it is not possible to find the exact solution to the equation used as the mathematical model for the description of the physical process of interest. A few analytical methods such as symmetry method based on Lie theory [1, 2], Prelle-Singer method [3], method based on Jacobi last multiplier [4], etc., analytical methods for approximate solution such as tanh method [5, 6], homotopy analysis method (HAM) [7, 8], Adomian decomposition method (ADM) [9–21], etc., numerical methods, viz., finite

P.K. Das (✉) · M.M. Panja
Department of Mathematics, Visva-Bharati, Santiniketan 731235, West Bengal, India
e-mail: prakashdas.das1@gmail.com

M.M. Panja
e-mail: madanpanja2005@yahoo.co.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_18

difference/element methods are used to find the solution of this problems. Among the approximation methods mentioned above, ADM is found to be the simplest one. Using ADM, Adomian and his collaborators [9–14], Wazwaz [15–21] as well as other researchers obtained the approximate solutions as the sum of finite number of terms with the leading term as the polynomial in independent variable involved in the problem. But in their approach, the boundary condition in case of infinite domain cannot be implemented in a straight forward way. Instead, it is desirable to express the successive terms in their approximate solution as a rational function with the help of Padé approximant to accommodate boundary conditions. Naturally, question arises whether straightforward method can be developed which is able to provide a rapidly convergent series approximation of the solution to the differential equation involving the physical processes that incorporate boundary conditions at $\pm\infty$ in a straightforward way in both cases of finite as well as infinite domain.

In this paper, we have addressed this problem and developed an recursive scheme for solving two-point nonlinear boundary value problems through a modification of the conventional ADM. Here we have introduced an operator associated with the linear part of the differential equation and derived a straightforward formula involving such operator for correction terms associated to the nonlinear part of the equation. We designate this method as the improved Adomian decomposition method (IADM), provides the solution in a series of exponentials instead of power of independent variable, appears in case of conventional ADM. Expansion in series of exponential perhaps is the source of accelerated convergence of the method proposed here.

The organization of this paper is as follows. The improved Adomian decomposition method (IADM) within finite domain has been discussed in Sect. 18.2. Its extension to infinite domain has been presented in Sect. 18.3. Our findings on utility of the proposed IADM developed in previous two sections have been illustrated in Sect. 18.4.

18.2 IADM in Finite Domain $[a, b]$

We consider here a two-point boundary value problem of the form

$$y''(x) - \lambda^2 y(x) = \mathscr{N}[y](x) + g(x), \; a \leq x \leq b \tag{18.1}$$

within finite domain $[a, b]$ subject to the Dirichlet boundary condition

$$y(a) = \alpha, \; y(b) = \beta \tag{18.2}$$

where $N[y]$ is an nonlinear term in y, and $g(x)$ is the inhomogeneous or source term, continuous over $[a, b]$. Instead of shifting the linear term $\lambda^2 y(x)$ of (18.1) into R.H.S in conventional ADM, we incorporate it into the operator $\hat{\mathscr{O}}[\cdot] \equiv \frac{d^2}{dx^2} - \lambda^2$, so that (18.1) can now be recast into the form

$$\hat{\mathscr{O}}[y](x) = \mathscr{N}[y](x) + g(x),\ a \le x \le b. \tag{18.3}$$

It is important to mention here that the linear operator $\hat{\mathscr{O}}[\cdot]$ can be written in the form

$$\hat{\mathscr{O}}[\cdot](x) = \mathrm{e}^{\lambda x} \frac{\mathrm{d}}{\mathrm{d}x} \left(\mathrm{e}^{-2\lambda x}\right) \frac{\mathrm{d}}{\mathrm{d}x} \left(\mathrm{e}^{\lambda x}[\cdot]\right) \tag{18.4}$$

which plays the fundamental role in expressing the solution in terms of rapidly convergent series of exponentials. One may reinterpret the inverse operator $\hat{\mathscr{O}}^{-1}$ as a twofold integral operator given by

$$\hat{\mathscr{O}}^{-1}[\cdot](x) = \mathrm{e}^{-\lambda x} \int_a^x \mathrm{e}^{2\lambda x'} \int_a^{x'} \mathrm{e}^{-\lambda x''}[\cdot](x'')\mathrm{d}x''\mathrm{d}x'. \tag{18.5}$$

Note that representing inverse operator by integrals for a linear operator with variable coefficient is also possible whenever it is factorizable. Application of $\hat{\mathscr{O}}^{-1}$ given in (18.5) to $y''(x) - \lambda^2 y(x)$, one gets

$$\begin{aligned}\hat{\mathscr{O}}^{-1}\left[y''(x) - \lambda^2 y(x)\right] &= \mathrm{e}^{-\lambda x} \int_a^x \mathrm{e}^{2\lambda x'} \int_a^{x'} \mathrm{e}^{-\lambda x''}\left(y''(x'') - \lambda^2 y(x'')\right)\mathrm{d}x''\mathrm{d}x' \\ &= \mathrm{e}^{-\lambda x} \int_a^x \mathrm{e}^{2\lambda x'} \\ &\quad \left(\mathrm{e}^{\lambda x'} y'(x') - \mathrm{e}^{-\lambda a} y'(a) + \lambda \mathrm{e}^{-\lambda x'} y(x') - \lambda \mathrm{e}^{\lambda a} y(a)\right)\mathrm{d}x' \\ &= y(x) - y(a)\mathrm{e}^{-\lambda(a-x)} - \mathrm{e}^{-\lambda a}\left(y'(a) + \lambda y(a)\right)\left(\frac{\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}}{2\lambda}\right).\end{aligned} \tag{18.6}$$

Operating $\hat{\mathscr{O}}^{-1}$ on both sides of (18.3) followed by using (18.6) one gets

$$\begin{aligned} y(x) = y(a)\mathrm{e}^{-\lambda(a-x)} &+ \mathrm{e}^{-\lambda a}\left(y'(a) + \lambda y(a)\right)\left(\frac{\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}}{2\lambda}\right) \\ &+ \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](x) + \hat{\mathscr{O}}^{-1}[g](x),\end{aligned} \tag{18.7}$$

which involve an unknown term $y'(a)$. To eliminate $y'(a)$, we substitute $x = b$ in Eq. (18.7) and solve for $e^{-\lambda a}\left(y'(a) + \lambda y(a)\right)$ to get

$$\mathrm{e}^{\lambda a}\left(y'(a) + \lambda y(a)\right) = \frac{2\lambda\left(y(b) - y(a)\mathrm{e}^{-\lambda(a-b)} - \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](b) - \hat{\mathscr{O}}^{-1}[g](b)\right)}{\mathrm{e}^{\lambda b} - \mathrm{e}^{-\lambda b}}. \tag{18.8}$$

Eliminating $\mathrm{e}^{-\lambda a}\left(y'(a) + \lambda y(a)\right)$ from (18.7) with the help of (18.8) gives the expression for $y(x)$ involving inverse operator

$$y(x) = y_0(x) - \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](b)\frac{\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}}{\mathrm{e}^{\lambda b} - \mathrm{e}^{-\lambda b}} + \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](x). \tag{18.9}$$

One can now apply the relevant steps of ADM for evaluating terms involving nonlinear operator $\mathscr{N}[y](x)$ where leading term $y_0(x)$ is given by

$$y_0(x) = y(a)\mathrm{e}^{-\lambda(a-x)} + \frac{y(b) - y(a)\mathrm{e}^{-\lambda(a-b)} - \hat{\mathscr{O}}^{-1}[g](b)}{\mathrm{e}^{\lambda b} - \mathrm{e}^{-\lambda b}}\left\{\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}\right\} + \hat{\mathscr{O}}^{-1}[g](x). \tag{18.10}$$

The successive corrections can be obtained recursively using the formula

$$y_{n+1}(x) = \hat{\mathscr{O}}^{-1}\left[\mathscr{A}_n\right](x) - \frac{\hat{\mathscr{O}}^{-1}\left[\mathscr{A}_n\right](b)}{\left(\mathrm{e}^{\lambda b} - \mathrm{e}^{-\lambda b}\right)}\left(\mathrm{e}^{\lambda x} - \mathrm{e}^{-\lambda x}\right), \ n \leq 0, \tag{18.11}$$

where $A_n(x), \ n \geq 0$ are Adomain polynomial for nonlinear term given by the formula

$$\mathscr{A}_m(x) = \frac{1}{m!}\left[\frac{d^m}{d\varepsilon^m}\mathscr{N}\left(\sum_{k=0}^{\infty} y_k \varepsilon^k\right)\right]_{\varepsilon=0}, \ m \geq 0. \tag{18.12}$$

18.3 IADM in Infinite Domain

Whenever the domain of independent variable become infinite, we write the inverse operator $\hat{\mathscr{O}}^{-1}$ as a twofold integral operator without limit given by

$$\hat{\mathscr{O}}^{-1}[\cdot](x) = \mathrm{e}^{-\lambda x}\int \mathrm{e}^{2\lambda x}\int e^{-\lambda x}[\cdot](x)\,\mathrm{d}x\,\mathrm{d}x. \tag{18.13}$$

In this case, operation of $\hat{\mathscr{O}}^{-1}$ on $y''(x) - \lambda^2 y(x)$ gives

$$\begin{aligned}\hat{\mathscr{O}}^{-1}\left(y''(x) - \lambda^2(x)\right) &= \mathrm{e}^{-\lambda x}\int \mathrm{e}^{2\lambda x}\int \mathrm{e}^{-\lambda x}\left(y''(x) - \lambda^2 y(x)\right)\,\mathrm{d}x\,\mathrm{d}x \\ &= \mathrm{e}^{-\lambda x}\int \mathrm{e}^{2\lambda x}\left(\mathrm{e}^{-\lambda x}y'(x) + \lambda\mathrm{e}^{-\lambda x}y(x) + c\right)\,\mathrm{d}x \\ &= y(x) + \frac{c}{2\lambda}\mathrm{e}^{\lambda x} - d\mathrm{e}^{-\lambda x}. \end{aligned} \tag{18.14}$$

involving two arbitrary constants c and d. Operating $\hat{\mathscr{O}}^{-1}$ on both sides of $\hat{\mathscr{O}}^{-1}[y](x) = N[y](x) + g(x)$ and use of (18.14), leads to

$$y(x) = -\frac{c}{2\lambda}\mathrm{e}^{\lambda x} + d\mathrm{e}^{-\lambda x} + \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](x) + \hat{\mathscr{O}}^{-1}[g](x). \tag{18.15}$$

Assuming $\lambda > 0$ and using the vanishing boundary condition $y(\infty) = 0$ for localized solution of (18.1) within $[0, \infty)$, we can obtain $c = 0$. Thus

$$y(x) = d\mathrm{e}^{-\lambda x} + \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](x) + \hat{\mathscr{O}}^{-1}[g](x),\ x \in [0, \infty). \tag{18.16}$$

The correction to the leading order due to presence of nonlinearities are obtained by executing steps followed in conventional ADM with

$$y_{n+1}(x) = \hat{\mathscr{O}}^{-1}\left[\mathscr{A}_n\right](x),\ n \geq 0 \tag{18.17}$$

with

$$y_0(x) = d\mathrm{e}^{-\lambda x} + \hat{\mathscr{O}}^{-1}[g](x), \tag{18.18}$$

where $\mathscr{A}_n(x),\ n \geq 0$ are Adomain polynomial for nonlinear term can be obtained using the formula (18.12). It is important to note that whenever the domain becomes $(-\infty, 0]$, instead of using vanishing boundary condition $y(\infty) = 0$, for localized solution (18.15) we use $y(-\infty) = 0$ and get

$$y(x) = -\frac{c}{2\lambda}\mathrm{e}^{\lambda x} + \hat{\mathscr{O}}^{-1}[\mathscr{N}[y]](x) + \hat{\mathscr{O}}^{-1}[g](x),\ x \in (-\infty, 0] \tag{18.19}$$

so that higher order corrections over leading order approximation

$$y_0(x) = -\frac{c}{2\lambda}\mathrm{e}^{\lambda x} + \hat{\mathscr{O}}^{-1}[g](x) \tag{18.20}$$

can obtained recursively form

$$y_{n+1}(x) = \hat{\mathscr{O}}^{-1}[\mathscr{A}]_n(x),\ n \geq 0. \tag{18.21}$$

In case of $\lambda < 0$, one has to proceed in the same way by retaining the term involving $\mathrm{e}^{\lambda x}$.

18.4 Illustrative Example

Our findings on getting approximate solution for nonlinear ODEs within finite and infinite domain by using IADM proposed here have been summarized in Tables 18.1 and 18.2, respectively.

Table 18.1 Solution of nonlinear ODE in finite domain by IADM

Problem	Linear term	Nonlinear term	Leading order approximation	Series solution	Exact solution
$y''(x) - 2y(x) = 2y(x)^3$ with boundary condition $y(0) = 0,\ y\left(\frac{\pi}{4}\right) = -1$	$y''(x) - 2y(x)$	$2y(x)^3$	$-\frac{\sinh\left(\sqrt{2}x\right)}{\sinh\left(\sqrt{2}\frac{\pi}{4}\right)}$	$-\frac{\sinh\left(\sqrt{2}x\right)}{\sinh\left(\sqrt{2}\frac{\pi}{4}\right)} + 0.0347465\sin\left(\sqrt{2}x\right)$ $-\frac{1}{32}\cosh\left(\frac{\pi}{2\sqrt{2}}\right)^3\left(-12\sqrt{2}x\operatorname{csch}\left(\sqrt{2}x\right)\right)$ $+9\sinh\left(\sqrt{2}x\right) + \sinh\left(3\sqrt{2}x\right) + \cdots$	$y(x) = -\tan(x)$

Table 18.2 Solution of nonlinear ODEs in infinite domain by IADM

Equations	Korteweg-de Varies equation $u_t + 6uu_x + u_{xxx} = 0$	Zakharov equation $iE_t + E_{xx} = \eta E$ $\eta_{tt} - \eta_{xx} = \left(\lvert E\rvert^2\right)_{xx}$	Camassa-Holm equation $u_t + 2ku_x - u_{xxt} + 3uu_x$ $-2u_x u_{xx} - uu_{xxx} = 0$
Similarity variables	$\xi = c(x - vt)$ $u(x,t) = U(\xi)$	$E(x,t) = \mathrm{e}^{iv}u(\xi)$ $\eta = \eta(\xi)$ $v = cx + \mathrm{d}t$ $\xi = x - 2ct$	$\xi = (x - ct),$ $u(x,t) = v(\xi) - k$
Reduced ODE	$U''(\xi) - \dfrac{v}{c^2}U(\xi) = -\dfrac{3}{c^2}U(\xi)^2$	$u''(\xi) - \left(c^2 + d\right)u(\xi) = \dfrac{u(\xi)^3}{\left(4c^2 - 1\right)}$	$v''(\xi) - v(\xi) =$ $\frac{1}{2(k+c)}\left(v'(\xi)^2 + 2v(\xi)v''(\xi) - 3v(\xi)^2\right)$
Leading order approximation	$de^{-\frac{\sqrt{v}}{c}\xi}$	$\gamma e^{-\sqrt{(c^2+d)\xi}}$	$de^{-\xi}$ when $\xi \succ 0$ $de^{+\xi}$ when $\xi \prec 0$
Sum of series	$u(x,t) = \frac{v}{2}\operatorname{sech}^2\left(\frac{\sqrt{v}}{2c}(x - vt) + m\right)$ where $m = -\frac{1}{2}\log\left(\frac{d}{2v}\right)$	$E(x,t) = \sqrt{2\left(1 - 4c^2\right)\left(c^2 + d\right)}\, e^{i(cx+\mathrm{d}t)}$ $\operatorname{sech}\left(\sqrt{c^2 + d}\,(x - 2ct) + m\right)$ $\eta(x,t) =$ $2\left(c^2 + d\right)\operatorname{sech}^2\left(\sqrt{c^2 + d}\,(x - 2ct) + m\right)$ where $m - \log\left(\frac{\gamma}{\sqrt{8}(1-4c^2)(c^2+d)}\right)$	$u(x,t) =$ $\begin{cases} de^{-(x-ct)} - k \text{ for } \xi \succ 0 \\ de^{+(x-ct)} - k \text{ for } \xi \prec 0 \end{cases}$

18.5 Conclusions

In this work, an improvement over conventional ADM has been proposed. The consequence is to get an approximate solution of nonlinear ODE in the series of exponential. As a result, the approximate solution become rapidly convergent and found to converges to the exact analytic solution for both kind of problems defined over bounded and unbounded domains. From this study, it also appears that conventional ADM can further be improved for problem consists of variable coefficient in their linear part in order to get rapidly convergent approximate solution of nonlinear ODEs used as mathematical models for physical processes.

References

1. P.J. Olver, *Applications of Lie Groups to Differential Equations*, 2nd edn. (Springer, New York, 1993)
2. N.H. Ibragimov, *CRC Handbook of Lie Group Analysis of Differential Equations*, vols. 1, 2, 3 (CRC Press, Boca Ratan, 1994, 1996)
3. A.G. Choudhury, P. Guha, B. Khanra, Solutions of some second order ODEs by the extended Prelle-Singer method and symmetries. J. Nonlin. Math. Phys. **15**, 365–382 (2008)
4. M.C. Nucci, P.G.L. Leach, Jacobis last multiplier and the complete symmetry group of the Ermakov-Pinney equation. J. Nonlinear Math. Phys. **12**, 305–320 (2005)
5. W. Malfliet, Solitary wave solutions of nonlinear wave equations. Am. J. Phys. **60** (1992)
6. W. Malfliet, W. Hereman, The Tanh method: l. Exact solutions of nonlinear evolution and wave equations. Phys. Scripta. **54**, 563–568 (1996)
7. S. Abbasbandy, E. Magyari, E. Shivanian, The homotopy analysis method for multiple solutions of nonlinear boundary value problems. Commun. Nonlinear Sci. Numer. Simulat. **14**, 3530–3536 (2009)
8. S. Xinhui, Z. Liancun, Z. Xinxin, S. Xinyi, Homotopy analysis method for the asymmetric laminar flow and heat transfer of viscous fluid between contracting rotating disks. Appl. Math. Model. **36**, 1806–1820 (2012)
9. G. Adomian, *Solving Frontier Problems of Physics: The Decomposition Method* (Kluwer, 1994)
10. J.S. Duan, R. Rach, A new modification of the Adomian decomposition method for solving boundary value problems for higher order nonlinear differential equations. Appl. Math. Comput. **218**, 4090–4118 (2011)
11. G. Adomian, R. Rach, Inversion of nonlinear stochastic operators. J. Math. Anal. Appl. **91**, 39–46 (1983)
12. G. Adomian, R. Rach, Analytic solution of nonlinear boundary-value problems in several dimensions by decomposition. J. Math. Anal. Appl. **174**, 118–137 (1993)
13. G. Adomian, R. Rach, A new algorithm for matching boundary conditions in decomposition solutions. Appl. Math. Comput. **58**, 61–68 (1993)
14. G. Adomian, R. Rach, Modified decomposition solution of linear and nonlinear boundary-value problems. Nonlinear Anal. **23**, 615–619 (1994)
15. A.M. Wazwaz, Approximate solutions to boundary value problems of higher order by the modified decom-position method. Comput. Math. Appl. **40**, 679–691 (2000)
16. A.M. Wazwaz, The modified Adomian decomposition method for solving linear and nonlinear boundary value problems of 10th-order and 12th-order. Int. J. Nonlinear Sci. Numer. Simul. **1**, 17–24 (2000)

17. A.M. Wazwaz, A reliable algorithm for obtaining positive solutions for nonlinear boundary value problems. Comput. Math. Appl. **41**, 1237–1244 (2001)
18. A.M. Wazwaz, The numerical solution of fifth-order boundary value problems by the decomposition method. J. Comput. Appl. Math. **136**, 259–270 (2001)
19. A.M. Wazwaz, The numerical solution of sixth-order boundary value problems by the modified decomposition method. Appl. Math. Comput. **118**, 311–325 (2001)
20. A.M. Wazwaz, A reliable algorithm for solving boundary value problems for higher-order integro-differential equations. Appl. Math. Comput. **118**, 327–342 (2001)
21. A.M. Wazwaz, The numerical solution of special fourth-order boundary value problems by the modified decomposition method. Int. J. Comput. Math. **79**, 345–356 (2002)

Chapter 19
Numerical Algorithm for Computation of Complete Theoretical Seismogram in Layered Half-Space Media

Ajit De

Abstract A numerically stable computational scheme, using method of quadrature, has been presented in this study to compute surface response or theoretical seismogram in an n-layered vertically stratified media overlying a half-space with constant layer parameters. The spherical shape has been ignored in the present earth model. Simple buried source model has been considered. The present result has been compared with the observed or previously computed seismograms. The overflow error appearing in the numerical computation has been prevented by approximating layer matrices suitably or using generalized R/T (Reflection and Transmission) coefficients. The numerical result has been represented here graphically. The present study can be considered as first step toward computation of hazard map of a seismic region.

Keywords Theoretical seismogram · Runge-Kutta method · Generalized reflection · Transmission coefficients

19.1 Introduction

An elastic half-space is a simple model of earth. But earth's interior is inhomogeneous and divided into various inhomogeneous layers, including the crust with nonuniform P- and S-wave velocities. So it is the target of the researchers to construct efficient models which include earth's inhomogeneity, source geometry, and travel time of seismic waves. Hisada [6] proposed an analytical method for effective computation of displacement and stress of static (i.e., circular frequency $\omega = 0$) and dynamic (i.e., $\omega \neq 0$) Green's functions in a viscoelastic layered half-space model. The generalized R/T (Reflection and Transmission) coefficients, as proposed by Apsel and Luco [1], have been modified in the model to overcome the problems of receiver-source close depths and overflow. Desceliers et al. [2] presented a fast-hybrid numerical method to simulate transient wave propagation due to given transient loads in a multilayered

A. De (✉)
Department of Mathematics, Siliguri College, Siliguri 734001, West Bengal, India
e-mail: ajit_math@rediffmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_19

semi-infinite media. The method is based on time domain formulation associated with a 2D-space Fourier transform for two infinite layer dimensions and uses finite element approximations. Watson [7] suggested a faster machine computation by introducing modified matrix formulas for modal solution in a layered elastic half-space.

In the present study, the spherical shape of the earth has been ignored and it has been considered as a system of n-parallel vertically stratified multilayered model overlying a half-space where each layer parameter is constant. The Laplace transformed surface displacement field, as proposed by Harkrider [5], has been evaluated here by considering a method of quadrature based on local sampling of the kernel with a quartic polynomial [1]. The result in the time domain has been obtained through Fourier synthesis. The overflow error in numerical computation has been avoided here using modified R/T coefficients [1] and approximating hyperbolic functions suitably [3]. The numerical computation has been represented here graphically.

19.2 Formulation of the Problem and Basic Equations

A vertically stratified n-layered media overlying a half-space has been considered. The origin of the reference system has been considered on the surface of the media with xy-plane horizontal and z-axis directed inside it. The layer parameters λ, μ (Lame's constants) and density ρ of each layer are constant. The displacement vectors $\vec{u} = \vec{u}(r, \theta; t)$ in each layer satisfy the differential equation

$$\vec{\nabla}\left[(\lambda + 2\mu)\,\vec{\nabla}\bullet\vec{u}\right] - \vec{\nabla}\times\left[\mu\,\vec{\nabla}\times\vec{u}\right] + 2\left[\left(\vec{\nabla}\,\mu\,\vec{\nabla}\right)\vec{u} + \vec{\nabla}\,\mu\times\left(\vec{\nabla}\times\vec{u}\right)\right] = \rho\frac{\partial^2\vec{u}}{\partial t^2} \tag{19.1}$$

A source has been considered at a depth "h" below the surface as a time-dependent stress discontinuity $\Delta(t)$at the source layer S as

$$\begin{pmatrix} U_p^{S+}(h) \\ D_p^{S+}(h) \end{pmatrix} = \begin{pmatrix} U_p^{S-}(h) \\ D_p^{S-}(h) \end{pmatrix} + \begin{pmatrix} 0 \\ \Delta(t) \end{pmatrix}, \quad (p = \text{PSV or SH}) \tag{19.2}$$

where S+ and S- are, respectively, the sublayers below and above the source.

The dynamic displacement-stress vectors $\left(U_p^j(z, h, k)\right.$ and $D_p^j(z, h, k)$, $p = $ PSV or SH$\left.\right)$ in the jth layer of a layered half-space media can be expressed in terms of down and up going P and S waves using modified R/T coefficients [1] as

$$\begin{pmatrix} U_p^j(z, h, k) \\ D_p^j(z, h, k) \end{pmatrix} = \begin{pmatrix} E_{11}^j & E_{12}^j \\ E_{21}^j & E_{22}^j \end{pmatrix} \begin{pmatrix} \Lambda_d^j(z) & 0 \\ 0 & \Lambda_u^j(z) \end{pmatrix} \begin{pmatrix} C_d^j(h) \\ C_u^j(h) \end{pmatrix} \tag{19.3}$$

where $C_d^j(h)$ and $C_u^j(h)$ are, respectively, the down and up going coefficients and E^j as layer matrix in the jth layer. Now using Eqs. (19.1) and (19.2) and considering the displacement-stress continuity at the layer boundaries other than the source, the Laplace transformed radial and tangential components of surface displacement (Harkrider [5], Hisada [6]) on the surface ($z = 0$) of the media can be expressed as

$$\overline{u}_r(r,\theta) = \int_0^\infty \left[U^1_{PSV}(0,h,k)\frac{dJ_1(kr)}{dkr} + U^1_{SH}(0,h,k)\frac{J_1(kr)}{kr} \right] dk \cos\theta \quad (19.4)$$

$$\overline{u}_\theta(r,\theta) = -\int_0^\infty \left[U^1_{PSV}(0,h,k)\frac{J_1(kr)}{kr} + U^1_{SH}(0,h,k)\frac{dJ_1(kr)}{d(kr)} \right] dk \sin\theta \quad (19.5)$$

where U^j_{PSV} (z,h,k) and U^j_{SH} (z,h,k) are, respectively, the components of displacement in the jth layer due to PSV and SH waves, obtained form the layer matrix product using modified R/T coefficients and (r,θ) is the polar coordinate of the receiver on the free surface.

19.3 Discussions

A numerically stable scheme which is free from overflow error has been presented here. It has been observed that the integrands U^1_{PSV} $(0,h,k)$ and U^1_{SH} $(0,h,k)$ are well behaved at large wave numbers and the tail ends of the wave number integrals can be evaluated without complexity. Now to evaluate the integrations in (19.4) and (19.5) numerically, the upper limit of the integrations must be truncated to a finite value K_m. As the source and the receiver are at different depths, the exponential decay of the kernels $U^1_{PSV}(0,h,k)$ and U^1_{SH} $(0,h,k)$ are sufficient to guarantee that the truncated integral gives an accurate estimate to the total integral if K_m is selected in such a way that $K_m h \gg 1$ or giving K_m a finitely large value so that integrals differ negligibly for neighboring values of K_m.

The method of quadrature is based on sampling the kernel U^1_{PSV} $(0,h,k)$ or U^1_{SH} $(0,h,k)$ in such a way that it can be represented locally by a quartic polynomial of the form

$$U^1(0,h,k) = \sum_{q=1}^{5} \sum_{l=1}^{5} C_{ql}\, U^1_l \left(\frac{k-k_2}{\Delta k}\right)^{q-1}, \; k_1 \le k \le k_5 \quad (19.6)$$

[U^1 represents U^1_{PSV} $(0,h,k)$ or U^1_{SH} $(0,h,k)$] where the normalization constant $\Delta k = k_4 - k_2$, U^1_l represents $U^1(0,h,k)$ and C_{ql} are functions of the sampling points $k_l (l = 1, 5)$ [1].

The contribution of the interval (k_2, k_4) to the total integral in (19.4) or (19.5) can be approximated by

$$\Delta\overline{u}_f = \Delta\overline{u}_{f\ PSV} + \Delta\overline{u}_{f\ SH},\ (f = r \text{ or } \theta), \tag{19.7}$$

where

$$\Delta\overline{u}_{r\ PSV} = \sum_{l=1}^{5} U_l^1 \left(\frac{dJ_1(kr)}{d(kr)}\right)_l \cos\theta \sum_{q=1}^{5} \frac{C_{ql}\ \Delta k}{q}$$

$$\text{and } \Delta\overline{u}_{r\ SH} = \sum_{l=1}^{5} U_l^1 \frac{J_1\ (k_l r)}{k_l r} \cos\theta \sum_{q=1}^{5} \frac{C_{ql}\ \Delta k}{q}.$$

Now to get the result in the time domain, the integrations described above in (19.4) and (19.5) for each frequency are followed by a Fourier synthesis or FFT

$$u_f^*(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{u}_f \exp(i\omega t) d\omega,\ (where\ f = r, \theta) \tag{19.8}$$

The theoretical SH-seismogram due to a triangular source model (Fig. 19.1) in a layered half-space medium has been found to represent the time-dependent response of Borrego mountain earthquake event and our result agrees with the result of Franssens [4] at a point (60, 0) on the surface in xz-plane (Fig. 19.2). The model (i.e., Fig. 19.1) consists of a sediment layer of thickness 2.9 km and a time-dependent triangular source is located at a depth 9 km below the earth's surface. Figure 19.3 represents a theoretical SH-seismogram due to the same triangular source model but in an inhomogeneous medium with depth dependent linear variation of layer parameters. The inhomogeneity in the above model has been modeled by subdividing the medium into a finite number of isotropic sublayers with small values of layer thickness and decreasing values of layer parameters upto the free surface, so that the result converges. But one disadvantage in further sublayering is that it increases the possibility of overflow error in numerical computation. Another alternative technique to model vertical inhomogeneity is to solve the following ordinary differential equation of first order by Runge–Kutta method of order 4 in the jth layer.

$$\frac{d}{dz}\begin{pmatrix} U_p^j(z,h,k) \\ D_p^j(z,h,k) \end{pmatrix} = A \begin{pmatrix} U_p^j(z,h,k) \\ D_p^j(z,h,k) \end{pmatrix},\ (p = \text{PSV or SH}) \tag{19.9}$$

where A is either a 4 × 4 or 2 × 2 matrix, respectively, representing PSV and SH-wave in the jth layer [8].

A complete theoretical or synthetic seismogram at the same point (60, 0) on the surface has been presented in Fig. 19.4 for the one layer half-space earth model as described in Fig. 19.1. The numerical algorithm has been presented in the study through simple computer programming.

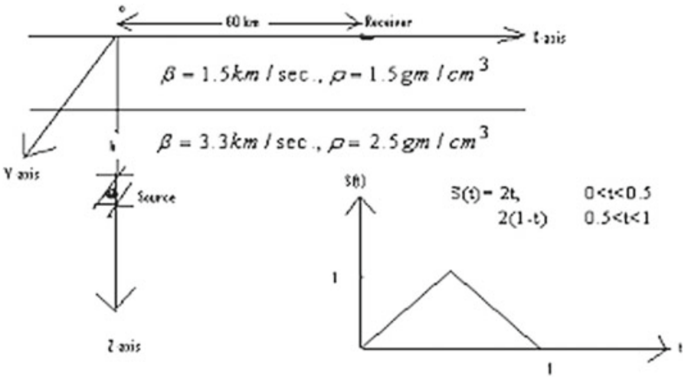

O
60 km
Receiver
X axis
$\beta = 1.5 km/sec., \rho = 1.5 gm/cm^3$
$\beta = 3.3 km/sec., \rho = 2.5 gm/cm^3$
Y axis
Source
Z axis
S(t)
S(t) = 2t, 0<t<0.5
2(1-t) 0.5<t<1
1
1
t

Fig. 19.1 One-layer Half-space model of Borrego mountain earthquake [4]

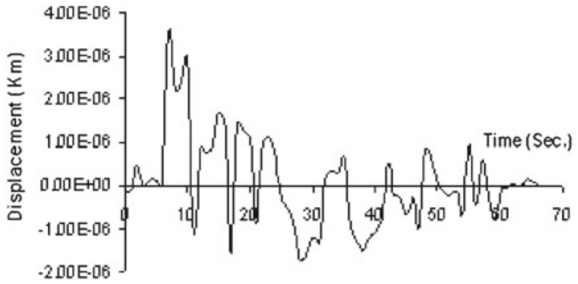


Fig. 19.2 Theoretical SH-seismogram due to a buried triangular source model (Fig. 19.1) at the receiver (60, 0), on the surface

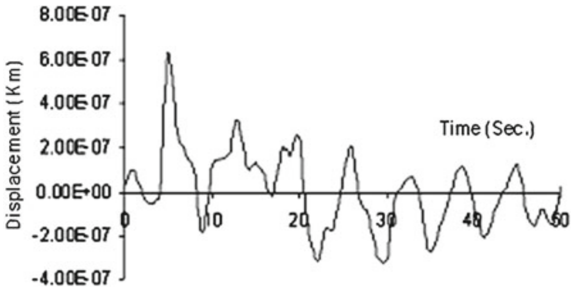


Fig. 19.3 Theoretical SH-seismogram due to the buried triangular source model at the same receiver (60, 0) on the surface of an inhomogeneous medium with linear variation of parameters

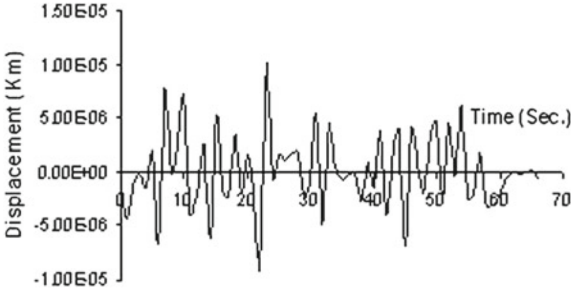


Fig. 19.4 Complete theoretical seismogram due to a buried triangular source model (Fig. 19.1) at the receiver (60, 0), on the surface

References

1. R.J. Apsel, J.E. Luco, On some Green's functions for a layered half-space. Bull. Seism. Soc. Am. **73**, 931–951 (1983)
2. C. Desceliers, C. Soize, Q. Grimal, G. Haiat, S. Naili, A time domain method to solve transient elastic wave propagation in a multilayer medium with a hybrid spectral-finite element space approximation. Wave Motion **45**(3), 383–399 (2008)
3. N. Florsch, D. Fah, P. Suhadolc, G. F. Panza, Complete synthetic seismogram for high-frequency multimode SH-waves. Pure Appl. Geophys. **136**(4), 529–560 (1991)
4. G.R. Franssens, Calculation of the elasto-dynamic Green's function in layered media by means of a modified propagator matrix method. Geophys **75**, 669–691 (1983)
5. D.G. Harkrider, Surface waves in multilayered elastic Snplacemedia SnI. Rayleigh and Love waves from buried sources in a multi-layerd elastic half-space. Bull. Seism. Soc. Am. **54**, 627–679 (1964)
6. Y. Hisada, An efficient method for computing Green's functions for a layered half-space with sources and receivers at close depths. Bull. Seism. Soc. Am. **84**, 1456–1472 (1994)
7. T.H. Watson, A note on fast computation of Rayleigh wave dispersion in the multilayered elastic half-space. Bull. Seism. Soc. Am. **60**(1), 161–166 (1970)
8. L. Zhu, L.A. Rivera, A note on the dynamic and static displacements from a point source in multilayered media. Geophys. J. Int. **148**, 619–627 (2002)

Chapter 20
Aleatory and Epistemic Uncertainty Quantification

Palash Dutta and Tazid Ali

Abstract In this paper, an effort has been made to combine aleatory and epistemic uncertainties in risk models. We have combined probabilistic distributions, generalized fuzzy numbers, and completely generalized interval valued fuzzy numbers.

