Trends in Mathematics

Vinai K. Singh
Yaroslav D. Sergeyev
Andreas Fischer
Editors

Recent Trends in Mathematical Modeling and High Performance Computing

M3HPCST-2020, Ghaziabad, India, January 9-11, 2020

Birkhäuser

Trends in Mathematics

Trends in Mathematics is a series devoted to the publication of volumes arising from conferences and lecture series focusing on a particular topic from any area of mathematics. Its aim is to make current developments available to the community as rapidly as possible without compromise to quality and to archive these for reference.

Proposals for volumes can be submitted using the Online Book Project Submission Form at our website www.birkhauser-science.com.

Material submitted for publication must be screened and prepared as follows:

All contributions should undergo a reviewing process similar to that carried out by journals and be checked for correct use of language which, as a rule, is English. Articles without proofs, or which do not contain any significantly new results, should be rejected. High quality survey papers, however, are welcome.

We expect the organizers to deliver manuscripts in a form that is essentially ready for direct reproduction. Any version of TEX is acceptable, but the entire collection of files must be in one particular dialect of TEX and unified according to simple instructions available from Birkhäuser.

Furthermore, in order to guarantee the timely appearance of the proceedings it is essential that the final version of the entire material be submitted no later than one year after the conference.

More information about this series at http://www.springer.com/series/4961

Vinai K. Singh • Yaroslav D. Sergeyev
Andreas Fischer
Editors

Recent Trends in Mathematical Modeling and High Performance Computing

M3HPCST-2020, Ghaziabad, India, January 9–11, 2020

Birkhäuser

Editors
Vinai K. Singh
Department of Applied Mathematics
Inderprastha Engineering College
Ghaziabad
UP, India

Yaroslav D. Sergeyev
DIMES
University of Calabria
Rende, Italy

Andreas Fischer
Faculty of Math and Natural Sciences
Technische Universität Dresden
Dresden
Sachsen, Germany

ISSN 2297-0215 ISSN 2297-024X (electronic)
Trends in Mathematics
ISBN 978-3-030-68280-4 ISBN 978-3-030-68281-1 (eBook)
https://doi.org/10.1007/978-3-030-68281-1

Mathematics Subject Classification: 37-XX, 68-XX, 65-XX, 41-XX

This book is published under the imprint Birkhäuser, www.birkhauser-science.com, by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The aim of this book is twofold. Firstly, it presents interesting problems from natural and engineering sciences and techniques to tackle them. Secondly, it demonstrates the deep and fruitful interaction between those problems, mathematical modeling and theory, computational methods as well as scientific and high performance computing.

The chapters we have selected do not only show that all of these areas indeed contribute to successfully solve a large variety of problems. Rather, they underline that success very often relies on strong links between problem understanding, mathematics, and computing. Without scientific and high performance computing the power of modern compute systems cannot be efficiently exploited. Vice versa, the design of new compute systems and architectures heavily benefits from the knowledge and ideas of knowledgeable users. Another important basis for the success of a solution technique is the link between the problem, its appropriate modeling, and the design of algorithms, taking into account available or potential computational capabilities. Of course, this process is closely connected with mathematical understanding and the development of deep mathematical insights.

The 32 chapters of the book are assigned to one of the following parts:

- Partial and Ordinary Differential Equations
- Optimization and Optimal Control
- High Performance and Scientific Computing
- Stochastic Models and Statistics

Many of the contributions are related to more than one of these chapters so that our assignment is just a first orientation. The focus of the contributions ranges from more theoretical topics—like the conceptualization of infinity, error bounds in complementarity, efficient domain decomposition, or the analysis of probability distributions—to various applied themes, among them high capacity wireless communication, the impact of a Tsunami, reactor safety, infectious disease modeling, or the improvement of human health. Let us also highlight some of the contributions from computational science: parallel simulations by means of linear hybrid automaton models for cyber-physical systems, challenges for high

performance computing caused by machine learning and artificial intelligence, the simulation of power saving concepts on circuits, or the parallelization of feature extraction techniques.

We are convinced that readers will find interesting and inspiring chapters and that the book will serve practitioners and researches as well as beginners and experts.

Our sincere thanks go to all authors who submitted a manuscript. Moreover, we greatly acknowledge the work of the reviewers. The editors would like to express their thanks to Springer for the support, especially for the valuable help by Mr. Raghupathy Kalyanaraman. We are thankful to our families and friends for their patience, support, and love during the preparation of this volume. Last but not least, we acknowledge the financial support by the All India Council of Technical Education (AICTE), New Delhi, and by the Dr. A. P. J. Abdul Kalam Technical University in Lucknow under the Visvesvaraya Research Promotion Scheme (VRPS).

Ghaziabad, Uttar Pradesh, India — Vinai Kumar Singh

Rende, Italy — Yaroslav D. Sergeyev

Dresden, Sachsen, Germany — Andreas Fischer

September, 2020

Acknowledgments

This work is financially supported by the All India Council of Technical Education (AICTE), New Delhi, and Dr. A. P. J. Abdul Kalam Technical University, Lucknow, under the Visvesvaraya Research Promotion Scheme (VRPS).

Contents

Contributors

Syed Fahad Anwer Department of Mechanical Engineering, ZHCET, Aligarh Muslim University, Aligarh, India

Shaila S. Benal Department of Mathematics, B.L.D.E.A's V. P. Dr. P. G. Halakatti College of Engineering and Technology, Vijayapur, Karnataka, India

Sumit Kaur Bhatia Department of Mathematics, Amity Institute of Applied Science, Amity University, Noida, UP, India

Mahadev Biradar Department of Mathematics, Basaveshvara Engineering College, Bagalkot, Karnataka, India

K. L. Bondar P. G. Department of Mathematics, Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India

S. P. A. Bordas Faculté des Sciences, de la Technologie et de la Communication, University of Luxembourg, Luxembourg, LuxembourgInstitute of Research and Development, Duy Tan University, Da Nang, VietnamSchool of Engineering, Cardiff University, Wales, UK

Simrat Chaudhary Department of Mathematics, Amity Institute of Applied Science, Amity University, Noida, UP, India

Sudipa Chauhan Department of Mathematics, Amity Institute of Applied Science, Amity University, Noida, UP, India

P. U. Chopade Department of Mathematics, D.S.M.'s Arts, Commerce and Science College, Jintur, Maharashtra, India

G. C. Done P. G. Department of Mathematics, N.E.S. Science College, Nanded, Maharashtra , India

S. M. Dsouza Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai, India

Ali Farajzadeh Department of Mathematics, Razi University, Kermanshah, Iran

Abdul Hannan Faruqi Department of Mechanical Engineering, ZHCET, Aligarh Muslim University, Aligarh, India

Andreas Fischer Faculty of Math and Natural Sciences, Technische Universität Dresden, Dresden, Sachsen, Germany

Souvik Ganguli Department of Electrical and Instrumentation Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

Tanweer Jalal National Institute of Technology, Srinagar, India

Mario Jelitte Faculty of Mathematics, Institute of Numerical Mathematics, Technische Universität Dresden, Dresden, Germany

Seifedine Kadry Department of Mathematics and Computer Science, Faculty of Science, Beirut Arab University, Beirut, Lebanon

Sonia Kalra Department of Statistics, Punjabi University, Patiala, Punjab, India

Kaushik Preeti Inderprastha Engineering College, Ghaziabad, India

T. Khajah Department of Mechanical Engineering, University of Texas at Tyler, Tyler, TX, USA

N. Gopalakrishna Kini Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India

Rohit Kumar Department of Physics, Panjab University, Punjab, Chandigarh, India

Surendra Kumar Faculty of Mathematical Sciences, Department of Mathematics, University of Delhi, Delhi, India

Vinod Kumar SRMIST - NCR Campus, Modinagar, UP, India

Ishwar Maharudrappa Department of Mathematics, Basaveshwar Engineering College, Bagalkot, Karnataka, India

Geetika Malik Amity Institute of Applied Sciences, Amity University, Noida, India

Ishfaq Ahmad Malik National Institute of Technology, Srinagar, India

Bankim C. Mandal School of Basic Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar, India

Jitendra Kumar Maurya Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi, India

Shashi Kant Mishra Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi, India

Mishra Upendra Amity University Rajasthan, Kant Kalwar, Jaipur, India

R. A. Muneshwar P. G. Department of Mathematics, N.E.S, Science College, Nanded, Maharashtra, India

S. Natarajan Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai, India

Noura Obeid Department of Mathematics and Computer Science, Faculty of Science, Beirut Arab University, Beirut, Lebanon

S. P. Pandey RBS Engineering Technical Campus, Bichpuri, Agra, India

Ishita Puri Department of Information Technology, UIET, Panjab University, Chandigarh, India

Rajeev K. Puri Department of Physics, Panjab University, Punjab, Chandigarh, India

David M. Ramsey Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wrocław, Poland

Sahar Ranjbar Department of Mathematics, Razi University, Kermanshah, Iran

B. Ashwath Rao Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India

M. Mohideen Abdul Razak Bhabha Atomic Research Centre Facilities (BARC-Facilities), Kalpakkam, Tamil Nadu, India

Michael M. Resch High Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Stuttgart, Germany

Abdus Samad Department of Ocean Engineering, IIT Madras, Chennai, Tamil Nadu, India

Neha Sam Department of Mathematics, Jesus and Mary College, University of Delhi, New Delhi, India

G. C. Sankad Department of Mathematics, Research Center Affiliated to Visvesvaraya Technological University, Belagavi, BLDEA's Vachana Pitamaha Dr. P. G. Halakatti College of Engg. and Tech., Vijayapur, India

Avanish Shahi Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi, India

Krishna Kumar Sharma School of Vocational Studies and Applied Sciences, Gautam Buddha University, Greater Noida, UP, India

M. L. Sharma Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India

M. Hamid Siddique Department of Mechanical Engineering, ADCET, Ashta, Maharashtra, India

A. K. Singh RBS Engineering Technical Campus, Bichpuri, Agra, India

Sanjeev Kumar Singh Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi, India

Vinai Kumar Singh Department of Applied Mathematics, Inderprastha Engineering College, Ghaziabad, UP, India

Neelam Singla Department of Statistics, Punjabi University, Patiala, Punjab, India

Sucheta Department of Physics, Panjab University, Panjab, Chandigarh, India

Jagadish V. Tawade Department of Mathematics, Bheemanna Khandre Institute of Technology, Bhalki, India

Krishna C. Tripathi Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India

Shweta Upadhyaya Amity Institute of Applied Sciences, Amity University, Noida, UP, India

Vidhi Vashishth Department of Mathematics, Jesus and Mary College, University of Delhi, New Delhi, Delhi, India

H. C. Vidya Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India
National Institute of Technology, Puducherry, India

M. Vinothkumar SRMIST – NCR Campus, Modinagar, UP, India

Yadav Shobha Department of Mathematics, University of Delhi, New Delhi, Delhi, India

Yukti Department of Mathematics, Jesus and Mary College, University of Delhi, New Delhi, Delhi, India

Part I
Partial and Ordinary Differential Equations

Some Significant Results on the Union Graph Derived from Topological Space

R. A. Muneshwar and K. L. Bondar

Abstract In the recent paper, authors introduced a concept of union graph $\mho(\tau)$ of a topological space (X, τ) and discussed some basic properties such as connectedness, diameter, and girth of union graph. They proved that $\omega(\mho(\tau))$ and $\chi(\mho(\tau))$ of union graph of (X, τ) are equal, if $|X| > 3$. Moreover, we discussed some results on domination number, independence number, degree, etc., of a union graph. In this paper, we discuss some important properties of $\mho(\tau)$. It is shown that if τ be any topology defined on X with $|X| = 3$, then the union graph $\mho(\tau)$ is the connected graph if and only if topology τ is discrete topology or $\tau = \{\phi, X, U, V = U^c\}$. Moreover, we show that if τ be any topology other than discrete topology defined on X with $|X| = n$ and if K_n or P_{n+1} is a subgraph of a simple connected graph G having at most $(2^n - 3)$ vertices, then G is not a union graph of (X, τ).

1 Introduction

Graph theory has wide range of applications in different fields. Beck [1] introduced a concept of zero divisor graph of ring. In the recent decades, graph of several algebraic structures are defined. Among these graphs, zero divisor graphs of ring and module are more attractive for many researchers because of their application in several areas such as Electrical Engineering, Computer Science, etc. Angsuman Das introduced some graphs of vector space that can be found in [2–4] and derived some properties of these graphs. Some authors discussed the graph of a vector space that can be found in [5, 9]. Some properties on incomparability graphs of lattices were discussed by Wasadikar, M. and Survase P. [10, 11]. We [6–8] also introduced some

R. A. Muneshwar (✉)
P. G. Department of Mathematics, N.E.S. Science College, Nanded, Maharashtra, India

K. L. Bondar
P. G. Department of Mathematics, Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_1

new concepts of graphs of (X, τ)] and discussed some important properties of these graphs.

Definition 1 (Union Graph of Topological Space [8]) Let τ be a topology defined on a finite set X, and then, a graph $\mho(\tau) = (V, E)$ is called as a union graph of a (X, τ), where $V = \{U \in \tau \mid U \neq \phi, U \neq X\}$ and $E = \{(U_1, U_2) \mid U_1 \cup U_2 = X, \forall U_1, U_2 \in V\}$.

Note If two vertices U_1, U_2 are adjacent in $\mho(\tau)$, then it is denoted by $U_1 \sim U_2$ or (U_1, U_2).

As the graph has a wide range of applications in various fields, this motivates to define a union graph of topological space and make it applicable into various fields. An attempt has been made to convert topology into graph and try to study various properties of topology by using graph theory. For undefined concepts and terms, the reader can be referred to [12].

2 Connectedness of Union Graph

Now we discuss some results on connectedness of a union graph of (X, τ). Throughout this section, even if it is not mentioned explicitly, we assume that the underlying graph G is simple connected graph.