Keywords Aleatory and epistemic uncertainty · Generalized fuzzy numbers · Interval valued fuzzy numbers · Risk assessment

20.1 Introduction

In some situations both aleatory and epistemic uncertainties co-exist in risk assessment. Then, it is important to develop special techniques, which can handle propagation of uncertainties (i.e., fuzzy and random), for carrying out risk assessment. Different attempts have been made by different researchers for joint propagation of aleatory and epistemic uncertainty in the same computation of risk viz., Guyonnet et al. [18, 19], Baudrit et al. [3–6], Kentel and Aral [24], Anoop et al. [1], Li et al. [25], Rao et al. [30], Baraldi and Zio [2], Helton and Oberkampf [23], Limbourg and de Rocquigny [26], Flage et al. [15, 16], Dutta and Ali [11], Haldar and Reddy [20], Pedroni et al. [28, 29]. Here, it is seen that more often representation of epistemic uncertainty is considered as Type-I fuzzy set. However, it is not always possible for a membership function of the type to precisely assign one point from [0, 1] so it is more realistic to assign interval value. According to Gehrke et al. [17] many people believe that assigning an exact number to experts opinion is too restrictive and the assignment of an interval valued is more realistic. Hence, it is necessary to go through interval valued fuzzy set (IVFS) to handle such situations. In May, 1975 Sambuc [31] presented in his doctoral research (thesis) the concept of IVFS named as fuzzy set. After development of IVFVs, different researchers have studied

P. Dutta (✉) · T. Ali
Department of Mathematics, Dibrugarh University, Dibrugarh 786004, India
e-mail: palash.dtt@gmail.com

T. Ali
e-mail: tazidali@yahoo.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_20

this issue and applied in different areas. An IVFS is a set in which every element has degree of membership in the form of an interval. One can say, IVFS consist of two membership function, one is upper membership function (UMF) and other is lower membership function (LMF). Dutta [12] presented an approach to combined probability distributions, type-I fuzzy set (normal fuzzy numbers) and generalized fuzzy numbers, Chutia [7] made an effort for combining probability distribution and interval valued fuzzy number and applied to environmental risk modeling with a case study, Dutta [13] also presented an approach to combine probability distributions, normal fuzzy numbers and generalized interval valued fuzzy numbers. In this paper, an approach has been proposed to combine probabilistic distributions, generalized fuzzy numbers, and completely generalized interval valued fuzzy numbers. A case study in risk assessment has been carried out in this setting.

20.2 Basic Concept of Fuzzy Set Theory

In this section, some necessary backgrounds and notions [10, 21, 22] of fuzzy set theory that will be required in the sequel are reviewed.

20.2.1 Let X be a universal set. Then, the fuzzy subset A of X is defined by its membership function

$$\mu_A : X \to [0, 1]$$

which assign a real number $\mu_A(x)$ in the interval [0, 1], to each element $x \in A$, where the value of $\mu_A(x)$ at x shows the grade of membership of x in A.

20.2.2 Given, a fuzzy set A in X and any real number $\alpha \in [0, 1]$. Then the α-cut of A, denoted by

$$\alpha A = \{x \in X : \mu_A(x) \geq \alpha\}$$

20.2.3 The support of a fuzzy set A defined on X is a crisp set defined as

$$\mathrm{Supp}(A) = \{x \in X : \mu_A(x) > 0\}$$

20.2.4 The height of a fuzzy set A, denoted by $h(A)$ is the largest membership grade obtain by any element in the set and it is denoted as

$$h(A) = \mathrm{Sup}_{x \in X}[\mu_A(x)]$$

20.2.5 Generalized fuzzy numbers (GFN): The membership function of GFN $A = [a, b, c, d; w]$ where $a \leq b \leq c \leq d$, $0 < w < 1$ is defined as [8, 9]

$$\mu_A(x) = \begin{cases} 0 & x < a \\ w\frac{x-a}{b-a} & x \in [a, b] \\ w & x \in [b, c] \\ w\frac{d-x}{d-c} & x \in [c, d] \\ 0 & x > d \end{cases} \tag{20.1}$$

If $w = 1$, then GFN A is a normal trapezoidal fuzzy number $A = [a, b, c, d]$. If $a = b$ and $c = d$, then A is a crisp interval. If $b = c$ then A is a generalized triangular fuzzy number. If $a = b = c = d$ and $w = 1$ then A is a real number. Compared to normal fuzzy number the GFN can deal with uncertain information in a more flexible manner because of the parameter w that represent the degree of confidence of opinions of decision makers.

20.2.6 An IVFS A defined in the universe of discourse X is represented by

$$A = \{(x, [\mu_A^L(x), \mu_A^L(x)]) : x \in X\}$$

where $0 \leq \mu_A^L(x) \leq \mu_A^U(x) \leq 1$ and the membership grade $\bar{\mu}_A(x)$ of elements of x to the IVFS A is represented by an interval $[\mu_A^L(x), \mu_A^L(x)]$, i.e., $(\bar{\mu}_A(x) = [\mu_A^L(x), \mu_A^U(x)])$.

20.2.7 If an IVFS A satisfies the following properties

◇ A is normal
◇ A is defined in a closed bounded interval
◇ A is convex set

Then, A is called an interval valued fuzzy number.
20.2.8 α-cut of IVFN: A generalization of α-cut of IVFS is

$$\alpha A = \left\{x : \mu_A^L(x) \geq \alpha, \mu_A^U(x) \geq \alpha\right\}$$

20.3 Approach to Quantify Aleatory and Epistemic Uncertainty

In uncertainty modeling in terms of fuzzy set theory it is observed that representation of uncertain parameters is Type-I fuzzy set where it is considered that membership function precisely assign a point from [0,1]. However, in certain situation it is not possible [17] so it is important to adopt IVFS to represent such uncertain situation. In this approach, probability distribution, generalized fuzzy numbers, and completely generalized IVFNs have been combined.

To depict the proposed approach, consider any arbitrary mathematical model

$$M = f(P_i, G_k, F_l) \tag{20.2}$$

where $i = 1, 2, \ldots, m$; $k = 1, 2, \ldots, s$; and $l = 1, 2, \ldots, n$ which is a function of parameters. Suppose P_is are m parameters presented by probabilistic distributions; G_ks as are s parameters presented by generalized fuzzy numbers with heights w_k and F_l are n parameters presented by completely generalized interval valued fuzzy numbers (IVFNs) with height of UMFs and LMFs are w_l^U and w_l^L, respectively.

The approach is explained below:

Step 1: Consider, all generalized fuzzy numbers G_k with heights w_k as well as UMF (upper fuzzy numbers) $F_1^u, F_2^u, \ldots, F_n^u$ of completely generalized interval valued fuzzy numbers (IVFNs) with height w_l^U of UMFs. As both types of fuzzy numbers have different heights, so to deal with the model, we consider $\alpha = [0, w]$ where $w = \min(w_k, w_l^U)$.

Step 2: Calculate α-cut for each fuzzy number (α can be taken stepwise from 0 to w). Then $s + n$ numbers of closed intervals (as α-cut gives closed intervals) will be obtained.

Step 3: Generate m number of uniformly distributed random numbers from [0, 1] and perform Monte Carlo simulation to obtain m numbers of random numbers by sampling probability distribution.

Step 4: Assign all m random numbers and all combination of initial and end points of the $n + s$ intervals in the model M and calculate

$$M_1^{\text{inf}} = \text{Inf}(M) \text{ and } M_1^{\text{sup}} = \text{Sup}(M).$$

Step 5: Repeat Steps 1–4 for 5000 times. Then 5000 minimum values ($M_1^{\text{inf}}, M_2^{\text{inf}}, \ldots, M_{5000}^{\text{inf}}$) and maximum values ($M_1^{\text{sup}}, M_2^{\text{sup}}, \ldots, M_{5000}^{\text{sup}}$) will be obtained.

Step 6: Plot cumulative distribution function (cdf) of ($M_1^{\text{inf}}, M_2^{\text{inf}}, \ldots, M_{5000}^{\text{inf}}$) and ($M_1^{\text{sup}}, M_2^{\text{sup}}, \ldots, M_{5000}^{\text{sup}}$), which will produce a pair of cdfs, i.e., lower probability and upper probability.

Step 7: Consider, other α levels to calculate α-cut of each fuzzy number.

Step 8: Repeat Steps 1–6.

If proceeded in this way a family of cdfs will be obtained.

Step 9: Consider, all generalized fuzzy numbers G_k as well as LMF $F_1^l, F_2^l, \ldots, F_n^l$ of completely generalized interval valued fuzzy numbers F_l with heights w_l^L, respectively. Here also heights of both types of fuzzy numbers are different, so, we consider that $\alpha = [0, h]$ where $h = \min(G_k, w_l^L)$.

Step 10: Repeat Steps 2–8. In step 7 it should be noted that $\alpha = [0, h]$. Then we shall have another family of cdfs.

From these families of cdfs, membership functions at different fractiles can be generated. It will be completely generalized trapezoidal type interval valued fuzzy number. First, family of cdfs will produce UMF and later family of cdfs will give LMF with height w and h, respectively, of the resulting completely generalized interval valued fuzzy number generated at different fractiles.

20.4 Hypothetical Case Study

The general form of a comprehensive food chain risk assessment model as provided by EPA [14], 2001 is follows

$$CDI = \frac{C_f \times FIR \times FR \times EF \times ED \times CF}{BW \times AT} \tag{20.3}$$

where CID = Chronic daily intake (mg/kg-day), FIR = fish ingestion rate (g/day), FR = fraction of fish from contaminated source, EF = exposure frequency (day/year), ED = exposure duration (years), CF = conversion factor ($=10^{-9}$), BW = body weight (kg), AT = averaging time (days), and C_f = chemical concentration of fish tissue (mg/kg). The chemical concentration in fish tissue (C_f) can be computed as

$$C_f = PEC \times BCF \tag{20.4}$$

where PEC = predicted environmental concentration (mg/l) and BCF is the chemical bioaccumulation factor in fish (l/kg).

The noncancer risk model for fish ingestion is expressed as:

$$\text{Risk}_{\text{Non-Cancer}} = \frac{CDI}{Rfd} \tag{20.5}$$

where Rfd is the reference dose.

In this study, representation of the parameters predicted environmental concentration (PEC), chemical bioaccumulation factors (BCF) are considered to be fuzzy number while fish ingestion rate (FIR) is taken as normal probability distribution and other parameters are taken to be constant. Values of the parameters for the calculation of noncancer risk are given in the Table 20.1.

The result of the noncancer human health risk assessment is performed using our proposed approach and which is depicted in Fig. 20.1.

The result of the risk assessment is obtained in the form of family of Cdfs (basically two families of cdfs, one in red colored and another in blue colored) at different α-values. Red colored Cdfs are obtained for UMF and blue colored Cdfs are obtained for LMF of the uncertain input parameter BCF. From these cdfs, risk at different fractiles [10, 24, 27] can be calculated and which are obtained in the form of completely generalized interval valued fuzzy number with height of UMF and LMF are 0.8 and 0.7, respectively. It is because any arithmetic operations between generalized fuzzy numbers and normal fuzzy numbers produces generalized fuzzy number. For instance, at 95th fractile, noncancer risk value lies in the completely generalized interval valued fuzzy number whose UMF is [2.639e-07, 4.136e-07, 4.346e-07, 6.22e-07; 0.8] and LMF is [3.016e-07, 4.135e-07, 4.347e-07, 5.655e-07; 0.8] and which is depicted in Fig. 20.2.

Table 20.1 Parameter values used in the risk assessment

Parameter	Units	Type of variable	Value/distribution
Average time (AT)	Days	Constant	25550
Body weight (BW)	Kg	Probabilistic	Normal(70,5)
Exposure duration (ED)	Years	Constant	30
Exposure frequency (EF)	Days/year	Constant	350
Fraction of contaminated	–	Constant	
Fish (FR)			0.5
Fish ingestion rate (FIR)	g/day	Constant	170
Conversion factor (CF)	–	Constant	1E-09
PEC for As	ug/l	Fuzzy	[4, 5, 6; 0.8]
BCF for As	l/kg	Fuzzy	[35, 45, 55; 0.9] UMF
			[40, 45, 50; 0.7] LMF
Oral Rfd for As	mg/(kg.day)	Constant	3.0E-04

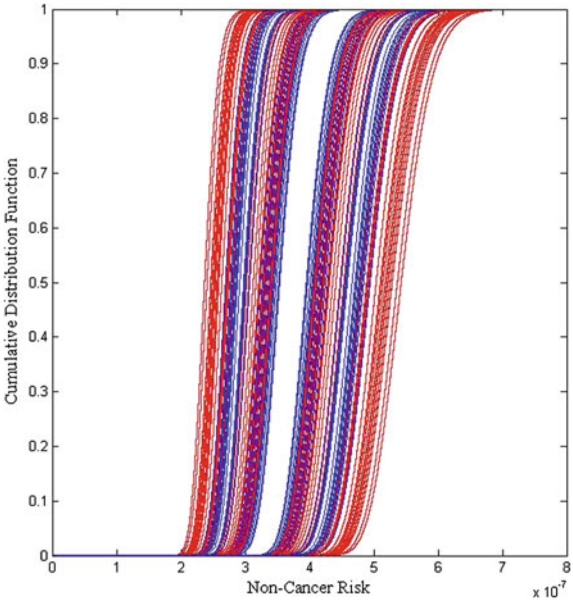


Fig. 20.1 Cumulative distribution functions of noncancer risk for different α values

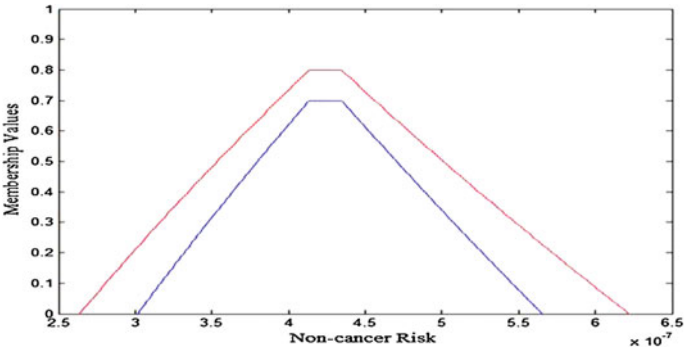


Fig. 20.2 Membership function of noncancer risk at 95th fractile

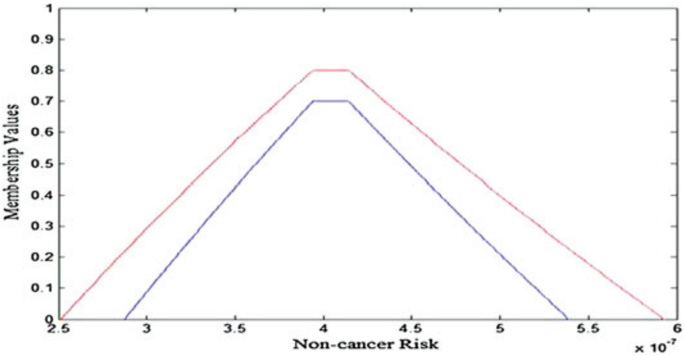


Fig. 20.3 Membership function of non cancer risk at 85th fractile

Similarly, at 85th fractile, risk values lie in the generalized fuzzy number [2.515e-07, 3.942e-07, 4.142e-07, 5.928e-07; 0.8] (UMF); [2.874e-07, 3.941e-07, 4.1432e-07, 5.389e-07; 0.7] (LMF). The graphical representation of the resulting noncancer risk value at 85th fractiles is depicted in Fig. 20.3.

20.5 Conclusion

In this paper, we have proposed a method to deal with situations where some possibilistic distributions are considered as normal interval valued fuzzy numbers together with generalized fuzzy numbers. We have discussed a hypothetical case study using the proposed approach. Risk is obtained in the form of Cdfs and from which, membership functions of the risk are generated at different fractiles. The membership functions of risk at different fractiles are completely generalized interval valued fuzzy numbers, since representation of at least one parameter is taken as generalized fuzzy number (IVFN). The upper and lower membership functions of the completely generalized interval valued fuzzy number is trapezoidal type generalized fuzzy number, because any arithmetic operation of generalized fuzzy numbers (also generalized fuzzy number and normal fuzzy number) produces trapezoidal type generalized fuzzy number.

References

1. M.B. Anoop, K.R. Balaji, N. Lakshmanan, Safety assessment of austenitic steel nuclear power plant pipelines against stress corrosion cracking in the presence of hybrid uncertainties. Int. J. Press. Vessels Pip. **85**(4), 238–247 (2008)
2. P. Baraldi, E. Zio, A combined monte carlo and possibilistic approach to uncertainty propagation in event tree analysis. Risk Anal. **28**(5), 1309–1326 (2008)
3. C. Baudrit, D. Dubois, D. Guyonnet, H. Fargier, Joint treatment of imprecision and randomness in uncertainty propagation, in *Proceedings of the Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (Perugia, 2004), pp. 873–880
4. C. Baudrit, D. Dubois, D. Guyonnet, Joint Propagation and exploitation of probabilistic and possibilistic information in risk assessment. IEEE Trans. Fuzzy Syst. **14**, 593–608 (2006)
5. C. Baudrit, D. Dubois, Practical representations of incomplete probabilistic knowledge. Comput. Stat. Data Anal. **51**(1), 86–108 (2006)
6. C. Baudrit, D. Dubois, N. Perrot, Representing parametric probabilistic models tainted with imprecision. Fuzzy Sets Syst. **159**, 1913–1928 (2008)
7. R. Chutia, Environmental risk modelling under probability-normal interval-valued fuzzy number. Fuzzy Inf. Eng. **3**, 359–371 (2013)
8. S.H. Chen, Operations on fuzzy numbers with function principal. Tamkang J. Manag. Sci. **6**(1), 13–25 (1985)
9. S.H. Chen, Ranking generalized fuzzy number with graded mean integration, in *Proceedings of the 8th International Fuzzy Systems Association World Congress*, vol. 2 (Taipei, Taiwan, Republic of China, 1999), pp. 899–902
10. P. Dutta, H. Boruah, T. Ali, Fuzzy arithmetic with and without using α-cut method: a comparative study. Int. J. Latest Trends Comput. **2**, 99–107 (2011)
11. P. Dutta, T. Ali, A hybrid method to deal with aleatory and epistemic uncertainty in risk assessment. Int. J. Comput. Appl. **42**, 37–44 (2012)
12. P. Dutta, Combined approach to propagate aleatory and epistemic uncertainty in risk assessment. Int. J. Math. Comput. Appl. Res. **3**(5), 2249–6955 (2013)
13. P. Dutta, An approach to deal with aleatory and epistemic uncertainty within the same framework: case study in risk assessment. Int. J. Comput. Appl. **80**(12), 40–45 (2013)
14. EPA U.S., Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part E, Supplemental Guidance for Dermal Risk Assessment). Office of Emergency and Remedial Response, EPA/540/R/99/005, Interim, Review Draft. United States Environmental Protection Agency. Sept 2001
15. R. Flage, P. Baraldi, E. Zio, T. Aven, *Possibility Probability Transformation in Comparing Different Approaches to the Treatment of Epistemic Uncertainties in a Fault Tree Analysis*, eds. by B. Ale, I.A. Papazoglu, E. Zio. Proceedings of the ESREL Reliability, Risk and Safety Conference (Rhodes, Greece, 2010), pp. 714–721
16. R. Flage, P. Baraldi, E. Zio, T. Aven, Probabilistic and Possibilistic treatment of epistemic uncertainties in fault tree analysis (2011) Submitted to Risk analysis
17. M. Gehrke, C. Walker, E. Walker, Some comments on interval valued fuzzy sets. Int. J. Intell. Syst. **11**, 751–759 (1996)

18. D. Guyonnet, B. Cme, P. Perrochet, A. Parriaux, Comparing two methods for addressing uncertainty in risk assessments. J. Environ. Eng. **125**(7), 660–666 (1999)
19. D. Guyonnet, B. Bourgine, D. Dubois, H. Fargier, B. Cme, J.P. Chil, Hybrid approach for addressing uncertainty in risk assessments. J. Environ. Eng. **126**, 68–78 (2003)
20. A. Haldar, R.K. Reddy, A random-fuzzy analysis of existing structures. Fuzzy Sets Syst. **48**, 201–210 (1992)
21. H. Hamrawi, S. Coupland, R. John, A novel a-cut representation for type-2 fuzzy sets, in *2010 IEEE International Conference Fuzzy Systems* (Barcelona, Spain, 2010)
22. H. Hamrawi, Type-2 fuzzy a-cuts. Ph.D. thesis. De Montfort University, 2011
23. J.C. Helton, W.L. Oberkampf, Alternative representations of epistemic uncertainty. Reliab. Eng. Syst. Saf. **85** 1–3 (2004)
24. E. Kentel, M.M. Aral, Probalistic-fuzzy health risk modeling. Stoch. Environ. Res. Risk Assess. **18**, 324–338 (2004)
25. J. Li, G.H. Huang, G.M. Zeng, I. Maqsood, Y.F. Huang, An integrated fuzzy-stochastic modeling approach for risk assessment of groundwater contamination. J. Environ. Manage. **82**, 173–188 (2007)
26. P. Limbourg, E. de Rocquigny, Uncertainty analysis using evidence theory—confronting level-1and level-2 approaches with data availability and computational constraints. Reliab. Eng. Syst. Saf. **95**, 550–564 (2010)
27. R.M. Maxwell, S.D. Pelmulder, A.F.B. Tompson, W.E. Kastenberg, On the development of a new methodology for groundwater-driven health risk assessment. Water Resour. Res. **34**, 833–847 (1998)
28. N. Pedronia, E. Zioa, E. Ferrariob, A. Pasanisid, M. Couplet, Propagation of aleatory and epistemic uncertainties in the model for the design of a flood protection dike, in *PSAM 11 and ESREL* (Helsinki, Finland, 2012)
29. N. Pedronia, E. Zioa, E. Ferrariob, A. Pasanisid, M. Couplet, Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model. Comput. Struct. 1–15 (2013)
30. K.D. Rao, H.S. Kushwaha, A.K. Verma, A. Srividya, Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies. Reliab. Eng. Syst. Saf. **92**, 947–956 (2007)
31. R. Sambuc, Function F-Flous, Application a laide au Diagnostic en Pathologie Thyroidienne (University of Marseille, These de Doctoraten Medicine, 1975)

Chapter 21
Bifurcation Analysis of SIR Model with Logistically Grown Susceptibles and Effect of Loss of Immunity of the Recovered Class

Rita Ghosh and Uttam Ghosh

Abstract In this paper, we have investigated an SIR model with logistic growth rate of susceptibles where rate of incidence is directly affected by the inhibitory factors or social or psychological factors. In our model three equilibrium points are obtained, one of them is endemic equilibrium point. One disease free equilibrium points is unstable in nature in all circumstance and the trajectories in the neighborhood of the endemic equilibrium point of our model undergoes a Hopf bifurcation subject to some critical value of the carrying capacity. Finally, numerical solution is done.

Keywords Inhibition effect · Hopf bifurcation · Logistic growth rate · Lose immunity

21.1 Introduction

The classical SIR model of Kermack and McKendrick [1] was a fundamental model in the study of epidemiological modeling of infectious diseases. The several models was analyzed considering different growth rate of susceptible and different incidence rate [2–7]. Wang and Ruan [8] studied epidemic model with constant birth rate of the susceptibles and constant removal rate of the infected class with standard incidence rate and they discuss existence of Hopf bifurcation. Kaddar [2] studied the dynamics of a delayed SIR epidemic model with a modified saturated incidence rate and established the existence of delay dependent Hopf bifurcation. Hopf bifurcation in an eco-epidemic model was studied considering that the prey population is infected with a microparasite and predator functional response is Holling type-I. The criterion for existence of Hopf type small periodic oscillation was reported [4]. Kar and Mondal [9] studied SIR epidemic model with logistic growth rate with saturated incidence

R. Ghosh
Department of Applied Mathematics, University of Calcutta,
92, APC Road, Kolkata 700009, West Bengal, India

U. Ghosh (✉)
Department of Mathematics, Nabadwip Vidyasagar College, Nadia, West Bengal, India
e-mail: uttam_math@yahoo.co.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_21

rate and they showed that two type of bifurcation occurs depending on value of delay. Stability and Hopf bifurcation for a delayed SIR epidemic model with Logistic Growth rate of susceptibles was investigated by Xue and Li [10]. In their model they assumed members of the recovered class will never be susceptible and consider the rate of infection of the form $\frac{\beta SI}{1+\alpha I}$ where α is the inhibitory factor. In this paper, we have proposed an SIR epidemic model with logistic growth rate where a percentage of recovered class will lose immunity and return back into the susceptible class. The rate of incidence is considered in the form $\frac{\beta SI}{1+\alpha I}$, which is directly affected by the inhibitory factors such as social awareness (α). Since the social awareness of the susceptibles will increase the factor $1+\alpha S$ will consequently decrease the term $\frac{\beta SI}{1+\alpha I}$. This implies due to social awareness of the susceptibles will decrease the transfer of individuals from susceptibles class to infected class. The paper is organized as follows, in first part of the paper we have formulate the model, in second part stability analysis of the equilibrium points and Hopf bifurcation criterion is analyzed. Finally, numerical simulation is

21.2 Mathematical Formulation

Here, we consider the model in which the newly appointed S—class has logistic growth rate and the rate of infection is directly affected by the inhibitory factors. Logistic growth of the susceptible in the SIR model is more realistic as the number of susceptible cannot grow exponentially. If $S(t)$ be the number of susceptible, $I(t)$ be the number of infected and $R(t)$ be the number of recovered individuals at time t. Then, the governing differential equation of the proposed model is done.

$$\frac{dS}{dt} = rS\left(1-\frac{S}{k}\right) - \frac{\beta SI}{1+\alpha S} - dS + \mu R \tag{21.1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{1+\alpha S} - (d+\gamma)I \tag{21.2}$$

$$\frac{dR}{dt} = \gamma I - (d+\mu)R \tag{21.3}$$

where, r = birth rate (intrinsic growth rate) of the susceptible class
k = carrying capacity
β = the transmission rate of infection
α = the parameter that measure the inhibitory factors
d = the natural death of the population
μ = rate at which the recovered class losses immunity and becomes susceptible
γ = rate at which the infected individuals recovered.

The model (21.2) shows that recovered class losses immunity at the rate of μ become susceptibles. So the term $+\mu R$ enters in the first equation of model.

21.3 Stability Analysis of the Model

The system admits three equilibrium points, one of them is endemic and the other two of them are disease free equilibrium points. The equilibrium points are $A_0 = (S_0, I_0, R_0) = (0, 0, 0)$, $A_1 = (S_1, I_1, R_1) = \left(k\left(1 - \frac{d}{r}\right), 0, 0\right)$ and $A_2 = (S_2, I_2, R_2)$ where

$$S_2 = \frac{1}{\alpha\,(R_{01} - 1)}, \quad I_2 = \frac{(d + \mu)\left(d + \frac{S_2 r}{k}\right) S_2\,(R_{02} - 1)}{d(d + \mu + \gamma)}$$
$$R_2 = \frac{\gamma I}{d + \gamma}, \quad R_{01} = \frac{\beta}{\alpha(d + \gamma)}, \quad R_{02} = \frac{\alpha r\,(R_{01} - 1)\,k}{\alpha d\,(R_{01} - 1)\,k + r}.$$

The disease free equilibrium point $A_1\,(S_1, I_1, R_1)$ will exists only when $r > d$ so we are interested only when $r > d$. The endemic equilibrium point $A_2\,(S_2, I_2, R_2)$ will exist if $R_{01} > 1$ and $R_{02} > 1$. Since the endemic equilibrium point $A_2\,(S_2, I_2, R_2)$ and the disease free equilibrium point $A_1\,(S_1, I_1, R_1)$ are directly affected by the carrying capacity k. As k increases then decreases and consequently number of susceptibles will increase.

Since physically $A_0(0, 0, 0)$ is not important because in this case all the individual population goes to extinction and so stability analysis about this point is not taken into consideration. It can be shown that this point is an unstable equilibrium point.

Theorem 21.1 *If* $r > d$ *and* $\frac{\beta S_1}{1+\alpha S_1} - (d + \gamma) < 0$, *then the second disease free equilibrium point* $A_1\,(S_1, I_1, R_1)$ *is stable in nature otherwise unstable.*

Proof The characteristic equation of the system (21.2) for this equilibrium point is

$$\begin{vmatrix} r - d - \lambda & -\dfrac{\beta S_1}{1 + \alpha S_1} & \mu \\ 0 & \dfrac{\beta S_1}{1 + \alpha S_1} - (d + \gamma + \lambda) & 0 \\ 0 & \gamma & -(d + \gamma + \mu) \end{vmatrix} = 0. \tag{21.4}$$

Solving Eq. (21.4) we get $\lambda = -(r - d)$, $\frac{\beta S_1}{1+\alpha S_1} - (d + \gamma)$, $-(d + \mu)$. Since all the roots will be negative when $r > d$ and $\frac{\beta S_1}{1+\alpha S_1} - (d + \gamma) < 0$. Hence, the solution in the neighbourhood of this disease free equilibrium point is stable in nature. Otherwise one root is always positive and other two roots will be negative and consequently, the solution in the neighborhood of this point will be unstable in nature. Hence, the theorem is proved.

Again if $r < d$ then A_1 does not exist.

Theorem 21.2 *If* $1 < R_{01} < 1 + \frac{2}{k\alpha}$, $R_{02} > 1$ *and* $r > d$. *Then the endemic equilibrium point* $A_2\,(S_2, I_2, R_2)$ *will be asymptotically stable.*

The characteristic equation about the point A_2 (S_2, I_2, R_2) is

$$\begin{vmatrix} r(1-\frac{2S_2}{k})-d-\frac{\beta I_2}{(1+\alpha S_1)^2}-\lambda & -\frac{\beta S_2}{1+\alpha S_2} & \mu \\ \frac{\beta I_2}{(1+\alpha S_1)^2} & -\lambda & 0 \\ 0 & \gamma & -(d+\lambda+\mu) \end{vmatrix} = 0. \qquad (21.5)$$

Solving the above, we get

$$\lambda^3 + C_1\lambda^2 + C_2\lambda + C_3 = 0, \qquad (21.6)$$

where

$$\begin{aligned} C_1 &= 2d+\mu-r+\frac{2rS_2}{k}+\frac{\beta I_2}{(1+\alpha S_1)^2} \\ C_2 &= \frac{\beta I_2(2d+\gamma+\mu)}{(1+\alpha S_1)^2}+(d+\mu)\left(d+\mu-r+\frac{2rS_2}{k}\right) \\ C_3 &= \frac{d\beta I_2}{(1+\alpha S_1)^2}(d+\mu+\gamma) \end{aligned}$$

and

$$\begin{aligned} C_1C_2-C_3 = {} & \frac{\beta^2 I_2^2(\gamma+2d+\mu)}{(1+\alpha S_1)^4}+\frac{\beta I_2}{(1+\alpha S_1)^2} \\ & \left\{\left(-r+\frac{2rS_2}{k}\right)(3d+2\mu+\gamma)+(2d+\mu)^2+\mu(d+\gamma)\right\} \\ & +(d+\mu)\left\{\left(-r+\frac{2rS_2}{k}\right)^2+\left(-r+\frac{2rS_2}{k}\right)(3d+\mu)+d(2d+\mu)\right\} \end{aligned}$$

C_1 and C_2 are positive under the conditions stated in the theorem, C_3 is always positive and $C_1C_2 - C_3 > 0$.

Therefore, Routh Hurwitz criterion is satisfied and the eigenvalues values must have negative real part. Hence, the solutions in the neighborhood of endemic equilibrium point A_2 (S_2, I_2, R_2) will be stable in nature. Hence, the theorem is proved.

It is obvious from the definition of C_1 and C_2 that sign of them can be controlled by changing the value of carrying capacity (k).

21.4 Hopf Bifurcation Around Positive Equilibrium

Since the expression of C_1, C_2, C_3, and $C_1C_2 - C_3$ depends on carrying capacity k. The sign of C_1, C_2, C_3 and $C_1C_2 - C_3$ can be controlled by changing the values of k. A Hopf bifurcation of the system is expected for some range of k where $C_1C_2 - C_3 = 0$.

Theorem 21.3 *The system (21.2) undergoes a Hopf bifurcation for $R_{01} > 1$ and $R_{02} > 1$ when the carrying capacity k passes through the critical value k_c with the restriction $r > 2d$ and $k > \frac{4rS_2}{2r-(3d+\mu)}$.*

Proof Hopf bifurcation will occur if $C_1(k)C_2(k) - C_3(k) = 0$ with $C_i(k) > 0$, $i = 1, 2, 3$ and $\frac{d(Re\lambda)}{dk} \neq 0$ at $k = k_c$.

As for $C_1C_2 = C_3$ with $C_i > 0$, then the characteristic equation becomes

$$\left(\lambda^2 + C_2\right)(\lambda + C_1) = 0$$

having roots $-C_1, \pm i\sqrt{C_2}$. So there are purely imaginary eigen values and one is strictly negative real eigenvalue. We assume that for k is in the neighborhood of $k = k_c$ the roots have the form $\lambda_1 = P_1(k) + P_2(k)$, $\lambda_2 = P_1(k) - P_2(k)$ and $\lambda_3 = -P_3(k)$ where $P_i(k)$, $i = 1, 2, 3$ are real. In view of the above roots, the corresponding characteristic equation will be

$$\lambda^3 + (P_3 - 2P_1)\lambda^2 + \left(P_1^2 + P_2^2 - 2P_1P_3\right)\lambda + P_3\left(P_1^2 + P_2^2\right) = 0 \tag{21.7}$$

comparing we get $C_1 = P_3 - 2P_1$, $C_2 = P_1^2 + P_2^2 - 2P_1P_3$, and $C_3 = P_3\left(P_1^2 + P_2^2\right)$.

Since $P_1(k) = 0$ at $k = k_c$ then from the above we get

$$(C_1 + 2P_1)C_2 = C_3 - 2P_1(C_1 + 2P_1)^2. \tag{21.8}$$

Differentiating both sides of (21.8) w.r.t. k we obtain

$$\begin{aligned}(C_1 + 2P_1)\frac{dC_2}{dk} &+ C_2\left(\frac{dC_1}{dk} + 2\frac{dP_1}{dk}\right)\\ &= \frac{dC_3}{dk} - \left\{2\frac{dP_1}{dk}(C_1 + 2P_1)^2 + 2P_1\frac{d(C_1 + 2P_1)^2}{dk}\right\}\end{aligned} \tag{21.9}$$

Using the condition at $k = k_c$, $P_1(k) = 0$ we get

$$\left(\frac{dP_1}{dk}\right)_{k=k_c} = -\left\{\frac{\frac{d(C_1C_2-C_3)}{dk}}{2C_1^2 + C_2}\right\}_{k=k_c}.$$

Using the values of C_1, C_2, and C_3 we get

$$\left(\frac{dP_1}{dk}\right)_{k=k_c} = -\frac{1}{2C_1^2 + C_2}\Bigg[\Bigg\{\frac{2\beta^2(2d+\mu+\gamma)I_2 r S_2^2(d+\mu)}{k^2 d(d+\mu+\gamma)\,(1+\alpha S_2)^4} + \frac{\beta}{(1+\alpha S_2)^2}$$
$$\left(4d^2 + 4d\mu + \mu^2 + (d+\mu)\gamma\right)\Bigg\}\frac{S_2^2 r(d+\mu)}{k^2 d(d+\mu+\gamma)}$$
$$+\frac{\beta(3d+2\mu+\gamma)S_2^2 r(d+\mu)}{k^2\,(1+\alpha S_2)^2\,d(d+\mu+\gamma)}\left\{\frac{4rS_2}{k} + r - 2d\right\}$$
$$+\frac{2rS_2(d+\mu)}{k^2}\left\{2r\left(1-\frac{2S_2}{k}\right) - 3d - \mu\right\}\Bigg]_{k=k_c} < 0.$$

Hence, when $k < k_c$ then the solution will be unstable and for $k > k_c$ the solution will be stable in nature in the neighbourhood of A_2. Thus, Hopf bifurcation occurs when the carrying capacity crosses the critical value $k = k_c$.

Hence the result.

21.5 Numerical Simulation

The numerical simulation is done considering several values of the parameters. First, we consider $r = 14.0$, $K = 30$, $\beta = 1$, $d = 2$, $\gamma = 0.001$, $\mu = 0.01$, $\alpha = 0.021$, and then $R_{01} = 23.79$, $R_{02} = 1.26$, the endemic equilibrium point is obtained $(2.09, 11.51, 0.01)$, all values are taken as correct up to two decimal place. The corresponding numerical solution of the differential equations is shown in the Fig. 21.1.

For the critical value $k_c = 60.43$ remaining all other parameter unchanged we obtain $R_{01} = 23.80$, $R_{02} = 2.08$ and the endemic equilibrium point (2.09, 12.02, 0.01) and the corresponding graph of the numerical solution is shown in Fig. 21.2. It is clear from the figure that a Hopf bifurcation of periodic solution occurs at $k = k_c$ and the solution oscillates about the endemic equilibrium point.

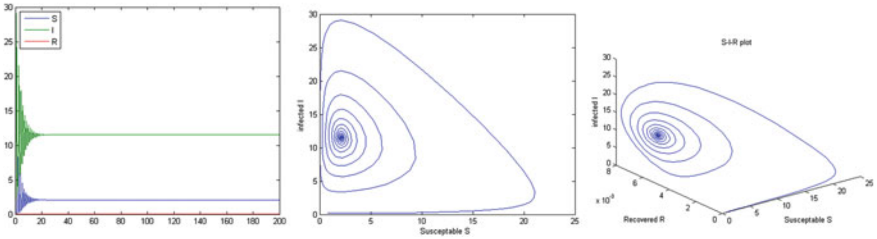


Fig. 21.1 This figure represents the stability behavior of the endemic equilibrium of system

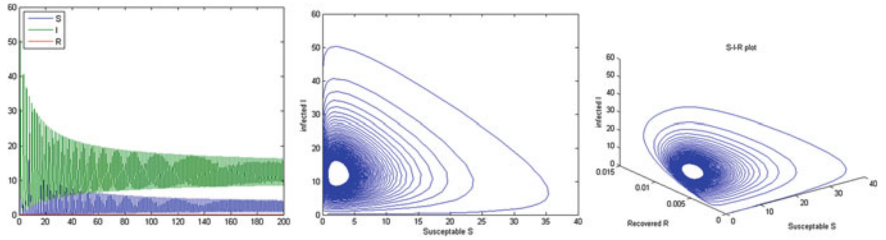


Fig. 21.2 This figure represents periodic solution

21.6 Conclusions

In this paper, we formulate the epidemic model with logistic growth rate of susceptibles and effect of loss of immunity of the recovered class and the inhibitory effect is also taken into consideration. Here, three equilibrium points obtained. Two of them are disease free and one is endemic equilibrium point. The solution in the neighborhood of disease free equilibrium point $A_0(0, 0, 0)$ is always unstable in nature. The solution in the neighborhood of other disease free equilibrium point $A_1(S_1, R_1, I_1)$ is stable if $r > d$ and $\frac{\beta S_1}{1+\alpha S_1} < 0$ other wise unstable in nature. The solution in the neighborhood of endemic equilibrium point trajectories undergoes a Hopf bifurcation depending on the values of carrying capacity k.

Acknowledgments Authors are grateful to Prof. Susmita Sarkar, Department of Applied Mathematics, University of Calcutta for her valuable guidance at every step of this work. Authors also thank to the reviewer for his valuable comments which has helped to improve the paper.

References

1. W.O. Kermack, A.G. McKendrick, Contribution to mathematical theory of epidemics. P. R. Soc. Lond. A Mat. **115**, 700–721 (1927)
2. A. Kaddar, On the dynamics of a dynamics of delayed SIR epidemic model with a modified saturated incidence rate. J. Differ. Equ. **133**, 1–7 (2009)
3. A. Kaddar, Stability analysis of delayed SIR epidemic model with saturated incidence rate. Nonlinear Anal. **3**, 299–306 (2010)
4. D. Mukherjee, Hopf bifurcation in an eco-epidemic model. Appl. Math. Comput. **217**, 2118–2124 (2010)
5. W. Wang, Epidemic models with nonlinear infection forces. Math. Biosci. Eng. **3**, 267–279 (2006)
6. U. Ghosh, S. Sarkar, D.K. Khan, Modelling of infectious disease in presence of vaccination and delay. Int. J. Epidemiol. Infect. **2**(3), 50–57 (2014)
7. D. Xiao, S. Ruan, Global analysis of an epidemic model with non-monotone incidence rate. Math. Biosci. **208**, 419–429 (2007)
8. W. Wang, S. Ruan, Bifurcation analysis of an epidemic model with constant removal rate of the infectives. J. Math. Anal. Appl. **291**, 775–793 (2004)

9. T.K. Kar, P.K. Mondal, Global dynamics and bifurcation in delayed SIR model. Nonlinear Anal.: Real World Appl. **12**, 2058–2068 (2011)
10. Y. Xue, T. Li, *Stability and Hopf Bifurcation for a delayed SIR Epidemic Model with Logistic Growth* (Hindawi Publishing Corporation, 2013), pp. 1–11

Chapter 22
SIR Epidemic Modelling in Presence of Inhibitory Effect and Delay

Uttam Ghosh and Rita Ghosh

Abstract Modelling of infectious diseases was investigated by several authors using mathematical methods. Here, we consider the model in which inhibition effect is taken into consideration in presence of delay of the infected to become infectious. In this model two equilibrium points are found, one is disease free equilibrium point and the other is the endemic equilibrium point. The endemic equilibrium point will exists under certain condition. The character of solutions in the neighbourhood of endemic equilibrium point is directly affected by inhibitory effect. The solutions in the neighbourhood of disease free equilibrium point will be asymptotically stable when the basic reproduction number less than one and the solution in the neighbourhood of endemic equilibrium point will be asymptotically stable when the basic reproduction number greater than one.

Keywords Equilibrium point · Inhibition effect · Asymptotically stable

22.1 Introduction

From the prehistory of civilization the human population is affected by different infectious diseases. It is most important to formulate the spreading mechanism of such diseases and finding mechanisms to control them. Mathematicians are using mathematical models to study the above disease mechanisms. Karmack and Mckendrick [1] was first given the formulation of SIR model. The history of epidemic and different epidemic models may be found in famous books Bailey [2], Murray [3], Ma and Li [4] and Anderson and May [5]. Recently Hetchote and Tunder [6], Liu et al. [7, 8], Hetchote et al. [9], Xiao and Ruan [10], Ghosh et al. [11] and many authors investigated different epidemic models. They studied different epidemic models

U. Ghosh (✉)
Department of Mathematics, Nabadwip Vidyasagar College,
Nadia, West Bengal, India
e-mail: uttam_math@yahoo.co.in

R. Ghosh
University of Calcutta, 92, APC Road, Kolkata 700009, West Bengal, India

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_22

considering different incidence rate of infection. One of them is bilinear incidence rate kSI, where S and I are, respectively, the number of susceptible and infected individuals in the population and k is the rate of infection is always positive [12]. Some authors used different saturated incidence rate [12–14].

In this paper, we consider two models in which the incidence rate of infection is of the form $\frac{kSI}{1+\alpha S+\beta I}$ where α, β are positive constants denotes the inhibitory effect (including all kind of saturation effects, taking of appropriate saturation effect or sociological and psychological effects).

In the first model, delay is not taken into consideration. The second model is analyzed in presence of delay. In both the cases two equilibrium points arises, one is the disease free equilibrium point and other is the endemic equilibrium point. The endemic equilibrium point exists depending under certain condition and is directly affected by the inhibitory effect.

22.2 Mathematical Formulation

Let $S(t)$ be the number of susceptible, $I(t)$ be the number of infected and $R(t)$ be the number of recovered individuals such that $N(t) = S(t) + I(t) + R(t)$. Here, incidence rate of infection is taken into consideration is $\frac{kSI}{1+\alpha S+\beta I}$. The corresponding differential equation (in absence of delay) is given in Eq. (22.1) and the same model in presence of delay is given in (22.2).

$$\frac{dS}{dt} = b - dS - \frac{kSI}{1+\alpha S+\beta I} + \gamma R \tag{22.1a}$$

$$\frac{dI}{dt} = \frac{kSI}{1+\alpha S+\beta I} - (\mu + d)\, I \tag{22.1b}$$

$$\frac{dR}{dt} = \mu I - (\gamma + d)\, R \tag{22.1c}$$

where b—is the birth rate, d—is the natural death rate of the population, μ—is the natural recovery rate of infected individuals, γ—is the rate at which recovered individuals lose immunity and return to the susceptible class, τ = time of delay of infected class to becomes infectious.

$$\frac{dS}{dt} = b - dS - \frac{kSI}{1+\alpha S+\beta I} + \gamma R \tag{22.2a}$$

$$\frac{dI}{dt} = \frac{kS(t-\tau)I(t-\tau)e^{-d\tau}}{1+\alpha S(t-\tau)+\beta I(t-\tau)} - (\mu + d)\, I \tag{22.2b}$$

$$\frac{dR}{dt} = \mu I - (\gamma + d)\, R \tag{22.2c}$$

22.3 Stability Analysis of First Model

To find disease free equilibrium points set

$$\frac{dS}{dt} = \frac{dI}{dt} = \frac{dR}{dt} = 0.$$

Solving the equations two equilibrium points are obtained. The equilibrium points are

$$(S_0, I_0, R_0) = \left(\frac{b}{d}, 0, 0\right), (S_1, I_1, R_1),$$

where

$$I_1 = \frac{(d+\gamma)(b\alpha + d)\,(R_{02} - 1)}{d\beta(d+\gamma) + d(d+\gamma+\mu)\alpha(R_{01} - 1)}, \quad S_1 = \frac{1 + \beta I_1}{\alpha\,(R_{01} - 1)}, \quad R_1 = \frac{\mu I_1}{(d+\gamma)},$$
$$R_{01} = \frac{k}{\alpha(d+\mu)}, \quad R_{02} = \frac{kb}{(b\alpha + d)(d+\mu)}.$$

The endemic equilibrium points will exists only when $R_{01} > 1$ and $R_{02} > 1$. Since biologically the model will be meaningful in the 1st octant only, i.e. in the region $\{(S, I, R), S \geq 0, I \geq 0$ and $R \geq 0\}$.

Theorem 22.1 *The system $S + I + R = \frac{b}{d}$ is a manifold of the system (22.1), which is attracting fixed point in the first octant.*

Proof Proof of the theorem is an immediate consequence of [12].

Theorem 22.2 *The disease free equilibrium point of (22.1) is locally asymptotically stable for all $R_{01} < 1$.*

Proof Linearizing the system about the disease free equilibrium point (S_0, I_0, R_0), put $S = \acute{S} + S_0$, $I = \acute{I} + I_0$, $R = \acute{R} + R_0$, and rewriting the system omitting the dot sign we obtain

$$\frac{dS}{dt} = -dS - \frac{kS_0 I}{1 + \alpha S_0} + \gamma R \tag{22.3a}$$
$$\frac{dI}{dt} = \frac{kS_0 I}{1 + \alpha S_0} - (\mu + d)\,I \tag{22.3b}$$
$$\frac{dR}{dt} = \mu I - (\gamma + d)\,R \tag{22.3c}$$

The corresponding characteristic equation of the above system (22.3) are $\lambda = -d, -(d+\mu)$ and $\lambda = (d+\mu)(R_{02} - 1)$. Since two of them is negative always and the third root will be negative if $R_{02} < 1$ then all the roots are negative and consequently, the solution in the neighbour of the disease free equilibrium point will be asymptotically stable in nature. Again when $R_{02} > 1$ then one the three roots will be positive and other two is negative and the disease free equilibrium point will be a saddle point. Therefore, solutions in the neighbourhood of this point will be unstable in nature. When $R_{02} > 1$ then other equilibrium point will exist.

Theorem 22.3 *The endemic equilibrium point of (22.1) is asymptotically stable for all $R_{01} > 1$ and $R_{02} > 1$.*

Proof Since the endemic equilibrium point will exists if $R_{02} > 1$ and $R_{01} > 1$. Linearizing the system about the endemic equilibrium point (S_1, I_1, R_1), put $S = \acute{S} + S_1$, $I = \acute{I} + I_1$, $R = \acute{R} + R_1$, and rewriting the system omitting the dot sign we obtain

$$\frac{dS}{dt} = -(d + A)S - BI + \gamma R \tag{22.4a}$$

$$\frac{dI}{dt} = AS + (B - \mu + d)\, I \tag{22.4b}$$

$$\frac{dR}{dt} = \mu I - (\gamma + d)\, R \tag{22.4c}$$

where

$$A = \frac{kI_1\,(1 + \beta I_1)}{(1 + \alpha S_1 + \beta I_1)^2}, \quad B = \frac{kS_1\,(1 + \alpha S_1)}{(1 + \alpha S_1 + \beta I_1)^2}$$

The corresponding characteristic equation of the above system (22.4) is

$$\lambda^3 + C_1\lambda^2 + C_2\lambda + C_3 = 0 \tag{22.5}$$

where $C_1 = 2d + A + \gamma + \dfrac{\beta I_1(d + \mu)}{(1 + \alpha S_1 + \beta I_1)}$.

$C_2 = (A + d)(d + \gamma) + \dfrac{\beta I_1(d + \mu)}{(1 + \alpha S_1 + \beta I_1)} + (d + \gamma)\dfrac{\beta I_1(d + \mu)}{(1 + \alpha S_1 + \beta I_1)} + AB.$

$C_3 = Ad(d + \gamma + \mu) + d(d + \gamma)\dfrac{\beta I_1(d + \mu)}{(1 + \alpha S_1 + \beta I_1)}$.

Since C_1, C_2, C_3, and $C_1 \cdot C_2 - C_3$ all are positive and therefore, all roots of the Eq. (22.5) have negative real part (Routh-Horwtz criteria). Thus disease free equilibrium point is asymptotically stable for $R_{01} > 1$ and $R_{02} > 1$. This concludes the proof.

Since R_{02} is directly affected inhibitory effect α. As α—increase R_{02} decreases and tends to zero when α tends to infinity. The graph of R_{02} is shown in the Fig. 22.1. Considering $d = 0.04$, $b = 5.0$, and $\mu = 0.05$ and taking k along y-axis and α along x-axis graph of R_{02} is plotted.

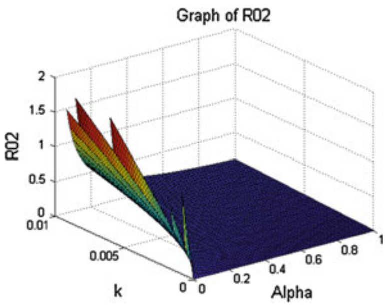


Fig. 22.1 Graph of R_{02}

From Fig. 22.1 it is clear that as α tends to 1 what ever k may be then R_{02} becomes less than 1 and then endemic equilibrium points does not exist. One can determine a critical value of α such that the R_{02} greater than 1.

22.4 Stability Analysis Second Model

In this model also two equilibrium points of the system are found, one is the disease equilibrium point $(S_0, I_0, R_0) = (\frac{b}{d}, 0, 0)$ which always exists and the other is the endemic equilibrium point (S_{11}, I_{11}, R_{11}) where

$$I_{11} = \frac{(d+\gamma)(b\alpha + d)\,(R_{04} - 1)}{d\beta(d+\gamma) + d(d+\gamma+\mu)\,(R_{03}-1)}, \quad S_{11} = \frac{1+\beta I_1}{\alpha\,(R_{03}-1)},$$
$$R_{11} = \frac{\mu I_{11}}{(d+\gamma)}, \quad R_{03} = \frac{ke^{-d\tau}}{\alpha(d+\mu)}, \quad R_{04} = \frac{kbe^{-d\tau}}{(b\alpha+d)(d+\mu)}.$$

Since the endemic equilibrium points will exists $R_{03} > 1$ and $R_{04} > 1$.

Theorem 22.4 *The disease free equilibrium point of (22.2) is asymptotically stable for $R_{04} < 1$ and unstable when $R_{04} > 1$.*

Proof Linearizing the system about the disease free equilibrium point (S_0, I_0, R_0), put $S = \acute{S} + S_0$, $I = \acute{I} + I_0$, $R = \acute{R} + R_0$ in (22.2) neglecting the higher order terms and rewriting the system omitting the dot sign we obtain

$$\frac{dS}{dt} = -dS - \frac{kS_0 I}{1+\alpha S_0} + \gamma R \tag{22.6a}$$
$$\frac{dI}{dt} = \frac{kS_0 I(t-\tau)e^{-d\tau}}{1+\alpha S_0} - (\mu + d)\,I \tag{22.6b}$$
$$\frac{dR}{dt} = \mu I - (\gamma + d)\,R \tag{22.6c}$$

The corresponding characteristic roots of the above system (22.6) are $\lambda = -d, -(d+\mu)$ and the other root satisfy the transcendental equation

$$\frac{kS_0e^{-((d+\lambda)\tau)}}{1+\alpha S_0} = \lambda + (d+\mu). \tag{22.7}$$

If $\tau = 0$ then $R_{03} = R_{01}$ and $R_{04} = R_{02}$ then $\lambda = (d+\mu)(R_{02}-1)$, Since in the first case we already establish the disease free equilibrium point will be asymptotically stable in absence of delay if $R_{02} < 1$. If possible let Eq. (22.7) has imaginary root of the form $\lambda = i\omega(\omega > 0)$. Hence putting $\lambda = i\omega$ in (22.7) and separating real and imaginary part we get

$$\left.\begin{array}{l} \mu + d = A_1e^{-d\tau}S_0 \cos\tau\omega \\ \omega = A_1e^{-d\tau}S_0 \sin\tau\omega \end{array}\right\}$$

where $A_1 = \frac{kb}{\alpha b+d}$, Squaring and adding the above two and writing in the simplest form we get $\omega^2 = (d+\mu)^2(R_{04}^2 - 1)$. Since if $R_{04} < 1$ then there exists no real value of ω, such that $i\omega$ is a root of equation (22.7). By Rouche's theorem all the eigenvalues have negative real part of equation (22.7) and consequently, the solutions will be stable in nature. If $R_{04} > 1$, then, the disease free equilibrium (S_0, I_0, R_0) is unstable for $\tau = 0$. By Kuang theorem the equilibrium point (S_0, I_0, R_0) is unstable for all $\tau \geq 0$.

This concludes the proof

Theorem 22.5 *The endemic equilibrium point of (22.2) is asymptotically stable for $R_{03} > 1$ and $R_{04} > 1$.*

Proof Since the endemic equilibrium points will exists only when $R_{03} > 1$ and $R_{04} > 1$. Now, Linearzing the system of equations about the endemic equilibrium point (S_{11}, I_{11}, R_{11}) we get the reduce system (putting $S = \acute{S} + S_{11}$, $I = \acute{I} + I_{11}$, $R = \acute{R} + R_{11}$ and writing omitting the dot sign)

$$\frac{dS}{dt} = -(d+A)S - BI + \gamma R \tag{22.8a}$$

$$\frac{dI}{dt} = e^{-d\tau}AS(t-\tau) + (B-\mu-d)\,I(t-\tau)e^{-d\tau} \tag{22.8b}$$

$$\frac{dR}{dt} = \mu I - (\gamma + d)\,R \tag{22.8c}$$

where

$$A = \frac{kI_{11}\,(1+\beta I_{11})}{(1+\alpha S_{11}+\beta I_{11})^2}, \quad B = \frac{kS_{11}\,(1+\alpha S_{11})}{(1+\alpha S_{11}+\beta I_{11})^2}.$$

The corresponding characteristic equation of the above system (22.8) is

$$\lambda^3 + d_1\lambda^2 + d_2\lambda + d_3 - e^{-(d+\lambda)\tau}\{B\lambda^2 + B(2d+\gamma)\lambda + BA\mu\gamma\} = 0 \tag{22.9}$$

where $d_1 = 3d + A + \mu + \gamma$, $d_2 = (d+A)(d+\gamma) + (d+A)(d+\mu) + (d+\mu)(d+\gamma)$ and $d_3 = (d+A)(d+\gamma)(d+\mu)$.

For $\tau = 0$ the equation reduce to Eq. (22.5) and then $R_{04} = R_{02}$ and from Theorem 22.3 the system will be asymptotically stable. Hence, instability occurs for a particular value $\tau = 0$. If possible let one root (say $\lambda = i\omega$) of Eq. (22.9) must exists lies on the imaginary axis. Putting $\lambda = i\omega$ in (22.9) we get after equating real and imaginary part and writing in the simplest form

$$\omega^6 + p_1\omega^4 + p_2\omega^2 + p_3 = 0 \tag{22.10}$$

where

$$p_1 = (d+A)^2 + (d+\gamma)^2 + (d+\mu+B)\frac{\beta I_{11}(d+\mu)}{(1+\alpha S_{11}+\beta I_{11})} + B^2\left(1 - e^{-2d\tau}\right),$$

$$p_2 = \left\{(d+B+\mu)\frac{\beta I_{11}(d+\mu)}{(1+\alpha S_{11}+\beta I_{11})} + d^2\right\}(d+\gamma)^2 + (d+\gamma)^2(d+A)^2$$
$$+ 2AB\mu\gamma e^{-2d\tau} + B^2\left\{d^2 + (d+\gamma)^2\right\}\left(1 - e^{-2d\tau}\right) + \left(A^2 + 2dA\right)(d+\mu)^2$$

and

$$p_3 = \{d(d+\gamma)B + A\mu\gamma\}^2\left(1 - e^{-2d\tau}\right) + \left\{d(d+\gamma)B + A\mu\gamma + d^3\right\}$$
$$\left\{d(d+\gamma)\frac{\beta I_{11}(d+\mu)}{(1+\alpha S_{11}+\beta I_{11})} + Ad(d+\mu+\gamma\right\}.$$

Since Eq. (22.10) is a cubic in ω^2 having all the coefficients are positive and consequently positive ω^2 cannot be found from (22.10). Thus no ω can be found such that $i\omega$ is a root of equation (22.10). By Rouche's theorem the real parts of all the eigenvalues have negative real part and consequently the solutions will be stable in nature. Therefore, the solutions in the neighbourhood of endemic equilibrium point (S_{11}, I_{11}, R_{11}) is locally asymptotically stable for $R_{04} > 1$ and $R_{03} > 1$, consequently by Kuang theorem, the endemic equilibrium point is asymptotically stable for $\tau \geq 0$. This concludes the proof.

22.5 Numerical Simulation

The numerical computation is done considering $b = 15.0, k = 0.008, d = 0.04, \gamma = 0.001, \mu = .01, \alpha = 0.12, \beta = 0.5$. The diseases free equilibrium point is $(375, 0, 0)$ in both the cases. For $\tau = 0$ the endemic equilibrium point is $A_{11}(343.32, 25.47, 6.21)$ for the first model. For $\tau = 2$ the endemic equilibrium point for the second model is $A_{21}(353.15, 17.56, 4.28)$. Here, the critical analysis is done considering the endemic equilibrium points only because it is important in real aspect. From Theorem 22.3 it is clear that $A_{11}(343.32, 25.47, 6.21)$ is global attractor of the first model and from Theorem 22.5 it is clear that $A_{21}(353.15, 17.56, 4.28)$ is a global attractor for the second model.

Since the number of susceptible increases more when delay is taken into consideration and compare to when delay is not taken into consideration. And consequently, the number of infected is increasing more when delay is taken into consideration which is clear from Figs. 22.2 and 22.3. It is also clear from the disease free equilibrium points that the number of susceptible will be more in delay considering case, i.e. those disease will spread fast which has no incubation period.

Again though the numbers R_{01}, R_{02}, R_{03} and R_{04} is not direct function of β but solution depends directly on β. When β decreases then the number of infected increases and when β increases then the number of infected is tending to zero and then the endemic equilibrium points reduces to disease free equilibrium point. Again since both S and I are depends on the another Inhibition parameter α, As α increases then S decreases and one time it becomes biologically invalid because then S becomes negative but I becomes a finite quantity.

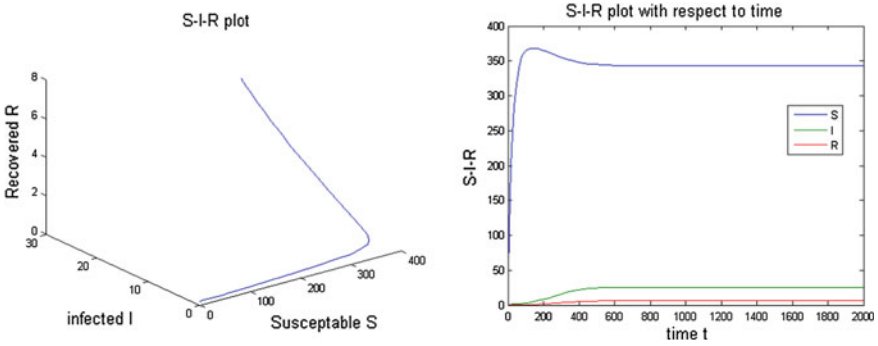


Fig. 22.2 Graph of S–I–R for $\tau = 0$

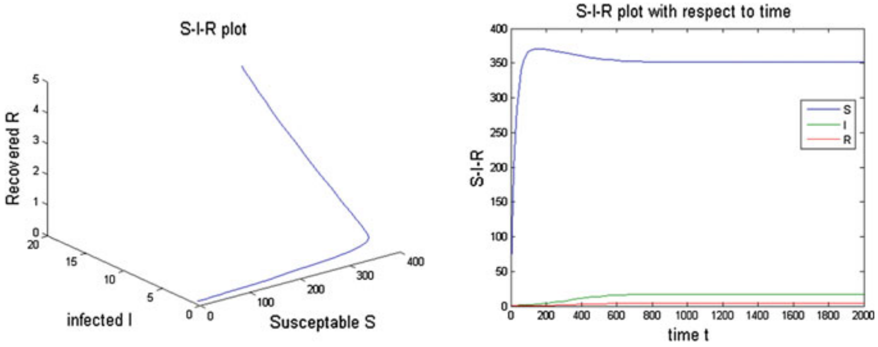


Fig. 22.3 Graph of S–I–R for $\tau = 2$

22.6 Discussion

Our goal of mathematical formulation of different epidemic modelling is to obtain some condition which will control processes. In the two models we obtain four numbers R_{01}, R_{02}, R_{03} and R_{04} depending on which we can conclude weather the solutions will be stable or unstable in nature. Since the numbers are directly dependent on different parameters such as birth rate, death rate, etc. controlling this parameter we can easily able to control the parameters but controlling some of the parameter are beyond of human capacity we avoid them uncontrolling parameters. In this problem, the inhibitory parameter (or the saturation affect) is directly controlling the stability of the solutions. From the numerical simulation it is clear that the number of infective individuals (22.1a) is directly affected by the parameters $\alpha$and β. Another important parameter $\tau$is also controlling the disease spreading mechanism. When τ increases the number of infected will decrease because of slow spreading of the disease.

Acknowledgments Authors are grateful to Prof. Susmita Sarkar, Department of Applied Mathematics, University of Calcutta for her valuable guidance at every step of this work. Authors also thank the reviewer for his valuable comments which has helped to improve the paper.

References

1. W.O. Kermack, A.G. McKendrick, Contribution to mathematical theory of epidemics. P. Roy. Soc. Lond. A Mat. **115**, 700–721 (1927)
2. N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases* (Griffin, London, 1975)
3. J.D. Murray, *Mathematical Biology* (Springer, New York, 1993)
4. Z. Ma, J. Li (eds.), *Dynamical Modeling and Analysis of Epidemics* (World Scientific, 2009)
5. R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1998)
6. H.W. Hethcote, D.W. Tudor, Integral equation models for endemic infectious diseases. J. Math. Biol. **9**, 37–47 (1980)
7. W.M. Liu, H.W. Hethcote, S.A. Levin, Dynamical behaviour of epidemiological models with nonlinear incidence rates. J. Math. Biol. **25**, 359–380 (1987)
8. W.M. Liu, S.A. Levin, Y. Iwasa, Influence of nonlinear incidence rates upon the behaviour of SIRS epidemiological models. J. Math. Biol. **23**, 187–204 (1986)
9. H.W. Hethcote, M.A. Lewis, P. van den Driessche, An epidemiological model with delay and a nonlinear incidence rate. J. Math. Biol. **27**, 49–64 (1989). ISSN: 1072-6691
10. D. Xiao, S. Ruan, Global analysis of an epidemic model with nonmonotone incidence rate. Math. Biosci. **208**, 419–429 (2007)
11. U. Ghosh, S. Chowdhury, D.K. Khan, *Mathematical Modelling of Epidemiology in Presence of Vaccination and Delay, Computer Science and Information Technology (CS and IT)*, 2013, pp. 91–98. doi:10.5121/csit.2013.3209
12. S. Pathak, A. Maiti, G.P. Samanta, Rich dynamics of an SIR epidemic model. Nonlinear Anal. Model. Control **15**(1), 71–81 (2010)
13. U. Ghosh, S. Sarkar, D.K. Khan, Modelling of infectious disease in presence of vaccination and delay. Int. J. Epidemiol. Infect. **2**(3), 50–57 (2014)
14. A. Kaddar, Stability analysis of delayed SIR epidemic model with saturated incidence rate. Nonlinear Anal. **3**, 299–306 (2010)

Chapter 23
Numerical Solutions of Incompressible Viscous Flows in a Double-Lid-Driven Cavity

Hemanta Karmakar and Swapan K. Pandit

Abstract Applications of the compact scheme based on 5-point stencil to spatial differencing of the streamfunction velocity formulation of the two-dimensional incompressible viscous flows governed by Navier-Stokes equations in a two-sided lid-driven rectangular cavity is presented. This cavity problem has multiple steady solutions for some aspect ratios. However, for the square cavity, the fluid flow problem produces only a single steady solution for both the parallel and antiparallel motion of the walls. The flow patterns are unlike to the one-sided lid-driven cavity flows. The transient solution involves different vortex structures and free share layers. The computed results show the accuracy, efficiency, and stability of the compact scheme even for higher Res. Results obtained are in well agreement with the numerical and experimental results available in the literature.

Keywords Transient solutions · Incompressible viscous fluid · Rotating secondary vortices

23.1 Introduction

The past few decades have seen the development of many numerical schemes [1–5]. Whenever, there is a new scheme developed for the study of computational fluid dynamics, it is used to study the benchmark problem of one-sided lid-driven cavity flow for code verification. Another classic example is the flow induced by the tangential movement of two facing cavity boundaries with uniform velocities [6, 7]. In the practical field, this is applied in several engineering situations, such as the flow over cutouts, designs, and repeated slots on the walls of heat exchangers or on the surface of aircraft bodies. If the two facing walls move in the same direction, it is termed

H. Karmakar (✉) · S.K. Pandit
Visva-Bharati, Integrated Science Education and Research Centre (ISERC),
Santiniketan 731235, West Bengal, India
e-mail: 1987hemanta@gmail.com

S.K. Pandit
e-mail: swapankumar.pandit@visva-bharati.ac.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_23

as parallel wall motion and if in the opposite direction, it is termed as antiparallel wall motion. Kuhlmann and other investigators [7] (see, the references therein) have done some experimental and computational work on two-sided lid-driven rectangular cavity with various aspect ratios. Furthermore, in contrast to the fairly large number of studies conducted for single-sided lid-driven cavities, only a few investigations have been carried out for flows in two-sided lid-driven cavities. Overall, no compact schemes have been found to solve the two-sided lid-driven cavity flows with high Reynolds number.

The aim of this paper is to study the transient solutions even for higher Res for both the parallel and antiparallel motion of the walls.

23.2 Problem

An incompressible viscous flow in a square cavity whose two walls, i.e., top and bottom side moves in a same or opposite direction with uniform velocity is the problem which we have focused in our present work. The other vertical walls are kept stationary. The boundary condition of above type motion are shown in the Fig. 23.1.

The governing equations describing the incompressible viscous flows in a two-sided lid-driven cavity are the Navier-Stokes equations (N-S) which can be written in terms of non-dimensional streamfunction (ψ)-vorticity (ζ) form as follows:

$$-\frac{\partial^2\psi}{\partial x^2}-\frac{\partial^2\psi}{\partial y^2}=\zeta, \tag{23.1}$$

$$Re\frac{\partial\zeta}{\partial t}-\frac{\partial^2\zeta}{\partial x^2}-\frac{\partial^2\zeta}{\partial y^2}+uRe\frac{\partial\zeta}{\partial x}+vRe\frac{\partial\zeta}{\partial y}=0 \tag{23.2}$$

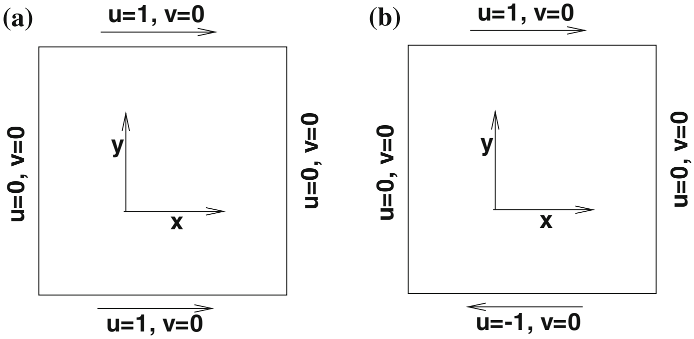


Fig. 23.1 Schematic picture of two-sided lid-driven cavity with boundary conditions **a** parallel and **b** antiparallel wall motion

where $V = (u, v)$ is the velocity vector and $Re \left(= \frac{V_0 L}{\nu}\right)$ is the Reynolds number with V_0, L, ν are, respectively, reference velocity, cavity length, and kinematic viscosity.

Now eliminating ζ from the above two equations, the pure streamfunction formulation of the full governing equations including the source term f can be written as

$$l\frac{\partial}{\partial t}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}\right)+a\frac{\partial^4\psi}{\partial x^4}+b\frac{\partial^4\psi}{\partial x^2\partial y^2}+c\frac{\partial^4\psi}{\partial y^4}+d\left(\frac{\partial^3\psi}{\partial x^3}+\frac{\partial^3\psi}{\partial x\partial y^2}\right)$$
$$+e\left(\frac{\partial^3\psi}{\partial y^3}+\frac{\partial^3\psi}{\partial x^2\partial y}\right)=f \tag{23.3}$$

where $l = Re, a = -1, b = -2, c = -1, d = uRe, e = vRe$, and $f = 0$.

Assuming the physical domain to be rectangular and constructing on it a uniform rectangular mesh of steps h and k in the x and y-directions, respectively, the discretized form of Eq. (23.3) at the (i, j)th node is given by

$$l\left(\delta_x^2\delta_t^+\psi_{i,j}^n+\delta_y^2\delta_t^+\psi_{i,j}^n\right)+\frac{12}{h^2}a_{i,j}\left(-\delta_x^2\psi_{i,j}-\delta_x v_{i,j}\right)$$
$$+\frac{1}{2}b_{i,j}\left(-\delta_x\delta_y^2 v_{i,j}+\delta_x^2\delta_y u_{i,j}\right)+\frac{12}{k^2}c_{i,j}\left(-\delta_y^2\psi_{i,j}+\delta_y u_{i,j}\right)+d_{i,j}\left(-\delta_x^2 v_{i,j}\right)$$
$$+e_{i,j}\delta_x^2 u_{i,j}+d_{i,j}\left(-\delta_y^2 v_{i,j}\right)+e_{i,j}\delta_y^2 u_{i,j}=f_{i,j}+O\left(\Delta t, h^2, k^2\right), \tag{23.4}$$

where we have used $u = \frac{\partial\psi}{\partial y}$ and $v = -\frac{\partial\psi}{\partial x}$. After having a second-order approximations in (23.4) for space, we now intend to discretize time derivative as accurately as possible and obtain a stable numerical scheme. Introducing weighted time average parameter μ such that $t_\mu = (1-\mu)t^{(n)} + \mu t^{(n+1)}$ for $0 \leq \mu \leq 1$, where n denote the nth time level, we can have a family of integrators; for example, forward Euler for $\mu = 0$, backward Euler for $\mu = 1$ and Crank-Nicholson for $\mu = 0.5$.