Theorem 1 *Let τ be any topology other than discrete topology defined on X with $|X| = 3$. If K_3 is a subgraph of a graph G with at most 6 vertices, then G is not union graph of (X, τ).*

Proof Suppose G is a union graph of (X, τ) and K_3 is a subgraph of G with vertex set $V(K_3) = \{A, B, C\}$ of K_3, as shown in Fig. 1.

From graph, we have $A \sim B$, $A \sim C$, and $B \sim C$, and hence, $A \cup B = X$, $A \cup C = X$, and $B \cup C = X$.

Case (i) If $|A| = |B| = |C| = 2$, then clearly $A \cap B$, $A \cap C$, and $B \cap C$ are the open subsets of X of cardinality 1. Therefore, τ is a discrete topology, a contradiction.

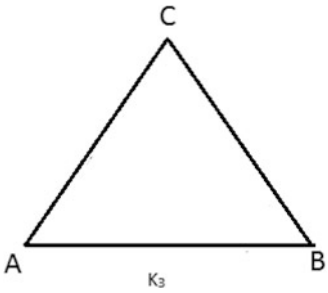


Fig. 1 Subgraph K_3 of graph G

Case (ii) If $|A| = |B| = 2, |C| = 1$, then $C \subset A$ or $C \subset B$, i.e., $C \not\subset A$ or $C \not\subset B$, a contradiction.

Case (iii) If $|A| = |B| = 1, |C| = 2$, then $A \subset C$ or $B \subset C$, i.e., $A \not\subset C$ or $B \not\subset C$, a contradiction.

Thus, if K_3 is a subgraph of G, then G is not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 5$. □

Theorem 2 *Let τ be any topology other than discrete topology defined on X with $|X| = 4$. If K_4 is a subgraph of a graph G with at most 12 vertices, then G is not a union graph of (X, τ).*

Proof Let τ be any topology defined on X with $X = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. Assume that G is a union graph of (X, τ) with at most 12 vertices and K_4 is a subgraph of G, with $V(K_4) = \{A, B, C, D\}$ is vertex set of K_4, as shown in Fig. 2.

From simple set theory and by Theorem 5.1[8], we have $A = \{\alpha_1, \alpha_2, \alpha_3\}$, $B = \{\alpha_1, \alpha_2, \alpha_4\}$, $C = \{\alpha_1, \alpha_3, \alpha_4\}$, and $D = \{\alpha_2, \alpha_3, \alpha_4\}$. Then, $A \cap B \cap C$, $A \cap B \cap D$, $A \cap C \cap D$, and $B \cap C \cap D$ are the 1-element open sets in X, a contradiction, since τ is other than discrete topology. Thus, if G is not a union graph of (X, τ) with at most 12 elements whose subgraph is K_4. □

Theorem 3 *Let τ be any topology other than discrete topology defined on X with $|X| = n$. If K_n is a subgraph of a graph G with at most $(2^n - 3)$ vertices, then $\{\alpha_1\}, \{\alpha_2\}, ..., \{\alpha_n\} \in \tau$.*

Proof Let τ be any topology other than discrete topology defined on X with $X = \{\alpha_1, \alpha_2, ..., \alpha_{n-1}, \alpha_n\}$. Assume that G is a union graph of (X, τ) with at most $2^n - 3$ vertices and K_n is a subgraph of G, and let $V = \{A_1, A_2, ..., A_n\}$ be a vertex set of K_n. Then by Theorem 5.1 [8], vertices $A_1, A_2, A_{(n-1)}, A_n$ of K_n are of cardinality $(n - 1)$. As $A_1, A_2, ..., A_{n-1}, A_n$ are open in (X, τ), then the finite intersection $B_k = \cap_{1, i \neq k}^{n} A_i = \{\alpha_k\}$ is open in X, $\forall\ k = 1, 2, ..., n$. Thus, all singleton subsets are open in (X, τ). □

Theorem 4 *Let τ be any topology other than the discrete topology defined on X with $|X| = 3$. If K_n is a subgraph of graph G with at most $2^n - 3$ vertices, then $\{\alpha_1\}, \{\alpha_2\}, ..., \{\alpha_n\} \in \tau$.*

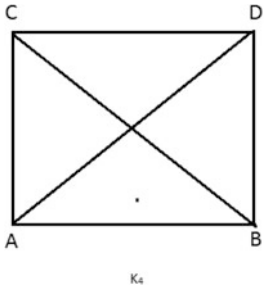


Fig. 2 Subgraph K_4 of graph G

Proof Let τ be any topology other than discrete topology defined on X with $X = \{\alpha_1, \alpha_2, ..., \alpha_{n-1}, \alpha_n\}$.

Step I: Let there is a topological space (X, τ), where τ is other than discrete topology with $|X| = 3$ whose graph is G. Suppose K_3 is a subgraph of G. By Theorem 5.1 [8], open sets A_1, A_2, A_3 of cardinality 2 are the vertices of K_3. Then clearly, $B_k = \cap_{1,i\neq k}^{3} A_i = \{\alpha_k\}$ are open in X, $\forall$ $k = 1, 2, 3$, a contradiction, since τ is other than the discrete topology.

Step II: Assume that the result is true for $m = n - 1$, i.e., if $K_{(n-1)}$ is a subgraph of G, then $\{\alpha_1\}$, $\{\alpha_2\}$, ..., $\{\alpha_{(n-1)}\}$ are open in (X, τ).

Let τ be any topology other than discrete topology defined on X with $|X| = n$ and K_n is a subgraph of G.

Claim: The subsets $\{\alpha_k\}$, $\forall\, k = 1, 2, ..., n$, are open in (X, τ).

If A_1,A_2, . . . , A_n are the vertices of K_n, then by Theorem 5.1 [8], we have $A_k = \{\alpha_i\}_{i=1,i\neq k}^{n}$, $\forall\, k = 1, 2, ..., n$, are the vertices of K_n. Suppose $Y = \{\alpha_i\}_{i=1}^{n-1} \in \tau$. As $A_1, A_2, ..., A_{(n-1)}, A_n \in \tau$ then $B_k = Y \cap A_k$, $\forall k = 1, 2, ..., (n-1)$, are the open subsets of Y of cardinality $(n-2)$.
Let $Z = \{B_k = Y \cap A_k \mid \forall k = 1, 2, ..., (n-1)\}$ be a collection of open subsets of Y of cardinality $(n-2)$, and then by Theorem 5.1[8], Z forms a clique in G_y of size $(n-1)$, where G_y is a graph of subspace Y that is a subgraph of the graph of $\mho(\tau)$. Then by mathematical induction, we have that $\{\alpha_1\}$, $\{\alpha_2\}$, ..., $\{\alpha_{(n-1)}\}$ are the open subsets of Y, and hence, they are open in X. Let $A = \cap_{k=1}^{n} A_k = \{\alpha_n\} \in \tau$. Thus, all singleton subsets $\{\alpha_1\}$, $\{\alpha_2\}$, ..., $\{\alpha_{(n-1)}\}$, $\{\alpha_n\}$ are open in (X, τ), a contradiction. Thus by mathematical induction, the result holds for all n. □

Theorem 5 *Let τ be any topology other than discrete topology defined on X with $|X| = n$. If K_n is a subgraph of graph G with at most $2^n - 3$ vertices, then G is not a union graph of (X, τ).*

Proof Let τ be any topology other than discrete topology defined on X with $X = \{\alpha_1, \alpha_2, ..., \alpha_{n-1}, \alpha_n\}$. Assume that, if K_n is a subgraph of G with at most $2^n - 3$ vertices. Then by Theorems 3 and 4, $\{\alpha_1\}$, $\{\alpha_2\}$, ..., $\{\alpha_n\} \in \tau$, a contradiction. Thus, if K_n is a subgraph of G, then G is not a union graph of topology τ on X. □

Theorem 6 *Let τ be any topology other than discrete topology defined on X with $|X| = 3$. If G is a graph with at most 5 vertices and P_3 is a subgraph of G, then G is not a union graph of (X, τ).*

Proof Let τ be any topology other than discrete topology defined on X. Assume that P_3 is a subgraph of G with $V(P_3) = \{A, B, C\}$ be the vertex set of P_3, as shown in Fig. 3.

From graph, we have $A \sim B$, $B \sim C$, and hence, $A \cup B = X$ and $B \cup C = X$.

Case (i) If $|A| = |B| = |C| = 2$, then $A \cup B = X$, $B \cup C = X$, and $A \cup C = X$, and hence, $A \sim B$, $A \sim C$, and $B \sim C$ a contradiction.

Case (ii) If $|A| = |B| = 2$ and $|C| = 1$, then $A \cup B = X$ and $A \cap B \neq \phi$. Let us denote $C = A \cap B$, and then $C \not\sim A$ and $C \not\sim B$, a contradiction.

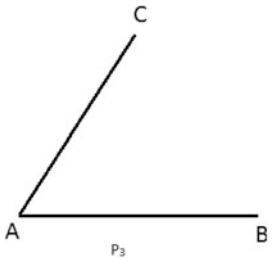


Fig. 3 The subgraph P_3 of graph G

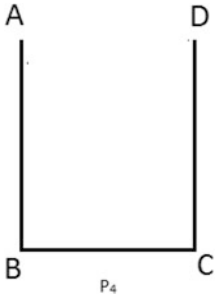


Fig. 4 The subgraph P_4 of graph G

Case (iii) If $|A| = |B| = 1$ and $|C| = 2$, then $A \subset C$ or $B \subset C$, and hence, $A \not\sim C$ or $B \not\sim C$ and $A \not\sim B$, a contradiction.

Thus, if P_3 is a subgraph of G, then G is not a union graph of (X, τ). □

Theorem 7 *Let τ be any topology other than discrete topology defined on X with $|X| = 3$. If P_4 is a subgraph of graph G with at most 5 vertices, then G is not a union graph of (X, τ).*

Proof Let τ be any topology other than discrete topology defined on X. Assume that P_4 is a subgraph G with the vertex set $\{A, B, C, D\}$ of P_4, as shown in Fig. 4. From graph, we have $A \not\sim C$ and $A \not\sim D$, and hence, $A \cup C \neq X$, $A \cup D \neq X$. Thus, $A \subset D$ or $D \subset A$ and $A \subset C$ or $C \subset A$.

Case (i) If $A \subset D$ and $A \subset C$, then $|A| = 1$, $|C| = |D| = 2$. Hence, $|C \cap D| = 1$ and $C \cap D = A$. From graph, we have $A \sim B$, and hence, $A \cup B = X$, $A \cap B = \phi$, and $|B| = 2$. As $|B| = |D| = 2$, then $B \sim D$, a contradiction, since $|X| = 3$. Thus, G is not a union graph of (X, τ).
Case (ii) If $A \subset D$ and $C \subset A$, then $C = \phi$, a contradiction.
Case (iii) If $D \subset A$ and $A \subset C$, then $D = \phi$, a contradiction.
Case (iv) If $D \subset A$ and $C \subset A$, then $C \cup D = A \neq X$, $C \not\sim D$, a contradiction.

Thus, if P_4 is a subgraph of G, then G is not a union graph of (X, τ). □

Theorem 8 *Let τ be any topology other than discrete topology defined on X with $|X| = 4$. If P_5 is a subgraph of graph G with at most 12 vertices, then G is not a union graph of (X, τ).*

Proof Let τ be any topology other than discrete topology defined on X. Assume that P_5 is a subgraph of G,, and let $V(P_5) = \{A, B, C, D, E\}$ be a vertex set of P_5, as shown in Fig. 5.

From graph, we have $A \not\sim C$, $A \not\sim D$, $A \not\sim E$, $B \not\sim D$, $B \not\sim E$, and $C \not\sim E$. Hence, $A \cup C \neq X$, $A \cup D \neq X$, $A \cup E \neq X$, $B \cup D \neq X$, $B \cup E \neq X$, and $C \cup E \neq X$.

Case I: If $|A| = 3$, as $A \not\sim C$, $A \not\sim D$, and $A \not\sim E$, then we have $C, D, E \subset A$, and $B = A^c$. Therefore, $C \cup D \subset A \neq X$ with $C \not\sim D$, a contradiction. Thus, $|A| \neq 3$. Similarly, we can prove that $|B|, |C|, |D|, |E| \neq 3$.

Case II: If $|A| = 2$, then $B = A^c$, since $A \sim B$. As $B \sim C$ and $|C| \neq 3$, then $C = A$, a contradiction. Hence, $|A| \neq 2$, and similarly, we can prove $|B|, |C|, |D|, |E| \neq 2$.

Also, if $|A| = 1$, then $|B| = 3$, a contradiction. Thus, if P_5 is a subgraph of G with at most 12 vertices, then G is not a union graph of (X, τ) with $|X| = 4$. □

Theorem 9 *Let τ be any topology other than discrete topology defined on X with $|X| = n$. If P_{n+1} is a subgraph of graph G with at most $2^n - 2$ vertices, then G is not a union graph of (X, τ).*

Proof Let τ be any topology other than discrete topology defined on X with $X = \{\alpha_1, \alpha_2, ..., \alpha_{n-1}, \alpha_n\}$. Assume that P_{n+1} is a subgraph of a G with vertex set $V(P_{n+1}) = \{A_1, A_2, ..., A_{n+1}\}$ of P_{n+1}, as shown in Fig. 6. From Fig. 6, we have

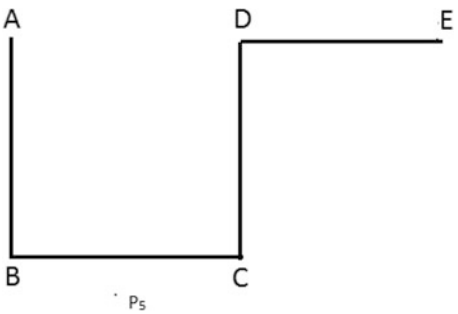


Fig. 5 The subgraph P_5 of graph G

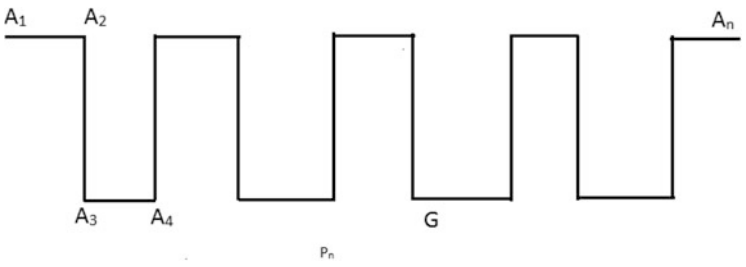


Fig. 6 The subgraph P_{n+1} of graph G

$A_1 \not\sim A_i$, for $i = 3, 4, ..., (n+1)$, $A_2 \not\sim A_j$, for $j = 4, 5, ..., (n+1)$, $A_{(n-1)} \not\sim A_{(n+1)}$. Hence, $A_1 \cup A_i \neq X$, for $i = 3, 4, ..., (n+1)$, $A_2 \cup A_j \neq X$, for $j = 3, 4, ..., (n+1)$, and $A_{(n-1)} \cup A_{(n+1)} \neq X$.