23.3 Results and Discussions

In Table 23.1, the vortex centres of the primary and secondary vortices have been presented and compared the results presented in [6] (values within the parenthesis). The transient flow pattern (see Fig. 23.2) of incompressible viscous fluid in a square cavity whose top and bottom lids are moving along the positive x-axis direction with the uniform velocity has been shown in Fig. 23.2. With the advancement of time, a pair of counter-rotating secondary vortices is seen to be appeared near the center of the right wall which is placed symmetrically about the horizontal centerline.

Table 23.1 Location of primary and secondary vortex center for parallel wall motion

Re	Primary vortex centre		Secondary vortex centre	
	Bottom		Bottom	
	x	y	x	y
100	0.6167(0.6145)	0.2001(0.2026)	–	–
400	0.5833(0.5845)	0.2416(0.2388)	0.9834(0.9873)	0.4833(0.4638)
1000	0.5333(0.5354)	0.2417(0.2452)	0.9583(0.9551)	0.4583(0.4570)
2000	0.5167(0.5132)	0.2425(0.2474)	0.9417(0.9400)	0.4584(0.4573)

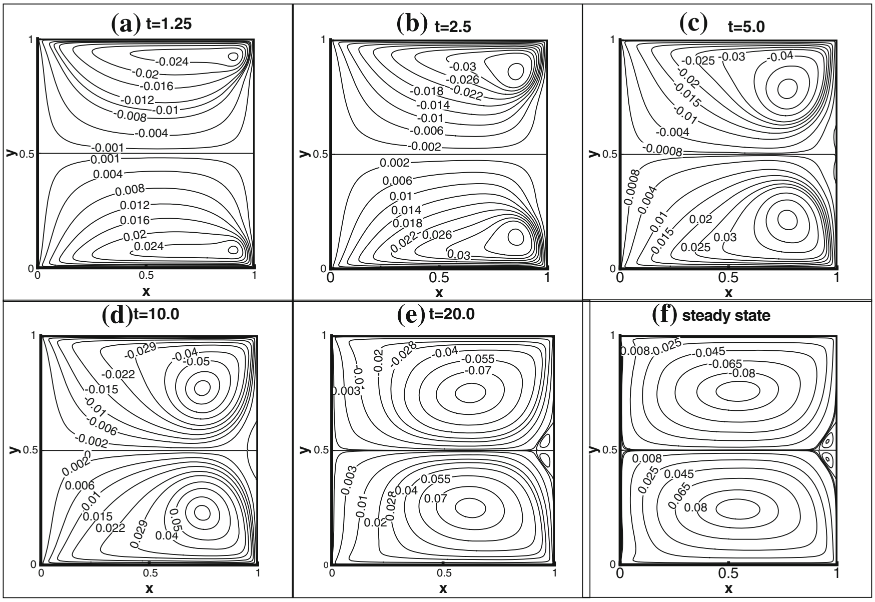


Fig. 23.2 Parallel wall motion of the cavity flow problem: evolution of streamlines at different time stations for the lid-driven cavity flow with aspect ratio 1.0 for Re = 1000

In Table 23.2, we have presented the primary and secondary vortex centres formed in antiparallel motion and compared the results presented in [6] (values within the parenthesis). The transient flow pattern of an incompressible viscous fluid is shown in Fig. 23.3 in which the top and bottom walls move into the opposite direction with uniform velocity. It is seen that in the steady-state only one primary vortex formed at the geometric center of the cavity. It is also seen that two secondary vortices formed at the top left and bottom right corners of the cavity and the position of the primary vortex center is same as the geometric center of the cavity.

In Fig. 23.4, we have shown the horizontal velocity profile along the vertical centerline for parallel wall motion from $Re = 100$ to 3200.

Table 23.2 Location of upper and lower vortex centre for antiparallel wall motion

Re	Primary vortex centre		Secondary vortex centre			
			Bottom		Top	
	x	y	x	y	x	y
100	0.5000(0.5001)	0.5000(0.5002)	–	–	–	–
400	0.5000(0.5002)	0.5000(0.4981)	–	–	–	–
1000	0.5000(0.5009)	0.5000(0.4980)	0.9583(0.9507)	0.1333(0.1319)	0.0417(0.0492)	0.8667(0.8663)
2000	0.5000(0.5002)	0.5000(0.5001)	0.9250(0.9227)	0.1083(0.1082)	0.075(0.0771)	0.8917(0.8920)

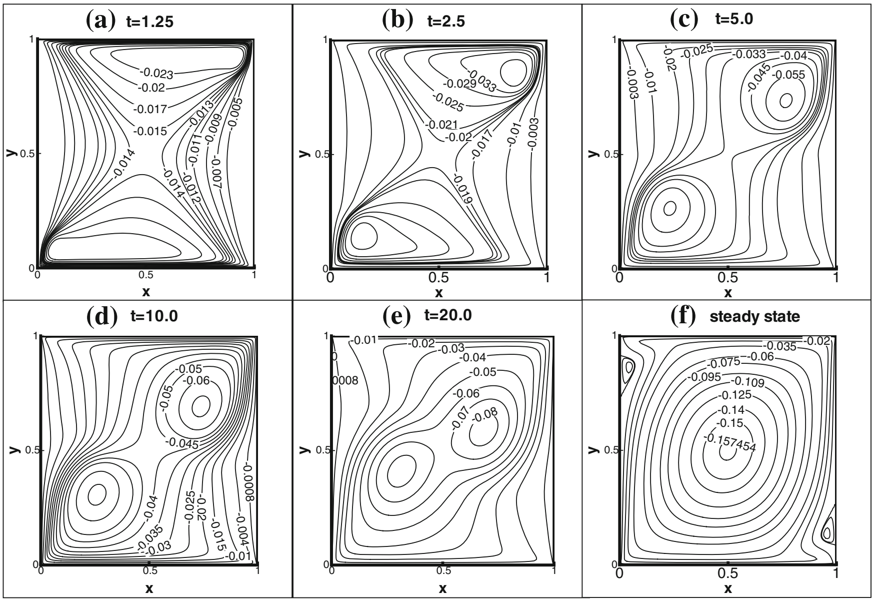


Fig. 23.3 Antiparallel wall motion of the cavity flow problem: evolution of streamlines at different time stations for the lid-driven cavity flow with aspect ratio 1.0 for Re = 1000

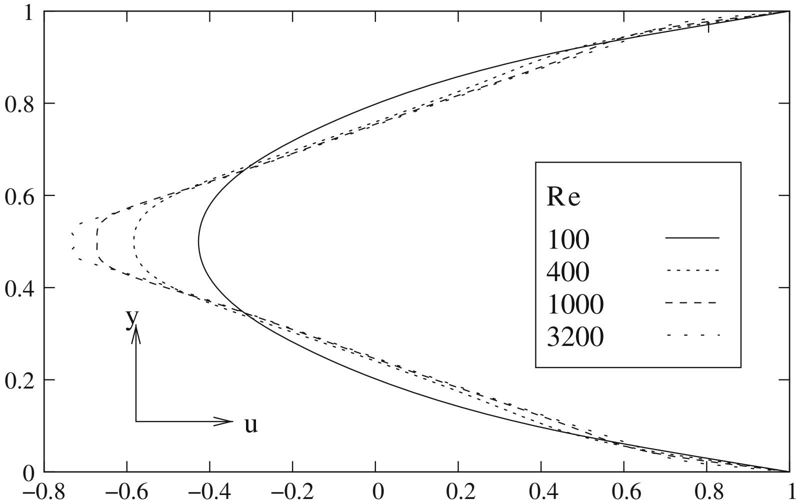


Fig. 23.4 Horizontal velocity *u* along the vertical centerline for parallel wall motion from Re = 100 to 3200

23.4 Conclusion

The present work involves the computation of incompressible flows in a two-sided lid-driven cavity using time dependent compact scheme based on 5-point stencil to spatial differencing of the streamfunction velocity formulation. We have investigated the transient flow for both parallel and antiparallel motion of the two facing walls. The transient solution reveals different vortex structures and free share layers which are unlike to the one-sided lid-driven cavity flows. Results obtained are in well agreement with the available numerical results.

References

1. M.M. Gupta, R.P. Manohar, J.W. Stephenson, A singel cell high order scheme for the convection-diffusion equation with variable coefficients. Int. J. Numer. Methods Fluids **4**, 641–651 (1984)
2. R.A. Kupferman, A central-difference scheme for a pure stream function formulation of incompressible viscous flows. SIAM J. Sci. Comp. **23**(1), 1–18 (2001)
3. J.D. Hoffman, Relationship between the truncation errors of centered finite-difference approximations on uniform and nonuniform meshes. J. Comput. Phys. **46**, 469–474 (1982)
4. M. Ben-Artzi, J.P. Croisille, D. Fishelov, S. Trachtenberg, A pure-compact scheme for the streamfunction formulation of Navier-Stokes equations. J. Comput. Phys. **205**, 640–664 (2005)
5. S.K. Pandit, J.C. Kalita, D.C. Dalal, A transient higher order compact scheme for incompressible viscous flows on geometries beyond rectangular. J. Comput. Phys. **225**, 1100–1124 (2007)
6. D.A. Perumal, A.K. Dass, Simulation of incompressible flows in two sided lid driven cavities Part I—FDM. CFD Lett. **2**(1), 13–24 (2010)
7. C.H. Blhom, H.C. Kulhmann, The two sided lid driven cavity: experiments on stationary and time dependent flows. J. Fluid Mech. **450**, 67–95 (2002)

Chapter 24
Propagation of SH-Type Wave in Anisotropic Layer Overlying an Anisotropic Viscoelastic Half-Space

S. Kumar and P.C. Pal

Abstract An analysis has been carried out for the propagation of SH-type wave in anisotropic layer overlying anisotropic viscoelastic half-space of higher order. The dispersion relation is obtained under certain boundary conditions. The numerical results are discussed through figures for a particular model by plotting the graph between phase velocity and wave number for different values of thickness of layer.

Keywords SH-type wave · Anisotropic layer · Inhomogeneity · Viscoelastic coefficient · Phase velocity

24.1 Introduction

Nowadays, for engineers, physicists and seismologists have become a great challenge to explore the interior of earth due to high demand of raw materials like minerals, crude oils, coal, natural gases, etc. for the industries and fulfil the needs of growing population. The study of wave propagation is making revolution for mankind. It helps in exploring or predicting the hidden resources in the earth. Also, the number of earthquakes increasing day by day around the world draws attention for seismologists to study the seismic waves, as we know that the earth is highly inhomogeneous and anisotropic and some materials exhibit viscoelastic behaviour. In order to describe the nature of wave propagation accurately, we have to consider anisotropy with viscoelastic properties of materials. When seismic waves propagate underground, then they are not only influenced by anisotropy of the medium but also by intrinsic viscosity of the medium [1]. Das and Sengupta [2] have discussed the surface-wave propagation in general viscoelastic media of higher order. They considered the general theory of surface waves in higher order viscoelastic solid containing time rate of strain and investigated the particular surface waves of Rayleigh, Love and Stoneley

S. Kumar (✉) · P.C. Pal
Department of Applied Mathematics, Indian School of Mines, Dhanbad 826004, India
e-mail: santosh453@gmail.com

P.C. Pal
e-mail: pcpal.ism@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_24

type. Kakar et al. [3] have studied the propagation of Rayleigh, Love and Stoneley waves in fibre-reinforced, general viscoelastic media of higher order (nth order) including time strain under the effect of gravity.

In the present investigation, an attempt has been made to study the behaviour of SH-type wave when upper boundary plane is considered as free surface. The layer is anisotropic and half-space is of anisotropic viscoelastic material of higher order. The dispersion relation is obtained in determinant form. The numerical results are discussed through figures for a particular model, and effects of thickness of layer on phase velocity are shown.

24.2 Formulation of the Problem and its Solution

Here, we consider an anisotropic elastic layer of finite thickness h lying over a half-space of anisotropic viscoelastic material of higher order. The interface of these two mediums is considered at $z = 0$, whereas free surface is at $z = -h$. Here, z axis is directed vertically downward and x axis is assumed in the direction of the propagation of wave with velocity c. For SH-type of waves, the displacement and body forces do not depend on y, and if (u, v, w) be the displacement at any point $P\ (x, y, z)$ into the medium, then $u = w = 0$ and v is the function of x, z and t. The two equations of motions are identically satisfied (Fig. 24.1).

For anisotropic layer, the equation of motion for SH-type wave without body forces is given by

$$\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yz}}{\partial z} = \rho_1 \frac{\partial^2 v_1}{\partial t^2} \tag{24.1}$$

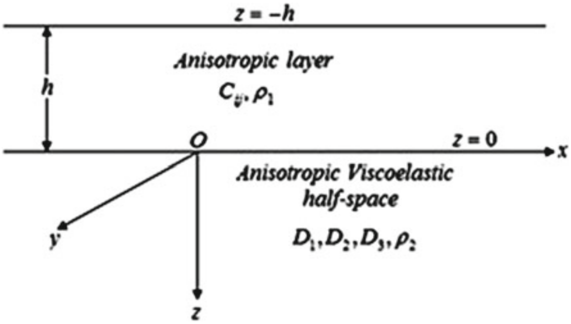


Fig. 24.1 Geometry of the problem

The stress–strain relations and density are taken as

$$\left.\begin{aligned}\tau_{xy} &= C_{46}\frac{\partial v_1}{\partial z} + C_{66}\frac{\partial v_1}{\partial x}\\ \tau_{yz} &= C_{44}\frac{\partial v_1}{\partial z} + C_{46}\frac{\partial v_1}{\partial x}\end{aligned}\right\} \tag{24.2}$$

Substituting (24.2) in (24.1), we have

$$C_{66}\frac{\partial^2 v_1}{\partial x^2} + 2C_{46}\frac{\partial^2 v_1}{\partial x \partial z} + C_{44}\frac{\partial^2 v_1}{\partial z^2} = \rho_1 \frac{\partial^2 v_1}{\partial t^2} \tag{24.3}$$

Assuming the solution as $v_1\,(x, z, t) = V_1\,(z)\, e^{ik(x-ct)}$ and substituting in (24.3), we have

$$\frac{d^2 V_1}{dz^2} + 2ik\alpha_1 \frac{dV_1}{dz} - k^2 \alpha_2 \left\{\frac{c^2}{\beta_1^2} - 1\right\} V_1 = 0 \tag{24.4}$$

where $\alpha_1 = \frac{C_{46}}{C_{44}}$, $\alpha_2 = \frac{C_{66}}{C_{44}}$ and $\beta_1^2 = \frac{C_{66}}{\rho_1}$.

The solution of equation (24.4) is given as

$$V_1\,(z) = Ae^{-iks_1 z} + Be^{iks_2 z} \tag{24.5}$$

where $s_1 = \alpha_1 + \sqrt{\alpha_1^2 + \alpha_2\left\{\frac{c^2}{\beta_1^2} - 1\right\}}$ and $s_2 = -\alpha_1 + \sqrt{\alpha_1^2 + \alpha_2\left\{\frac{c^2}{\beta_1^2} - 1\right\}}$.

Hence, the displacement and stress component for anisotropic layer are given by

$$v_1\,(x, z, t) = V_1\,(z)\, e^{ik(x-ct)} = \left(Ae^{-iks_1 z} + Be^{iks_2 z}\right) e^{ik(x-ct)} \tag{24.6}$$

$$\left(\tau_{yz}\right)_I = C_{44}\frac{\partial v_1}{\partial z} + C_{46}\frac{\partial v_1}{\partial x} \tag{24.7}$$

For anisotropic viscoelastic half-space, the equation of motion for SH-type wave without body forces is given by

$$\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yz}}{\partial z} = \rho_2 \frac{\partial^2 v_2}{\partial t^2} \tag{24.8}$$

The stress–strain relations as considered by Flugge [4] are

$$\left.\begin{aligned}\tau_{xy} &= D_2 \tfrac{\partial v_2}{\partial z} + D_3 \tfrac{\partial v_2}{\partial x}\\ \tau_{yz} &= D_1 \tfrac{\partial v_2}{\partial z} + D_2 \tfrac{\partial v_2}{\partial x}\end{aligned}\right\} \tag{24.9}$$

where $D_1 = \sum_{\lambda=0}^{n} C_{44}^{\lambda} \frac{\partial^\lambda}{\partial t^\lambda}$, $D_2 = \sum_{\lambda=0}^{n} C_{46}^{\lambda} \frac{\partial^\lambda}{\partial t^\lambda}$ and $D_3 = \sum_{\lambda=0}^{n} C_{66}^{\lambda} \frac{\partial^\lambda}{\partial t^\lambda}$.

Substituting (24.9) in (24.8), we have

$$D_3 \frac{\partial^2 v_2}{\partial x^2} + 2D_2 \frac{\partial^2 v_2}{\partial x \partial z} + D_1 \frac{\partial^2 v_2}{\partial z^2} = \rho_2 \frac{\partial^2 v_2}{\partial t^2} \tag{24.10}$$

Assuming the solution as $v_2\,(x, z, t) = V_2\,(z)\,e^{ik(x-ct)}$ and substituting in (24.10), we have

$$\frac{d^2 V_2}{dz^2} + 2ik\alpha_4 \frac{dV_2}{dz} - k^2 \alpha_5 \left\{ 1 - \frac{c^2}{\beta_2^2} \right\} V_2 = 0 \tag{24.11}$$

where $\alpha_4 = \frac{D_2'}{D_1'}, \alpha_5 = \frac{D_3'}{D_1'}$, $\beta_2^2 = \frac{D_3'}{\rho_2}$ and $D_1' = \sum_{\lambda=0}^{n} C_{44}^{\lambda}\,(-ikc)^{\lambda}$, $D_2' = \sum_{\lambda=0}^{n} C_{46}^{\lambda}\,(-ikc)^{\lambda}$, $D_3' = \sum_{\lambda=0}^{n} C_{66}^{\lambda}\,(-ikc)^{\lambda}$.

The solution of equation (24.11) is given as

$$V_2\,(z) = Ce^{-ikm_1 z} + De^{ikm_2 z} \tag{24.12}$$

where $m_1 = \alpha_4 + \sqrt{\alpha_4^2 + \alpha_5 \left\{ \frac{c^2}{\beta_2^2} - 1 \right\}}$ and $m_2 = -\alpha_4 + \sqrt{\alpha_4^2 + \alpha_5 \left\{ \frac{c^2}{\beta_2^2} - 1 \right\}}$.

Hence, the displacement and stress component for half-space are given by

$$v_2\,(x, z, t) = Ce^{-ik(m_1 z - x + ct)} \tag{24.13}$$

$$\left(\tau_{yz}\right)_{II} = D_1 \frac{\partial v_2}{\partial z} + D_2 \frac{\partial v_2}{\partial x} \tag{24.14}$$

24.3 Boundary Conditions

We assume that anisotropic layer and the half-space are in welded contact. Therefore, the boundary conditions are the continuity of displacement and stress at the interface. Mathematically, these boundary conditions can be expressed as follows: (i) $v_1 = v_2$ at $z = 0$. (ii) $\left(\tau_{yz}\right)_I = \left(\tau_{yz}\right)_{II}$ at $z = 0$. (iii) $\left(\tau_{yz}\right)_I = 0$ at $z = -h$ (upper boundary as free surface).

Substituting (24.6), (24.7), (24.13) and (24.14) in the above boundary conditions, we have three homogeneous equation in A, B and C, and eliminating A, B and C from these equations, we have

$$\tanh\left[ikh\sqrt{\alpha_1^2+\alpha_2\left\{\frac{c^2}{\beta_1^2}-1\right\}}\right]=\frac{\alpha_2\left\{\frac{c^2}{\beta_1^2}-1\right\}-2\sqrt{\alpha_1^2+\alpha_2\left\{\frac{c^2}{\beta_1^2}-1\right\}}}{2\alpha_1\left(\alpha_4+\sqrt{\alpha_4^2+\alpha_5\left\{\frac{c^2}{\beta_2^2}-1\right\}}\right)} \tag{24.15}$$

Equation (24.15) is the dispersion relation of SH-type wave propagation in anisotropic layer overlying anisotropic viscoelastic half-space of higher order.

24.4 Numerical Results and Discussion

In order to show the dependency of phase velocity on wave number, we have taken data for anisotropic medium from Rasolofosaon and Zinszner [5].

$C_{44} = 25.97$ Gpa, $C_{46} = 0.43$ Gpa, $C_{66} = 37.82$ Gpa and $\rho_1 = 2727\,\text{kg/m}^3$.

For anisotropic viscoelastic half-space, the viscoelastic coefficients are considered up to second order and are taken as follows:

$\mathrm{C}_{44}^0 = 324$ Gpa, $\mathrm{C}_{44}^1 = 198$ Gpa, $\mathrm{C}_{44}^2 = 248$ Gpa, $\mathrm{C}_{46}^0 = 59$ Gpa, $\mathrm{C}_{46}^1 = 78$ Gpa, $\mathrm{C}_{46}^2 = 79$ Gpa, $\mathrm{C}_{66}^0 = 79.7$ Gpa, $\mathrm{C}_{66}^1 = 66.1$ Gpa, $\mathrm{C}_{66}^2 = 81$ Gpa, $\rho_2 = 3320\,\text{kg/m}^3$.

The graphs are plotted separately for both real and imaginary parts for phase velocity against wave number. In Fig. 24.2, the graph is plotted for real part of phase velocity, i.e. Re $\left(\frac{c}{\beta_1}\right)$, against non-dimensional wave number kh for different values of thickness of layer. The figure reveals the fact that the phase velocity decreases for increasing values of kh, but as we increase the thickness of layer, the magnitude of phase velocity decreases for all values of kh. In Fig. 24.3, the graph is plotted for imaginary part of phase velocity, i.e. Im $\left(\frac{c}{\beta_1}\right)$, against non-dimensional wave number

Fig. 24.2 Variation of phase velocity, i.e. Re $\left(\frac{c}{\beta_1}\right)$, against non-dimensional wave number kh

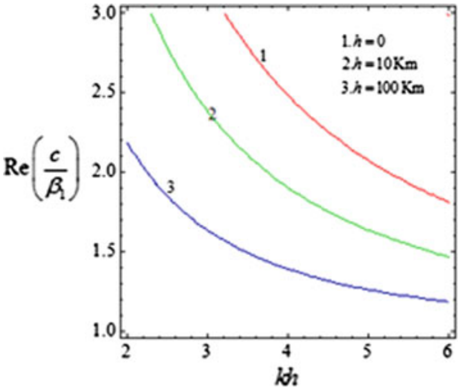
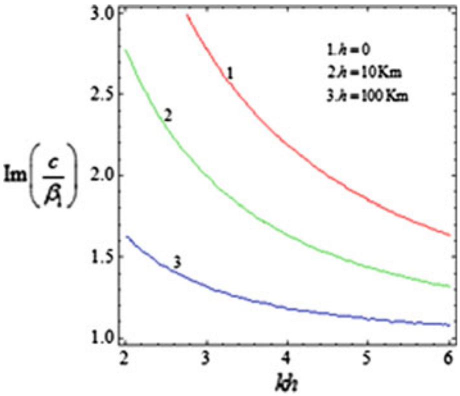


Fig. 24.3 Variation of phase velocity, i.e. Im $\left(\frac{c}{\beta_1}\right)$, against non-dimensional wave number kh



kh for different values of thickness of layer. It can be observed from the figure that the rate at which phase velocity decreases is little higher than for real phase velocity, i.e. the magnitude of phase velocity is different but the nature of curve is almost same for both the graphs.

24.5 Conclusions

The propagation of SH-type surface waves is investigated in anisotropic layer lying over anisotropic viscoelastic half-space of higher order. The solutions for layer and half-space are obtained analytically and dispersion relation is obtained. The numerical results are discussed through figures. From figures, it can be concluded that the thickness of layer has great impact on the phase velocity. So the thickness of layers within the earth plays a vital role in seismic wave propagation.

Acknowledgments One of the authors, Mr. S. Kumar, is grateful to the I.S.M., Dhanbad authorities for financial support and facilitating us with best facility.

References

1. J.M. Carcione, Wave propagation in anisotropic linear viscoelastic media: theory and simulated wavefields. Geophys. J. Int. **101**(3), 739–750 (1990)
2. T.P. Das, P.R. Sengupta, Surface waves in general visco-elastic media of higher order. Indian J. pure Appl. Math. **21**(7), 661–675 (1990)
3. R. Kakar, S. Kakar, K. Kaur, Rayleigh, love, stonely waves in fibre-reinforced, general viscoelastic media of higher order under gravity. Int. J. Phys. Math. Sci. **4**(1), 53–61 (2013)
4. W. Flugge, *Visco-Elasticity* (Blaisdell Publishing Co., Waltham, 1967)
5. P.N.J. Rasolofosaon, B.E. Zinszner, Comparison between permeability anisotropy and elasticity anisotropy of reservoir rocks. Geophysics **67**, 230–240 (2002)

Chapter 25
Global Stability and Chaos-Control in Delayed N-Cellular Neural Network Model

Amitava Kundu and Pritha Das

Abstract In this paper stability, bifurcation, and chaotic behavior of a cellular neural network (CNN) model which is a regular array of n (≥ 3) cells with continuous activation function are presented. In the delayed cellular neural network model (DCNN) criteria for the global asymptotic stability of the equilibrium point is presented by constructing suitable Lyapunov functional. Numerical simulations are given to verify the analytical results. The role of delay in chaos control of the CNNs has been shown numerically.

Keywords Time delay · Global asymptotic stability · Chaos control · Cellular neural network

25.1 Introduction

In this paper, the generalization to cellular neural network (CNN) model (introduced by Chua and Yang [2]) for neurons with two-way (bidirectional) time delayed connections between the neurons and itself using delay-differential equations is studied. In recent years, neural networks (especially, Hopfield type, cellular, and bidirectional associative memory, recurrent neural networks) have been applied successfully in many areas, such as signal processing, pattern recognition, associative memories [6, 8]. Processing of moving images requires the introduction of delay in the signal transmitted among the cells [7]. Brain areas are assumed to be bidirectionally (BAM) coupled forming delayed feedback loops. The malfunctioning of the neural system is often related to changes in the delay parameter causing unmanageable shifts in the phases of the neural signals [4].

We are motivated to study effectiveness of time delay as well as synaptic weights in changing the dynamics of n-dimensional BAM cellular neural network. With this

A. Kundu · P. Das (✉)
Department of Mathematics, IIEST, Shibpur, Botanic Garden, Howrah 711103, India
e-mail: prithadas01@yahoo.com

A. Kundu
e-mail: akbesu@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_25

motivation, we studied the global stability analysis of the system and obtained sufficient criteria with respect to synaptic weights. We investigated the system numerically without time delay showing complex dynamics including chaos but the chaotic behavior of the system is controlled as the interconnection transmission delay is introduced.

This paper is structured as follows: In Sect. 25.2 a sufficient condition for the global asymptotic stability of the equilibrium point in the delayed model is found by using Lyapunov functional method. In Sect. 25.3, numerical simulations are presented to verify the analytical results. Besides numerical results are discussed showing changes of dynamics of the system from unstable to stable (chaos control [1, 3, 5, 10]) due to time delay. Finally, some concluding remarks have been drawn on the implication of our results in the context of related work mentioned above in Sect. 25.4.

25.2 Mathematical Model with Time Delay and Global Stability

We consider, an artificial n-neuron network model of cellular neural networks time delayed connections between the neurons by the delay differential equations:

$$\frac{dx_i}{dt} = -c_i x_i(t) + a_{ii} f\,[x_i(t)] + \sum_{j=1, i\neq j}^{n} b_{ij} f\left[x_j(t-\tau_j)\right],\ c_j > 0,\ i = 1, 2, 3, \ldots, n \tag{25.1}$$

with $x_i(t)$ is the activation state of ith neuron at time t, $f[x_i(t)]$ is the output state of the ith neuron at time t, a_{ii} is self-synaptic weight, b_{ij} is the strength of the jth neuron on the ith neuron at time $(t - \tau_j)$, τ_j is the signal transmission delay along the axon of the jth unit and is nonnegative constant and c_i (decay rate) is the rate with which the ith neuron will reset its potential to the resting state in isolation when disconnected from the network. In the following, we assume that each of the relation between the output of the cell f and the state of the cell possess following properties:

(H1) f is bounded on $\mathscr{R}$.
(H2) There is a number $\mu > 0$ such that $\mid f(u) - f(v) \mid \leq \mu \mid u - v \mid$ for any $u, v \in \mathscr{R}$.

It is easy to find from (H2) that f is a continuous function on $\mathscr{R}$. In particular, if output state of the cell is described by $f(x_i) = \tanh(x_i)$, then it is easy to see that the function f, clearly satisfy the hypotheses (H1) and (H2). To clarify our main results, we present following two lemmas.

Lemma 25.1 *For the DCNN (1), suppose that the output of the cell f satisfy the hypotheses (H1) and (H2) above. Then all solutions of the DCNN (1) remain bounded for [0,+∞).*

Lemma 25.2 *Let f(t) be a nonnegative function defined on [0,+∞) such that f(t) is integrable (that is* $\int_0^\infty f(t)\,dt < +\infty$*) and uniformly continuous on [0,+∞). Then* $\lim_{t\to\infty}$ *f(t) = 0.*

Theorem 25.1 *For the DCNN (1), suppose that the outputs of the cell [f(*$x_i(t)$*)] satisfy the hypotheses (H1) and (H2) above and there exists constants* $\omega_i > 0$, $\omega_j > 0$ $(i, j = 1, 2, \ldots, n)$ *such that*

$$\mu^2\left\{\left(2 \mid a_{ij} \mid^2\right) + \sum_{j=1, i\neq j}^{n}\left(\mid b_{ij} \mid^2 + \frac{\omega_j}{\omega_i} \mid b_{ij} \mid^2\right)\right\} < 2c_i \tag{25.2}$$

$i = 1, 2, \ldots, n$, *in which* μ *is a constant number of the hypothesis (H2) above. Then the equilibrium x* of the DCNN (1) is also globally asymptotically stable independent of delays.*

Applying Theorem 25.1 above, we can easily establish the following corollary.

Corollary 25.1 *For the DCNN (1), suppose output of the cell [f(*$x_i(t)$*)] satisfy the hypotheses (H1) and (H2) above and there exists constant* μ *such that*

$$\mu^2\left\{\left(2 \mid a_{ii} \mid^2\right) + \sum_{j=1, i\neq j}^{n}\left(\mid b_{ij} \mid^2\right)\right\} < 2c_i \tag{25.3}$$

$i = 1, 2, \ldots, n$, *in which* μ *is a constant number of the hypothesis (H2) above. Then the equilibrium* x^* *of the DCNN (1) is also globally asymptotically stable independent of delays.*

25.3 Numerical Results

In this section, the analytical results obtained above are verified by using numerical examples given below. We used Matlab 7.10 for simulation. Besides, in numerical portion we assume $n = 5$ and excitatory self-connections but excitatory and inhibitory interconnecting synaptic weights.

25.3.1 Example 1

$$\begin{aligned}
\dot{x}_1 &= -10x_1(t) + 2.1\tanh[x_1(t)] + 2.17\tanh[x_2(t-\tau)]\\
\dot{x}_2 &= -20x_2(t) - 3.5\tanh[x_1(t-\tau)] + \tanh[x_2(t)] + 3.11\tanh[x_3(t-\tau)]\\
\dot{x}_3 &= -20x_3(t) - 1.425\tanh[x_2(t-\tau)] + 3.4\tanh[x_3(t)] - 1.1\tanh[x_4(t-\tau)]
\end{aligned}$$

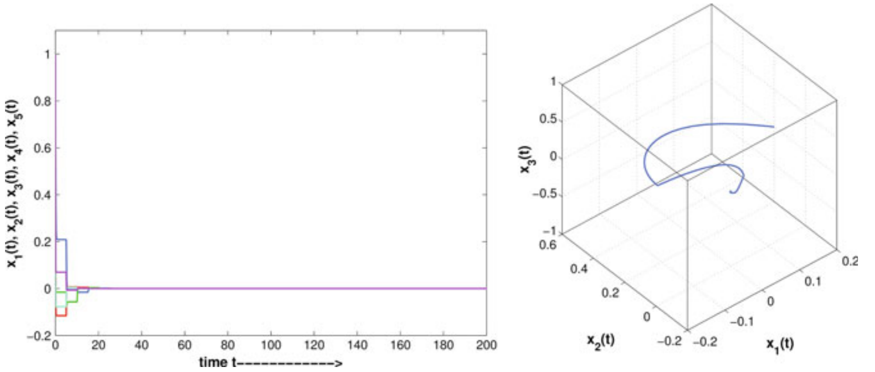


Fig. 25.1 Solution trajectory and corresponding phase portrait in (x_1, x_2, x_3) space showing stable behavior satisfying globally asymptotically stable conditions (25.3)

$$\begin{aligned}
\dot{x}_4 &= -30x_4(t) - 1.75\tanh[x_3(t-\tau)] + 2\tanh[x_4(t)] - 1.1\tanh[x_5(t-\tau)]\\
\dot{x}_5 &= -25x_5(t) + 2\tanh[x_4(t-\tau)] + 3\tanh[x_5(t-\tau)]\,.
\end{aligned}$$

Time series plot and phase portrait of Example 1 showing stable behavior satisfying conditions (25.3) is illustrated in Fig. 25.1.

But without any time delay and violating the restriction of weight parameters for global asymptotic stability, we consider the next example.

25.3.2 Example 2

$$\begin{aligned}
\dot{x}_1 &= -1.1x_1(t) + 2.1\tanh[x_1(t)] + 2.17\tanh[x_2(t)]\\
\dot{x}_2 &= -1.5x_2(t) - 2.51\tanh[x_1(t)] + \tanh[x_2(t)] + 3.11\tanh[x_3(t)]\\
\dot{x}_3 &= -2.5x_3(t) - 1.75\tanh[x_2(t)] + 3.4\tanh[x_3(t)] - 7\tanh[x_4(t)]\\
\dot{x}_4 &= -15x_4(t) - 6.75\tanh[x_3(t)] + 16.9\tanh[x_4(t)] - 10.1\tanh[x_5(t)]\\
\dot{x}_5 &= -25x_5(t) + 10\tanh[x_4(t)] + 30\tanh[x_5(t)]\,.
\end{aligned}$$

The phase portraits (Fig. 25.2) show the evidence that there exists chaotic attractors with the above set of parameter values in Example 2.

This chaotic nature is controlled as interconnection transmission delay is introduced (see Example 3) of the previous example with same parameter values and periodic behavior is illustrated in Fig. 25.3.

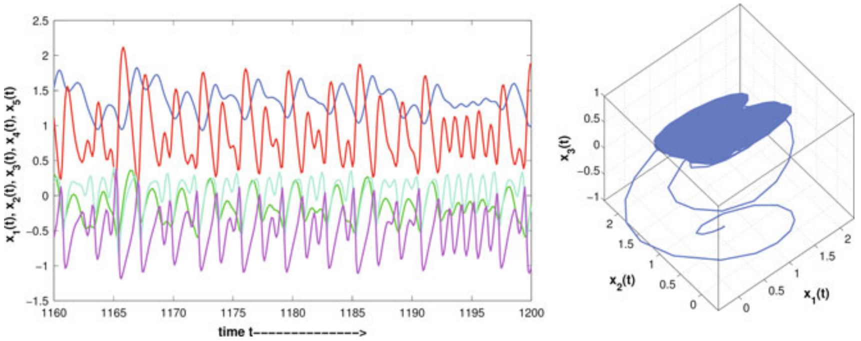


Fig. 25.2 Solution trajectory showing chaotic behavior and corresponding phase portrait in (x_1, x_2, x_3) space

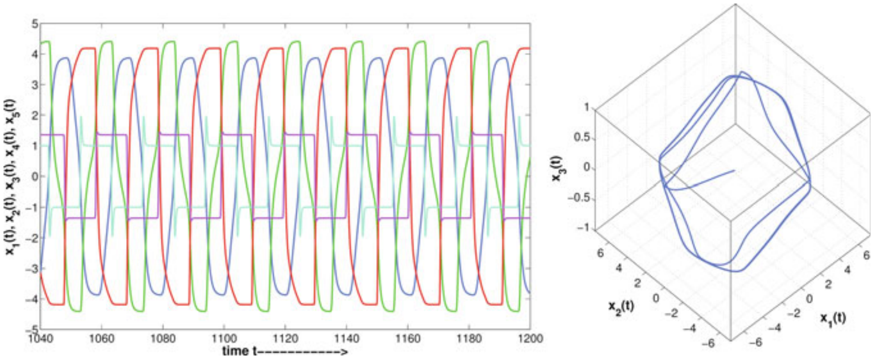


Fig. 25.3 Control of chaos: Solution trajectory showing periodic behavior corresponding phase portrait in (x_1, x_2, x_3) space

25.3.3 Example 3

$$
\begin{aligned}
\dot{x}_1 &= -1.1x_1(t) + 2.1\tanh\left[x_1(t)\right] + 2.17\tanh\left[x_2(t-\tau)\right] \\
\dot{x}_2 &= -1.5x_2(t) - 2.51\tanh\left[x_1(t-\tau)\right] + \tanh\left[x_2(t)\right] + 3.11\tanh\left[x_3(t-\tau)\right] \\
\dot{x}_3 &= -2.5x_3(t) - 1.75\tanh\left[x_2(t-\tau)\right] + 3.4\tanh\left[x_3(t)\right] - 7\tanh\left[x_4(t-\tau)\right] \\
\dot{x}_4 &= -15x_4(t) - 6.75\tanh\left[x_3(t-\tau)\right] + 16.9\tanh\left[x_4(t)\right] - 10.1\tanh\left[x_5(t-\tau)\right] \\
\dot{x}_5 &= -25x_5(t) + 10\tanh\left[x_4(t-\tau)\right] + 30\tanh\left[x_5(t)\right].
\end{aligned}
$$

25.4 Discussion

A set of sufficient conditions for the global asymptotic stability of the equilibrium point in the delayed model have been shown (proofs can be established by using two lemmas, Lyapunov functional method, and combining with the techniques of inequality of DCNNs). The results obtained show some restrictions on synaptic weights and decay parameters for the global stability of the system are independent of any delay. Here, we did not assume any symmetry of the connection matrix $(b_{ij})_{n\times n}$, excitatory and inhibitory connections did not influence the sufficient criteria for global stability. Besides, we considered the output state $[f(x_i(t)]$ only satisfying hypotheses (H1) and (H2) above, not requiring them to be differentiable.

This paper also deals with chaos control by introducing time delay (see Fig. 25.3) in CNNs. Furthermore, different from the work of Yang and Huang [9] where adjustable parameter lies off the main diagonal (self-connection weights) and they have showed chaotic dynamics for particular values of weight parameter, we are able to control that type of dynamical behavior of CNNs in more generalized format. Moreover, the methods of this paper may be applicable in more complicated systems also.

References

1. L. Chen, K. Aihara, Global searching ability of chaotic neural networks. IEEE Trans. Circuits Syst. I **46**(8), 974–993 (1999)
2. L.O. Chua, L. Yang, Cellular neural networks: Theory. IEEE Trans. Circuits Syst. **35**(10), 1257–1272 (1998)
3. Y.S. Huang, C.W. Wu, Stability of cellular neural network with small delays. Discrete Contin. Dyn. Syst. **2005**, 420–426 (2005)
4. A. Kundu, P. Das, A.B. Roy, Complex dynamics of a four neuron network model having a pair of short-cut connections with multiple delays. Nonlinear Dyn. **72**(3), 643–662 (2013)
5. E. Ott, C. Grebogi, J.A. Yorke, Controlling chaos. Phys. Rev. Lett. **64**, 1196–1199 (1990)
6. T. Roska, T. Boros, P. Thiran, L.O. Chua, Detecting simple motion using cellular neural networks, in *Proceedings of the IEEE International Workshop on Cellular Neural Network Application* (1990), pp. 127–138
7. T. Roska, L.O. Chua, Cellular neural networks with non-linear and delay-type template elements and non-uniform grids. Int. J. Circuit Theor. Appl. **20**, 469–481 (1992)
8. P.L. Venetianter, T. Roska, Image compression by cellular neural networks. IEEE Trans. Circuits Syst. I **45**(3), 205–215 (1998)
9. X.S. Yang, Y. Huang, Chaos and two-tori in a new family of 4-CNNs. Int. J. Bifur. Chaos. **17**(3), 953–963 (2007)
10. Q. Zhang, X. Wei, J. Xu, Stability of delayed cellular neural networks. Chaos, Solitons Fractals **31**, 514–520 (2007)

Chapter 26
Scattering of Water Wave by Undulating Porous Bed Topography in an Ice-Covered Ocean

Sandip Paul and Soumen De

Abstract Water-wave scattering by porous bottom undulations of an ocean with an ice cover is investigated by perturbation analysis. The first-order reflection and transmission coefficients were obtained using Green's integral theorem and Fourier transform technique. It is shown that the expressions for the first-order reflection and transmission coefficients are same for both the techniques. The first-order reflection coefficient is computed numerically and it is observed that the porosity of the ocean bottom has an effect on the reflection and transmission coefficients. The problem is also studied when the porosity parameter is a complex number. Numerical results are depicted graphically in a number of figures for different values of parameters.

Keywords Water-wave scattering · Porous bed · Ice cover · Perturbation analysis · Green's integral theorem

26.1 Introduction

In polar region, ocean surface covered by a thin ice plate plays a critical role in study of ice-wave interaction problem. The infinitely large ice plate acts as insulator in transforming heat from the water beneath and the air by solar radiation. The study of water wave travelling beneath the ice sheet is important because it may cause cracks in the ice sheet. Linear water-wave interaction with thin floating ice cover is modelled as a thin elastic sheet has been studied by a number of researchers Chakrabarti [1], Fox and Squire [2] under the assumptions of the linearized theory. The scattering of water waves by the edge of a semi-infinite ice sheet in a finite depth ocean using residue calculus technique has been studied by Linton and Chung [3]. Evans and

S. Paul (✉)
Department of Mathematics, Adamas Institute of Technology,
Kolkata 700126, India
e-mail: sandippaulmac@gmail.com

S. De
Department of Applied Mathematics, University of Calcutta,
Kolkata 700009, India
e-mail: soumenisi@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_26

Davies [4] considered the problem of scattering of obliquely incident waves on a semi-infinite thin elastic plate in water of finite depth using Wiener–Hopf technique. Chung and Fox [5] investigated the problem of interaction between ocean-going waves and a semi-infinite ice sheet, focusing on the calculation of the reflection of incident waves. Water-wave scattering by a strip of an ice cover floating on the surface of deep water is considered by Gayen et al. [6].

The literature concerning a study of ocean wave interaction with an ice cover in the presence of a undulating porous bottom of some special types has taken attention to the researchers (see Martha and Bora [7], Zhu [8], Silva et al. [9]). Earlier on, Evans and Linton [10] considered the problem of scattering of water waves by a varying bottom topography an used mapping method in which the problem was first transferred into a uniform strip resulting in a variable-free surface boundary condition. Mandal and Basu [11] studied that the diffraction of water waves by a small cylindrical elevation of the bottom of a laterally unbounded ocean covered by an ice sheet is investigated by the perturbation analysis. Mase and Takeba [12], Zhu [8] and Silva et al. [9] investigate the wave scattering problem involving porous bed. Martha et al. [13] considered the problem of Oblique water-wave scattering by small undulation on porous sea-bed. They obtain the first-order reflection and transmission coefficients. The problem of oblique wave propagation over a small deformation in a channel flow consisting of two layers was considered by Mahapatra and Bora [14].

In the present paper, we consider the problem of scattering of an incoming wave train by porous bottom undulation of an ocean of finite depth which is covered by a thin sheet of ice instead of having a free surface. The bed is composed of some specific kind of rigid porous material which is characterized by a known porosity parameter η, whose dimension is inverse of length. Porosity parameter considered here is either real or complex. The motion inside the porous bottom has been neglected. In deriving the linearized ice-cover condition, it has been assumed that the waves are long compared to the thickness of the ice sheet (see Gol'dshtein and Marchenko [15], Chakrabarti [1]). However, Mandal and Basu [11] shows that this assumption is not necessary. The incoming wave train is partially reflected and partially transmitted through the ocean. A simplified perturbation technique is employed to reduce the original boundary value problem coupled one up to first order. This problem is solved here by two methods, based on the Green's integral theorem and Fourier transform technique, to obtain the first-order reflection and transmission coefficients in terms of integrals involving the shape function describing the bottom undulations. The first-order coefficients are depicted graphically against the wave number for a sinusoidal-shape function for its physical importance. The effects of flexural rigidity of the ice sheet and porosity on the reflection and transmission coefficients are investigated numerically and corresponding graphs are plotted against the wave number of the incident wave for different values of the porous parameter and for a fixed value of the flexural rigidity.

26.2 Formulation of the Problem

We consider an incompressible, inviscid, homogeneous water of finite depth h and a two-dimensional Cartesian co-ordinate system with y-axis is taken vertically downward and x-axis (i.e. $y = 0$) is the position of the infinite ice sheet. Here, the fluid is assumed to be irrotational. The undulating porous bottom is given by $y = h + \varepsilon c(x)$, where c(x) is a bounded and continuous function describing the shape of the bottom topography such that $c(x)$ vanishes at infinity and ε being a small non-dimensional number denotes the measure of smallness of the bottom elevation. Assuming the time dependence of the dependent variables to be of the form $e^{-i\omega t}$, the velocity potential can be written as $\text{Re}\{\psi(x, y)e^{-i\omega t}\}$, where $\psi(x, y)$ satisfies the Laplace's equation

$$\nabla^2\psi = 0 \text{ in } 0 \le y < h + \varepsilon c(x), -\infty \le x \le \infty. \tag{26.1}$$

The water is assumed to be covered by thin infinitely extent ice sheet of very small thickness d. Here, the ice sheet is considered as a thin elastic plate floating over the fluid extending to infinity. Then the linearized boundary condition at the thin ice sheet [c.f. Mandal and Basu [11]] is

$$K\psi + \left(1 + D\frac{\partial^4}{\partial x^4}\right)\frac{\partial\psi}{\partial y} = 0 \text{ on } y = 0, \tag{26.2}$$

where $D = \frac{Ed^3}{12(1-\nu^2)\rho g}$ is the flexural rigidity of the plate, $K = \frac{\omega^2}{g}$, ω being the angular frequency of the incident wave; E, ν, d, ρ and ρ_e are the Young modulus, Poisson's ratio, the thickness of the ice sheet, the density of the fluid and the density of the ice sheet, respectively; and g being the acceleration due to gravity.

The linearized condition on undulating porous bottom is

$$\frac{\partial\psi}{\partial n} - \eta\psi = 0 \text{ on } y = h + \varepsilon c(x) \tag{26.3}$$

In the boundary condition (26.3), the normal derivative $\frac{\partial}{\partial n}$ has been involved. The porous effect parameter corresponding to the sea-bed under consideration is denoted by η, which is taken to be real.

The above BVP suggests us to assume the progressive wave $\psi(x, y)$ defined in the infinite strip $-\infty < x < \infty$, $0 \le y \le h$; propagating just below the ice sheet is given by

$$\psi_0(x, y) = \frac{1}{K}\left\{K \sinh k_0 y - \left(1 + Dk_0^4\right)k_0 \cosh k_0 y\right\} e^{ik_0 x} \text{ in} -\infty < x < \infty, 0 \le y \le h. \tag{26.4}$$

If $c(x) = 0, -\infty < x < \infty$, i.e. if the bottom be the plane $y = h$ without any undulation, the above boundary value problem suggests the propagation of a plane water wave, whose velocity potential is $\psi_0(x, y)e^{-i\omega t}$, where the wave number k_0 satisfies the dispersion equation

$$\varDelta\,(u) = 0, \tag{26.5}$$

where

$$\varDelta\,(u) \equiv (1 + Du^4)u^2 \sinh uh - u\left\{K + (1 + Du^4)\eta\right\} \cosh uh + \eta K \sinh uh. \tag{26.6}$$

The dispersion equation (26.5) has exactly two real roots $\pm k_0 (k_0 > 0)$, two conjugate complex roots $\pm\mu, \pm\bar{\mu}$ with positive real and imaginary parts and an infinite number of purely imaginary roots $\pm ik_n, n = 1, 2, 3, \ldots$ satisfying the transcendental equation

$$(1 + Dk_n^4)k_n^2 \sin k_n h + k_n\left\{K + (1 + Dk_n^4)\eta\right\} \cos k_n h - \eta K \sin k_n h = 0. \tag{26.7}$$

Although we have considered η as a real quantity, it is possible to find the dispersion relation for complex η too. Suppose $\eta = \eta_1 + i\eta_2$, with $\eta_1, \eta_2 \neq 0$ are real and the dispersion equation has a non-zero real root.

When a train of waves from negative infinity (i.e. $x \longrightarrow -\infty$) propagates below the thin ice sheet with mode k_0 incident upon the undulating porous sea-bed, the wave energy will partially reflected by and partially transmitted over the bottom with mode k_0 and then ψ will satisfy the following infinite requirements:

$$\begin{array}{lll} \psi(x, y) \to \psi_0(x, y) + R\psi_0(-x, y) & \text{as} & x \longrightarrow -\infty \\ \psi(x, y) \to T\psi_0(x, y) & \text{as} & x \longrightarrow \infty \end{array}, \tag{26.8}$$

where R denotes the reflection coefficient corresponding to the reflected wave and T is the transmission coefficient corresponding to the transmitted wave.

For very small undulation, (26.3) can be approximated up to the first order of smallness, which is

$$\psi_y(x, y) - \varepsilon\frac{d}{dx}\{c(x)\psi_x(x, h)\} - \eta\{\psi(x, y) + \varepsilon c(x)\psi_y(x, y)\} + O\left(\varepsilon^2\right) = 0 \text{ on } y = h. \tag{26.9}$$

26.3 The Perturbation Technique

In the absence of bottom undulation, there is no energy reflection and total transmission occurred. As a result, we express $\psi(x, y)$, R and T as

$$\begin{aligned}&\psi(x,y)=\psi_0(x,y)+\varepsilon\psi_1(x,y)+O(\varepsilon^2)\\&R(x,y)=\varepsilon R_1(x,y)+O(\varepsilon^2)\\&T(x,y)=1+\varepsilon T_1(x,y)+O(\varepsilon^2)\end{aligned}\tag{26.10}$$

Using the expressions in Eq. (26.10) in Eqs. (26.1)–(26.3) and Eqs. (26.8) and (26.9), we found that $\psi_1(x, y)$ satisfies the following B.V.P.:

$$\nabla^2\psi_1 = 0 \text{ in } \quad 0 \le y \le h, \quad -\infty < x < \infty, \tag{26.11}$$

$$K\psi_1 + \left(1 + D\frac{\partial^4}{\partial x^4}\right)\psi_{1y} = 0 \text{ on } y = 0, \tag{26.12}$$

$$\psi_{1y} - \eta\psi_1 = V(x) \text{ on } y = h, \tag{26.13}$$

where

$$\begin{aligned}V(x) &\equiv \frac{d}{dx}\{c(x)\psi_{0x}(x,h)\} + \eta c(x)\psi_{0x}(x,h)\\&= -\frac{ik}{K}A(k)\frac{d}{dx}\left\{c(x)\,e^{ikx}\right\} - \frac{\eta k}{K}B(k)\,c(x)\,e^{ikx},\end{aligned}\tag{26.14}$$

and

$$\begin{aligned}A(k) &= \left(1 + Dk^4\right)k\cosh kh - K\sinh kh,\\B(k) &= \left(1 + Dk^4\right)k\sinh kh - K\cosh kh,\end{aligned}\tag{26.15}$$

The infinite conditions are

$$\psi_1(x,y) \sim \begin{cases} R_1\psi_0(-x,y) & \text{as } x \to -\infty,\\ T_1\psi_0(x,y) & \text{as } x \to \infty.\end{cases}\tag{26.16}$$

26.4 Solution by Using Green's Integral Theorem

The boundary value problem given in Eqs. (26.11)–(26.13) can be solved by constructing a Green's function $G(x, y; \alpha, \beta)$ satisfying the following B.V.P.:

$$\nabla^2 G = 0 \text{ in } 0 \le y \le h, \text{except at } (\alpha,\beta), \text{where } 0 < \beta < h \tag{26.17}$$

$$KG + \left(1 + D\frac{\partial^4}{\partial x^4}\right)G_y = 0 \text{ on } y = 0, \tag{26.18}$$

$$G_y - \eta G = 0 \text{ on } y = h, \tag{26.19}$$

$$G \sim \log r \;\; as \;\; r = \left\{(x-\alpha)^2 + (y-\beta)^2\right\}^{\frac{1}{2}} \longrightarrow 0, \tag{26.20}$$

$$G \sim \text{ multiple of } \frac{1}{K}\left\{K \sinh k_0 y - \left(1 + Dk_0^4\right) k_0 \cosh k_0 y\right\} e^{ik_0|x-\alpha|}$$
$$\text{as } |x-\alpha| \to \infty. \tag{26.21}$$

The condition (26.18) describes the fact that G represents an outgoing wave as $|x-\alpha| \to \infty$.

To solve the above boundary value problem, we used the method used by Rhodes-Robinson [16] and we get

$$\begin{aligned} G(x, y; \alpha, \beta) = {} & \frac{2\pi i\{(k_0 \cosh k_0 h - \eta \sinh k_0 h)\cosh k_0(h-y) + \eta \sinh k_0 y\}}{k_0 \cosh k_0 h \Delta'(k_0)} \\ & \times \{K \sinh k_0\beta - (1 + Dk_0^4)k_0 \cosh k_0\beta\}e^{ik_0|x-\alpha|} \\ & + \sum_{n=1}^{\infty} \frac{2\pi\{(k_n \cos k_n h - \eta \sin k_n h)\cos k_n\,(h-y) + \eta \sin k_n y\}}{k_n \cos k_n h \Delta'(k_n)} \\ & \times \{K \sin k_n\beta - (1 + Dk_n^4)k_n \cos k_n\beta\}e^{-k_n|x-\alpha|} \\ & + \frac{2\pi i\{(\lambda \cosh \lambda h - \eta \sinh \lambda h)\cosh \lambda\,(h-y) + \eta \sinh \lambda y\}}{\lambda \cosh \lambda h \Delta'(\lambda)} \\ & \times \left\{K \sinh \lambda\beta - \left(1 + D\lambda^4\right)\lambda \cosh \lambda\beta\right\} e^{i\lambda|x-\alpha|} \\ & + \frac{2\pi i\left\{\left(\bar\lambda \cosh \bar\lambda h - \eta \sinh \bar\lambda h\right)\cosh \bar\lambda\,(h-y) + \eta \sinh \bar\lambda y\right\}}{\bar\lambda \cosh \bar\lambda h \Delta'\left(\bar\lambda\right)} \\ & \times \left\{K \sinh \bar\lambda\beta - \left(1 + D\bar\lambda^4\right)\bar\lambda \cosh \bar\lambda\beta\right\} e^{-i\bar\lambda|x-\alpha|}. \end{aligned} \tag{26.22}$$

For outgoing, nature of $G(x, y; \alpha, \beta)$ given in (26.18) suggests us to consider

$$\begin{aligned} G(x, y; \alpha, \beta) \to {} & \frac{2\pi i\{(k_0 \cosh k_0 h - \eta \sinh k_0 h)\cosh k_0(h-y) + \eta \sinh k_0 y\}}{k_0 \cosh k_0 h \Delta'(k_0)} \\ & \times \{K \sinh k_0\beta - (1 + Dk_0^4)k_0 \cosh k_0\beta\}e^{ik_0|x-\alpha|}. \end{aligned} \tag{26.23}$$

Apply Green's integral theorem on $\psi_1(x, y)$ and $G(x, y; \alpha, \beta)$ in the form

$$\int_C \left(\psi_1 \frac{\partial G}{\partial n} - G\frac{\partial \psi_1}{\partial n}\right) ds = 0, \tag{26.24}$$

where C is the closed contour taken in positive sense bounding common region of the interior of the rectangle formed by the line segments $y = 0, h(-X \le x \le X)$ and $x = \pm X(0 \le y \le h)$ and the exterior of the circle of radius ρ and centre at (α, β) and ultimately making $X \to \infty$ and $\rho \to 0$.

Due to the outgoing nature of $\psi_1(x, y)$ and $G(x, y; \alpha, \beta)$, the only contribution to the integral (26.24) is from the line $y = 0(-X \leq x \leq X)$ and from the circle as $\rho \to 0$, and finally we obtain the solution of the B.V.P. as

$$\psi_1(\alpha, \beta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(x, h; \alpha, \beta) V(x) dx. \tag{26.25}$$

where $V(x)$ is given in (26.14).

To get the first-order transmission and reflection coefficients, we take $\alpha \to \pm\infty$ in the infinite conditions given in (26.16) and (26.23) above and using in (26.25).

Taking $\alpha \to \infty$ in Eqs. (26.16) and (26.23), we have

$$\psi_1(\alpha, \beta) \to T_1 \psi_0(\alpha, \beta) \tag{26.26}$$

$$G(x, h; \alpha, \beta) \to \frac{2\pi i K}{\Delta'(k_0)} e^{-ik_0 x} \psi_0(\alpha, \beta). \tag{26.27}$$

Using (26.27) in (26.25), we obtain T_1 as

$$T_1 = \frac{ik_0\{k_0 A(k_0) - \eta B(k_0)\}}{\Delta'(k_0)} \int_{-\infty}^{\infty} c(x) dx. \tag{26.28}$$

Similarly, taking $\alpha \to -\infty$ in Eqs. (26.16) and (26.23), we have

$$\psi_1(\alpha, \beta) \to R_1 \psi_0(-\alpha, \beta), \tag{26.29}$$

$$G(x, h; \alpha, \beta) \to \frac{2\pi i K}{\Delta'(k_0)} e^{-ik_0 x} \psi_0(-\alpha, \beta). \tag{26.30}$$

Using (26.30) in (26.25), we obtain R_1 as

$$R_1 = -\frac{ik_0\{k_0 A(k_0) + \eta B(k_0)\}}{\Delta'(k_0)} \int_{-\infty}^{\infty} c(x) e^{2ik_0 x} dx, \tag{26.31}$$

where

$$A(k_0) = \left(1 + Dk_0^4\right) k_0 \cosh k_0 h - K \sinh k_0 h,$$

and

$$B(k_0) = \left(1 + Dk_0^4\right) k_0 \sinh k_0 h - K \cosh k_0 h.$$

26.5 Solution by Using Fourier Transform Technique

The boundary value problem described by Eqs. (26.11)–(26.13) can be solved easily by Fourier transform technique to get ψ_1, by defining

$$\Psi_1(\xi, y) = \int_{-\infty}^{\infty} \psi_1(x, y) e^{-i\xi x} dx. \tag{26.32}$$

We note that the wave number k is assumed to have a very small imaginary part so that ψ_1 decreases exponentially as $x \to \infty$.

Then $\Psi_1(\xi, y)$ satisfies the following B.V.P.:

$$\frac{d^2\Psi_1}{dy^2} - \xi^2\Psi_1 = 0 \text{ on } 0 \le y \le h, \tag{26.33}$$

$$K\Psi_1 + \left(1 + D\xi^4\right)\Psi_{1y} = 0 \text{ on } y = 0, \tag{26.34}$$

$$\Psi_{1y} - \eta\Psi_1 = \bar{V}(\xi) \text{ on } y = h, \tag{26.35}$$

where

$$\bar{V}(\xi) = \int_{-\infty}^{\infty} V(x)\, e^{-i\xi x} dx. \tag{26.36}$$

Solving the above B.V.P., we have the following expression for $\Psi_1(\xi, y)$:

$$\Psi_1(\xi, y) = \frac{\bar{V}(\xi)}{\Delta(\xi)} \left\{\left(1 + D\xi^4\right)\xi \cosh \xi y - K \sinh \xi y\right\}. \tag{26.37}$$

It is clear that the dispersion relation $\Delta(\xi)$ is an even function of ξ and so, by taking inverse Fourier transform on $\Psi_1(\xi, y)$ defined by

$$\psi_1(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_1(\xi, y) e^{i\xi x} d\xi, \tag{26.38}$$

we will get the first-order potential function as

$$\psi_1(x, y) = \frac{1}{2\pi} \int_0^{\infty} \frac{\left\{\left(1 + D\xi^4\right)\xi \cosh \xi y - K \sinh \xi y\right\}}{\Delta(\xi)} \left\{\bar{V}(-\xi)e^{-i\xi x} + \bar{V}(\xi)e^{i\xi x}\right\} d\xi. \tag{26.39}$$

The integrand on the right side has poles at the zeros of the dispersion equation given by Eq. (26.6) and the expression for $\psi_1(x, y)$ can be obtained by finding the residue at the poles. We consider a contour in the fluid region just below the poles of the integrand. On rotation of the contour as used in Martha et al. [13], we get the behaviour of $\psi_1(x, y)$ as $x \to \infty$ and we find

$$\psi_1\left(x,y\right) \longrightarrow i\frac{\left(1+Dk_0^4\right)k_0\cosh k_0y - K\sinh k_0y}{\varDelta'\left(k_0\right)}\bar{V}\left(k_0\right)e^{ik_0x}, \tag{26.40}$$

where

$$\bar{V}\left(k_0\right) = \frac{\left\{k_0A\left(k_0\right)-\eta B\left(k_0\right)\right\}}{K}\int_{-\infty}^{\infty}c(x)dx. \tag{26.40}$$

Comparing with the infinite condition given in (26.16), the first-order transmission coefficient is given by

$$T_1 = \frac{ik_0\left\{k_0A\left(k_0\right)-\eta B\left(k_0\right)\right\}}{\varDelta'\left(k_0\right)}\int_{-\infty}^{\infty}c(x)dx. \tag{26.41}$$

Similarly as $x \to -\infty$, we get the first-order reflection coefficient as

$$R_1 = -\frac{ik_0\left\{k_0A\left(k_0\right)+\eta B\left(k_0\right)\right\}}{\varDelta'\left(k_0\right)}\int_{-\infty}^{\infty}c(x)e^{2ik_0x}dx, \tag{26.42}$$

where

$$A\left(k_0\right) = \left(1+Dk_0^4\right)k_0\cosh k_0h - K\sinh k_0h,$$

and

$$B\left(k_0\right) = \left(1+Dk_0^4\right)k_0\sinh k_0h - K\cosh k_0h.$$

26.6 Particular Case

Example 1 For progressive wave of mode k_0, we consider the shape function $c(x)$ as

$$c(x) = \begin{cases} a\sin lx, & \text{for } \dfrac{-n\pi}{l} \le x \le \dfrac{n\pi}{l}; \\ 0, & \text{otherwise}; \end{cases} \tag{26.43}$$

where a and l are the amplitude of the sinusoidal ripple on the bottom surface and the ripple wave number, respectively, n being the number of ripple.

On substitution (26.43) in Eqs. (26.41) and (26.42), the first-order reflection and transmission coefficients are given as follows:

$$R_1 = (-1)^n\frac{2iak_0l\left\{k_0A\left(k_0\right)+\eta B\left(k_0\right)\right\}}{\left(l^2-4k_0^2\right)\varDelta'\left(k_0\right)}\sin\frac{2k_0n\pi}{l}, \tag{26.44}$$

$$T_1 = 0, \tag{26.45}$$

where $\varDelta'\left(k_0\right)$ is the derivative of $\varDelta\left(k_0\right)$ given in (26.6).

Example 2 For Progressive wave of mode k_0, we consider the shape function $c(x)$ in the form of an exponentially decaying bottom as

$$c(x) = ae^{-\mu|x|}, \quad (\mu > 0) \;\; for \quad -\infty < x < \infty \tag{26.46}$$

where a and μ are the amplitude of the ripple on the bottom surface and the ripple wave number, respectively. In this case, the top of the elevation lies at the point $(0, H)$ on either side where it decreases exponentially.

On substitution (26.46) in Eqs. (26.41) and (26.42), the first-order reflection and transmission coefficients are given as follows:

$$R_1 = -\frac{2iak_0\mu\left\{k_0 A\left(k_0\right) + \eta B\left(k_0\right)\right\}}{\left(l^2 + 4k_0^2\right) \Delta'\left(k_0\right)} \tag{26.47}$$

$$T_1 = \frac{2iak_0\left\{k_0 A\left(k_0\right) - \eta B\left(k_0\right)\right\}}{\mu \Delta'\left(k_0\right)} \tag{26.48}$$

Example 3 For progressive wave of mode k_0, we consider the shape function $c(x)$ as

$$c(x) = ae^{-\kappa x^2}, \quad (a, \kappa > 0) \text{ for } \quad -\infty < x < \infty \tag{26.49}$$

In this case, the bottom undulations are in the form of elevation with Gaussian profile, the maximum elevation occurring at $(0, H)$.

On substitution (26.49) in Eqs. (26.41) and (26.42), the first-order reflection and transmission coefficients are given as follows:

$$R_1 = -\frac{iak_0\left\{k_0 A\left(k_0\right) + \eta B\left(k_0\right)\right\}}{\Delta'\left(k_0\right)} \left(\frac{\pi}{\kappa}\right)^{1/2} e^{-k_0^2/\kappa} \tag{26.50}$$

$$T_1 = \frac{iak_0\left\{k_0 A\left(k_0\right) - \eta B\left(k_0\right)\right\}}{\Delta'\left(k_0\right)} \left(\frac{\pi}{\kappa}\right)^{1/2} \tag{26.51}$$

26.7 Numerical Results

In the previous section, we have considered three special forms of the bottom undulation. Since a sinusoidal ripple has a significant importance in ocean research, we studied a sinusoidal patch in the ocean bed for graphical representation of the first-order reflection coefficient $|R_1|$ as a function of the wave number Kh.

In Fig. 26.1, $|R_1|$ depicted against the wave number Kh for $D/h^4 = 0.3, a/h = 0.1, lh = 1, n = 2$ and $\eta h = 0.0, 0.08, 0.15$. It is clear that $|R_1|$ is an oscillating function of Kh and the peak value increases with the porous effect parameter. For a particular porous effect parameter ηh, if the bed wave number is twice the wave number along the x-axis ($l = 2k_0$), the theory predicts a resonant interaction between

Fig. 26.1 $|R_1|$ as a Function of Kh for Different ηh

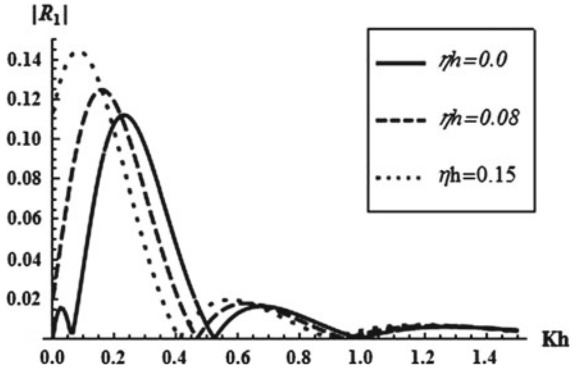
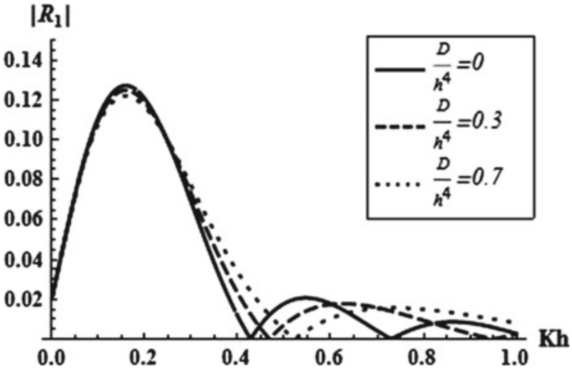


Fig. 26.2 $|R_1|$ as a Function of Kh for Different D/h^4

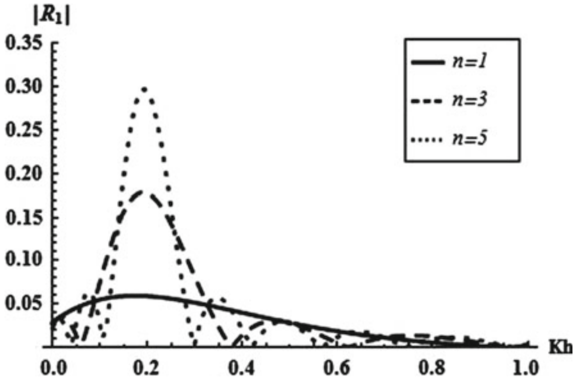


the bed and free surface validating the theoretical result obtained by Mandal and Basu [11].

In Fig. 26.2, the graph shows the effect of the flexural rigidity of the ice sheet on propagation of wave for a specific type of porous bottom with $\eta h = 0.08$; $a/h = 0.1$; $lh = 1, n = 2$ for four different values of $D/h^4 (= 0, 0.3, 0.7)$ against Kh. When $D/h^4 = 0$, i.e. when the upper surface of the ocean is free, the first-order reflection coefficient $|R_1|$ is more oscillatory than in the presence of ice sheet, i.e. when $D/h^4 \neq 0$ and the peak value decreases with increasing D/h^4. One more thing can be observed from this figure that, as the value of D/h^4 increases, the number of poles in $|R_1|$ increases.

Figure 26.3 depicted here shows the first-order reflection coefficient $|R_1|$ against Kh for $\eta h = 0.05$, $D/h^4 = 0.3$, $a/h = 0.1$, $lh = 1$ and $n = 1, 3, 5$. It is clearly observed that the oscillating nature of $|R_1|$ increases with n. The peak value increases as the number of patches at the bottom increases showing the physical nature of the reflected wave obtained in (26.44) and the peak value become unbounded for indefinite value of n.

Fig. 26.3 $|R_1|$ as a Function of Kh for Different n



26.8 Conclusion

Scattering of surface waves by porous bottom undulation in a ocean of finite depth is investigated. Using a simplified perturbation analysis, the problem is reduced up to first order. The boundary value problem is solved by two methods: Green's integral theorem and Fourier transform technique. First-order reflection and transmission coefficients are obtained in terms of computable integrals and found that for sinusoidal bottom, the first-order transmission coefficient is identically zero. The first-order reflection coefficient presented graphically in a number of figures and it is observed that the reflection coefficient increases with increasing porous effect for an ice sheet of specific rigidity and the same type of result obtained for different types of ice cover with various rigidities where the bottom having a fixed porosity. Also for the sinusoidal bottom, the wave transmission and reflection increases as the ripple number increases. So far we have considered the real porosity coefficient only, but in case of complex porosity coefficient $\eta = \eta_1 + i\eta_2$ with $\eta_1, \eta_2 \neq 0$, it is found that the dispersion equation has only one real root as zero, and hence there is no progressive wave.