Case I: If $|A_1| = (n-1)$. As $A_1 \not\sim A_i$, for $i = 3, 4, ..., (n+1)$, then $A_i \subset A_1$, for $i = 3, 4, ..., (n+1)$. Therefore, $A_i \cup A_j \neq X$ for $i \neq j = 3, 4, ..., (n+1)$, and hence, $A_i \not\sim A_j$ for $i \neq j = 3, 4, ..., (n+1)$, a contradiction. Thus, $|A_1| \neq (n-1)$, similarly $|A_i| \neq (n-1)$, for $i = 2, 3, 4, ..., (n+1)$.

Case II: If $|A_1| = (n-2)$.

As $A_1 \not\sim A_i$, for $i = 3, 4, ..., (n+1)$, then $A_i \subset A_1$, for $i = 3, 4, ..., (n+1)$ or A_i is the singleton set. If $A_i \subset A_1$, for $i = 3, 4, ..., (n+1)$, or A_i is the singleton set, then $A_i \not\sim A_j$ for $i \neq j = 3, 4, ..., (n+1)$, a contradiction. Thus, $|A_i| \neq (n-2)$, similarly $|A_i| \neq (n-2)$, for $i = 2, 3, 4, ..., (n+1)$.

Case III: If $|A_1| = (n-k)$, for simplicity suppose $A_1 = \{a_1, a_2, ..., a_{(n-k)}\}$.

As $A_1 \sim A_2$, then $A_1^c \subset A_2$, and hence $\{a_{(n-k+1)}, a_{(n-k+2)}, ..., a_n\} \subset A_2$. Again as $A_2 \sim A_3$, then $A_2^c \subset A_3$ and $A_3^c \subset A_2$, that is, $\{a_1, a_2, ..., a_{(n-k)}\} \subset A_3$. Also, as $A_3 \sim A_4$, then $A_3^c \subset A_4$, that is, $\{a_{(n-k+1)}, a_{(n-k+2)}, ..., a_n\} \subset A_4$. As $A_1 \not\sim A_i$, for $i = 3, 4, ..., (n+1)$, then $A_1^c \not\subset A_i$, for $i = 3, 4, ..., (n+1)$, that is, $\{a_{(n-k+1)}, a_{(n-k+2)}, ..., a_n\} \not\subset A_i$, for $i = 3, 4, ..., (n+1)$, which is a contradiction. Thus, $|A_1| \neq (n-k)$, similarly $|A_i| \neq (n-k)$, for $i = 2, 3, 4, ..., (n+1)$.

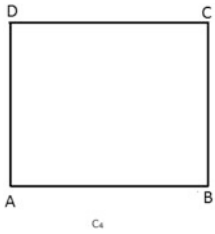
Thus, if $P_{(n+1)}$ is G with at most $2^n - 2$ vertices, then G is not a graph of (X, τ) with $|X| = n$. □

Theorem 10 *Let τ be any topology other than discrete topology defined on X with $|X| = 3$. If G is a graph with at most 5 vertices and C_4 is a subgraph of G, then G is not a union graph of (X, τ).*

Proof Suppose G is a union graph of (X, τ) and C_4 is a subgraph of G with $V(C_4) = \{A, B, C, D\}$ is the vertex set of C_4, as shown in Fig. 7.

From graph, we have $C \sim D$, $A \sim D$, $A \sim B$, and $B \sim C$. Therefore, $C \cup D = X$, $A \cup D = X$, $A \cup B = X$, $B \cup C = X$, and hence, $(A \cap C) \cup D = X$, $(A \cap C) \cup B = X$, $A \cup (B \cap D) = X$, $C \cup (B \cap D) = X$. This implies $|A \cap C| = 1$, $|D| = 2$, $|A \cap C| = 1$, $|B| = 2$, $|B \cap D| = 1$, $|A| = 2$, $|B \cap D| = 1$, $|C| = 2$, a contradiction. Thus, if C_4 is a subgraph of G with at most 5 vertices, then G is not a union graph of (X, τ), with $|X| = 3$. □

Fig. 7 The subgraph C_4 of G



Theorem 11 *If τ be any topology defined on X with $|X| = 3$, then $\mho(\tau)$ is a connected graph if and only if τ is the discrete topology or $\tau = \{\phi, X, U, V = U^c\}$.*

Proof Let τ be any topology other than discrete topology defined on X with $|X| = 3$ and $|\tau| = 6$. Suppose G is connected with vertex set $V(G) = \{A, B, C, D\}$, then possible simple connected graphs with four vertices are as shown in Figs. 8 and 9.

As K_3 is a subgraph of graphs G_3, G_4, G_5, G_7 and G_8, then by Theorem 1, they are not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 6$. Also, P_4 is a subgraph of G_1, then by Theorem 7, G_1 is not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 6$. As C_4 is a subgraph of G_2, then by Theorem 10, C_4 is not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 6$. Since P_3 is a subgraph of G_6, then by Theorem 6, G_6 is not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 6$.
Suppose $\mho(\tau)$ is connected with vertex set$V(\mho(\tau)) = \{A, B, C\}$, then possible connected graph with three vertices is as shown in Fig. 10.

As K_3 is a subgraph of graph G_9, then by Theorem 1, G_9 is not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 5$. Also, P_3 is a subgraph of G_{10}, then by Theorem 6, G_{10} is not a union graph of (X, τ) with $|X| = 3$ and $|\tau| = 5$.

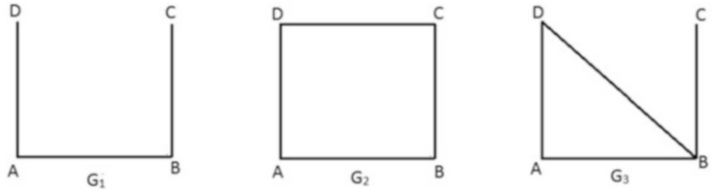


Fig. 8 Some connected graph with four vertices

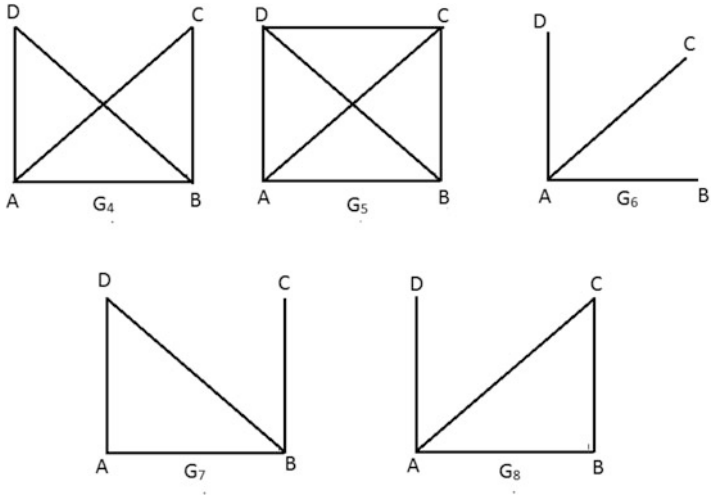


Fig. 9 Some connected graph with four vertices

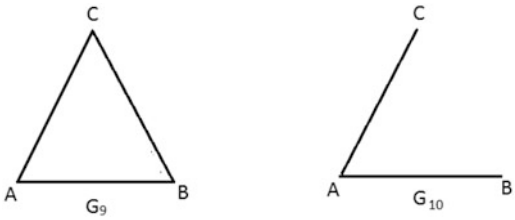


Fig. 10 Some connected graph with three vertices

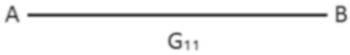


Fig. 11 Some connected graph with two vertices

Let τ be any topology defined on X with $|X| = 3$ and $|\tau| = 4$. Suppose G_{11} is connected with $V(G_{11}) = \{A, B\}$, then possible connected graph with two vertices is as shown in Fig. 11.

Suppose graph of a topology (X, τ) is $G_8 = P_2$ with vertex set $V(G) = \{A, B\}$. From graph, we have $A \sim B$; hence, $A \cup B = X$. Thus, $B = A^c$ (otherwise, $A \cap B \neq \phi$, a contradiction).

Conversely, suppose $\tau = \tau_d$ be the discrete topology, then by Theorem 3.8 [8], $\mho(\tau)$ is the connected graph. Also, if $\tau = \{\phi, X, U, V = U^c\}$, then clearly $U \cup V = X$, and hence, $U \sim V$. Thus, graph $\mho(\tau)$ is the connected graph with two vertices U and V. □

3 Conclusion

In this paper, we study connectedness and some important results of union graph of (X, τ). It is shown that if τ is any topology defined on X with $|X| = 3$ then the corresponding union graph $\mho(\tau)$ is connected if and only if τ is discrete topology or $\tau = \{\phi, X, U, V = U^c\}$. Moreover, we show that if P_{n+1} or K_n is a subgraph of a simple connected graph G with at most $(2^n - 2)$ vertices then G is not a union graph of (X, τ) with $|X| = n$ and τ is other than discrete topology. The main goal of this study is to discuss different properties of (X, τ) and $\mho(\tau)$. The present study also tries to establish relationship between $\mho(\tau)$ and (X, τ).

Acknowledgments The authors are thankful to Mr. Krishnath Masalkar and Dr. Pradnya Survase for fruitful discussions and their helpful suggestions in this work.

References

1. I. Beck, Coloring of commutative rings. J. Algebra **116**(1), 208–226 (1988)
2. A. Das, Non-zero component graph of a finite dimensional vector space. Commun. Algebra **44**(9), 3918–3926 (2016)

3. A. Das, Subspace inclusion graph of a vector space. Commun. Algebra **44**(11), 4724–4731 (2016)
4. A. Das, Non-zero component graph of a finite dimensional vector space. Commun. Algebra. Available at http://arxiv.org/abs/1506.04905
5. N. Jafari Rad, S.H. Jafari, Results on the intersection graphs of subspaces of a vector space. http://arxiv.org/abs/1105.0803v1
6. R.A. Muneshwar, K.L. Bondar, Open subset inclusion graph of a topological space. J. Discrete Math. Sci. Cryptography **22**(6), 1007–1018 (2019)
7. R.A. Muneshwar, K.L. Bondar, Some significant properties of the intersection graph derived from topological space using intersection of open sets. Far East J. Math. Sci. (FJMS) **11991**, 29–48 (2019)
8. R.A. Muneshwar, K.L. Bondar, Some properties of the union graph derived from topological space using union of open sets. Far East J. Math. Sci. (FJMS) **121**(2), 101–121 (2019)
9. Y. Talebi, M.S. Esmaeilifar, S. Azizpour, A kind of intersection graph of vector space. J. Discrete Math. Sci. Cryptography **12**(6), 681–689 (2009)
10. M. Wasadikar, P. Survase, Incomparability graphs of lattices, in *ICMMSC 2012*, ed. by P. Balasubramaniam, R. Uthayakumar. CCIS, vol. 283 (Springer, Heidelberg, 2012), pp. 78–85
11. M. Wasadikar, P. Survase, Incomparability graphs of lattices II, in *IWOCA 2012*, ed. by S. Arumugam, B. Smyth. Lecture Notes in Computer Science, vol. 7643 (Springer, Berlin, Heidelberg, 2012), pp. 148–161
12. D.B. West, *Introduction to Graph Theory* (Prentice Hall, 2001)

Existence and Uniqueness Results of Second Order Summation–Difference Equations in Cone Metric Space

G. C. Done, K. L. Bondar, and P. U. Chopade

Abstract In this paper, we investigate the existence and uniqueness results for summation–difference type equations in cone metric spaces. The results are obtained by using some extensions of Banach's contraction principle in a complete cone metric space.

1 Introduction

The study of difference equations is found to be more useful in the field of numerical, engineering, as well as social sciences. Agrawal [2] and Kelley and Peterson [13] had developed a theory of difference equation and their inequalities. Later K. L. Bondar et al. [4–7] studied existence, uniqueness, and comparison results for some difference equations and summation equations. G. C. Done, K. L. Bondar, and P. U. Chopade investigated the existence and uniqueness results for summation–difference type equations and nonhomogeneous first order nonlinear difference equation with nonlocal condition in cone metric spaces, which can be found in [8, 9].

The aim of this paper is to study the existence and uniqueness of solutions for the summation–difference equations of second order and the existence of unique common solution of the summation equations.

In Section 3, we consider the following summation–difference equation of second order of the form:

G. C. Done (✉)
P. G. Department of Mathematics, N.E.S. Science College, Nanded, Maharashtra, India

K. L. Bondar
P. G. Department of Mathematics, Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India

P. U. Chopade
Department of Mathematics, D. S. M.'s Arts, Commerce and Science College, Jintur, Maharashtra, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_2

$$\Delta^2 x(t-1) = Ax(t) + \sum_{s=0}^{t-1} k(t,s,x(s)) + \sum_{s=0}^{b-1} h(t,s,x(s)), \quad t \in J = [0,b], \tag{1}$$

$$x(0) = x_0, \quad \Delta x(0) = y_0, \tag{2}$$

where A is the infinitesimal generator of a strongly continuous cosine family $\{C(t) : t \in \mathbb{R}\}$ in a Banach space X, the functions $k, h : J \times J \times X \to Z$ are continuous, and the given x_0, y_0 are the elements of X.