References

1. A. Chakrabarti, On the solution of the problem of scattering of surface-water waves by the edge of an ice cover. Proc. R. Soc. Lond. A **456**, 1087–1099 (2000)
2. C. Fox, V.A. Squire, On the oblique reflection and transmission of ocean waves at shore fast sea ice. Phil. Trans. R. Soc. Lond. A **347**, 185–218 (1994)
3. C.M. Linton, H. Chung, Reflection and transmission at the ocean/sea-ice boundary. Wave Motion **38**, 43–52 (2003)
4. D.V. Evans, T.V. Davies, Wave-ice interaction. Report no. 1313, Davidson Laboratory, Stevents Institutes of Technology, New Jersey, USA (1968)
5. H. Chung, C. Fox, Calculation of wave-ice interaction using the Wiener-Hopf technique. N. Z. J. Math. **31**, 1–18 (2002)

6. R. Gayen, B.N. Mandal, A. Chakrabarti, Water-wave scattering by an ice-strip. J. Eng. Math. **53**, 21–37 (2005)
7. S.C. Martha, S.N. Bora, Reflection and transmission coefficients for water wave scattering by a sea-bed with small undulation. ZAMM, Zeitschrift fiir Angewandte Mathematik und Mechanik **87**, 314–321 (2007)
8. S. Zhu, Water waves within a porous medium on an undulating bed. Coast. Eng. **42**, 87–101 (2001)
9. R. Silva, P. Salles, A. Palacio, Linear wave propagating over a rapidly varying finite porous bed. Coast. Eng. **44**, 239–260 (2002)
10. D.V. Evans, C.M. Linton, On step approximations for water-wave problems. J. Fluid Mech. **278**, 229–249 (1994)
11. B.N. Mandal, U. Basu, Wave diffraction by a small elevation of the bottom of an ocean with an ice-cover. Arch. Appl. Mech. **73**, 812–822 (2004)
12. H. Mase, K. Takeba, Bragg scattering of waves over porous rippled bed, in *Proceedings of the 24th ICCE'94* (1994), pp. 635–649
13. S.C. Martha, S.N. Bora, A. Chakrabarti, Oblique water-wave scattering by small undulation on a porous sea-bed. Appl. Ocean Res. **29**, 86–90 (2007)
14. S. Mohapatra, S.N. Bora, Oblique water wave scattering by bottom undulation in a two-layer fluid flowing through a channel. J. Mar. Sci. Appl. **11**, 276–285 (2012)
15. R.V. Gol'dshtein, A.V. Marchenko, The diffraction of plane gravitational waves by the edge of an ice cover. PMM USSR **53**, 731–736 (1989)
16. P.F. Rhodes-Robinson, Fundamental singularities in the theory of water waves with surface tension. Bull. Aust. Math. Soc. **2**, 317–333 (1970)

Chapter 27
Numerical Simulations of Natural Convection and Entropy Generation in a Square Cavity with an Adiabatic Body

Swapan K. Pandit

Abstract The present work deals with the numerical simulations of natural convection and entropy generation in a two-sided differentially heated square cavity with an adiabatic square body located at the centre of the cavity. We have applied our recently proposed higher order compact scheme [1] based on nine-point compact stencil to spatial differencing of the streamfunction-vorticity formulation of the two-dimensional incompressible viscous flows governed by Navier–Stokes equations including the energy transport equations. In addition, local entropy generation distributions are computed using the steady-state values of velocity and temperature. The present results are compared with numerical results available in the literature and excellent match is observed in all the cases.

Keywords Natural convection · Entropy generation · Square cavity · Adiabatic body

27.1 Introduction

Over the past few decades, there have been considerable interests to study the heat transfer in a thermally driven square cavity due to large number of technical applications such as packed sphere beds, chemical catalytic reactors, grain storage, geothermal reservoirs, solidification of casting, crude oil production, etc. The review of the natural convection reveals that the available studies could be classified into two categories. They are natural convection in square cavities with and without the presence of a block (block may be sink, source, or adiabatic) [2, 3]. Most of the previous studies on natural convection considered the classic case of thermal convection between the

S.K. Pandit (✉)
Integrated Science Education and Research Centre (ISERC), Visva-Bharati,
Santiniketan 731235, West Bengal, India
e-mail: swapankumar.pandit@visva-bharati.ac.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_27

bottom hot and top cold walls, without any obstacles between them [4–6]. This study was cited by several authors later as a benchmark exercise for validation purposes.

It should be noted that many numerical methods, including finite difference, finite element, finite volume, and lattice Boltzman methods, have been used to investigate the steady natural convection in a square cavity. Most of these numerical schemes are either first-order or second-order accurate in space, particularly the central difference ones have been used in a large number because of their straight forwardness in application. Also, whenever there have been attempts to solve for the flows using higher order [7], they are confined invariably to uniform space grids.

The aim of the present work is to study the hydrodynamic, thermal, and entropy generation characteristics of a differentially heated square cavity in the presence of single block using our recently proposed higher order compact scheme.

27.2 Problem

The schematic of a two-dimensional rectangular enclosure with an adiabatic square body and the coordinate system is shown in Fig. 27.1a. The system consists of a square enclosure with sides of length L, within which another square solid body with sides of length W is centered. The bottom wall is kept at a constant high temperature of T_h, whereas the top wall at a constant low temperature of T_c. The left and right side walls along the horizontal direction are adiabatic. A multi-domain methodology (see Fig. 27.1b) is used to consider the square body at the center of the computational domain.

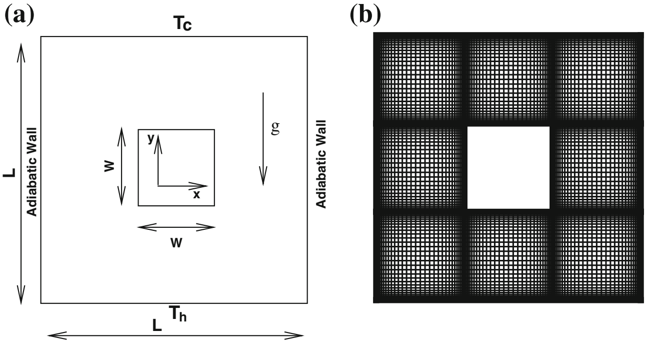


Fig. 27.1 **a** Geometry and **b** Mesh structure of the problem

27.3 Governing Equations

The governing equations for this problem are the incompressible N–S equations for the unsteady two-dimensional flows in terms of variable vorticity (ζ) and stream-function (ψ), which are supplemented by the energy equation for the nondimensional temperature T:

$$-\frac{\partial^2\psi}{\partial x^2}-\frac{\partial^2\psi}{\partial y^2}=\zeta, \tag{27.1}$$

$$\frac{1}{Pr}\frac{\partial\zeta}{\partial t}-\frac{\partial^2\zeta}{\partial x^2}-\frac{\partial^2\zeta}{\partial y^2}+\frac{u}{Pr}\frac{\partial\zeta}{\partial x}+\frac{v}{Pr}\frac{\partial\zeta}{\partial y}+\frac{1}{Da}\zeta=Ra\frac{\partial T}{\partial x}. \tag{27.2}$$

$$\frac{\partial T}{\partial t}-\frac{\partial^2 T}{\partial x^2}-\frac{\partial^2 T}{\partial y^2}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}=0. \tag{27.3}$$

If the reference temperature T_0 is taken as being equal to T_c, the dimensionless boundary conditions for the present problem are specified as follows:

$$u=v=\psi=0,\ \frac{\partial T}{\partial y}=0 \text{ at left wall,}$$
$$u=v=\psi=0,\ \frac{\partial T}{\partial y}=0 \text{ at right wall}$$

and

$$u=v=\psi=0,\ T=1 \text{ at bottom wall and}$$
$$u=v=\psi=0,\ T=0 \text{ at top wall.}$$

The different entropy parameters, i.e., Local entropy generation due to heat transfer (S_T) and Local entropy generation due to fluid friction (S_F), can be written in nondimensional form based on the local thermodynamic equilibrium of linear transport theory [8] as follows:

$$S_T=\left(\frac{\partial T}{\partial x}\right)^2+\left(\frac{\partial T}{\partial y}\right)^2 \tag{27.4}$$

$$S_F=\gamma\left[2\left(\frac{\partial u}{\partial x}\right)^2+2\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)^2\right] \tag{27.5}$$

where the parameter γ in Eq. (27.5) is called irreversibility distribution ratio and is defined as

$$\gamma=\left[\frac{\alpha_m}{L\Delta T}\right]^2\left(\frac{\mu T_0}{k}\right). \tag{27.6}$$

27.4 Results and Discussions

To access the accuracy of the present numerical approach, we have studied the benchmark problem for the differentially heated square cavity with the hot left wall and cold right wall in the presence of adiabatic top and bottom walls, similar to the case reported by Ilis et al. [9]. We have computed the maximum value of local entropy generation due to heat transfer (l.h.t.max) and the maximum value of local entropy generation due to fluid friction (l.f.f.max). The results in Fig. 27.2 in terms of entropy generation due to heat transfer and fluid friction are in excellent agreement with the work [9].

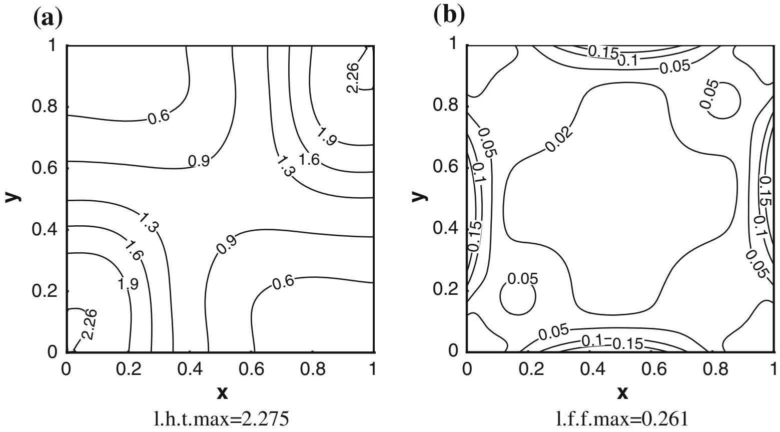


Fig. 27.2 Local entropy generation due to **a** heat transfer (S_T) and **b** fluid friction (S_F) for a cavity with a hot left wall and cold right wall with adiabatic *top* and *bottom* walls at $Ra = 10^3$ for $Pr = 0.7$ (benchmark problem)

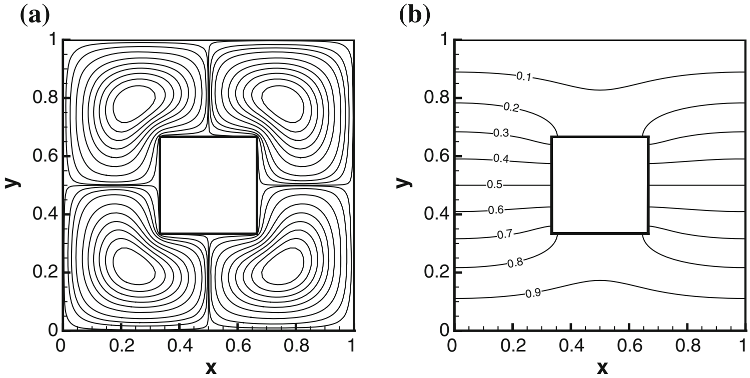


Fig. 27.3 **a** Streamlines and **b** isotherms at the steady state for $Ra = 10^3$

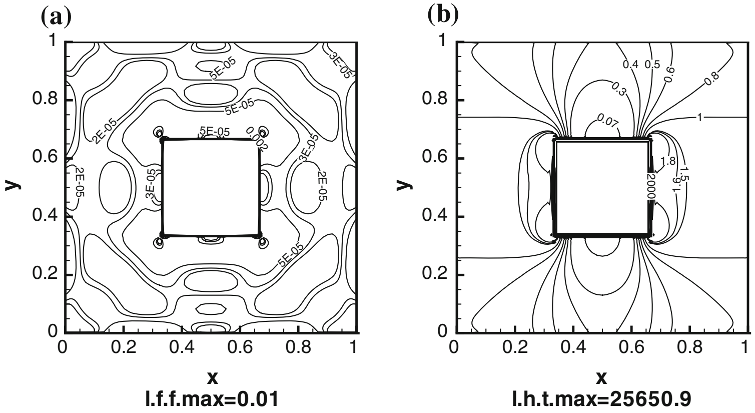


Fig. 27.4 Local entropy generation due to **a** fluid friction and **b** heat transfer in a cavity with an adiabatic body at center for $Ra = 10^3$

Figure 27.3 shows the streamlines and isotherms corresponding to the steady-state solution. The streamlines form four symmetric vortices circulating in the clockwise and counter-clockwise directions. The direction of rotation of these cells is uniquely determined by the thermal boundary condition. The values of isotherms at the lower part ($0.5 \leq y \leq 0$) are in the range of 0.5–1, and those at the upper part ($0 \leq y \leq 0.5$) are in the range of 0–0.5. The thermal gradient in the upper half of the enclosure is symmetric to that in the lower half and there is also a leftright symmetry about the vertical centerline. In Fig. 27.4, we have shown the computed values of entropy generation due to fluid friction and heat transfer. It is seen that the maximum value of local entropy generation due to heat transfer (l.h.t.max) is much more than l.f.f.max. These maximum values occur at the corner and near the adiabatic body.

References

1. S.K. Pandit, J.C. Kalita, D.C. Dalal, A transient higher order compact scheme for incompressible viscous flows on geometries beyond rectangular. J. Comput. Phys. **225**, 1100–1124 (2007)
2. P.S. Mahapatra, S. De, K. Ghosh, N.K. Manna, A. Mukhopadhyay, Heat transfer enhancement and entropy generation in a square enclosure in the presence of adiabatic and isothermal blocks. Numer. Heat Trans. Part A **64**, 577–596 (2013)
3. M.Y. Ha, I.K. Kim, H.S. Yoon, S. Lee, Unsteady fluid flow and temperature fields in a horizontal enclosure with an adiabatic body. Phys. Fluids **14**(9), 3189–3202 (2002)
4. G. De Vahl Davis, Natural convection in a square cavity: a benchmark numerical solution. Int. J. Numer. Methods Fluids **3**, 249–264 (1983)
5. G. De Vahl Davis, I.P. Jones, Natural convection in a square cavity: a comparison exercise. Int. J. Numer. Methods Fluids **3**(1), 227–248 (1983)
6. J.C. Kalita, D.C. Dalal, A.K. Dass, Fully compact higher-order computation of steady-state natural convection in a square cavity. Phys. Rev. E **64**(6), 066703–13 (2001)

7. Z. Tian, Y. Ge, A fourth-order compact fnite difference scheme for the steady stream function vorticity formulation of the NavierStokes/Boussinesq equations. Int. J. Numer. Meth. Fluids **41**, 495–518 (2003)
8. R. Anandalakshmi, T. Basak, Numerical simulations for the analysis of entropy generation during natural convection in porous rhombic enclosures. Numer. Heat Trans. A **63**, 257–284 (2013)
9. G.G. Ilis, M. Mobedi, B. Sunden, Effect of aspect ratio on entropy generation in a rectangular cavity with differentially heated vertical walls. Int. Commun. Heat Mass Trans. **35**, 696–703 (2008)

Chapter 28
On an Interface Elliptic Crack

Tushar Kanti Saha and Arabinda Roy

Abstract The three-dimensional problem of an elliptic crack located at the interface between two bonded dissimilar elastic half-spaces and crack faces subjected to normal pressure equal in magnitude and opposite in direction is considered here. Considering a Cartesian coordinate system with the xOy-plane coinciding with the crack plane and origin O coinciding with the crack centre, the mixed boundary conditions on the $z = 0$ plane give rise to three pairs of dual integral equations. This typical mixed boundary value problem is solved here analytically for the first time for normal pressure prescribed on the crack faces. With uniform normal pressure, the three pairs of dual integral equations are reduced to two sets of dual integral equations, which further reduce to a Cauchy singular integral equation that is solved using Plemelj formula. The present work opens up the possibility of further research work in the field of interface elliptic crack located at the interface of bonded elastic or piezoelectric solids.

Keywords Interface elliptic crack · Analytical solution · Dual integral equation · Cauchy singular integral equation · Plemelj formula

28.1 Introduction

The problem of two semi-infinite dissimilar elastic bodies joined along the interface plane with a crack embedded in the interface is of immense practical importance. The problem represents the idealization of two dissimilar elastic solids welded together

T.K. Saha (✉)
Department of Mathematics, Surendranath College,
24/2 M.G. Road,
Kolkata 700009, India
e-mail: tushar0303@yahoo.com

A. Roy
Department of Applied Mathematics, University of Calcutta,
92 A.P.C. Road,
Kolkata 700009, India
e-mail: roy_arabinda1@yahoo.co.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_28

with cracks developed along the weld plane owing to faulty joining techniques. The motivation of such study lies in its application to the fracture of layered composites, as such materials are being used in wide range of engineering field in recent years.

The study of 2D interface crack problems was initiated by Williams [1]. Thereafter, Mossakovsky and Rykba [2] initiated the study of 3D interface crack problems. Several studies have been carried out since then in the field of interface crack problems both in 2D and in 3D (see [3–18]). In the field of 3D interface crack problems, most of the works carried out so far are confined to the case of a penny-shaped crack (see e.g. [7–12]). Other than the work of Shifrin et al. [18], which gives an analytical–numerical solution to the problem, no other work is found in the literature regarding interface elliptic crack. In the present study, we have attempted to give an exact analytical solution to the problem considered.

The problem considered here is that of determining the displacement field in the vicinity of an elliptic crack situated at the interface of two half-spaces of different elastic materials bonded together along their common plane boundary. The deformation in the two half-spaces is a result of the application of a symmetrically distributed pressure to the faces of the crack. The approach of Roy and Saha [20] has been adopted to reduce the mixed boundary value problem to three pairs of dual integral equations. On suitable transformations from Cartesian to polar coordinates, these three pairs of dual integral equations reduce to two pairs for the special case of the crack faces subjected to uniform normal pressure. These two pairs of dual integral equations are similar to those obtained earlier by Lowengrub and Sneddon [12]. Hence, following the same method, the analytical solution for the displacement field is obtained by reducing the pair of integral equations to a Cauchy singular integral equation and using Plemelj formula (see [19] for solution of such singular integral equations).

28.2 Formulation of the Problem

Let the elliptic crack occupies the region (see Fig. 28.1)

$$S : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1;\ z = 0\ (a \text{ is the semi-major axis and } b \text{ is the semi-minor axis of the ellipse}) \tag{28.1}$$

The equation of elastic equilibrium is as follows:

$$(\lambda_0 + \mu)\,\nabla\nabla \cdot \mathbf{u} + \mu\nabla^2\mathbf{u} = \mathbf{0} \tag{28.2}$$

where λ_0, μ are the elastic constants and $\mathbf{u}$ is the displacement vector. Solution of this equation can, in general, be expressed in terms of three harmonic potentials and are unique and complete. However, the choices of the harmonic functions are not unique and are often dictated by the nature of the problem. Consequently, a number

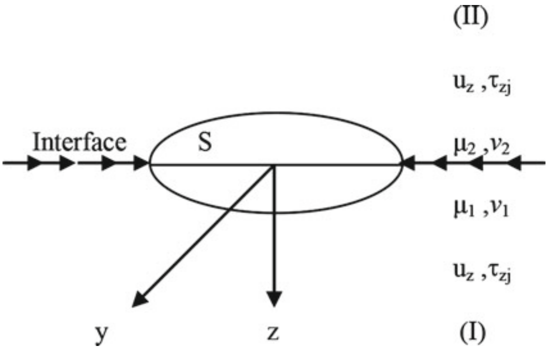


Fig. 28.1 Geometry of the problem

of representations of **u** are possible. Here, we take the displacement field in terms of three scalar potentials ϕ, ψ, χ [21, 22] as

$$\mathbf{u} = \nabla\phi - z\nabla\psi + (3 - 4\nu)\mathbf{k}\psi + \nabla \times (\mathbf{k}\chi) \tag{28.3}$$

where ν is the Poisson's ratio, and ϕ, ψ and χ satisfy the three-dimensional Laplace's equation

$$\nabla^2(\phi, \psi, \chi) = 0 \tag{28.4}$$

Let $z > 0$ be occupied by the medium (to be denoted by (I)) with elastic constants λ_{01}, μ_1, ν_1 and the displacement field in this medium be denoted by $\mathbf{u}_1 = (u_1, v_1, w_1)$. Similarly, let $z < 0$ be occupied by the medium (II) with elastic constants λ_{02}, μ_2, ν_2 and the corresponding displacement field be denoted by $\mathbf{u}_2 = (u_2, v_2, w_2)$.

Similarly, let the scalar potentials in the two media be denoted by (ϕ_1, ψ_1, χ_1) and (ϕ_2, ψ_2, χ_2), respectively.

The solutions of the Laplace's equation (28.4) in the two media are as follows:

$$\left(\phi_j, \psi_j, \chi_j\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(P_j, Q_j, R_j\right)\exp\left[i(\xi x + \eta y) - \lambda z\right] d\xi d\eta \tag{28.5}$$

where $j = 1$ for medium I and 2 for medium II and P_j, Q_j and R_j are the functions of ξ, η and $\lambda = \sqrt{\xi^2 + \eta^2}$.

We assume that the lower and upper faces of the crack are subjected to a prescribed pressure τ^0_{zz}. The conditions satisfied inside the crack area are

$$\left.\begin{array}{ll}\text{(i) } \tau^{(1)}_{zz}(x, y, 0+) + \tau^0_{zz} = 0, & \text{(ii) } \tau^{(2)}_{zz}(x, y, 0-) + \tau^0_{zz} = 0 \\ \text{(iii) } \tau^{(1)}_{zx} = \tau^{(2)}_{zx} = 0, & \text{(iv) } \tau^{(1)}_{zy} = \tau^{(2)}_{zy} = 0\end{array}\right\} \forall (x, y) \in S \tag{28.6a}$$

The conditions satisfied on the crack plane ($z = 0$) outside the crack area are

$$\left.\begin{array}{lll} \text{(i) } \tau_{zx}^{(1)} - \tau_{zx}^{(2)} = 0, & \text{(ii) } \tau_{zy}^{(1)} - \tau_{zy}^{(2)} = 0, & \text{(iii) } \tau_{zz}^{(1)} - \tau_{zz}^{(2)} = 0 \\ \text{(iv) } u_x^{(1)} - u_x^{(2)} = 0, & \text{(v) } u_y^{(1)} - u_y^{(2)} = 0 & \text{(vi) } u_z^{(1)} - u_z^{(2)} = 0 \end{array}\right\} \forall (x, y) \notin S \tag{28.6b}$$

In addition to these conditions, the stress and displacement should vanish at infinity. Satisfying the boundary conditions, we get

$$\begin{aligned} 2\lambda\mu_1 P_1 &= \mu_1(3 - 4\nu_1)Q_1 + \mu_2 Q_2 \\ 2\lambda\mu_2 P_2 &= -\mu_1 Q_1 - \mu_2(3 - 4\nu_2)Q_2 \\ R_2(\xi, \eta) &= -\frac{\mu_1}{\mu_2} R_1(\xi, \eta) \end{aligned} \tag{28.7}$$

Using the integral representations (28.5) in (28.3), and utilizing the boundary conditions (28.6a, 28.6b) together with (28.7), we arrive at three pairs of dual integral equations. We recast these integral equations to a desired format by the following substitutions step by step:

$$(1)\ \kappa_1 = 3 - 4\nu_1,\ \kappa_2 = 3 - 4\nu_2,\ \Gamma = \tfrac{\mu_2}{\mu_1} \tag{28.8}$$

$$\left.\begin{aligned} (2)\ A_1(\xi, \eta) &= -\lambda P_1(\xi, \eta) \\ B_1(\xi, \eta) - A_1(\xi, \eta) &= \kappa_1 Q_1(\xi, \eta) \\ \kappa_1^{-1} C_1(\xi, \eta) &= -\lambda R_1(\xi, \eta) \end{aligned}\right\} \tag{28.9}$$

$$\left.\begin{aligned} (3)\ W(\xi, \eta) &= \tfrac{1}{2}(\kappa_1\kappa_2 - 1)A_1 + \left\{\kappa_1\Gamma + \tfrac{1}{2}(\kappa_1\kappa_2 + 1)\right\} B_1 \\ U(\xi, \eta) &= \left\{\kappa_1\Gamma + \tfrac{1}{2}(\kappa_1\kappa_2 + 1)\right\} A_1 + \tfrac{1}{2}(\kappa_1\kappa_2 - 1) B_1 \\ V(\xi, \eta) &= (1 + \Gamma)\, C_1(\xi, \eta) \end{aligned}\right\} \tag{28.10}$$

$$\left.\begin{aligned} \alpha &= (\kappa_1 - 1)\Gamma - (\kappa_2 - 1) \\ \beta &= (\kappa_1 + 1)\Gamma - (\kappa_2 + 1) \\ \gamma &= \tfrac{(\kappa_2 + \Gamma)(1 + \kappa_1\Gamma)}{1 + \Gamma} \\ t(x, y) &= (\kappa_2 + \Gamma)(1 + \kappa_1\Gamma)\kappa_1\mu_1^{-1}\tau_{zz}^0(x, y) \end{aligned}\right\} \tag{28.11}$$

$$\left.\begin{array}{ll} \text{so that, } & \alpha U(\xi, \eta) + \beta W(\xi, \eta) = (\kappa_2 + \Gamma)(1 + \kappa_1\Gamma)[(\kappa_1 - 1)A_1 + (\kappa_1 + 1)B_1] \\ \text{and, } & \beta U(\xi, \eta) + \alpha W(\xi, \eta) = (\kappa_2 + \Gamma)(1 + \kappa_1\Gamma)[(\kappa_1 + 1)A_1 + (\kappa_1 - 1)B_1] \end{array}\right\} \tag{28.12}$$

Finally, we have the following three pairs of dual integral equations:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda^{-1}[i\xi U(\xi, \eta) + i\eta V(\xi, \eta)]\exp[i(\xi x + \eta y)]\, d\xi d\eta = 0,\ \ \forall (x, y) \notin S \tag{28.13a}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[i\xi\left\{\beta U\left(\xi,\eta\right)+\alpha W\left(\xi,\eta\right)\right\}+i\eta\gamma V\left(\xi,\eta\right)\right]\exp\left[i\left(\xi x+\eta y\right)\right]d\xi d\eta=0,$$
$$\forall(x,y)\in S \qquad (28.13b)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\lambda^{-1}\left[i\eta U\left(\xi,\eta\right)-i\xi V\left(\xi,\eta\right)\right]\exp\left[i\left(\xi x+\eta y\right)\right]d\xi d\eta=0,\ \ \forall(x,y)\notin S \qquad (28.14a)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[i\eta\left\{\beta U\left(\xi,\eta\right)+\alpha W\left(\xi,\eta\right)\right\}-i\xi\gamma V\left(\xi,\eta\right)\right]\exp\left[i\left(\xi x+\eta y\right)\right]d\xi d\eta=0,$$
$$\forall(x,y)\in S \qquad (28.14b)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W\left(\xi,\eta\right)\exp\left[i\left(\xi x+\eta y\right)\right]d\xi d\eta=0,\ \ \forall(x,y)\notin S \qquad (28.15a)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\lambda\left[\alpha U\left(\xi,\eta\right)+\beta W\left(\xi,\eta\right)\right]\exp\left[i\left(\xi x+\eta y\right)\right]d\xi d\eta=t\left(x,y\right),\ \ \forall(x,y)\in S \qquad (28.15b)$$

Transforming to cylindrical polar coordinate system through the following transformations

$$(\xi a,\eta b)=\left(k\cos\chi,k\sin\chi\right),\ k\in(0,\infty),\ \chi\in[0,2\pi];$$

and

$$(x,y)=\left(ar\cos\theta,br\sin\theta\right),\ r\in[0,1],\ \theta\in[0,2\pi];$$

$$k=\left(a^2\xi^2+b^2\eta^2\right)^{1/2},\ \chi=\tan^{-1}\left(\frac{\eta b}{\xi a}\right),\ r=\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)^{1/2},$$

$$\lambda^2=\left(\xi^2+\eta^2\right)=\frac{k^2}{b^2}\left(1-k_0^2\cos^2\xi\right),\ k_0^2=\left(1-\frac{b^2}{a^2}\right),$$

$$\text{or, }\lambda=\frac{k}{b}\Delta\left(k_0\right),\ \text{ where }\Delta\left(k_0\right)=\left(1-k_0^2\cos^2\chi\right)^{1/2};$$

and using the following result of Bessel functions $J_n(\cdot)$

$$\exp\left(\pm iz\cos\theta\right)=\sum_{n=0}^{\infty}\varepsilon_n\left(\pm i\right)^n J_n\left(z\right)\cos n\theta\ \text{ where, }\varepsilon_n=\begin{cases}1,\ n=0\\ 2,\ n<0\end{cases}$$

we have, for the special case of crack faces subjected to uniform pressure, the following reduced pairs of integral equations:

$$\int_0^\infty kW_0^C(k)\,J_0(kr)\,dk = 0, \quad (r > 1) \tag{28.16a}$$

$$\int_0^\infty k^2\left(\alpha U_0^C(k) + \beta W_0^C(k)\right) J_0(kr)\,dk = \frac{\pi ab^2 t_0}{I_{0,0}^C}, \quad (0 \le r \le 1) \tag{28.16b}$$

and,

$$\int_0^\infty kU_0^C(k)\,J_1(kr)\,dk = 0, \quad (r > 1) \tag{28.17a}$$

$$\int_0^\infty k^2\left(\beta U_0^C(k) + \alpha W_0^C(k)\right) J_1(kr)\,dk = 0, \quad (0 \le r \le 1) \tag{28.17b}$$

where U_0^c and W_0^c are the first cosine terms in the Fourier expansion of $U(\xi,\eta)$ and $W(\xi,\eta)$, respectively; $I_{0,0}^C = \frac{1}{2}\int_0^{2\pi} \Delta(k_0)\,d\chi$; and t_0 is the equivalent form of $t(x,y)$ for uniform pressure τ_0.

28.3 Analytical Solution of the Pairs of Integral Equations

The pairs of integral equations (28.16a, 28.16b) and (28.17a, 28.17b) are similar in form to those of a penny-shaped crack at the interface of two dissimilar semi-infinite elastic bodies (see Eqs. (2.19)–(2.22) of [12]). Hence, we adopt the technique of Lowengrub and Sneddon [12] for the solution of the above pairs of integral equations. We assume the solutions of $U_0^C(k)$ and $W_0^C(k)$ as

$$\left\{\begin{array}{rl} & kU_0^C(k) = \displaystyle\int_0^1 \psi_0^C(\rho)\cos(k\rho)\,d\rho = \sqrt{\frac{\pi}{2}}\int_0^1 \psi_0^C(\rho)(k\rho)^{1/2} J_{-1/2}(k\rho)\,d\rho \\ \text{and,} & kW_0^C(k) = \displaystyle\int_0^1 \phi_0^C(\rho)\sin(k\rho)\,d\rho = \sqrt{\frac{\pi}{2}}\int_0^1 \phi_0^C(\rho)(k\rho)^{1/2} J_{1/2}(k\rho)\,d\rho \end{array}\right. \tag{28.18}$$

where the functions are to be such that $\phi_0^C(-\rho) = -\phi_0^C(\rho)$ and $\psi_0^C(-\rho) = \psi_0^C(\rho)$ if

$$\int_{-1}^1 \psi_0^C(\rho)\,d\rho = 0.$$

Now, putting $\lambda_0^C(x) = \psi_0^C(x) + i\phi_0^C(x)$ and writing $\frac{2}{\pi}C + i\frac{ab^2 t_0}{E(k_0)}x = Q(x)$, the pairs (28.16a, 28.16b) and (28.17a, 28.17b) reduce to the following singular integral equation:

$$\beta\lambda_0^C(x) + \frac{1}{i\pi}\int_1^{-1}\frac{\alpha\lambda_0^C(\rho)}{\rho - x}d\rho = Q(x) \quad (-1 \leq x \leq 1) \tag{28.19}$$

$$Q(-x) = \overline{Q(x)} \tag{28.20}$$

The Cauchy singular integral equation (28.19) can be solved using Plemelj formula (see Muskhelishvili [19]). In the present case, we obtain the solution as follows:

$$\lambda_0^C(x) = \frac{ab^2\beta t_0}{\left(\beta^2 - \alpha^2\right)E(k_0)}Z(x)(-\omega + ix) \tag{28.21}$$

where the function $Z(t)$ is determined by the formula

$$Z(t) = \sqrt{\beta^2 - \alpha^2}\,[R(t) + iF(t)] = (\alpha + \beta)X^+(t) = (\beta - \alpha)X^-(t) \tag{28.22}$$

$$R(t) = \cos\left(\omega \ln\frac{1-t}{1+t}\right), \quad F(t) = \sin\left(\omega\ln\frac{1-t}{1+t}\right) \tag{28.23}$$

$$X(z) = (z-1)^{i\omega}(z+1)^{-i\omega} \tag{28.24}$$

$$\omega = \frac{1}{2\pi}\ln\frac{\beta+\alpha}{\beta-\alpha} = \frac{1}{2\pi}\ln\frac{\mu_2\kappa_1 + \mu_1}{\mu_1\kappa_2 + \mu_2} \tag{28.25}$$

Once λ_0^c is obtained, ψ_0^c and ϕ_0^c are obtained as real part and imaginary part of λ_0^c, respectively. Substituting these in (28.18), we get $U_0^c(k)$ and $W_0^c(k)$ which give back $U(\xi, \eta)$ and $W(\xi, \eta)$. By back substitution in equations (28.10), (28.9), (28.7) and (28.5), one can then get back the potentials ϕ_j and ψ_j step by step.

28.4 Conclusion

The problem of an elliptic crack lying at the interface of two bonded dissimilar elastic solids has been formulated as three pairs of dual integral Eqs. (28.13a, 28.13b) to (28.15a, 28.15b). For the special case of the crack faces subjected to uniform normal pressure, the coefficients associated to the Fourier cosine term $V_0^c(k)$, responsible for the solution of the potential λ_j, vanishes and these three pairs of integral equations have been reduced to two pairs of integral equations that are similar to those obtained earlier in the problem of a penny-shaped crack lying at a similar interface. The two pairs of dual integral equations reduce to Cauchy singular integral equation following [12]. This singular integral equation is solved through Plemelj formula.

Thus the solution to the displacement field has been obtained analytically which has opened up the field of research in interface elliptic crack problems. The quantities of physical interest, e.g. crack-opening displacement, stress intensity factor and crack energy release rate, are under consideration.

Acknowledgments The first author (T.K. Saha) acknowledges the University Grants Commission of India for awarding a Minor Research Project (PSW-149/11-12 (ERO)). The authors also gratefully acknowledge the reviewer for the constructive comments.