In Sect. 4, we study the existence of unique common solution of the summation equations of the following type:

$$x(t) = \sum_{s=a}^{t-1} k_1(t,s,x(s)) + \sum_{s=a}^{b-1} h_1(t,s,x(s)) + g_1(t), \qquad t \in [a,b], \tag{3}$$

$$x(t) = \sum_{s=a}^{t-1} k_2(t,s,x(s)) + \sum_{s=a}^{b-1} h_2(t,s,x(s)) + g_2(t), \qquad t \in [a,b], \tag{4}$$

where $x, g_1, g_2 : [a,b] \to X$; the functions $k_i, h_i : [a,b] \times [a,b] \times X \to X (i = 1, 2)$ are the continuous functions.

Finally in Sect. 5, we give example to illustrate the application of our results.

2 Preliminaries

Let us recall the concepts of the cone metric space, and we refer the reader to [1, 10, 11, 14, 16] for the more details.

Definition 1 Let E be a real Banach space and P is a subset of E. Then, P is called a cone if and only if:

1. P is closed, nonempty, and $P \neq 0$;
2. $a, b \in \mathbb{R}, a, b \geq 0, x, y \in P \Rightarrow ax + by \in P$;
3. $x \in P$ and $-x \in P \Rightarrow x = 0$.

For a given cone $P \in E$, we define a partial ordering relation $\leq$ with respect to P by $x \leq y$ if and only if $y - x \in P$. We shall write $x < y$ to indicate that $x \leq y$ but $x \neq y$, while $x \ll y$ will stand for $y - x \in int\, P$, where int P denotes the interior of P. The cone P is called normal if there is a number $K > 0$ such that $\leq x \leq y$ implies $\|x\| \leq k\|y\|$, for every $x, y \in E$. The least positive number satisfying above is called the normal constant of P.

In the following, we always suppose E is a real Banach space, P is cone in E with int $P \neq \phi$, and $\leq$ is partial ordering with respect to P.

Definition 2 ([10]) Let X be a nonempty set. Suppose that the mapping $d : X \times X \to E$ satisfies

(d_1) $0 \leq d(x, y)$ for all $x, y \in X$ and $d(x, y) = 0$ if and only if $x = y$;
(d_2) $d(x, y) = d(y, x)$, for all $x, y \in X$;
(d_3) $d(x, y) \leq d(x, z) + d(z, y)$, for all $x, y \in X$.

Then, d is called a cone metric on X and (X, d) is called a cone metric space.

Example 1 ([10]) Let $E = \mathbb{R}^2$, $P = \{(x, y) \in E : x, y \geq 0\}$, $x = \mathbb{R}$, and $d : X \times X \to E$ such that $d(x, y) = (|x - y|, \alpha|x - y|)$, where $\alpha \geq 0$ is a constant and then (X, d) is a cone metric space.

Definition 3 Let X be an ordered space. A function $\Phi : X \to X$ is said to a comparison function if every $x, y \in X, x \leq y$ implies that $\Phi(x) \leq \Phi(y)$, $\Phi(x) \leq x$ and $\lim_{n\to\infty} \|\Phi^n(x)\| = 0$, for every $x \in X$.

Example 2 Let $E = \mathbb{R}^2$, $p = \{(x, y) \in E : x, y \geq 0\}$, and it is easy to check that $\Phi : E \to E$ with $\Phi(x, y) = (ax, ay)$, for some $a \in (0, 1)$, is a comparison function. Also if Φ_1, Φ_2 are the two comparison functions over $\mathbb{R}$, then $\Phi(x, y) = (\Phi_1(x), \Phi_2(y))$ is also a comparison function over E.

3 Existence and Uniqueness of Solutions

Let X be a Banach space with norm$\|.\|$. Let $B = C(J, X)$ Banach space of all continuous functions from J into X endowed with supremum norm

$$\|x\|_\infty = sup\{\|x(t)\| : t \in J\}.$$

Let $P = (x, y) : x, y \geq 0 \subset E = \mathbb{R}^2$, and define

$$d(f, g) = (\|f - g\|_\infty, \alpha\|f - g\|_\infty)$$

for every $f, g \in B$, and then it is easily seen that (B, d) is a cone metric space.

In many cases, it is advantageous to treat second ordered difference equations directly rather than to convert first order systems. We can study second order equations in the theory of the strongly continuous cosine family. If $\{C(t) : t \in \mathbb{R}\}$ is a strongly continuous cosine family in X, then $\{S(t) : t \in \mathbb{R}\}$ associated to the given strongly continuous cosine family is defined by

$$S(t)x = \sum_{s=0}^{t-1} C(s)x, \qquad x \in X, \qquad t \in \mathbb{R}.$$

The infinitesimal generator $A : X \to X$ of a cosine family $\{C(t) : t \in \mathbb{R}\}$ is defined by

$$Ax = \Delta^2 C(t)x|_{v=0}, \qquad x \in D(A),$$

where $D(A) = \{x \in X : C(.)x \in C^2(\mathbb{R}, X)\}$. Let $M \geq 1$ and N be the two positive constants such that $\|C(t)\| \leq M$ and $\|S(t)\| \leq N$ for all $t \in J$.

Definition 4 The function $x \in B$ that satisfies the summation equation

$$x(t) = C(t)x_0 + S(t)y_0 + \sum_{s=0}^{t-1} S(t-s)\Big[\sum_{\tau=0}^{s-1} k(s, \tau, x(\tau)) + \sum_{\tau=0}^{b-1} h(s, \tau, x(\tau))\Big], \quad t \in J$$

is called the mild solution of the initial value problem (1)–(2).

We need the following Lemma for further discussion.

Lemma 1 ([15]) *Let (X, d) be a complete cone metric space, where P is a normal cone with normal constant K. Let $f : X \to X$ be a function such that there exists a comparison function $\Phi : P \to P$ such that*

$$d(f(x), f(y)) \leq \Phi(d(x, y))$$

for very $x, y \in X$. Then, f has a unique fixed point.

We list the following hypotheses for our convenience:

(H1): There exist continuous functions $p_1, p_2 : J \times J \to \mathbb{R}^+$ and a comparison function $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$(\|k(t, s, u) - k(t, s, v)\|, \alpha\|k(t, s, u) - k(t, s, v)\|) \leq p_1(t, s)\Phi(d(u, v)),$$

and

$$(\|h(t, s, u) - h(t, s, v)\|, \alpha\|h(t, s, u) - h(t, s, v)\|) \leq p_2(t, s)\Phi(d(u, v)),$$

for every $t, s \in J$ and $u, v \in Z$.

(H2):

$$N \sum_{t=0}^{b-1} \sum_{s=0}^{b-1} [p_1(t, s) + p_2(t, s)] \leq 1.$$

Theorem 1 *Assume that hypotheses $(H1) - (H2)$ hold. Then, the initial value problem (1)–(2) has a unique solution x on J.*

Proof The operator $F : B \to B$ is defined by

$$Fx(t) = C(t)x_0 + S(t)y_0 + \sum_{s=0}^{t-1} S(t-s)[\sum_{\tau=0}^{s-1} k(s,\tau,x(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,x(\tau))]. \tag{5}$$

By using the hypothesis $(H1)-(H2)$, we have

$$\begin{aligned}
&(\|Fx(t)-Fy(t)\|, \alpha\|Fx(t)-Fy(t)\|) \\
&= \Big(\Big\| \sum_{s=0}^{t-1} S(t-s)[\sum_{\tau=0}^{s-1} k(s,\tau,x(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,x(\tau))] \\
&\qquad - \sum_{s=0}^{t-1} S(t-s)[\sum_{\tau=0}^{s-1} k(s,\tau,y(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,y(\tau))] \Big\|, \\
&\quad \alpha \Big\| \sum_{s=0}^{t-1} S(t-s)[\sum_{\tau=0}^{s-1} k(s,\tau,x(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,x(\tau))] \\
&\qquad - \sum_{s=0}^{t-1} S(t-s)[\sum_{\tau=0}^{s-1} k(s,\tau,y(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,y(\tau))] \Big\| \Big) \\
&\le \sum_{s=0}^{t-1} N\Big(\Big\| \sum_{\tau=0}^{s-1} k(s,\tau,x(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,x(\tau)) - \sum_{\tau=0}^{s-1} k(s,\tau,y(\tau)) - \sum_{\tau=0}^{b-1} h(s,\tau,y(\tau)) \Big\|, \\
&\qquad \alpha \Big\| \sum_{\tau=0}^{s-1} k(s,\tau,x(\tau)) + \sum_{\tau=0}^{b-1} h(s,\tau,x(\tau)) - \sum_{\tau=0}^{s-1} k(s,\tau,y(\tau)) - \sum_{\tau=0}^{b-1} h(s,\tau,y(\tau)) \Big\| \Big) \\
&\le \sum_{s=0}^{t-1} N\Big[\Big(\sum_{\tau=0}^{s-1} \|k(s,\tau,x(\tau)) - k(s,\tau,y(\tau))\|, \alpha \sum_{\tau=0}^{s-1} \|k(s,\tau,x(\tau)) - k(s,\tau,y(\tau))\| \Big) \\
&\qquad + \Big(\sum_{\tau=0}^{b-1} \|h(s,\tau,x(\tau)) - h(s,\tau,y(\tau))\|, \alpha \sum_{\tau=0}^{b-1} \|h(s,\tau,x(\tau)) - h(s,\tau,y(\tau))\| \Big) \Big] \\
&\le \sum_{s=0}^{t-1} N\Big[\sum_{\tau=0}^{s-1} P_1(t,s)\Phi(\|x-y\|_\infty, \alpha\|x-y\|_\infty + \sum_{\tau=0}^{b-1} P_2(t,s)\Phi(\|x-y\|_\infty, \alpha\|x-y\|_\infty \Big] \\
&\le \sum_{s=0}^{b-1} N\Big[\sum_{\tau=0}^{b-1} P_1(t,s)\Phi(\|x-y\|_\infty, \alpha\|x-y\|_\infty + \sum_{\tau=0}^{b-1} P_2(t,s)\Phi(\|x-y\|_\infty, \alpha\|x-y\|_\infty \Big]
\end{aligned}$$

$$\leq \Phi(\|x-y\|_\infty, \alpha\|x-y\|_\infty) N \sum_{s=0}^{b-1}\sum_{\tau=0}^{b-1}[P_1(t,s)+P_2(t,s)]$$

$$\leq \Phi(\|x-y\|_\infty, \alpha\|x-y\|_\infty) \tag{6}$$

for every $x, y \in B$. This implies that $d(Fx, Fy) \leq \Phi(d(x, y))$, for every $x, y \in B$. Now an application of Lemma 1 gives that the operator F has a unique point in B. This means that Eqs. (1)–(2) have a unique solution. □

4 Existence of Common Solutions

Let X be a Banach space with norm$\|.\|$. Let $Z = C([a, b], X)$ be a Banach space of all continuous functions from J into X endowed with supremum no

$$\|x\|_\infty = sup\{\|x(t)\| : t \in [a, b]\}.$$

Let $P = (x, y) : x, y \geq 0 \subset E = \mathbb{R}^2$ be a cone and define $d(f, g) = (\|f - g\|_\infty, \alpha\|f - g\|_\infty)$ for every $f, g \in Z$, and then it is easily seen that (Z, d) is a cone metric space.

Definition 5 ([12]) A pair (S, T) of self-mappings X is said to be weakly compatible if they commute at their coincidence point (i.e., $STx = TSx$ whenever $Sx = Tx$). A point $y \in X$ is called point of coincidence of a family T_j, $j = 1, 2, \ldots$ of self-mappings on X if there exists a point $x \in X$ such that $y = T_j x$ for all $j = 1, 2, \ldots$.

Lemma 2 ([3]) *Let (X, d) be a complete cone metric space and P be an order cone. Let $S, T, f : X \to X$ be such that $S(X) \cup T(X) \subset f(X)$. Assume that the following conditions hold:*

(i) $d(Sx, Ty) \leq \alpha d(fx, Sx) + \beta d(fy, Ty) + \gamma d(fx, fy)$, *for all $x, y \in X$, with $x \neq y$, where α, β, γ are the non-negative real numbers with $\alpha + \beta + \gamma < 1$.*

(ii) $d(Sx, Tx) < d(fx, Sx) + d(fx, Tx)$, *for all $x \in X$, whenever $Sx \neq Tx$. If $f(X)$ or $S(X) \cup T(X)$ is a complete subspace of X, then S, T, and f have a unique point of coincidence. Moreover, if (S, f) and (T, f) are weakly compatible, then S, T, and f have a unique common fixed point.*

We list the following hypotheses for our convenience:

(H3): Assume that

$$(F)x(t) = \sum_{s=a}^{t-1} k_1(t, s, x(s)) + \sum_{s=a}^{b-1} h_1(t, s, x(s))$$

and

$$(G)x(t) = \sum_{s=a}^{t-1} k_2(t, s, x(s)) + \sum_{s=a}^{b-1} h_2(t, s, x(s))$$

for all $t, s \in [a, b]$.

(H4): There exist $\alpha, \beta, \gamma, p \geq 0$ such that

$$\begin{aligned}(|Fx(t) - Gy(t) + g_1(t) - g_2(t)|, \alpha|Fx(t) - Gy(t) + g_1(t) - g_2(t)|) \\ \leq \alpha(|Fx(t) + g_1(t) - x(t)|, p|Fx(t) + g_1(t) - x(t)|) \\ +\beta(|Gy(t) + g_2(t) - y(t)|, p|Gy(t) + g_2(t) - y(t)|) \\ +\gamma(|x(t) - y(t)|, p|x(t) - y(t)|), \quad ,\end{aligned}$$

where $\alpha + \beta + \gamma < 1$, for every $x, y \in Z$ with $x \neq y$ and $t \in [a, b]$.

(H5): Whenever $Fx + g_1 \neq Gx + g_2$

$$\begin{aligned}\sup_{t\in[a,b]} (|Fx(t) - Gy(t) + g_1(t) - g_2(t)|, \alpha|Fx(t) - Gy(t) + g_1(t) - g_2(t)|) \\ < \sup_{t\in[a,b]} \alpha(|Fx(t) + g_1(t) - x(t)|, p|Fx(t) + g_1(t) - x(t)|) \\ +\beta(|Gx(t) + g_2(t) - y(t)|, p|Gx(t) + g_2(t) - y(t)|)\end{aligned}$$

for every $x \in Z$.

Theorem 2 *Assume that hypotheses* $(H3) - (H5)$ *hold. Then, the summation equations* (3)–(4) *have a unique common solution* x *on* $[a, b]$.

Proof Define $S, T : Z \to Z$ by $S(x) = Fx + g_1$ and $T(x) = Gx + g_2$. Using hypothesis, we have

$$\begin{aligned}(|Sx(t) - Ty(t)|, \alpha|Sx(t) - Ty(t)|) \leq \alpha(|Sx(t) - x(t)|, p|Sx(t) - x(t)|) \\ +\beta(|Ty(t) - y(t)|, p|Ty(t) - y(t)|) \\ +\gamma(|x(t) - y(t)|, p|x(t) - y(t)|),\end{aligned}$$

for every $x, y \in Z$ and $x \neq y$. Hence,

$$\begin{aligned}(\|S - T\|_\infty, \alpha\|S - T\|_\infty) \leq \alpha(\|Sx - x\|_\infty, p\|Sx - x\|_\infty) \\ +\beta(\|Ty - y\|_\infty, p\|Ty - y\|_\infty) \\ +\gamma(\|x - y\|_\infty, p\|x - y\|_\infty).\end{aligned}$$

Next, if $s(x) \neq T(x)$, we have

$$(\|S - T\|_\infty, \alpha\|S - T\|_\infty) \leq \alpha(\|Sx - x\|_\infty, p\|Sx - x\|_\infty) + \beta(\|Tx - x\|_\infty, p\|Tx - x\|_\infty)$$

for every $x \in Z$. By Lemma 2, if f is the identity map on Z, the summation equations (3)–(4) have a unique common solution. □

5 Application

In this section, we give an example to illustrate the usefulness of our results. In Eqs. (1)–(2), we define

$$k(t, s, x) = ts + \frac{xs}{6}, \quad h(t, s, x) = (ts)^2 + \frac{tsx^2}{6}, \quad s, t \in [0, 2], \quad x \in C([0, 2], \mathbb{R}),$$

and consider metric $d(x, y) = (\|x - y\|_\infty, \alpha\|x - y\|_\infty)$ on $C([0, 2], \mathbb{R})$ and $\alpha \geq 0$. Then clearly, $C([0, 2], \mathbb{R})$ is a complete cone metric space.
Now, we have

$$\begin{aligned}
&(|k(t, s, x(s)) - k(t, s, y(s))|, \alpha|k(t, s, x(s)) - k(t, s, y(s))|) \\
&\quad = (|ts + \frac{xs}{6} - (ts + \frac{ys}{6})|, \alpha|ts + \frac{xs}{6} - (ts + \frac{ys}{6})|) \\
&\quad = (|ts + \frac{xs}{6} - ts - \frac{ys}{6}|, \alpha|ts + \frac{xs}{6} - ts - \frac{ys}{6}|) \\
&\quad = (\frac{s}{6}|x - y|, \alpha\frac{s}{6}|x - y|) \\
&\quad = \frac{s}{6}(\|x - y\|_\infty, \alpha\|x - y\|_\infty) \\
&\quad = p_1^* \Phi_1^*(\|x - y\|_\infty, \alpha\|x - y\|_\infty),
\end{aligned}$$

where $p_1^*(t, s) = \frac{s}{3}$, which is a function of $[0, 2] \times [0, 2]$ into $\mathbb{R}^+$ and a comparison function $\Phi_1^* : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\Phi_1^*(x, y) = \frac{1}{2}(x, y)$. Also, we have

$$\begin{aligned}
&(|h(t, s, x(s)) - h(t, s, y(s))|, \alpha|h(t, s, x(s)) - h(t, s, y(s))|) \\
&\quad = (|(ts)^2 + \frac{tsx^2}{6} - ((ts)^2 + \frac{tsy^2}{6})|, \alpha|(ts)^2 + \frac{tsx^2}{6} - ((ts)^2 + \frac{tsy^2}{6})|) \\
&\quad = (|(ts)^2 + \frac{tsx^2}{6} - (ts)^2 - \frac{tsy^2}{6}|, \alpha|(ts)^2 + \frac{tsx^2}{6} - (ts)^2 - \frac{tsy^2}{6}|)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{ts}{6}|x^2 - y^2|, \alpha\frac{ts}{6}|x^2 - y^2|\right) \\
&\leq \frac{ts}{6}(\|x^2 - y^2\|_\infty, \alpha\|x^2 - y^2\|_\infty) \\
&\leq \frac{ts}{6}(\|x - y\|_\infty, \alpha\|x - y\|_\infty) \\
&= p_2^*\Phi_1^*(\|x - y\|_\infty, \alpha\|x - y\|_\infty),
\end{aligned}
$$

where $p_2^*(t, s) = \frac{ts}{3}$, which is a function of $[0, 2] \times [0, 2]$ into $\mathbb{R}^+$.
Moreover,

$$\sum_{s=0}^{1}[p_1^*(t, s) + p_2^*(t, s)] = \sum_{s=0}^{1}[\frac{s}{3} + \frac{ts}{3}] = \frac{1}{3}(1 + t)$$

$$\sup_{t\in[0,2]} \frac{1}{3}(1 + t) = 1.$$

Also,

$$\sum_{t=0}^{1}\sum_{s=0}^{1}[p_1^*(t, s) + p_2^*(t, s)] = \sum_{t=0}^{1}\sum_{s=0}^{1}[\frac{s}{3} + \frac{ts}{3}] = \sum_{t=0}^{1}\frac{1}{3}(1 + t) \leq 1.$$

With these choices of functions, all requirements of Theorem 1 are satisfied; hence, the existence and uniqueness are verified.

6 Conclusion

In this paper, the existence and uniqueness of solutions for second order summation–difference type equations and the existence of unique common solution of the summation equations in cone metric spaces have been studied. Moreover, an application is discussed.

References

1. M. Abbas, G. Jungck, Common fixed point results for noncommuting mappings without continuity in cone metric spaces. J. Math. Anal. Appl. **341**(1), 416–420 (2008)
2. R. Agarwal, *Difference Equations and Inequalities: Theory, Methods and Applications* (Marcel Dekker, New York, 1991)
3. M. Arshad, A. Azam, P. Vetro, Some common fixed point results in cone metric spaces. Fixed Point Theory Appl. **2009**(Article ID 493965), 11 p. (2009)

4. K.L. Bondar, Existence and uniqueness of results for first order difference equations. J. Modern Methods Numer. Math. **2**(1,2), 16–20 (2011)
5. K.L. Bondar, Existence of solutions to second order difference boundary value problems. Vidarbh J. Sci. **6**(1–2), 1–4 (2011)
6. K.L. Bondar, Some scalar difference inequalities. Appl. Math. Sci. **5**(60), 2951–2956 (2011)
7. K.L. Bondar, V.C. Borkar, S.T. Patil, Some existence and uniqueness results for difference boundary value problems. Bull. Pure Appl. Sci. Math. Stat. **29**(2), 291–296 (2010)
8. G.C. Done, K.L. Bondar, P.U. Chopade, Existence and uniqueness of solution of first order nonlinear difference equation. J. Math. Comput. Sci. **10**(5), 1375–1383 (2020)
9. G.C. Done, K.L. Bondar, P.U. Chopade, Existence and uniqueness of solution of summation-difference equation of finite delay in cone metric space. Commun. Math. Appl. **11**(3), 1–10 (2020)
10. L.G. Huang, X. Zhang, Cone metric spaces and fixed point theorems of contractive mappings. J. Math. Anal. Appl. **332**(2), 1468–1476 (2007)
11. D. Ilic, V. Rakocevic, Common fixed points for maps on cone metric space. J. Math. Anal. Appl. **341**(3), 876–882 (2008)
12. G. Jungck, B.E. Rhoades, Fixed point for set valued functions without continuity. Ind. J. Pure Appl. Math. **29**(3), 771–779 (1998)
13. W.G. Kelley, A.C. Peterson, *Difference Equation* (Academic Press, 2001)
14. M.K. Kwong, On Krasnoselskii's cone fixed point theorems. Fixed Point Theory Appl. **2008**, Article ID 164537, 18 p.
15. P. Raja, S.M. Vaezpour, Some extensions of Banach's contraction principle in complete cone metric spaces. Fixed Point Theory Appl. **2008**, Article ID 768294, 11 p.
16. S.H. Vaezpour, R. Hamlbarani, Some notes on the paper cone metric spaces and fixed point theorems of contractive mappings. J. Math. Anal. Appl. **345**(2), 719–724 (2008)

Influence of Radiant Heat and Non-uniform Heat Source on MHD Casson Fluid Flow of Thin Liquid Film Beyond a Stretching Sheet

Jagadish V. Tawade, Mahadev Biradar, and Shaila S. Benal

Abstract In the present paper, we are exploring the movement of Casson thin liquid film fluid with thermal conduction and impacting radiant heat and non-uniform heat source/sink above time dependent stretching plane. Using appropriate method, the governed non-linear PDE renewed in to ODE. The confluence method has been digitally revealed. The effect of the f' (skin friction) and temperature profile on thin film movement has been discussed numerically. Furthermore, the concept and physical parameters such as Pr, S, Nr, and M", and Casson fluid parameter had been conversed diagrammatically. Pr, S, Nr, and M denote Prandtl number, unsteadiness, thermal radiation, and magnetic field, respectively.

Keywords Casson fluid flow · Non-uniform heat source · Stretching sheet · Nusselt number · Prandtl number

1 Introduction

Uniqueness and study of flow of heat convey of thin films have fascinated attention of many researchers' outstanding prosperous applications in the last two decades. Refer to their multiple implementations in engineering such as food refining, reactor liquefaction, and representation polymeric amide; smooth prominence of artificial section; and tempering. Protuberance process is appreciatively worth for maintenance of the surface grade of transfer. All shell procedures require refined

J. V. Tawade (✉)
Department of Mathematics, Bheemanna Khandre Institute of Technology, Bhalki, India

M. Biradar
Department of Mathematics, Basaveshwar Engineering College, Bagalkot, Karnataka, India

S. S. Benal
Department of Mathematics, B.L.D.E.A's V. P. Dr. P. G. Halakatti College of Engineering and Technology, Vijayapur, Karnataka, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_3

plane for the finest product of exterior properties such as small abrasion, strength, and clearness. The standard product of bump swelling depends greatly on the film flow and heat transfer predictable on a lean liquid above a stretching plate, investigate the study of energy and heat transfer in such processes is essential.

The examination flow of a thin liquid layer on an unstable stretching sheet was studied by Wang [1], and later it was enlarged by Dandapat et al. [2] as well as heat transfer properties. It is reported that since fluids have slower friction the speed generates a considerable amount of heat generated in case of flexible sheet extrusion, and thus, the heat transfer speed can change significantly, due to viscous dissipation. Sarma and Rao [3] studied the heat transfer in a viscoelastic fluid on a stretching sheet analytically in the existence of viscous dissipation and interior heat generation. Sarojamma et al. [4] introduced a numerical model of unsteady MHD flux convinced by an integrated stretching plane turn round Casson fluid with magnetohydrodynamic radiant heat. Abel et al. [5] inspected the result of asymmetrical heat source on MHD heat transfer in a liquid film on a digitally unstable stretching sheet with viscous dissipation. Many authors [6–12] have executed a mathematical analysis and results of various thermo-physical characteristics on the flow of fluid film on a stretching plane in the presence of different physical parameters to show the velocity and heat transfer. Casson fluid is initiated by Casson and has been examined by the flow of curves of printing inks and later explained and represented by blood, honey, jelly, tomato sauce, polymers, etc. Lately, Megahed et al. [13], Kalyani et al. [14], Vijaya et al. [15], Eldabe et al. [16], etc. inspected the influence of flow on heat transfer of a Casson fluid in a thin film on a drawing sheet. In this article, we have investigated the flow of Casson liquid film fluid with transference of heat and having the effect of thermal radiation and a non-uniform heat source/sink on a time dependent stretch plane. Almost all the abovementioned studies unnoticed, the collective consequences of thermal and non-uniform radiation on heat convey which is principle point of sight of preferred properties of result. In the existing study, we put in the same for the heat transfer Casson thin liquid film from an unsteady stretching sheet. The conclusion acquired in comparison individual with Wang et al. [1], Megahed et al. [13], Kalyani et al. [14]. Without a query from the bench top, our outcomes are in admirable agreement with that of the results mentioned above in certain borderline cases.