References

1. M.L. Williams, The stresses around a fault or crack in dissimilar media. Bull. Seismol. Soc. Am. **49**, 199–204 (1959)
2. V.I. Mossakovsky, M.T. Rykba, Generalization ot the Griffith-Sneddon criterion for the case of a non-homogeneous body. Prikl. Mat. Mekh. **28**, 1061–1069 (1964) (English translation in PMM J. Appl. Math. Mech. **28**, 1277–1286)
3. R.L. Salganik, The brittle fracture of cemented bodies. J. Appl. Math. Mech. **27**, 1468–1478 (1963)
4. A.H. England, A crack between dissimilar media. J. Appl. Mech. **32**, 400–402 (1965)
5. F. Erdogan, Stress distribution in bonded dissimilar materials with cracks. J. Appl. Mech. **32**, 403–410 (1965)
6. J.R. Rice, G.C. Sih, Plane problems of cracks in dissimilar media. J. Appl. Mech. **32**, 418–423 (1965)
7. K. Arin, F. Erdogan, Penny-shaped crack in an elastic layer bonded to dissimilar half spaces. Int. J. Eng. Sci. **9**, 213–232 (1971)
8. F. Erdogan, K. Arin, Penny-shaped interface crack between an elastic layer and a half space. Int. J. Eng. Sci. **10**, 115–125 (1972)
9. M. Lowengrub, I.N. Sneddon, The effect of shear on a penny-shaped crack at the interface of an elastic half-space and a rigid foundation. Int. J. Eng. Sci. **10**, 899–913 (1972)
10. J.R. Willis, The penny-shaped crack on an interface. Q. J. Mech. Appl. Mech. **25**, 367–385 (1972)
11. M.K. Kassir, A.M. Bregman, The stress-intensity factor for a penny-shaped crack between two dissimilar materials. J. Appl. Mech. **39**, 308–310 (1972)
12. M. Lowengrub, I.N. Sneddon, The effect of internal pressure on a penny-shaped crack at the interface of two bonded dissimilar elastic half-spaces. Int. J. Eng. Sci. **12**, 387–396 (1974)
13. R.V. Goldstein, V.M. Vainshelbaum, Axisymmetric problem of a crack at the interface of layers in a multi-layered medium. Int. J. Eng. Sci. **14**, 335–352 (1976)
14. M. Comninou, The interface crack. J. Appl. Mech. **44**, 631–636 (1977)
15. R. Calhoun, M. Lowengrub, Stress in the vicinity of a Griffith crack at the interface of a layer bonded to a half plane an approximate method. Int. J. Eng. Sci. **16**, 423–441 (1978)
16. A.K. Gautesen, J. Dundurs, The interface crack in a tension field. J. Appl. Mech. **54**, 93–98 (1987)
17. J.R. Rice, Elastic fracture mechanics concepts for interfacial cracks. J. Appl. Mech. **55**, 98–103 (1988)
18. E.I. Shifrin, B. Brank, G. Surace, Analytical-numerical solution of elliptical interface crack problem. Int. J. Fract. **94**, 201–215 (1998)
19. N.I. Muskhelishvili, *Singular Integral Equations* (Dover Publications INC., New York, 1992)
20. A. Roy, T.K. Saha, Weight function for an elliptic crack in an infinite medium. Part-I. Normal Loading. Int. J. Fract. **103**, 227–241 (2000)

21. C.K. Youngdahl, On the completeness of a set of stress functions appropriate to the solution of elastic problems in general cylindrical coordinates. Int. J. Eng. Sci. **7**, 61–79 (1969)
22. K. Marguerre, Ansätzezur Lösung der Grundgleichungen der Elastizitätstheorie. Z. Angew. Math. Mech. **35**, 242–263 (1955)

Chapter 29
Dynamical Complexity of a Ratio-Dependent Predator-Prey Model with Strong Additive Allee Effect

Pallav Jyoti Pal and Tapan Saha

Abstract In this paper, a predator-prey systems of two species is proposed where prey population is subjected to a strong additive Allee effect and predator population consumes the prey according to the ratio-dependent Holling type-II functional response. We use the blow-up technique in order to explore the local structure of orbits in the vicinity of origin. We have determined the conditions for extinction/survival scenarios of species. Some basic dynamical results; the stability; phenomenon of bi-stability and the existence of separatrix curves; Hopf bifurcation; saddle-node bifurcation; homoclinic bifurcation, and Bogdanov-Takens bifurcation of the system are studied. Numerical simulation results that complement the theoretical predictions are presented. A discussion of the consequences of additive Allee effect on the model along with the ecological implications of the analytic and numerical findings is presented.

Keywords Predator-prey model · Allee effect · Stability and bifurcation · Hopf bifurcation · Saddle-node bifurcation · Bogdanov-Taken bifurcation

29.1 Introduction

The modeling of predator-prey interactions incorporating Allee effect [2, 3, 11] in prey population growth has become a broad field of research in ecology for the understanding of population dynamics. The originator and namesake of Allee effect was Warder Clyde Allee (1885–1955), an US zoologist and ecologist, who observed that many animal and plant species suffer a decrease of the per capita rate of increase as their populations reach small sizes or low densities. In particular, the population exhibits a "critical size or density," below which the per capita growth rate is negative

P.J. Pal
Department of Mathematics, Krishna Chandra College, Hetampur,
Birbhum 731124, India
e-mail: pallav.pjp@gmail.com

T. Saha (✉)
Department of Mathematics, Presidency University, Kolkata 700073, India
e-mail: tapan.maths@presiuniv.ac.in

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_29

and the population declines on average, and above which the per capita growth rate is positive and the population increases on average yielding convergence to the carrying capacity. This ecological phenomenon is termed as strong Allee effect. The Allee effect can be caused by difficulties in finding mating partners for sexual reproduction at small densities, inbreeding depression, demographic stochasticity, or a reduction in cooperative interactions.

Several algebraic forms to describe the Allee effect are available in the literature, see Table 1 of [2] or Table 3.1 of [3]. In this present study, we consider the equation

$$\frac{dx}{dt} = x\left[r(1-\frac{x}{K}) - \frac{m}{x+b}\right] \tag{29.1}$$

which is commonly known as an additive Allee effect, where K is the carrying capacity, r denotes the intrinsic per capita growth rate of the population, m and b are the Allee effect constants such that $K > b$. The term subtracted from the logistic growth term is proportional to $\frac{m}{x+b}$ in Eq. (29.1) is to represent the reduction of the per capita growth rate of a population due to Allee effect.

We have considered the ratio-dependent functional response where the consumption rate of the predator is a function of the prey-to-predator ratio, not on the absolute numbers of prey only or both species. There are growing explicit biological and physiological evidence (cf. [1, 6]) that in many situations when predators have to search for food (and, therefore, have to share or compete for food), a more suitable general predator-prey theory should be based on the ratio-dependent theory. To the best of our knowledge, the effect of additive Allee on a ratio-dependent [5–7, 12, 13] predator-prey model is entirely unaddressed in the literature to date. However, the effect of multiplicative Allee effect (with single and multiple mechanism) on ratio-dependent predator-prey model have recently been described in [4, 9]. In this paper, we offer a contribution toward addressing this major research gap by establishing complete study of the dynamics including a detailed bifurcation analysis of our proposed model.

This paper is organized as follows: The model is proposed in Sect. 29.2 along with some basic results. Section 29.3 deals with the mathematical analysis including existence of equilibria, stability, and Hopf bifurcation analysis of the model. This section also discusses the stability analysis of the origin (a complicated equilibrium point). In Sect. 29.4, we prove the existence of a Bogdanov-Takens bifurcation of codimension 2 including a series of other bifurcations, such as saddle-node bifurcation, Hopf bifurcation, and Homoclinic bifurcations. In Sect. 29.5, we perform numerical simulation in support of our analytical results and discuss the main results of the paper.

29.2 Model Description and Basic Results

In this paper, we consider a predator-prey model where the prey growth is damped by the strong additive Allee effect given by Eq. (29.1) and the functional response of predator to prey abundance is ratio-dependent given by $\frac{cx}{x+\vartheta y}$ where c is the capturing rate of the predator and ϑ is the half-saturation constant of the predator functional response. Accordingly, we are concerned with the following ratio-dependent Holling-type II predator-prey model

$$\frac{dx}{dt} = \left[r\left(1-\frac{x}{K}\right) - \frac{m}{x+b}\right]x - \frac{cxy}{x+\vartheta y}, \tag{29.2a}$$

$$\frac{dy}{dt} = \frac{c_1xy}{x+\vartheta y} - dy. \tag{29.2b}$$

such that $x(0) > 0$, $y(0) > 0$. In system (29.2), $x(t)$ and $y(t)$ stands for prey and predator density at time t, and c_1, d are positive constants that stand for conversion rate of prey into predators biomass, death rate of predator, respectively.

Non-dimensionalization of this model (29.2) can be performed by using the transformation $x = K\widehat{x}$, $y = \frac{K}{\vartheta}\widehat{y}$, $t = \frac{\widehat{t}}{r}$ and dropping the hats for notational convenience, we derive

$$\frac{dx}{dt} = x(1-x) - \frac{\gamma x}{x+\rho} - \frac{\alpha xy}{x+y} = f(x,y), \tag{29.3a}$$

$$\frac{dy}{dt} = \frac{\beta xy}{x+y} - \delta y = g(x,y), \tag{29.3b}$$

where $\alpha = \frac{c}{r\vartheta}$, $\beta = \frac{c_1}{r}$, $\gamma = \frac{m}{r}$, $\rho = \frac{b}{K}$ and $\delta = \frac{d}{r}$ are the dimensionless parameters with the following initial conditions

$$x(0) = x_0 > 0, \quad y(0) = y_0 > 0.$$

The model system (29.3) is not well-defined at the origin and for this we define $f(0,0) = g(0,0) = 0$. To illustrate the types of Allee effect (cf. [10]) on the prey population in the absence of predator, we present Fig. 29.1. In this study, we only consider strong Allee effect on the prey population and we aim to discuss the complex interplay between the strong additive Allee effect and the predation on the deterministic population dynamics in continuous time. For strong Allee effect, we

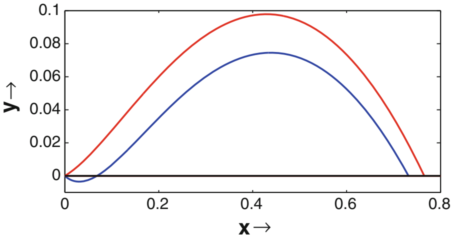


Fig. 29.1 *Blue curve* Strong Allee effect for $\gamma > \rho$, $\rho < 1$ and $(\rho+1)^2 > 4\gamma$. The parameter values are $\alpha = 0.3$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.2$. *Red curve* Weak Allee effect for $\gamma < \rho$. Parameter values are $\alpha = 0.3$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.3$

have $\gamma > \rho$, $\rho < 1$ and $(\rho + 1)^2 > 4\gamma$. Based on the standard methods we shall present some preliminary results like positivity and the boundedness of solutions of the system (29.3) for the case of a strong Allee effect without proof.

Lemma 29.1 *Solutions of model (29.3) corresponding to initial conditions (29.4) are defined on* $[0, +\infty)$ *and remain positive for all* $t \geq 0$.

Theorem 29.1 *All the solutions of system (29.3) that initiate in* R^2_+ *are uniformly bounded with an ultimate bound.*

29.3 Stability and Hopf Bifurcation Results

We now find all biological feasible equilibria admitted by system (29.3). For all parameter values, (0, 0) is an equilibrium point (controversial equilibrium point) of the system. The equilibria on the positive x-axis are $E_1(x_1, 0)$ and $E_2(x_2, 0)$ where

$$x_1 = \frac{1 - \rho - \sqrt{D_1}}{2} \text{ and } x_2 = \frac{1 - \rho + \sqrt{D_1}}{2}$$

such that $D_1 = (1+\rho)^2 - 4\gamma > 0$. If $\gamma = \frac{1}{4}(1+\rho)^2$, both the axial equilibria collides to $\left(\frac{1}{2}(1-\rho), 0\right)$ and if $\gamma > \frac{1}{4}(1+\rho)^2$, there exists no axial equilibria on the positive x-axis. The other equilibria, if exists, are the interior equilibrium point(s). Assume $A = (1-\rho)\beta - (\beta - \delta)\alpha$, $B = \alpha\rho\ (\beta - \delta) + \beta\ (\gamma - \rho) > 0$ and $D_2 = A^2 - 4\beta B$, then we have the following three cases:

1. If $D_2 > 0$, then there exists two interior equilibrium points namely, $E_i^* \equiv (x_i^*, y_i^*)$, where $x_1^* = \frac{A-\sqrt{D_2}}{2\beta}$, $x_2^* = \frac{A+\sqrt{D_2}}{2\beta}$, $y_i^* = \frac{x_i^*(\beta-\delta)}{\delta}$, $i = 1, 2$ provided $A > 0$ and $\beta > \delta$.
2. If $D_2 = 0$, $\beta > \delta$ and $A > 0$ then the two positive equilibrium points E_1^* and E_2^* coincide to an unique interior equilibrium point $E^*(x^*, y^*) = \left(\frac{A}{2\beta}, \frac{A(\beta-\delta)}{2\delta\beta}\right)$.
3. If either $D_2 < 0$, or $A < 0$, the system (29.3) has no interior equilibrium point (Fig. 29.2).

29.3.1 Qualitative Property of Solutions Near (0, 0)

We note that system (29.3) is not well defined at $E_0 \equiv (0, 0)$. Thus system (29.3) cannot be linearized at (0, 0) and the standard linear stability analysis method for (0, 0) is not applicable. In Jost et al. [6] have studied the analytical behavior at (0, 0) for a common ratio-dependent model by blow-up method. Following [14], we have studied crucially all possible topological structures of a small neighborhood of (0, 0) where the trajectories approach the origin along characteristic directions. We redefine

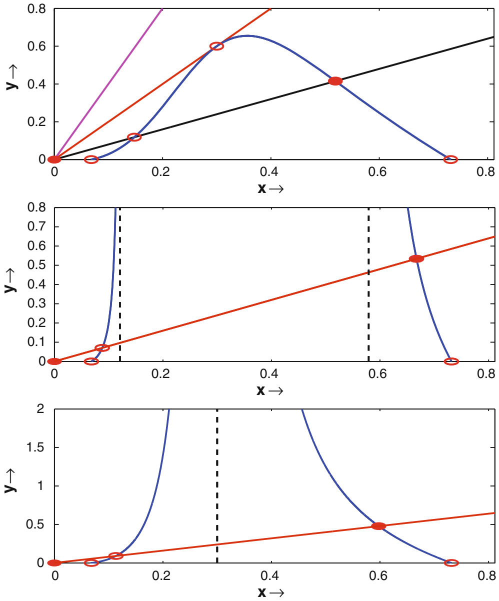


Fig. 29.2 Graphical illustration of nullclines. Here, the *red* (*blue*) curves represent the predator(prey) nontrivial nullclines. Equilibria are represented by *small red circles*. The *black lines* represent the vertical asymptotes that exists when the nontrivial prey nullcline become an unbounded curve. *Top panel* $\alpha = 0.3$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.2$, $\beta = 0.9$ (*black*), 1.5 (*red*) and 2.5 (*magenta*). *Middle panel* $\alpha = 0.1$, $\beta = 0.9$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.2$. *Lower panel* $\alpha = 0.2$, $\beta = 0.9$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.2$

the derivative as $\frac{dx}{dt} = \frac{dy}{dt} = 0$ when $(x, y) = (0, 0)$. To be compatible with ecological significance, we analyze the behavior of trajectories near $E_0(0, 0)$ in presence of all other critical points. Through time rescaling $dt \rightarrow (x + \rho)(x + y)\, d\tau$, we obtain a polynomial differential equations system topologically equivalent to original one in the interior of first quadrant. We introduce polar coordinates (r, θ), setting $x = r \cos\theta$, $y = r \sin\theta$, and the polynomial differential equations system reduces to:

$$\frac{dr}{dt} = r^2 \left(H(\theta) + o(1)\right), \tag{29.4a}$$

$$\frac{d\theta}{dt} = r \left(G(\theta) + o(1)\right), \tag{29.4b}$$

where H and G are homogeneous trigonometric polynomials in the variables $\cos\theta$ and $\sin\theta$ such that

$$\begin{aligned}H(\theta) &= -\rho\,\delta\,\sin^3\theta - \cos\theta\sin\theta(-\rho\,(\beta-\delta)\,\sin\theta \\ &\quad + (\gamma-\rho+\alpha\,\rho)\cos\theta) + (\rho-\gamma)\cos^3\theta, \\ G(\theta) &= \cos\theta\sin\theta((\alpha\,\rho-\rho\,\delta-\rho+\gamma)\sin\theta \\ &\quad + (\rho\,\beta-\rho\,\delta+\gamma-\rho)\cos\theta).\end{aligned}$$

Then, the characteristic equation is given by $G(\theta) = 0$, i.e.,

$$\cos\theta\sin\theta\,((\alpha\,\rho-\rho\,\delta-\rho+\gamma)\sin\theta + (\rho\,\beta-\rho\,\delta+\gamma-\rho)\cos\theta) = 0. \quad (29.5)$$

By [14], any trajectory that will tends to origin must tend to it either spirally or along a fixed direction. This can be characterized from the characteristic equation. Clearly, $\rho\ (\beta-\delta)+\gamma-\rho > 0$. Then, the following three cases arise:

29.3.1.1 Case 1: $\alpha\,\rho-\rho\,\delta-\rho+\gamma\ >0$

In this case, the characteristic equation (29.5) has two roots in $0 \le \theta \le \frac{\pi}{2}$, namely $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{2}$.

Theorem 29.2 *Suppose that* $(\alpha\,\rho-\rho\,\delta-\rho+\gamma) > 0$. *Then*

1. *there exists* $\varepsilon_1 > 0$ *and* $r_1 > 0$ *such that there exists a unique orbit of the system in* $\{(\theta, r) : 0 \le \theta < \varepsilon_1,\ \ 0 < r < r_1\}$ *tends to* $(0, 0)$ *along* $\theta_1 = 0$ *as* $t \to +\infty$*,*
2. *there exists* $\varepsilon_2 > 0$ *and* $r_2 > 0$ *such that all orbits of the system in* $\left\{(\theta, r) : 0 \le \frac{\pi}{2} - \theta < \varepsilon_2,\ \ 0 < r < r_2\right\}$ *that tend to* $(0, 0)$ *along* $\theta_2 = \frac{\pi}{2}$ *as* $t \to +\infty$.

29.3.1.2 Case 2: $\alpha\,\rho-\rho\,\delta-\rho+\gamma=0$

In this case Eq. (29.5) has two roots in $0 \le \theta \le \frac{\pi}{2}$, namely $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{2}$ with θ_2 being a real multiple root of multiplicity two of $G(\theta) = 0$.

Theorem 29.3 *Suppose that* $(\alpha\,\rho-\rho\,\delta-\rho+\gamma) = 0$. *Then*

1. *there exists* $\varepsilon_3 > 0$ *and* $r_3 > 0$ *such that there exists a unique orbit of the system in* $\{(\theta, r) : 0 \le \theta < \varepsilon_3,\ \ 0 < r < r_3\}$ *tends to* $(0, 0)$ *along* $\theta_1 = 0$ *as* $t \to +\infty$*,*
2. *there exists* $\varepsilon_4 > 0$ *and* $r_4 > 0$ *such that all orbits of the system in* $\left\{(\theta, r) : 0 \le \frac{\pi}{2} - \theta < \varepsilon_4,\ \ 0 < r < r_4\right\}$ *that tend to* $(0, 0)$ *along* $\theta_2 = \frac{\pi}{2}$ *as* $t \to +\infty$.

29.3.1.3 Case 3: $\alpha\,\rho-\rho\,\delta-\rho+\gamma\ <0$

In this case, (29.5) has three simple roots, namely $\theta_1 = 0$, $\theta_2 = \frac{\pi}{2}$ and $\theta_3 = \arctan\frac{-(\rho\,\beta-\rho\,\delta+\gamma-\rho)}{\alpha\,\rho-\rho\,\delta-\rho+\gamma}$. We have exactly the same results as stated in the above theorems for characteristic directions θ_1 and θ_2 and we have to study for the other characteristic direction θ_3 only. We apply Briot-Bouquet transformation [14] to prove the following theorem.

Theorem 29.4 *Suppose* $(\alpha\rho - \rho\delta - \rho + \gamma) < 0$. *Then there exist* $\varepsilon_5 > 0$ *and* $r_5 > 0$, *such that all orbits of the system in* $\{(\theta, r) : 0 \leq |\theta - \theta_3| < \varepsilon_5,\ 0 < r < r_5\}$ *that tends to* $(0, 0)$ *along* θ_3 *as* $t \to \infty$.

29.3.2 Local Stability of Equilibria and Bifurcation Results

In this section, we focus on investigating the local asymptotic stability of the boundary equilibria E_1 and E_2 and interior equilibria E_i^*, $i = 1, 2$, whenever they exists, by studying the eigenvalues of the Jacobian matrix evaluated at each equilibrium points. Furthermore, we also study the existence of Hopf bifurcation around the interior equilibrium point E_2^* with α as bifurcation parameter arising when E_2^* looses its stability.

$E_1(x_1, 0)$ is a saddle point having the x-axis as an unstable manifold if interior equilibria does not exits, otherwise it is an unstable node provide $\beta \neq \delta$. If $\beta = \delta$, the system (29.3) is reduced to the following system by suitable transformation

$$\dot{z}_1 = \lambda_{11} z_1 + ||z||^2, \quad \dot{z}_2 = -\frac{\beta}{x_1} {z_2}^2 + ||z||^3,$$

where $\lambda_{11} = \frac{x_1\sqrt{D_1}}{x_1+\rho} > 0$. It indicates that $E_1(x_1, 0)$ is a saddle-node (repelling).

$E_2(x_2, 0)$ is stable if interior equilibria does not exits, otherwise it is a saddle having stable manifold along x-axis provided $\beta \neq \delta$. If $\beta = \delta$, $E_2 \equiv (x_2, 0)$ of system (29.3) is a saddle-node (attracting).

The trace and determinant of the Jacobian matrix J_i^* of the system (29.3) at E_i^* are given by

$$Tr(J_i^*)|_{(x_i^*, y_i^*)} = -\frac{x_i^*\sqrt{D_2}}{\beta\left(x_i^* + \rho\right)} - \frac{x_i^* y_i^* (\beta - \alpha)}{\left(x_i^* + y_i^*\right)^2} \text{ and}$$

$$\det J_i^*|_{(x_i^*, y_i^*)} = \frac{\left(2\beta x_i^* - A\right) x_i^{*2}\, y_i^*}{\left(x_i^* + \rho\right)\left(x_i^* + y_i^*\right)^2} = \frac{(-1)^i\sqrt{D_2} x_i^{*2}\, y_i^*}{\left(x_i^* + \rho\right)\left(x_i^* + y_i^*\right)^2}.$$

It clearly shows that, the critical point (x_1^*, y_1^*) is always a saddle point, where as, the locally asymptotic stability of the critical point (x_2^*, y_2^*) is determined by sign of trace of $J_2^*|_{(x_2^*, y_2^*)}$. For $\alpha < \beta$, $Tr(J_2^*)|_{(x_2^*, y_2^*)} < 0$. Therefore, the system (29.3) will be locally asymptotically stable around the interior equilibrium point $E_2^*(x_2^*, y_2^*)$ if $\alpha < \beta$.

29.3.2.1 Hopf Bifurcation and Its Degeneracy

Consider that $\exists\ \alpha = \alpha^*$ such that $Tr(J_2^*)|_{(x_2^*, y_2^*)} = 0$. Consequently, since $\det J_2^*|_{(x_2^*, y_2^*)} > 0$, both the eigenvalues of J_2^* at (x_2^*, y_2^*) are purely imaginary given by $\pm i\sqrt{\det J_2^*|_{(x_2^*, y_2^*, \alpha^*)}}$. It has been observed that the transversality condition for Hopf bifurcation holds, therefore, the system experiences a Hopf bifurcation at the critical value $\alpha = \alpha^*$. Further, $E_2^*(x_2^*, y_2^*)$ is unstable if $Tr(J_2^*)|_{(x_2^*, y_2^*)} > 0$. It is to be noted that the computation of explicit expression for α^* in terms of system parameters other than α is a very cumbersome task and is not carried out here. However, it may be observed that, whenever $\alpha < \alpha^*$, the positive interior equilibrium E_2^* of system (29.3) is a locally asymptotically stable node and for $\alpha > \alpha^*$, E_2^* is an unstable focus through a Hopf bifurcation that occurs around E_2^* due to the stability changes from stable to unstable at the critical value $\alpha = \alpha^*$. We will employ a numerical example to illustrate the fact discussed above.

Degeneracy of Hopf bifurcation point can be determined by computing Lyapunov coefficients or by deriving normal form with the help of central manifold argument. If it is nondegenerate then we have only one limit cycle around E_2^* in the vicinity of $\alpha = \alpha^*$ and if it is degenerate then we have to compute the multiplicity of the focus E_2^* at $\alpha = \alpha^*$. We have observed numerically that the first Lyapunov coefficient is positive.

29.3.2.2 Saddle-Node Bifurcation

Theorem 29.5 *The system (29.3) undergoes a saddle-node bifurcation around* $E^* \equiv (x^*, y^*)$ *when* $\rho = \rho^*$ *where* $\rho^* = \frac{-\beta+\alpha\,\beta-\alpha\,\delta+2\,\beta\,\sqrt{\gamma}}{\beta}$ *and* $A = (1-\rho)\beta-(\beta-\delta)\alpha$ *provided* $A > 0$, $\alpha\,(\beta-\delta)+2\,\beta\,\sqrt{\gamma} > \beta$ *and* $\beta > \alpha$.

Proof One of the eigenvalues of the Jacobian matrix (J^*, say) evaluated at $E^*(x^*, y^*)$ will be zero iff $\det J^*|_{(x^*, y^*)} = 0$, which gives $\rho = \frac{-\beta+\alpha\,\beta-\alpha\,\delta+2\,\beta\,\sqrt{\gamma}}{\beta} = \rho^*$, say. The other eigenvalue is given by $Tr\ (J^*) = -\frac{x_1^* y_1^* (\beta-\alpha)}{(x_1^*+y_1^*)^2}$ which will be negative in order to ensure a saddle-node bifurcation implying $\beta > \alpha$. The eigenvectors of J^* and $(J^*)^T$ associated to the eigenvalue 0 is given by $\Lambda_{21} = \left(\frac{\delta}{\alpha-\delta}, 1\right)^T$ and $\Lambda_{22} = \left(-\frac{\beta\,(\beta-\delta)}{\alpha\,\delta}, 1\right)^T$, respectively. Now, $\Lambda_{22}^T[F_\rho(E^*, \rho^*)] = \frac{-2\,(\beta-\delta)\gamma\,A\beta^2}{\alpha\,\delta\,(A+2\,\rho^*\,\beta)^2} \neq 0$ and $\Lambda_{22}^T[D^2F(E^*, \rho^*)(\Lambda_{21}, \Lambda_{21})] \neq 0$. Thus by using Sotomayor's theorem, we conclude that, the system experiences a saddle-node bifurcation around E^* at the bifurcation value $\rho = \rho^*$. This means that, there are no equilibria for $\rho < \rho^*$ and there are two equilibria namely $E_i^* \equiv (x_i^*, y_i^*)$, $i = 1, 2$ for $\rho > \rho^*$, one of which is saddle point and the other is a node.

29.4 Bogdanov-Takens Bifurcation

In this section, we discuss the Bogdanov-Takens bifurcation of the model system (29.3) by using the methods in [8]. We assume that the conditions $\beta > \delta$, $A > 0$, $A^2 = 4\beta B$ hold for which the two interior equilibria E_1^* and E_2^* merge into the nonhyperbolic critical point $E^*\left(\frac{A}{2\beta}, \frac{A(\beta-\delta)}{2\beta\delta}\right)$. Under these conditions it can be shown that E^* is a saddle node whenever $\alpha \neq \beta$; attracting if $\beta > \alpha$ and repelling if $\beta < \alpha$. We assume $\alpha = \beta = \alpha^*$. In this case, the Jacobian matrix corresponding to the linearization of (29.3) at E^* has two zero eigenvalues. Our first task is to investigate the nature of the critical point E^* under the conditions $\alpha = \beta = \alpha^*$ and $\delta = \delta^*$.

Using the following transformation $x_1 = x - x^*$, $y_1 = y - y^*$, $x^* = A/2\beta$, and $y^* = A(\beta - \delta)/2\beta\delta$, we get

$$\dot{x}_1 = \bar{p}_{10}x_1 + \bar{p}_{01}x_2 + \bar{p}_{20}x_1^2 + \bar{p}_{11}x_1x_2 + \bar{p}_{02}x_2^2 + O(||x||^3) \tag{29.6}$$

$$\dot{x}_2 = \bar{q}_{10}x_1 + \bar{q}_{01}x_2 + \bar{q}_{20}x_1^2 + \bar{q}_{11}x_1x_2 + \bar{q}_{02}x_2^2 + O(||x||^3) \tag{29.7}$$

where $\bar{p}_{ij} = \frac{1}{i!j!}\frac{\partial^{i+j} f}{\partial x_1^i \partial x_2^j}$, $\bar{q}_{ij} = \frac{1}{i!j!}\frac{\partial^{i+j} g}{\partial x_1^i \partial x_2^j}$ at E^* and $1 \leq i + j \leq 2$. Using a series of transformations, we reduce the system (29.3) to

$$\frac{d\omega_1}{dt} = \omega_2, \tag{29.8}$$

$$\frac{d\omega_2}{dt} = \rho_1\omega_1^2 + \rho_2\omega_1\omega_2 + O(||\omega||^3), \tag{29.9}$$

where $\rho_1 = \bar{p}_{01}\bar{q}_{20} + \bar{p}_{10}\ (\bar{p}_{20} - \bar{q}_{11}) - \frac{\bar{p}_{10}^2(\bar{p}_{11}-\bar{q}_{02})}{b_1} + \frac{\bar{p}_{02}\,\bar{p}_{10}^3}{\bar{p}_{01}^2}$ and $\rho_2 = -\frac{\bar{p}_{10}}{\bar{p}_{01}}(\bar{p}_{11} + 2\bar{q}_{02}) + 2\bar{p}_{20} + \bar{q}_{11}$ when $\rho_1\rho_2 \neq 0$. Hence, the critical point E^* is a cusp of codimension 2, i.e., a Bogdanov-Takens singularity. This shows that for parameters (α, δ) in a neighborhood of (α^*, δ^*), the model system (29.3) undergoes BT bifurcation at E^*.

Now our task is to derive the generic normal unfolding of BT singularity. Consider $\alpha = \beta = \alpha^* + \lambda_1$, $\delta = \delta^* + \lambda_2$ where λ_1 and λ_2 is very small. Then in a sufficiently small neighborhood of (x^*, y^*, λ^*), there exists a parameter dependent nonlinear smooth invertible variable transformations, smooth invertible parameter changes, and a direction preserving time reparametrization, which together reduce the system (29.3) to the following normal form

$$\frac{d\xi_1}{d\tau} = \xi_2, \tag{29.10}$$

$$\frac{d\xi_2}{d\tau} = \mu_1 + \mu_2\xi_1 + \xi_1^2 + s\xi_1\xi_2, \tag{29.11}$$

where $s = \pm 1$. The expressions of μ_1, μ_2 and the transversality condition of a Bogdanov-Takens bifurcation are not presented here for the lack of space.

Assume that $s = -1$. There exists a neighborhood of $(\mu_1, \mu_2) = (0, 0)$ in R^2 so that the bifurcation plane is divided into four regions by the following curves

1. $SN^+ = \{(\mu_1, \mu_2) : \mu_2^2 = 4\mu_1, \mu_2 < 0\}$,
2. $SN^- = \{(\mu_1, \mu_2) : \mu_2^2 = 4\mu_1, \mu_2 > 0\}$,
3. $H = \{(\mu_1, \mu_2) : \mu_1 = 0, \mu_2 < 0\}$,
4. $HL = \{(\mu_1, \mu_2) : \mu_1 = -\frac{6}{25}\mu_2^2 + O(\mu_2^2), \mu_2 < 0\}$,

where SN represents a saddle-node bifurcation curve having two branches SN^+ and SN^- corresponding to $\mu_2 < 0$ and $\mu_2 > 0$ respectively, H is the Hopf bifurcation curve and HL is the Homoclinic bifurcation curve. For the case $s = +1$, the local representations of bifurcation curves in a small neighborhood of $(\mu_1, \mu_2) = (0, 0)$ will be obtained by using the linear transformation of coordinates $(\xi_1, \xi_2, t, \mu_1, \mu_2) \rightarrow (\xi_1, -\xi_2, -t, \mu_1, -\mu_2)$.

29.5 Conclusion

The Allee effect has been shown to be very common in population dynamics. In this paper we have proposed a ratio-dependent predator-prey model with a strong additive Allee effect in prey population growth. We have shown that the trajectories approach the origin along characteristic directions divide a neighborhood of the origin into a finite number sectors. We have observed that the origin is always a point of attraction (cf. Fig. 29.3). For a certain set of parameters, the total extinction,

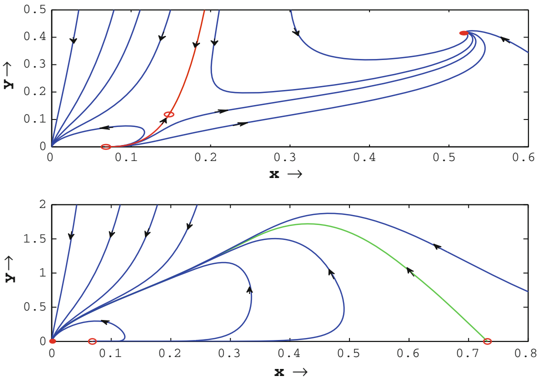


Fig. 29.3 *Top panel* Origin is an attractor for $\alpha = 0.3$ $((\alpha\rho - \rho\delta - \rho + \gamma) > 0)$, $\alpha = 0.25$ $((\alpha\rho - \rho\delta - \rho + \gamma) = 0)$ and $\alpha = 0.2$ $((\alpha\rho - \rho\delta - \rho + \gamma) < 0)$ with other parameter values $\beta = 0.9$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.2$. *Lower panel* For $\alpha = 0.3$, $\beta = 2.5$, $\gamma = 0.25$, $\delta = 0.5$, $\rho = 0.2$, origin is a global attractor

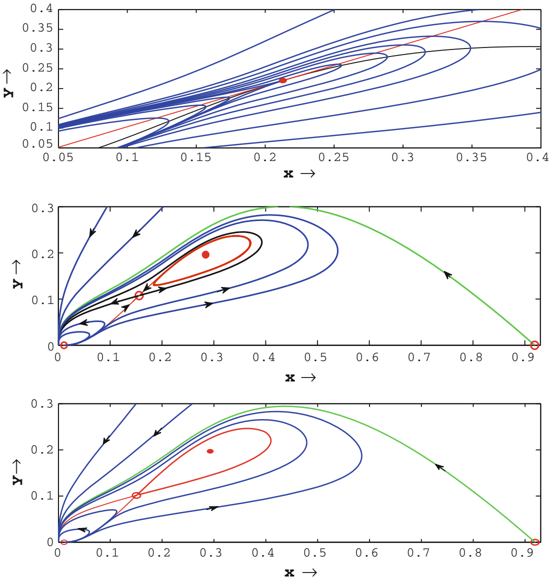


Fig. 29.4 Phase portraits in the (x, y) plane. *Top panel* When $\lambda_1 = 0, \lambda_2 = 0$, the unique degenerate interior equilibrium point E^* is a cusp of codimension 2. *Middle panel* When $\lambda_1 = 0.190685425, \lambda_2 = -0.151314575$, there exists a limit cycle. *Lower panel* When $\lambda_1 = 0.200385425, \lambda_2 = -0.0164314575$, there is a homoclinic orbit (shown by *red solid curve*). The parameter values are $\alpha = 1.009314575 = \beta, \gamma = 0.08, \delta = 0.505, \rho = 0.07$

population coexistence or the oscillating coexistence of population are observed. The bi-stability scenario is detected. Two singularities (0, 0) and E_2^* can be local attractor at the first quadrant, or a limit cycle coexists around E_2^* with a locally asymptotically stable point (0, 0). Both the basins of attractions are separated by a separatrix and the trajectories near the separatrix curve are extremely sensitive to the choice of initial condition. We have shown that the model exhibit codimension two bifurcations near a Bogdanov-Takens singularity, which produces a series of bifurcation like Hopf bifurcation, saddle-node bifurcation, Homoclinic curve when two parameters vary near the interior equilibrium point for some specific parameter values (cf. Fig. 29.4).