2 Theory/Calculation

We consider non-Newtonian liquid Casson thin film of thickness $h(t)$ over a heated stretching foil that emerges from a contracted slit at the beginning of the Cartesian coordinate system as shown schematically in Fig. 1. The movement of the fluid inside the film is outstanding in the stretching sheet. Unbroken plane equivalent to x-axis also proceeds in its individual plane with a velocity

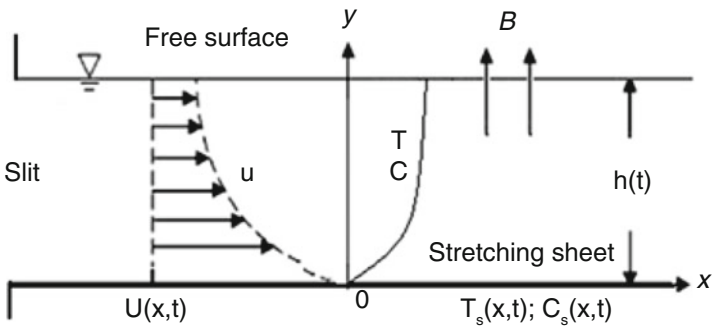


Fig. 1 Constitutes model of coordinate system

$$U(x,t) = \frac{bx}{(1-\alpha t)}. \tag{1}$$

α and b are the fixed constants with dimensions per moment. Stretching plate, temperature, and concentration T are implicit to be different with space x from the slit as

$$\mathrm{T_s(x,t) = T_0 - T_{ref}}\left[\frac{\mathrm{bx^2}}{2\nu}\right](1-\alpha \mathrm{t})^{-\frac{3}{2}}. \tag{2}$$

The stress–strain relation of non-Newtonian Casson fluid can be written as

$$\tau_{ij} = \begin{cases} 2(\mu_B + \frac{P_Y}{\sqrt{2\pi}})e_{ij}, & \pi > \pi_c \\ 2(\mu_B + \frac{P_Y}{\sqrt{2\pi_c}})e_{ij}, & \pi < \pi_c. \end{cases} \tag{3}$$

$\pi = e_{ij}e_{ij}$, and e_{ij} $(i,j)^t h$ μ_B is the plastic dynamic viscosity of the non-Newtonian fluid, P_y yields the stress of the fluid, π is the product of the component of deformation rate with itself, namely, $\pi = e_{ij}e_{ij}$, e_{ij} $(i,j)^t h$ is the component of deformation rate, and π_c is the critical value of π that depends on the non-Newtonian model. Under these presumptions, equations of the flow in the liquid film are given by

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{4}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = v\left(1+\frac{1}{\lambda}\right)\frac{\partial^2 u}{\partial y^2} - \frac{\sigma B^2}{\rho}u \tag{5}$$

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \frac{k}{\rho C_p}\frac{\partial^2 T}{\partial y^2} + \frac{\mu}{\rho C_p}\left(1+\frac{1}{\lambda}\right)\left(\frac{\partial u}{\partial y}\right)^2$$

$$+\frac{16\sigma^* T_0^3}{3\rho C_p k^*}\frac{\partial^2 T}{\partial y^2}+\frac{q'''}{\rho C_p}. \tag{6}$$

velocity components of fluid in x direction along with u and y directions along with v, T represents the temperature, μ denotes the dynamic viscosity, σ is the electrical conductivity, $\gamma = \mu_B \frac{\sqrt{2\pi_c}}{P_y}$ is the Casson parameter, ρ is the density, C_p is the specific heat at constant pressure, k is the thermal conductivity, σ^* Stefan–Boltzmann constant, k^* is the absorption coefficient. q'' "non-uniform heat source/sink" is modeled as

$$\mathrm{q}''' = \frac{\mathrm{Ku_{w(x)}}}{x\upsilon}\left[A^*(T_s - T_\infty)f' + (T - T_\infty)B^*.\right] \tag{7}$$

A^* and B^* are the coefficients of "space & temperature dependent internal heat generation/absorption respectively." There arise two cases: $A^* > 0$, $B^* > 0$ communicate to interior heat creation and $A^* < 0$, $B^* < 0$ communicate to interior heat immersion. In advance, it is implicit that the induced magnetic field is insignificantly small. The corresponding boundary conditions are

$$u = U,\ v = 0,\ \ T = T_s\ at\ y = 0 \tag{8}$$

$$\frac{\partial u}{\partial y} = \frac{\partial T}{\partial y} = 0\ at\ \ y = h \tag{9}$$

$$v = \frac{\partial h}{\partial t}\ \ at\ \ y = h. \tag{10}$$

At this phase, we construct a mathematical problem that perfectly worked out only for $x \geq 0$. In addition, the established surface planar liquid film is flat, therefore to keep away from the difficulty due to plane effects. The impact of port shear is due to the idle atmosphere, and especially the outcome of exterior tension is assumed irrelevant. The viscous shear stress

$$\tau = \mu\left(\frac{\partial u}{\partial y}\right)$$

& heat flux

$$q = -k\left(\frac{\partial T}{\partial y}\right)$$

evaporates action of free plane (at $y = h$). The following uniformity transformations are introduced:

$$\eta = \left[\frac{b}{\nu\,(1-\alpha t)}\right]^{\frac{1}{2}} y, \quad \psi = x\left[\frac{b}{1-\alpha t}\right]^{\frac{1}{2}} f(\eta) \tag{11}$$

$$T = T_0 - T_{ref}\left[\frac{bx^2}{2\nu\,(1-\alpha t)^{\frac{3}{2}}}\right]\theta(\eta), \quad \theta(\eta) = \frac{T-T_0}{T_s - T_0} \tag{12}$$

$$u = \frac{\partial \psi}{\partial y} = \frac{bx}{1-\alpha t} f'(\eta) \;\; and \;\; v = -\frac{\partial \psi}{\partial x} = -\left[\frac{b}{1-\alpha t}\right]^{\frac{1}{2}} f(\eta), \tag{13}$$

where ψ is the stream function and u and v are the velocity components.

2.1 Method of Solution

Eqs. (5)–(7) are converted into the following non-linear boundary value problem as:

$$\left(1+\frac{1}{\lambda}\right) f''' + \left[ff'' - s\left(f' + \frac{\eta}{2}f''\right) - (f')^2 - Mf'\right] = 0 \tag{14}$$

$$\left(1+\frac{4}{3}Nr\right)\theta'' + \Pr\left[f\theta' - 2f'\theta - \frac{S}{2}\left(\eta\theta' + 3\theta\right) + Ec\left(1+\frac{1}{\lambda}\right)(f'')^2 + A^* f' + B^*\theta\right] = 0. \tag{15}$$

The boundary conditions are

$$f'(0) = 1, \quad f(0) = 0, \quad \theta(0) = 1 \tag{16}$$

$$f''(\beta) = 0, \quad \theta'(\beta) = 0 \tag{17}$$

$$f(\beta) = \frac{S\beta}{2}, \tag{18}$$

where S $\equiv \frac{\alpha}{b}$ "unsteady parameter"; $M = \frac{\sigma B_0^2}{\rho b}$ "magnetic field parameter"; $\Pr = \frac{\rho C_p \nu}{k}$; $Nr = \frac{4\sigma^* T_\infty^3}{kk^*}$ "thermal radiation"; $Ec = \frac{U^2}{C_p(T_s - T_0)}$ "Eckert number", so that Eq. (11) gives

$$\beta = \left[\frac{b}{\nu(1-\alpha t)}\right]^{\frac{1}{2}} h. \tag{19}$$

Since β denotes the indefinite constant that should be determined, the complete position presents the boundary value problem. The velocity density is obtained by

$$\frac{dh}{dt} = -\frac{\alpha\beta}{2}\left[\frac{\nu}{b(1-\alpha t)}\right]^{\frac{1}{2}}. \tag{20}$$

Consequently, the kinematic limit at $y = h(t)$ given by (11) modifies into the open shell condition (20). The surface drag coefficient C_{fx} and "Nusselt number Nu_x" play an important role in estimating the surface drag force, the rate of heat transfer

$$C_f \mathrm{Re}_x^{\frac{1}{2}} = -2\left(1 + \frac{1}{\lambda}\right) f''(0), \tag{21}$$

and

$$Nu_x \mathrm{Re}_x^{-\frac{1}{2}} = \theta'(0), \tag{22}$$

where $\mathrm{Re}_x = \frac{Ux}{\upsilon}$ "local Reynolds number." Eqs. (14) and (15) corresponding to boundary conditions (16) to (18) are solved numerically by the shooting technique. These equations are renewed into a set of first order differential equations as follows:

$$\frac{df_0}{d\eta} = f_1, \quad \frac{df_1}{d\eta} = f_2, \quad \left(1 + \frac{1}{\lambda}\right)\frac{df_2}{d\eta} = S\left(f_1 + \frac{\eta}{2} f_2\right) + f_1^2 - f_0 f_2 + M f_1 \tag{23}$$

$$\frac{d\theta_0}{d\eta} = \theta_1, \quad \left(1 + \frac{4}{3}Nr\right)\frac{d\theta_1}{d\eta} = \mathrm{Pr}\left[\frac{S}{2}(3\theta_0 + \eta\theta_1) + 2f_1\theta_0 - \theta_1 f_0 - Ec\left(1 + \frac{1}{\lambda}\right)(f_2)^2 - A^* f_1 - B^*\theta_0.\right] \tag{24}$$

The associated periphery conditions take the form:

$$f_1(0) = 1, \quad f_0(0) = 0, \quad \theta_0(0) = 1 \tag{25}$$

$$f_0(\beta) = \frac{S\beta}{2}, \quad f_2(\beta) = 0, \quad \theta_1(\beta) = 0. \tag{26}$$

Here, $f_0(\eta) = f(\eta)$ and $\theta_0(\eta) = \theta$. This requires the initial values $f_2(0)$ *and* $\theta_1(0)$, and hence, the appropriate values are preferred and later integration is approved. A step size is $\Delta\eta = 0.01$. The value of β so obtained will satisfy the boundary condition $f_0(\beta) = \frac{S\beta}{2}$ with an error of tolerance 10^{-8}.

3 Results and Discussion

Current research was decided " to study the flow of non-Newtonian Casson liquid film flux in a time dependent stretching sheet with the consequences of MHD (non-uniform heat source/sink), A^*, and B^* & N_r thermal radiation." The determination of physical outcomes of different coating parameters on the velocity $f(\eta)$ and

temperature $\theta(\eta)$ profiles, displayed in Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11. Tables 1, 2, and 3 constitute evaluation of numerical results published earlier. In this article, the author explores the collaboration sound effects of "viscous dissipation, non-uniform heat source/sink," and "thermal radiation" for Casson fluid that has loads of scope in the heat exchange processes. However, film thickness β sink unsteady parameter β enhances, i.e., "by stretching the sheet, heat flux, and skin friction." Boosts quantity of boundary shrinks later. The obtained results are tabulated and exhibited graphically (Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11).

Figure 2a,b exhibits the impact of the liquid film thickness β, keeping $S = 0.8$ and $S = 1.2$ fixed. As β increases, the flow velocity of the liquid film drops. On the

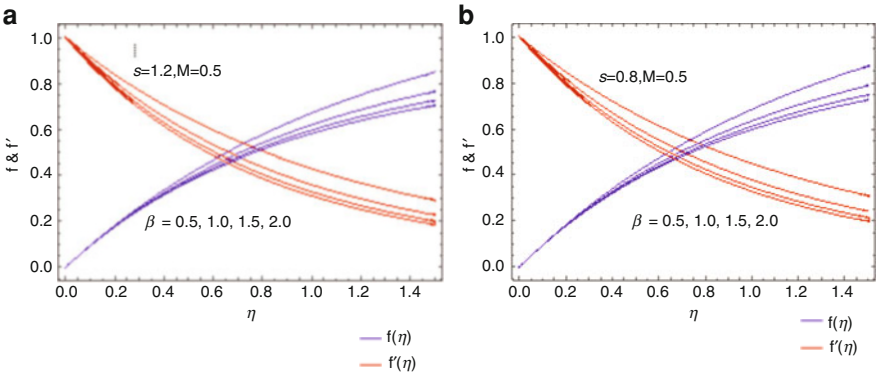


Fig. 2 (**a**) and (**b**) represents the results of β on transverse velocity (f) and horizontal velocity (f')

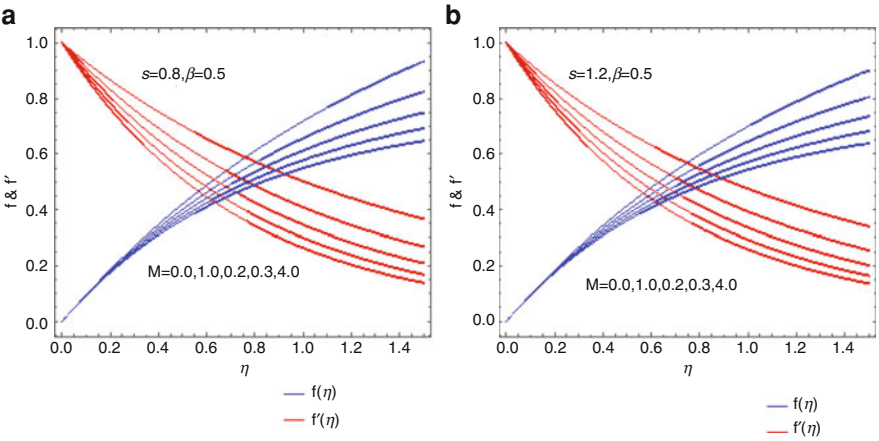


Fig. 3 (**a**) and (**b**) represents "magnetic field (M)" on transverse velocity (f) and horizontal velocity (f')

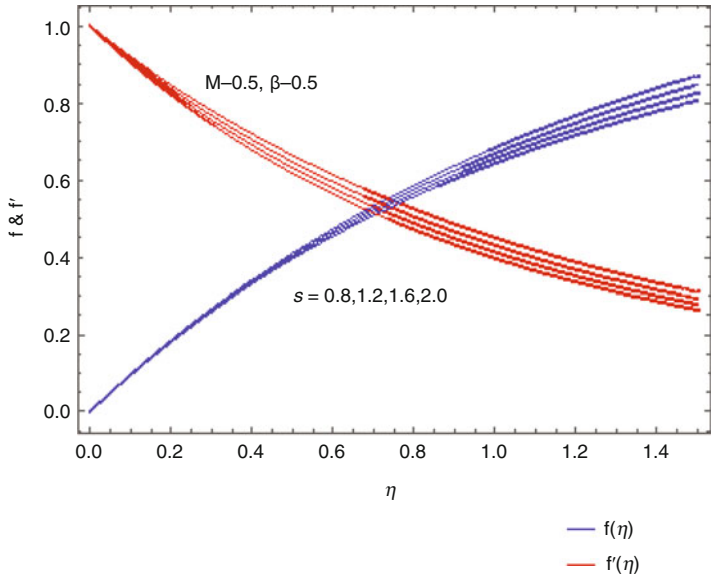


Fig. 4 This figure reveals for a fixed value of the "unsteady parameter (S)," the transverse velocity (f) enhances monotonically from the surface, and increasing values of S increase the transverse velocity for $\eta \geq 0.2$, while horizontal velocity (f')

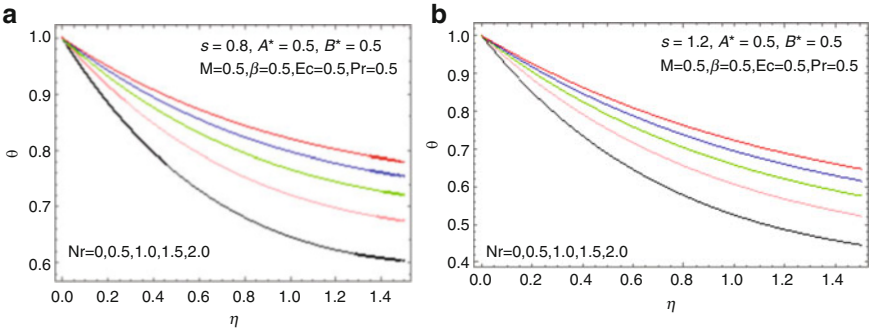


Fig. 5 (**a**) and (**b**) illustrates consequence "thermal radiation parameter Nr" on Casson fluid in for $S = 0.8$ and $S = 1.2$, respectively

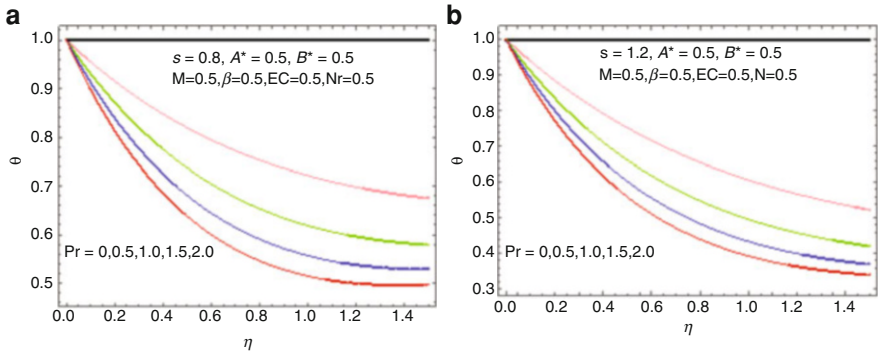


Fig. 6 (**a**) and (**b**) *plots* $0(\eta)$ for $S = 0.8$ and $S = 1.2$ for different values of Pr

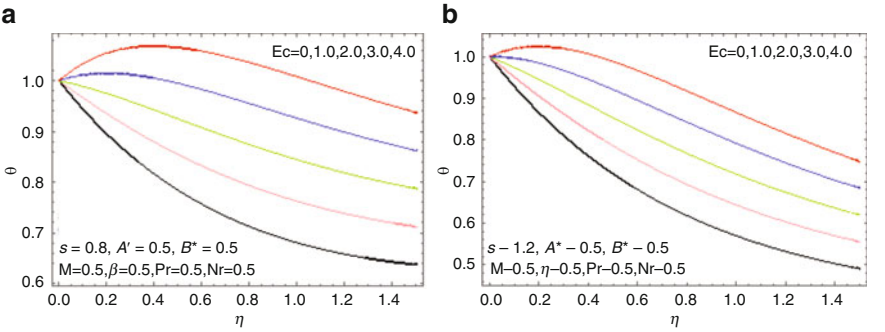


Fig. 7 (**a**) and (**b**) represents the temperature profile for $S = 0.8$ and $S = 1.2$, respectively, for variety of Ec, with $M = 0.5$ and Pr=0.5

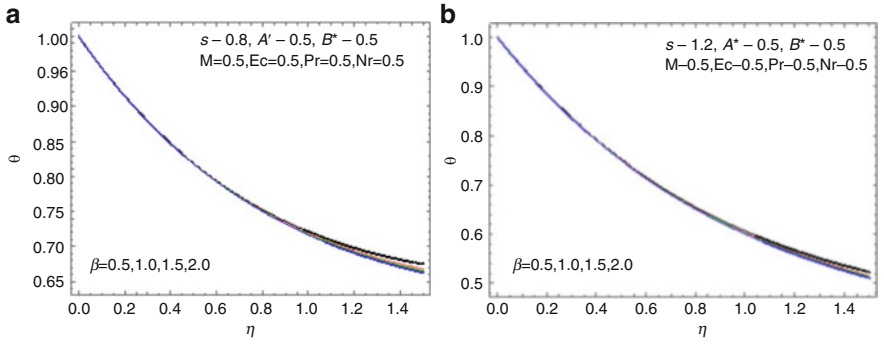


Fig. 8 (**a**) and (**b**) represents the liquid film thickness β for temperature profile

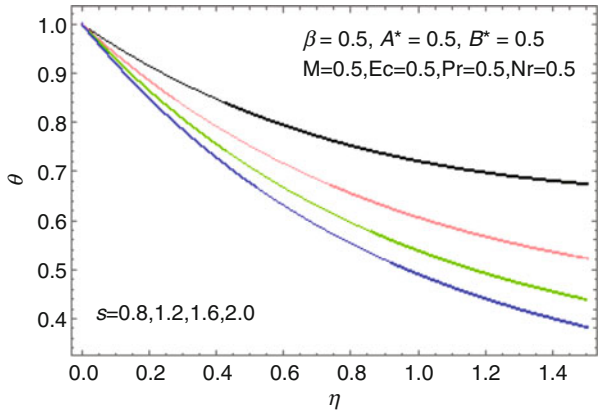


Fig. 9 Illustrates for various values of unsteadiness parameters for temperature profile

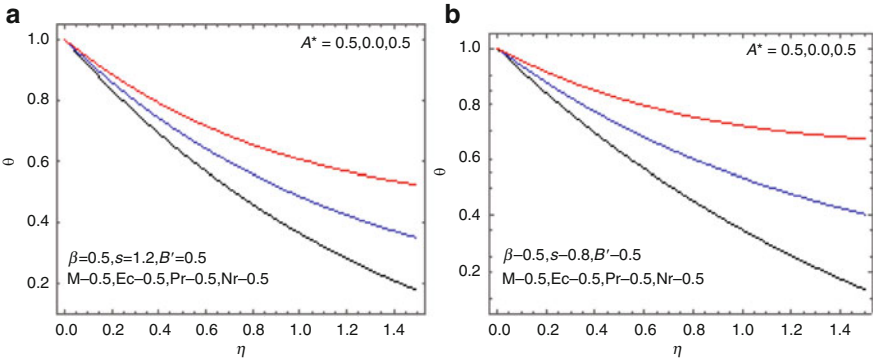


Fig. 10 (**a**) and (**b**) plots $0(\eta)$ for A^*

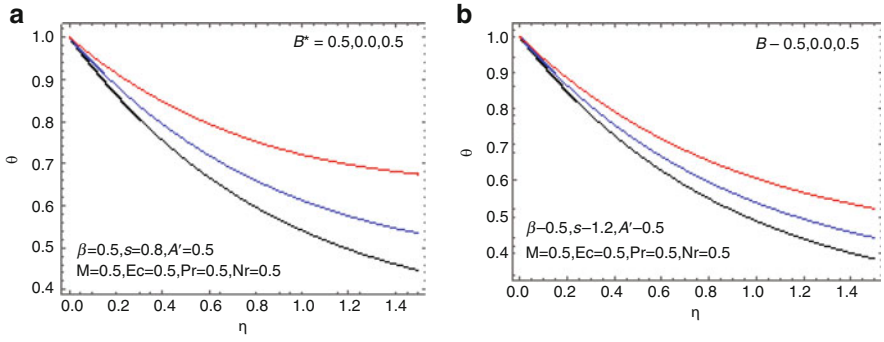


Fig. 11 (**a**) and (**b**) illustrates $0(\eta)$ for different values of B^*

Table 1 Comparing results of β and f'' with accessible text for various values of S

S	Wang[1]		Megahed [13]		Kalyani et al. [14]		Present study	
	β	$-\frac{f''(0)}{\lambda}$	β	$-f''(0)$	β	$-f''(0)$	β	$-f''(0)$
0.4	5.122490	1.307785	4.981450	1.134096	4.981455	1.134098	4.981448	1.19677904
0.6	3.131250	1.195155	3.131710	1.195126	3.131710	1.195128	3.131710	1.21744369
0.8	2.151990	1.245795	2.151994	1.245806	2.151990	1.245805	2.151990	1.23784971
1.0	2.543620	1.277762	1.543616	1.277769	1.543617	1.277769	1.543616	1.25799643
1.2	1.127780	1.279177	1.127781	1.279172	1.127780	1.279171	1.127780	1.27788890
1.4	0.821032	1.233549	0.821032	1.233545	0.821033	1.233545	0.821033	1.29753250
1.6	0.576173	1.491137	0.576173	1.114938	0.576176	1.114937	0.576176	1.31693527
1.8	0.356383	0.867414	0.356389	0.867414	0.356390	0.867416	0.356390	1.33609566

Table 2 Comparing results of $-(1+\frac{1}{\lambda})\ f''$ and $-\theta'(0)$" for variety of values

S	β	M	Kalyani et al. [14]		Present results	
			$-\left(1+\frac{1}{\lambda}\right)f''(0)$	$-\theta'(0)$	$-\left(1+\frac{1}{\lambda}\right)f''(0)$	$-\theta'(0)$
0.8	0.5	0.5	2.473527	1 355718	2.50341116	0.47073398
1.0			2.504972	1.409314	2.59394915	0.56141444
1.2			2.478698	1.450178	2.68156421	0.63299825
1.4			2.365328	1.463985	2.76655711	0.69337347
0.5	1.0	0.5	1.577477	1.289227	2.04553902	0.47226884
	2.0		1.585218	1.276834	1.77172679	0.47292167
	3.0		1.591939	1.269177	1.67043460	0.47300301
	4.0		1.596655	1.264258	1.61740280	0.47299459
0.5	0.5	0.0	2.157798	1.365686	2.17840898	0.50714984
		1.0	2.752908	1.346290	2.78884729	0.44032952
		2.0	3.239791	1.327678	3.28431898	3899588552
		3.0	3.662303	1.309009	3.71354914	0.34795885

other hand, velocity profile reduces with higher values of β. However, for the same variation of β, the temperature is found to increase as conveyed in Fig. 8a,b.

Figure 3a,b represents influence of magnetic field (M) on transverse velocity (f) and horizontal velocity (f'). The existence of magnetic field shows a decline in velocity in the momentum boundary layer up to the special point $\eta = 0.5$, where velocity attains a minimum and next starts enhancing. Expanding the magnetic field reduces velocity up to special point as higher values of M propose more conflicts in the fluid region due to Lorentz force, and after crossing the special point, the velocity enhances attaining its maximum at the free shell.

Table 3 Comparison of $-\theta'(0)$ considering variety of values of Pr, Ec, and Nr and variations of $A*$ and $B*$

"Pr"	"Ec"	"Nr"	"A*"	"B*"	"$-\theta'(0)$Kalyani et al. [14] $(A* = B* = 0)$"	"$-\theta'(0)$Present results"
0.7	0.1	0.5	0.5	0.5	1.108487	0.90460640
1.0					1.355718	1.11860029
2.0					1.982332	1.66981740
3.0					2.465460	2.09648657
1.0	0.0	0.5	0.5	0.5	1.315220	1.16438288
	1.0				0.910217	0.70655697
	2.0				0.505218	0.24873105
	3.0				0.100220	0.20909488
1.0	1.0	0.5	0.0	0.0	1.355718	1.11860029
		1.0			1.121379	0.91565716
		1.5			0.969474	0.78650605
		2.0			0.860352	0.69547944
1.0	1.0	0.5	−0.5	0.5	-	1.1047458586867287
			0.0		-	1.1047458585644556
			0.5		-	1.1047458585088834
			1.0		-	1.104745858492762
1.0	1.0	0.5	0.5	−0.5	-	1.104745858857642
				0.0	-	1.1047458586760088
				0.5	-	1.1047458585088834
				1.0	-	1.1047458583587242

Figure 4 discloses fixed value of the unsteady parameter (S), transverse velocity (f) enhances monotonically from the surface, and increasing values of S increases the transverse velocity for $\eta \geq 0.2$, while horizontal velocity (f') increases throughout the boundary layer with major increase at free surface. Raise of unsteadiness parameter enhances temperature in the boundary layer shown in Fig. 9.

Figure 5a,b illustrates the outcomes of thermal radiation parameter N_r on Casson fluid for $S = 0.8$ and $S = 1.2$, respectively. Practically, increasing values of Nr increases the temperature, which can be easily seen as the presence of thermal radiation releases higher thermal energy.

Figure 6a,b explains the results of Pr scheduled $\theta(\eta)$(temperature profile) for $S = 0.8$ and 1.2, respectively. From both figures, thermal boundary layer thickness and temperature decline as the Pr enhances. For higher ideals of Pr, temperature decreases rapidly. Figure 7a,b represents the consequences of Ec on temperature for $S = 0.8$ and $S = 1.2$, respectively. As temperature distribution fluid raises,

Eckert number increases with frictional heating. Figures 10a,b and 11a,b depict temperature profile on variety of "non-uniform heat source/sink parameters A^* and B^*." Increase in the values of A^* and B^* increases the values of $\theta(\eta)$.

4 Conclusion

Our aim is to find momentum and heat transfer presentation of Casson fluid in excess of extending plane in the presence of oblique M (magnetic field), "non-uniform heat source/sink," and Nr in k ("porous medium").

Inferences are as follows:

- As the heat transmit rate improves, the suction parameter attracts.
- In the current study, we have seen that as the magnetic field increases the resistance decreases.
- Non-uniform heat source increases for different values, and the"thermal boundary layer thickness also increases."
- The $\theta(\eta)$ "temperature profiles increases for superior values of radiant parameter."