References

1. R. Arditi, L. Ginzburg, Coupling in predator-prey dynamics: ratio-dependence. J. Theor. Biol. **139**(3), 311–326 (1989)
2. D. Boukal, L. Berec, Single-species models of the Allee effect: extinction boundaries, sex ratios and mate encounters. J. Theor. Biol. **218**(3), 375–394 (2002)
3. F. Courchamp, L. Berec, J. Gascoigne, *Allee Effects in Ecology and Conservation* (Oxford University Press, Oxford, 2009)

4. Y. Gao, B. Li, Dynamics of a ratio-dependent predator-prey system with a strong Allee effect. Discret. Contin. Dyn. Syst.-Ser. B **18**(9) (2013)
5. S. Hsu, T. Hwang, Y. Kuang, Rich dynamics of a ratio-dependent one-prey two-predators model. J. Math. Biol. **43**(5), 377–396 (2001)
6. C. Jost, O. Arino, R. Arditi, About deterministic extinction in ratio-dependent predator-prey models. Bull. Math. Biol. **61**(1), 19–32 (1999)
7. Y. Kuang, E. Beretta, Global qualitative analysis of a ratio-dependent predator-prey system. J. Math. Biol. **36**(4), 389–406 (1998)
8. Y. Kuznetsov, *Elements of Applied Bifurcation Theory*, vol. 112. (Springer, Berlin, 1998)
9. M. Sen, M. Banjerjee, A. Morozov, Bifurcation analysis of a ratio-dependent pre-predator model with the Allee effect. Ecol. Complex. **11**, 12–27 (2012)
10. M. Wang, M. Kot et al., Speeds of invasion in a model with strong or weak Allee effects. Math. Biosci. **171**(1), 83 (2001)
11. D. Xiao, S. Ruan, Global analysis in a predator-prey system with non monotonic functional response. SIAM J. Appl. Math. **61**(4), 1445–1472 (2001)
12. D. Xiao, S. Ruan, Global dynamics of a ratio-dependent predator-prey system. J. Math. Biol. **43**(3), 268–290 (2001)
13. R. Xu, M. Chaplain, F. Davidson, Persistence and global stability of a ratio-dependent predator-prey model with state structure. Appl. Math. Comput. **158**(3), 729–744 (2004)
14. Z. Zhi-Fen, D. Tong-Ren, H. Wen-Zao, D. Zhen-xi, *Qualitative Theory of Differential Equations*, vol. 101. (American Mathematical Society, Providence, 2006)

Chapter 30
A Simple Theoretical Approach to the Fermi Energy Under Size Quantization with Quantum Mathematical Modelling in Nanostructured Materials

Subhamoy Singha Roy

Abstract In modern days, with the advent of MBE, MOCVD, FLL and other experimental techniques, low-dimensional structures having quantum confinement in one, two and three dimensions such as ultrathin films, inversion layers, quantum wires and dots, have attracted much attention, not only for their potential in uncovering new phenomena in nanostructured electronics but also for their interesting devices application in heterostructure-based various materials, that are being currently studied because of the enhancement of carrier mobility and such quantum confined systems find ex-digital networks, optical modulators and also in other devices. In this paper, an effort is made to study the Fermi-Diracs distribution function in degenerate semiconductors forming band tails (f_s) on the basis of a newly formulated electron dispersion law (f_0 is the well Fermi-Dirac function) and also it will be of much more interest, to investigate the Fermi-Dirac distribution function under the condition of carrier degeneracy, since it will help my revise in transport coefficients and electron dynamics in electronic devices made of degenerate semiconductors (n-type GaAs as an example).

Keywords Fermi energy · Nanostructured materials · Fermi-Dirac distribution

30.1 Introduction

In this context, I wish to note that the formation of band tails in degenerate semiconductors is an experimental fact and often explained by the overlapping of the impurity band with the conduction and valance bands. Kane and Bonch-Bruevich have independently developed a semi-classical theory of band tailing in semiconductors having unperturbed parabolic energy bands. Kane's semi-classical model, considering the parabolic density of states (DOS), was used to explain the experimental findings of tunnelling and the optical absorption edges in heavily doped semiconductors. Unlike

S.S. Roy (✉)
Department of Physics, Department of Nanoscience and Nanotechnology,
JIS College of Engineering (Autonomous), Kalyani, Nadia 741 235, India
e-mail: ssroy.science@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_30

Kane, in this paper I have used the realistic picture of the variation of kinetic energy of the electron with the local point in space coordinates. This kinetic energy is then averaged over the entire region of variation presumptuous a Gaussian-type potential energy and furthermore the investigational results for the Fermi energy [1–10].

30.2 Theoretical Background

The Fermi-Dirac probability density function provides the probability that an energy level is occupied by a fermion which is in thermal equilibrium with a large reservoir. Fermions are by definition particles with half-integer spin (1/2, 3/2, 5/2, …). The conservative Fermi-Dirac distribution function is given by [11–13]

$$f_0(E) = \frac{1}{1 + \exp\left[\left(E - E_f\right)/k_B T\right]}, \tag{30.1}$$

where E is the total energy of electron as measured from the edge of the conduction band in the vertically upward direction, E_f is the corresponding Fermi energy, k_B is the Boltzmann constant and T is the temperature and the general expression of the carrier density in a semiconductor is obtained by integrating the product of the density of states with probability density function over all possible states. For electrons in the conduction band the integral is taken from the bottom of the conduction band, labelled E_c to the top of the conduction band.

Now the concentration in n-type non-degenerate semiconductor can be written as

$$n = \int_{E_C}^{\text{top of the conduction band}} n(E)\,\mathrm{d}E \quad = \int_{E_C}^{\text{top of the conduction band}} g_c(E) f(E)\,\mathrm{d}E \tag{30.2}$$

$f(E)$ is the Fermi function.

$$n = \int_{E_C}^{\infty} g_c(E) f(E)\,\mathrm{d}E. \tag{30.3}$$

Now the three-dimensional n-type and p-type non-degenerate semiconductor can be written as

$$n = \int_{E_C}^{\infty} \frac{8\pi\sqrt{2}}{h^3}\, m_e^{*3/2}\, \sqrt{E - E_C}\, \frac{1}{1 + \exp[(E - E_F)/kT]}\,\mathrm{d}E \tag{30.4}$$

$$p = \int_{-\infty}^{E_V} g_v(E)\,[1 - f(E)]\,\mathrm{d}E \tag{30.5}$$

$$\text{and } p = \int_{-\infty}^{E_V} \frac{8\pi\sqrt{2}}{h^3}\, m_h^{*3/2}\, \sqrt{E_V - E}\, \frac{1}{1 + \exp\left[(E_F - E)/kT\right]}\,\mathrm{d}E. \tag{30.6}$$

Now in the degenerate semiconductors, forming band tails, the electron energy at a particular point (**r**) is given as

$$E = \hbar^2 k^2/2m_c^* + V\left(\bar{r}\right), \tag{30.7}$$

where $\hbar = h/2\pi$, h is Planck constant, k is the wave vector of the electron and m_c^* is the effective electron mass at the edge of the conduction band and $V(r)$ is the impurity potential at a local point (**r**).

The potential energy can be written as [5–7]

$$F(V) = \frac{1}{\sqrt{\pi\eta_e^2}}\ \exp\left(-\frac{V^2}{\eta_e^2}\right), \tag{30.8}$$

where η_e is the screening potential. I wish to note that the Gaussian function for the impurity potential distribution has been derived by many authors [4–6].

The average kinetic energy of the whole system is obtained by averaging the local kinetic energy fluctuation as represented by

$$\int_{-\infty}^{E} (E - V)\, F(V)\, \mathrm{d}V = \left(\frac{\hbar^2 k^2}{2m_c}\right) \int_{-\infty}^{E} F(V)\, \mathrm{d}V. \tag{30.9}$$

As the function $F(V)$ is the Gaussian distribution with limits V of extending from $-\infty$ to $+\infty$, so in the right-hand side of (30.9), I extend the upper limit of V from $V \to E$ to $V \to \infty$.

Thus from (30.9), I can write

$$\left(\frac{\hbar^2 k^2}{2m_c^*}\right) = \gamma\left(E,\ \eta_e\right), \tag{30.10}$$

where

$$\gamma\left(E,\ \eta_e\right) = \left(\frac{\eta_e}{2\sqrt{\pi}}\right)\exp\left(-\frac{E^2}{\eta_e^2}\right) + \left[\frac{1}{2E}\right]\left[1 + Erf\left(\frac{E}{\eta_e}\right)\right]$$

in which $Erf\ (E/\eta_e)$ is the error function.

The average effect of $V\ (\bar{r})$ on $E\ \left(\langle E'\rangle\right)$ can be expressed as

$$\langle E'\rangle = \frac{\displaystyle\int_{-\infty}^{E} E' F(V)\, \mathrm{d}V}{\displaystyle\int_{-\infty}^{\infty} F(V)\, \mathrm{d}V}. \tag{30.11}$$

From (30.10) and (30.11) can be written as

$$\langle E' \rangle = \gamma \left(E, \ \eta_e \right). \tag{30.12}$$

I have shown in (30.10) that $\gamma \left(E, \ \eta_e \right)$ shows the band tailing effects so the $\langle E' \rangle$ also exists for negative values of E.

The impurity screening potential, η_e is given as

$$\eta_e = \frac{e^2}{\varepsilon_d} \left(4\pi \cdot N_i / K_D \right)^{1/2}, \tag{30.13}$$

where

$$N_i = \frac{1}{3\pi^2} \left(\frac{2m_c^*}{\hbar^2} \right) \gamma^{3/2} \ \left(\bar{E}_f, \ \eta_e \right) \tag{30.14}$$

in which $\bar{E}_f$ is the Fermi energy corresponding to the average energy $\langle E' \rangle$ in degenerate semiconductors forming band tails and

$$K_D^2 = \frac{e^2}{\varepsilon_d} \cdot \frac{1}{4\pi^2} \left(\frac{2m_c^*}{\hbar^2} \right)^{3/2} \left[1 + Erf \left(\frac{\bar{E}_f}{\eta_e} \right) \right]. \tag{30.15}$$

In the absence of band tails [11–13] I have

$$N_i = \frac{1}{3\pi^2} \left(\frac{2m_c^*}{\hbar^2} \right)^{3/2} E_f^{3/2}. \tag{30.16}$$

Comparing (30.13) and (30.16), find as

$$\eta_e \to 0, \ \gamma \left(\bar{E}_f, \ \eta_e \right) = E_f. \tag{30.17}$$

Thus the Fermi-Dirac statistics for degenerate semiconductors having Gaussian band tails can be written as

$$f_p \left(\langle E' \rangle, \bar{E}_f, T \right) = \frac{1}{1 + \exp \left[\left(\langle E' \rangle - \bar{E}_f \right) \Big/ k_B T \right]}, \tag{30.18}$$

where

$$\langle E' \rangle = \frac{1}{2\pi^{1/2}} \, \eta_e \ \exp \left(-E^2 / \eta_e^2 \right) + \frac{1}{2} E \left[1 + Erf \left(E / \eta_e \right) \right]. \tag{30.19}$$

It appears from (30.19) that in the absence of band tails $\eta_e \to 0$, $\langle E' \rangle = E$ and $\bar{E}_f = E_f$ respectively. Under this case, (30.18) gets simplified to (30.1) for Fermi-Dirac distribution function in the absence of band tails. It may be noted from

(30.19) that $\langle E' \rangle$ and E are not same. Furthermore, due to the screening of the impurity potential, η_e, the average electron energy $\langle E' \rangle$ of the conduction band electron is quite different from E. As a result, we find from (30.1) and (30.18) that $f_0\left(E, E_f, T\right)$ and $f_p\left(\langle E' \rangle,\ \bar{E}_f,\ T\right)$ are not same. I can, therefore, infer that the Fermi-Dirac distribution function for degenerate semiconductors forming band tails possesses additional concentration dependence through screening potential in addition to the usual E, E_f and T dependences.

30.3 Conclusion

For the purpose of numerical computations, Fig. 30.1 shows the parabolic density of states function with $E_c = 0$ and the density of states function, the Fermi function as well as the product of both which is the density of electrons per unit volume and per unit energy, $n(E)$. The integral corresponds to cross-hatched area under the curve.

Taking n-type GaAs as example, together with the parameters $m^* = 0.067m_0$, $\varepsilon_d = 99.061\varepsilon_0$, $E = 50\,\text{MeV}$. $Erf(x) = 1 - \left(at + bt^2 + ct^3\right)\ \exp\left(-x^2\right)$ [14, 15] and $T = 4.2K$, where $a = 0.34802$, $b = -0.09587$, $c = 0.74785$, $t = (1 + px)^{-1}$ and $p = 0.47047$. I have plotted in Fig. 30.2, f_0 and f_p as functions of carrier concentration.

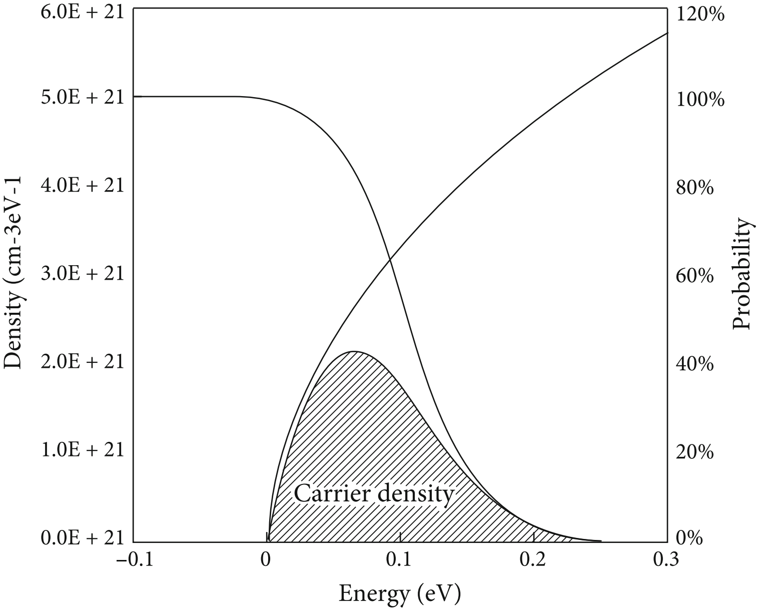


Fig. 30.1 Exposed are the density per unit energy, $n(E)$, and the probability of occupancy, $f(E)$. The carrier density equals the cross-hatched area under the curve

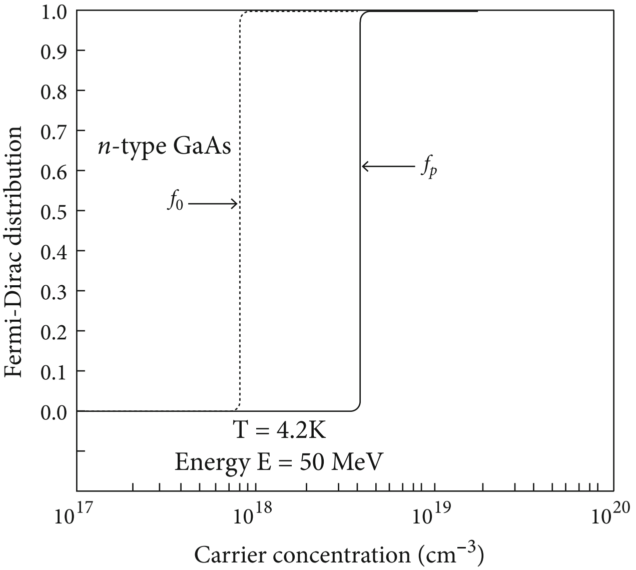


Fig. 30.2 The *solid plot* exhibits the dependence of f_p on the carrier concentration for n-GaAs at 4.2 K. The *dotted plot* exhibits the same dependence for f_0

The Fig. 30.2 indicates that f_0 and f_p show almost a step function varying between the values 0 and 1, with steps occurring at a lower N_i for f_0 and at higher values of carrier concentration for f_p. This implies that effects of carrier concentration on the distribution functions f_0 and f_p are more significant in case of f_p than that of f_0. Thus, I can conclude that the Fermi-Dirac distribution function f_p is more effective than f_0 at higher a value of carrier concentration, when the semiconductor becomes degenerate as a consequence of heavy doping and forms band tails. Thus I wish to remark that for degenerate semiconductors in the presence of band tails, the Fermi-Dirac distribution function f_p as given by (30.18) should be used than f_0 as given by (30.1). I can conclude that the experiment should be performed very carefully in order to obtain the accurate values of the distribution function comparable with the present theoretical value in heavily doped degenerate semiconductors forming band tails and it may be noted that the $E-k$ dispersion relation as formulated in this paper. Since the experimental results are not available in the literature to the best of our knowledge, I cannot compare our analysis with experiments although the theoretical results as given here would be useful in analyzing the experimental data when they appear and can also be used as the technique for probing the band structure in heavily doped non-parabolic semiconductors [14–20].

Acknowledgments The author is grateful to Kaberi Singha Roy and Arika Singha Roy for helpful comments that lead to the improvement of the paper.

References

1. O. Aina, M. Mattingly, F.Y. Juan, P.K. Bhattacharya, Photoluminescence characterisation of quantum well structures. Appl. Phys. Lett. **50**, 43 (1987)
2. J.W. Rowe, J.L. Shay, Phys. Rev. **3D**, 451 (1973)
3. H. Kildal, Band structure of CdGeAs near $k = 0$. Phys. Rev. **10**, 5082–5087 (1974)
4. R.K. Willardson, A.C. Beer (eds.), *Semiconductors and Semimetals*, vol. 1. (Academic, New York, 1966), p. 102
5. E.O. Kane, Phys. Rev. **131**, 79 (1963)
6. V.L. Bonch-Bruevich, Sov. Phys. Solid State **4**, 1953 (1963)
7. E.O. Kane, Solid State Electron **28**, 3 (1985)
8. R.A. Logan, A.G. Chenoweth, Phys. Rev. **131**, 89 (1963)
9. C.J. Hwang, J. Appl. Phys. **40**, 3731 (1969)
10. J.I. Pankove, Phys. Rev. A **130**, 2059 (1965)
11. B.R. Nag, *Electron Transport in Compound Semiconductors* (Springer, Berlin, 1980)
12. R. Dornhaus, G. Nimitz, *Springer Tracks in Modern Physics*, vol. 78. (Springer, Berlin Hiedelberg, 1976), p. 1
13. W. Zawadzki, *Handbook of semiconductor physics*, ed. by W. Paul, vol. 1. (Amsterdam, North Holland, 1982), p. 719
14. S. Singha Roy, Ph.D. Thesis, On some Electronic and Optical Properties of Non-Linear Optical and Optoelectronic Materials, Jadavpur University, India, 2005
15. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (Wiley, New York, 1964)
16. S. Singha Roy, On the optical absorption coefficient in optoelectronic compounds. Phys. Semicond. Device **2**, 932–34 (2003)
17. V.K. Arora, High-field distribution and mobility in semiconductors. J. Phys. C **18**, 3011–16 (1985)
18. P.T. Landsberg, Activity coefficient and the Einstein relation. Phys. Rev. B **33**, 8321 (1986)
19. S.S. Roy, Determination of the density of states function in highly degenerate semiconductors in the existence of electric field strength, SPIE Proc. **8542** (2012). doi:10.1117/12.970544
20. S.S. Roy, The simple theoretical analysis of quantum well wires superlattice (QWSL) of communication technology. *SPIE 8773, Photon Counting Applications IV; and Quantum Optics and Quantum Information Transfer and Processing*, 877314, 2013

Chapter 31
Entropy Generation During Mixed Convection in a Porous Double-Lid-Driven Cavity

Swapan K. Pandit, Anirban Chattopadhyay and Sreejata Sen Sarma

Abstract The issue of entropy generation in a vertically two-sided lid-driven square cavity filled with porous medium for mixed convection heat transfer is analysed by solving numerically the mass, momentum and energy balance equations, using Darcy's law and Boussinesq-incompressible approximation. Two opposite vertical walls are kept at different temperatures, while the bottom as well as the top walls is adiabatic. We have used Higher Order Compact (HOC) scheme [1] to discretize the governing equations. Entropy generation terms involving thermal and velocity gradients are evaluated accurately based on the elemental basis set via the Pade approximation method. We have first solved benchmark problem given in [2]. Excellent agreement was obtained between benchmark results and the results that validate our used computer code.

Keywords Mixed convection · Entropy generation · Double-Lid-Driven

31.1 Introduction

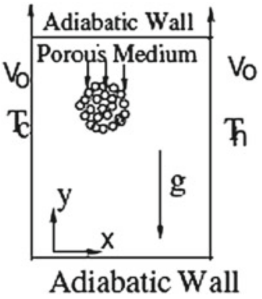
In recent years, the problem of mixed convection in enclosures with various thermal boundary conditions has been analysed in a number of studies by several researchers. In addition, the analysis of convective flow and heat transfer in fluid saturated porous media has also attracted the attention of many researchers during the past few decades. This type of flow can be found in grain storage, chemical catalytic reactors, solar collectors, heat exchangers, solidification of casting, separation processes in chemical industries, etc.

Very recently, Sivasankaran et al. [3] studied mixed convection flow and heat transfer in a square cavity with top lid moving filled with fluid saturated porous medium with sinusoidal temperature distributions on both side walls. They conclude that the non-uniform heating of both walls is beneficial for improving heat transfer, as compared to the case of uniform heating. Perusal of the literature reveals that only

S.K. Pandit · A. Chattopadhyay · S. Sen Sarma (✉)
Integrated Science Education and Research Centre (ISERC), Visva-bharati,
Santiniketan, Bolpur 731235, India
e-mail: sreejatasensarma@gmail.com

S. Sarkar et al. (eds.), *Applied Mathematics*, Springer Proceedings
in Mathematics & Statistics 146, DOI 10.1007/978-81-322-2547-8_31

Fig. 31.1 Physical configuration and coordinate system



few studies have been reported on entropy generation during convection in enclosures filled with porous medium. The analysis of entropy generation is a relatively modern method to evaluate the performance of a thermal system and to arrive at optimum design criteria based on the principle of "entropy generation minimization" [4]. Heat transfer processes are inherently an irreversible process and hence some amount of useful energy is destroyed in the process, leading to a decrease in efficiency of the system. The "loss" of useful energy due to irreversibility is given in terms of "entropy generation" based on second law of thermodynamics.

The present study describes numerically the entropy generation due to mixed convection flow in a square cavity filled with fluid saturated porous medium. The left and right moving wall have, respectively, the cold and sinusoidal temperature distribution, while both the top and bottom walls are adiabatic (see Fig. 31.1).

We have used fourth-order compact scheme on non-uniform grids presented in [1] to discretize the stream function–vorticity formulation of Navier–Stokes equations with the consideration of Darcy-Forchheimer model. The approach also involves discretizing the entropy generation equations using not only the nodal values of the unknown transport variable but also the values of its first derivatives. In turn first derivatives are discretized by using Pade approximation.

31.2 Governing Equations

The governing equations describing the incompressible viscous flows in a two-sided lid-driven cavity is in terms of non-dimensional stream function (ψ)-vorticity (ζ) formulation as follows:

$$-\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}\right)=\zeta \tag{31.1}$$

$$\frac{\partial\zeta}{\partial t}-\frac{1}{\mathrm{Re}}\cdot\frac{\partial^2\zeta}{\partial y}+u\frac{\partial\zeta}{\partial x}+v\frac{\partial\zeta}{\partial y}+\frac{1}{\mathrm{Re}Da}\zeta=\frac{Gr}{\mathrm{Re}^2}\frac{\partial T}{\partial x} \tag{31.2}$$

$$\frac{\partial T}{\partial t}-\frac{1}{\mathrm{Re}\,\mathrm{Pr}}\frac{D^2T}{\partial x^2}-\frac{1}{\mathrm{Re}\,\mathrm{Pr}}\frac{\partial^2 T}{\partial x}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}=0 \tag{31.3}$$

where $V = (u, v)$ is the velocity vector and T, the temperature.

$$Ra\left(=\frac{g\beta_T L^3\left(T_h - T_c\right)}{v\alpha}\right),\ Gr\left(=\frac{Ra}{Pr}\right),\quad Re\left(=\frac{V_0 L}{v}\right),$$
$$Pr\left(=\frac{v}{\alpha}\right),\qquad Da=\left(=\frac{K}{L^2}\right),\ Ri\left(=\frac{Gr}{Re^2}\right)$$

are, respectively, the Rayleigh number, Grasshof number, Reynolds number, Prandlt number, Darcy number and Richardson number with V_0, L, β_T, K, v are respectively reference velocity, cavity length, thermal expansion coefficient, permeability of the porous medium and kinematic viscosity. The dimensionless boundary conditions are as follows:

$u = 0, v = 1$ and $T = 0$ for $x = 0$ and $0 \le y \le 1$
$u = 0, v = 1$ and $T = \sin(\pi y)$ for $x = 1$ and $0 \le y \le 1$
$u = 0, v = 0$ and $\dfrac{\partial T}{\partial y} = 0$ for $y = 0$ and $0 \le x \le 1$; $u = 0, v = 0$ and $\dfrac{\partial T}{\partial y} = 0$ for $y = 0$ and $0 < x < 1$; $u = 0, v = 0$ and $\dfrac{\partial T}{\partial y} = 0$ for $y = 1$ and $0 \le x \le 1$.

In addition, entropy parameters such as local entropy generation due to heat transfer (S_{T1}) and local entropy generation due to fluid flow $\left(S_{f1}\right)$ can be written in nondimensional form based on the local thermodynamic equilibrium of linear transport theory [5], as follows:

$$S_{T_1} = \left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 \tag{31.4}$$

$$S_{f1} = \Gamma\left[\left\{u^2 + v^2\right\} + Da\left[2\left\{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2\right\} + \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2\right]\right] \tag{31.5}$$

where the parameter Γ in Eq. (31.5) is called irreversibility distribution ratio and is defined as

$$\Gamma = \left[\frac{\alpha^2}{k(\Delta T)^2}\right]\left(\frac{\mu T_0}{k}\right). \tag{31.6}$$

where, k is the effective thermal conductivity of the porous medium. α is the thermal diffusivity, T_0 is the bulk temperature, and μ is the dynamic viscosity. ΔT is the difference between maximum and minimum temperature of the hot wall. It may be noted here that in present study Γ is taken as 10^{-2}.

31.3 Discretization and Solution Procedure

We have discretised the transformed form of governing Eqs. (31.2)–(31.3) compactly using our recently proposed higher order compact scheme [1] designed for incompressible viscous fluid flows on non-uniform grids. It is worthwhile mentioning that we have obtained the entropy generation due to heat transfer and fluid friction by calculating the fourth-order accurate thermal and velocity gradients in the computational $(\xi - \eta)$ plane using classical Pade approximation which is given as follows:

$$\begin{aligned} \frac{1}{6}\left(\phi_{\xi}\right)_{(i-1,j)} + \frac{4}{6}\left(\phi_{\xi}\right)_{(i,j)} + \frac{1}{6}\left(\phi_{\xi}\right)_{(i+1,j)} &= \frac{\phi_{(i+1,j)} - \phi_{(i-1,j)}}{2h} \\ \frac{1}{6}\left(\phi_{\eta}\right)_{(i,j-1)} + \frac{4}{6}\left(\phi_{\eta}\right)_{(i,j)} + \frac{1}{6}\left(\phi_{\eta}\right)_{(i,j+1)} &= \frac{\phi_{(i,j+1)} - \phi_{(i,j-1)}}{2k} \end{aligned}$$

where ϕ stands for temperature and velocity variables. Here, h and k are, respectively, the uniform step lengths along horizontal and vertical directions in the computational plane.

The system of algebraic equations resulting from discretization of temperature equation, vorticity equation and stream function equation are solved in sequence using a decoupled algorithm in an outer–inner iteration procedure. In all of these computations, we have used biconjugate gradient stabilized method (BiCGStab) without preconditioning.

31.4 Results and Discussions

To assess the numerical accuracy of our computer code, we have compared the results of the problem described in Ilis et al. [2]. We have computed the maximum value of local entropy generation due to heat transfer (l.h.t.max) and the maximum value of local entropy generation due to fluid friction (l.f.f.max). An excellent match is seen (see Fig. 31.2).

The working fluid is chosen with Prandlt number Pr = 0.7 and is fixed throughout the study. The physical system consists of both the moving vertical walls along upward direction in which the shear and buoyancy forces are aiding each other on the right wall, whereas they are opposite on the left wall. So the circulation of the eddy is based on the dominant one. We have computed the results for three different

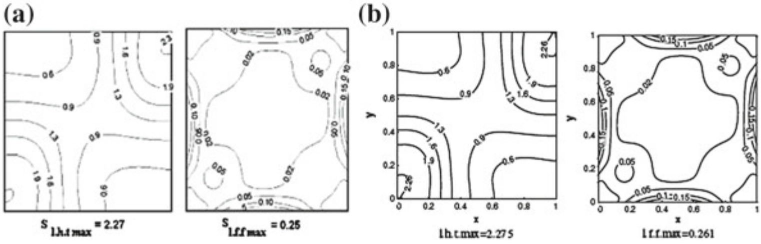


Fig. 31.2 Comparison of local entropy generation due to heat transfer and fluid friction. **a** Ilis et al. [2]. **b** Present

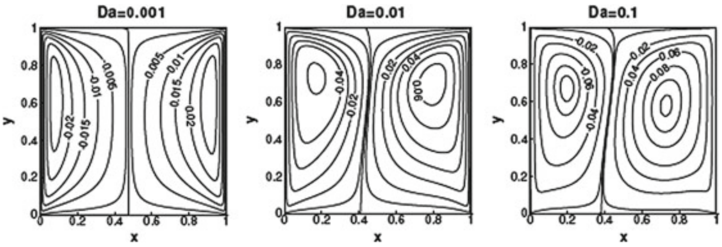


Fig. 31.3 Stream line contour for Ri = 0.1

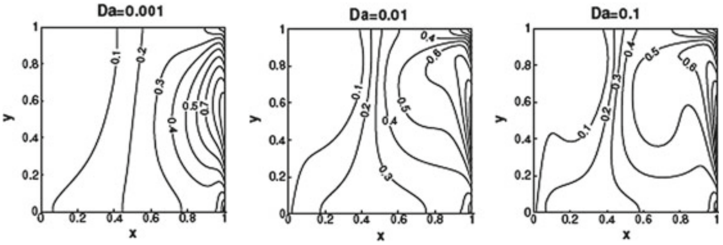


Fig. 31.4 Temperature contour for Ri = 0.1

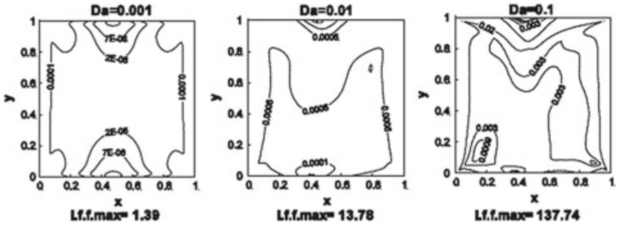


Fig. 31.5 Entropy generation due to fluid friction for Ri = 0.1

Das in both the cases of lower and higher Ris. Streamlines show two circulating cells in which a clockwise rotating cell is induced by shear force near the left wall and an anticlockwise rotating cell near the right wall. At lower Da, Da = 0.001 for Ri = 0.1, two symmetrical vortex formed while that symmetricity breaks down with the increase of Da (see Fig. 31.3). It is seen from the isotherms that the horizontal thermal gradients exist in the upper half mid-plane of the cavity. For higher Da, it clustered near the central zone. It is also seen (see Figs. 31.4 and 31.8) that the steeper thermal gradients in the mid-plane of the cavity disappear with the increase of the Richardson number. There is a significant change in streamline contours with the increase of Ri, i.e. Ri = 100. At Richardson number Ri = 100, the right cell occupies the majority of the cavity (see Figs. 31.5, 31.6, and 31.7).

From the entropy generation contour it is seen that the maximum value of local entropy generation occurs considerably higher due to heat transfer (l.h.t.max) for lower Ri in all the three cases. It is also observed that for a given Ri, the fluid friction value (l.f.f.max) is increasing with the increasing values of Da (Figs. 31.9 and 31.10).

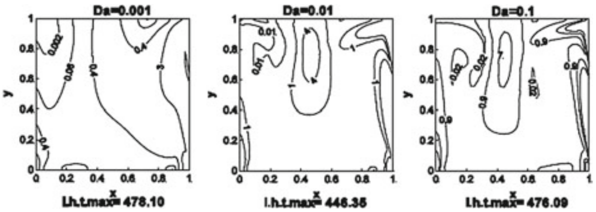


Fig. 31.6 Entropy generation due to heat transfer for Ri = 0.1

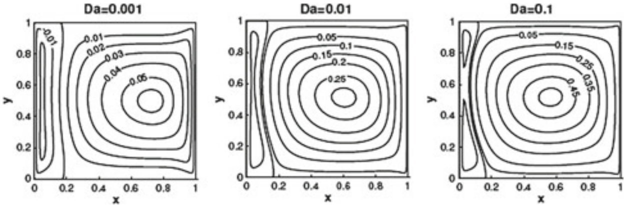


Fig. 31.7 Streamline contour for Ri = 100

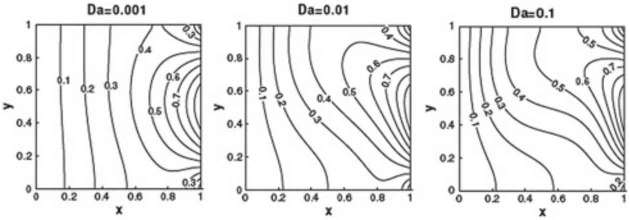


Fig. 31.8 Temperature contour for Ri = 100

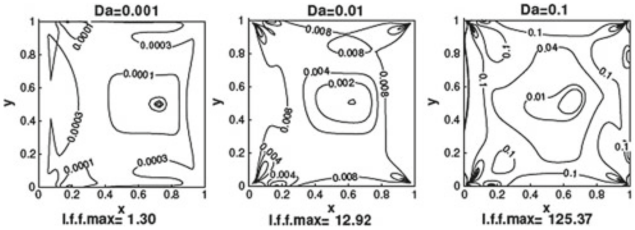


Fig. 31.9 Entropy generation due to fluid friction for Ri = 100

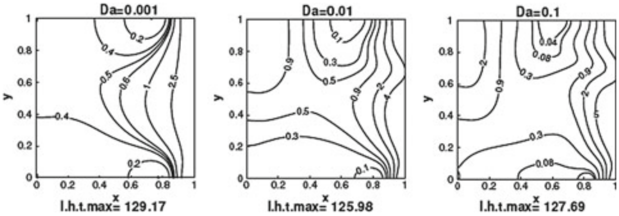


Fig. 31.10 Entropy generation due to heat transfer for Ri = 100

31.5 Summary

The present work involves the computation of incompressible flows in a two-sided lid-driven cavity using time-dependent compact scheme based on 9-point stencil to spatial differencing of the stream function–vorticity formulation of the Darcy-Forchheimer model including the energy transport equations. We have investigated the steady-state solutions for both parallel motion of the two vertical walls. The solutions reveal that there is a significant change in increasing Ri values.

References

1. S.K. Pandit, J.C. Kalita, D.C. Dalal, A transient higher order compact schemes for incompressible viscous flows on geometries beyond rectangular. J. Comput. Phys. **225**, 1100–1124 (2007)
2. G.G. Ilis, M. Mobedi, B. Sunden, Effect of aspect ratio on entropy generation in a rectangular cavity with differentially heated vertical walls. Int. Comm. Heat Mass Transfer **35**, 696–703 (2008)
3. S. Sivasankaran, V. Sivakumar, P. Prakash, Numerical study on mixed convection in a lid driven cavity with non-uniform heating on both side walls. Int. J. Heat Mass Transfer **53**, 4304–4315 (2010)
4. A. Bejan, *Entropy Generation Minimization* (CRC Press, Boca Raton, 1982)
5. T. Basak, R. Satish Kaluri, A.R. Balakrishnan, Effect of thermal boundary conditions on entropy generation during natural convection. Numer. Heat Transfer Part A **59**, 372–402 (2011)