4.1 Conflict of Interest

On behalf of all the co-authors, I am declaring that there is no conflict of interest.

Acknowledgments We heartily thank reviewers for responding to our work and giving positive comments, also address the concerns that were raised during the manuscript, and believe significantly to improve the effects of this process.

References

1. C. Wang, Analytic solutions for a liquid film on an unsteady stretching surface. Heat Mass Transfer **42**, 759–766 (2006)
2. B.S. Dandapat, B. Santra, H.I. Andersson, Thermocapillarity in a liquid film on an unsteady stretching surface. Int. J. Heat Mass Transfer **46**, 3009–3015 (2003)
3. M.S. Sarma, B.N. Rao, Heat transfer in a viscoelastic fluid over a stretching sheet. J. Math. Anal. Appl. **1**(1), 268–275 (1998)
4. G. Sarojamma, K. Sreelakshmi, B. Vasundhara, Mathematical model of MHD unsteady flow induced by a stretching surface embedded in a rotating Casson fluid with thermal radiation. IEEE, 1590–1595 (2016). 978-9-3805-44212/16/$31.00_c
5. M.S. Abel, J. Tawade, M.M. Nandeppanavar, Effect of nonuniform heat source on MHD heat transfer in a liquid film over an unsteady stretching sheet. Int. J. Nonlinear Mech. **44**(9), 990–998 (2009)

6. C. Wang, I. Pop, Analysis of the flow of a power-law fluid film on an unsteady stretching surface by means of homotopy analysis method. J. Non-Newtonian Fluid Mech. **138**(2), 161–172 (2006)
7. K. Vajravelu, K.V. Prasad, B.T. Raju, Effects of variable fluid properties on the thin film flow of Ostwald – de Waele fluid over a stretching surface. J. Hydrodyn. **25**(1), 10–19 (2013)
8. K.V. Prasad, K. Vajravelu, P.S. Datti, B.T. Raju, MHD flow and heat transfer in a power-law liquid film at a porous surface in the presence of thermal radiation. J. Appl. Fluid Mech. **6**(3), 385–395 (2013)
9. T. Khademinejad, M.R. Khanarmuei, P. Talebizadeh, A. Hamidi, On the use of the homotopy analysis method for solving the problem of the flow and heat transfer in a liquid film over an unsteady stretching sheet. J. Appl. Mech. Tech. Phys. **56**(4), 654–666 (2015)
10. M.A.A. Mahmoud, A.M. Megahed, MHD flow and heat transfer in a non-Newtonian liquid film over an unsteady stretching sheet with variable fluid properties. Can. J. Phys. **87**(10), 1065–1071 (2009)
11. M.M. Khader, A.M. Megahed, Numerical simulation using the finite difference method for the flow and heat transfer in a thin liquid film over an unsteady stretching sheet in a saturated porous medium in the presence of thermal radiation. J. King Saud Univ. Eng. Sci. **25**(1), 29–34 (2013)
12. C.H. Chen, Heat transfer in a power-law film over an unsteady stretching sheet. Heat Mass Transfer **39**(8), 791–796 (2003)
13. A.M. Megahed, Effect of slip velocity on Casson thin film flow and heat transfer due to unsteady stretching sheet in presence of variable heat flux and viscous dissipation. Appl. Math. Mech. Engl. Ed. **36**, 1273–1284 (2015)
14. K. Kalyani, K. Sreelakshmi, G. Sarojamma, Effect of thermal radiation on the Casson thin liquid film flow over a stretching sheet. Global J. Pure Appl. Math. **13**(6), 1575–1592 (2017)
15. N. Vijaya, K. Sreelakshmi, G. Sarojamma, Effect of magnetic field on the flow and heat transfer in a Casson thin film on an unsteady stretching surface in the presence of viscous and internal heating. Open J. Fluid Dyn. **6**(4), 303–320 (2016)
16. N.T.M. Eldabe, M.G.E. Salwa, Heat transfer of MHD non-Newtonian Casson fluid flow between two rotating cylinders. J. Phys. Soc. Jpn. **64**, 41 (1995)

MHD Boundary Layer Flow of Casson Fluid with Gyrotactic Microorganisms over Porous Linear Stretching Sheet and Heat Transfer Analysis with Viscous Dissipation

G. C. Sankad and Ishwar Maharudrappa

Abstract A study associated with bioconvection flow and heat transfer analysis due to Casson fluid and gyrotactic microorganisms over a linear stretching sheet through a porous medium in the presence of magnetic field is considered. The related governing equations of the physical situation are deformed into a system of nonlinear ordinary differential equations using Oberbeck–Boussinesq approximations and similarity transformation .The obtained coupled equations are solved with the help of differential transform method to get Taylor's series solutions for the momentum, energy, concentration of nanoparticles, and density of microorganisms. The combined effects of distinct nondimensional parameters on the solutions are represented through graphs.

Keywords Boundary layer · Stretching sheet · Bioconvection · Porous · Gyrotactic microorganisms · Differential transform method

1 Introduction

Influence of porous media and magnetic field on the boundary region of nanofluids near stretching surface has huge applications in mechanical industries as well as in physics. Investigation on Casson fluid flow and heat exchange on boundary region has shown great importance in industries. Sakiadis [1] examined the boundary layer performance on continuous solid surface moving on both flat and cylindrical plane.

G. C. Sankad (✉)
Department of Mathematics, Research Center Affiliated to Visvesvaraya Technological University, Belagavi, BLDEA's Vachana Pitamaha Dr. P. G. Halakatti College of Engg. and Tech., Vijayapur, India
e-mail: math.gurunath@bldeacet.ac.in

I. Maharudrappa
Department of Mathematics, Basaveshwar Engineering College, Bagalkot, Karnataka, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_4

Poo et al. [2] explored the consequence of variable viscosity on flow and heat exchange to a continuous moving flat plate. Mukhopadhyay et al. [3] studied MHD boundary region flow and warm exchange over a stretching sheet with variable viscosity. Abel et al. [4] analyzed viscoelastic fluid flow and heat transfer near the stretching sheet with variable viscosity. Study on bioconvection induced due to nanofluid and microorganisms has driven tremendous attention for its importance in improving quality of the product in bioindustries. Within the last few decades and in advance, lot of research work was done by the well-known scholars [5–14] on boundary region flow and heat exchange induced by the porous media near the stretching surface influenced by magnetic field.

The present work is on the study of bioconvection occurred due to Casson nanofluid and gyrotactic microorganisms in the MHD boundary region through porous media above the stretching sheet. Further, we have explored influence of Brownian motion and thermophoresis on the fluid flow and heat flow. It is revealed in the literature survey that no work is done on this physical circumstance where Casson model is characterized as non-Newtonian fluid containing gyrotactic microorganisms through porous media near the linear stretching surface introduced under the influence of magnetic field. Presently, we have tried to solve the arising modeled equations utilizing differential transform method and obtained the Taylor's series solution for a system of coupled nonlinear differential equations associated with boundary conditions. The outcomes of considered BVP are good in agreement with the boundary conditions.

2 Mathematical Formulation

Let us consider that unsteady non-Newtonian fluid including gyrotactic microorganisms is allowed to flow along the x-axis above the linear stretching sheet through the porous media. Magnetic field is applied uniformly normal to the surface of boundary region and neglecting the impact of induced magnetic field. It is supposed that fluid is water based so that microorganisms are alive and there is not much effect of the nanoparticles on activity of microorganisms. Fluid dilution avoids volatility of bioconvection due to microorganisms and nanoparticles. Based on these assumptions, we can form the following governing equations for the present problem:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{1}$$

$$u\frac{\partial u}{\partial x} + v\frac{\partial v}{\partial y} = \nu\left(1 + \frac{1}{\beta}\right)\frac{\partial^2 u}{\partial y^2} - \left(\frac{\sigma B_0^2}{\rho} + \frac{\nu}{K}\right)u - \left(\frac{1-C_\infty}{\rho_f}\right)\rho_\infty g\alpha(T - T_\infty) - \left(\frac{\rho_p - \rho_\infty}{\rho_f}\right)g(C - C_\infty) - \left(\frac{\rho_m - \rho_f}{\rho_f}\right)g\gamma(N - N_\infty) \tag{2}$$

$$u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}=\frac{K}{\rho C_p}\frac{\partial^2 T}{\partial y^2}+\tau\left(D_B\frac{\partial C}{\partial y}\frac{\partial T}{\partial y}+\frac{D_T}{T_\infty}\left[\frac{\partial T}{\partial y}\right]^2\right)+\frac{\nu}{C_p}\left[\frac{\partial u}{\partial y}\right]^2 \quad (3)$$

$$u\frac{\partial C}{\partial x}+v\frac{\partial C}{\partial y}=D_B\frac{\partial^2 C}{\partial y^2}+\frac{D_T}{T_\infty}\frac{\partial^2 T}{\partial y^2} \quad (4)$$

$$u\frac{\partial N}{\partial x}+v\frac{\partial N}{\partial y}+\frac{bW_c}{C_w-C_\infty}\left(\frac{\partial}{\partial y}\left[N\frac{\partial C}{\partial y}\right]\right)=D_m\frac{\partial^2 N}{\partial y^2}, \quad (5)$$

where ρ is the Casson fluid density, γ is an average volume of microorganisms, D_B is the Brownian diffusion coefficient, D_T is the thermophoresis diffusion coefficient, $\tau=\frac{(\rho C)_p}{(\rho C)_f}$ is the ratio of effective heat capacity of the fluid with ρ_f and ρ_p being density of the fluid and density of the particle, and β is the Casson fluid parameter.
The boundary conditions are

$$\left.\begin{array}{l} v=0,\quad u=bx,\quad T=T_w,\quad C=C_w,\quad N=N_w,\quad as\quad y\to 0,\\ u\to 0,\quad T\to T_\infty,\quad C\to C_\infty,\quad N\to N_\infty,\quad as\quad y\to\infty.\ , \end{array}\right\} \quad (6)$$

where $b>0$ is the stretching rate. Using the similarity transformation into the governing equations:

$$\left.\begin{array}{l} \eta=\frac{y}{x}Ra_x^{\frac{1}{4}}f(\eta),\quad \psi=mRa_x^{\frac{1}{4}}f(\eta),\quad \theta(\eta)=\frac{T-T_\infty}{T_w-T_\infty},\quad \phi(\eta)=\frac{C-C_\infty}{C_w-C_\infty},\\ \chi(\eta)=\frac{N-N_\infty}{N_w-N_\infty},\quad Ra_x=\frac{(1-C_\infty)\alpha g\Delta T_f}{m\nu}x^3. \end{array}\right\} \quad (7)$$

We form the following coupled nonlinear ordinary differential equations:

$$\left(1+\frac{1}{\beta}\right)f'''-\left(\frac{1}{2P_r}\right)f''+\left(\frac{3}{4P_r}\right)ff''-(M+\kappa)f'+\theta-N_r\phi-R_b\chi=0, \quad (8)$$

$$\theta''+\left(\frac{3}{4}\right)f\theta'+N_b\theta'\phi'+N_t\theta'^2+P_rE_cf''^2=0, \quad (9)$$

$$\phi''+\left(\frac{3}{4}\right)L_ef\phi'+\left(\frac{N_t}{N_b}\right)\theta''=0, \quad (10)$$

$$\chi''+\left(\frac{3}{4}\right)S_cf\chi'-P_e\left[\phi'\chi'+\phi''(\chi+\sigma)\right]=0. \quad (11)$$

The associated dimensionless boundary conditions are

$$\left.\begin{array}{l} f(0)=0, \quad f^{'}(0)=\lambda, \quad \theta(0)=1, \quad \phi(0)=1, \quad \chi(0)=1, \quad as \quad \eta \to 0, \\ f^{'}(\infty)=0, \quad \theta(\infty)=0, \quad \phi(\infty)=0, \quad \chi(\infty)=0, \quad as \quad \eta \to \infty. \end{array}\right\} \tag{12}$$

The dimensionless parameters used in Eqs. (9) to (12) are given by

$$\left.\begin{array}{l} P_r=\dfrac{\nu}{m}, M=\dfrac{\sigma B_0^2 x^2}{\rho \nu Ra_x^{\frac{1}{2}}}, \kappa=\dfrac{1}{Ra_x^{\frac{1}{2}} D_a}, N_r=\dfrac{(\sigma_p-\sigma_\infty)\Delta C_w}{\sigma_f(1-C_\infty)\alpha \Delta T_f}, \\ R_b=\dfrac{\gamma \Delta N_w \Delta \rho}{\sigma_f \alpha(1-C_\infty)\Delta T_w}, N_b=\dfrac{\tau D_B(C_w-C_\infty)}{m}, E_c=\dfrac{m^2 Ra_x}{x^2 C_p D T_f}, \\ L_e=\dfrac{m}{D_B}, S_c=\dfrac{m}{D_m}, P_e=\dfrac{bW_c}{(C_w-C_\infty)}, \sigma=\dfrac{N_\infty}{(N_w-N_\infty)}, \lambda=\dfrac{ax^2}{mR_x^{\frac{1}{2}}}, \end{array}\right\} \tag{13}$$

where M is the modified magnetic parameter, P_e is the Peclet number, L_e is the conventional Lewis number, N_r is the buoyancy ratio parameter, R_b is the bioconvection Rayleigh number, N_b is the Brownian motion parameter, N_t is the thermophoresis parameter, S_c is the Schmidt number, σ is the (dimensionless) bioconvection constant, and λ is the slip parameter.

3 Solution of the Problem

Using DTM [11], Eqs. (8)–(13) can be transformed into the following differential forms:

$$\begin{array}{l} \left(1+\frac{1}{\beta}\right)(s+1)(s+2)(s+3)F[s+3]=(M+\kappa)(s+1)F[s+1]-\theta[s]+ \\ N_r\phi[s]+R_b\chi[s]+\frac{1}{2}P_r\sum\limits_{r=0}^{s}(s-r+1)F[s-r+1](s+1)F[r+1]- \\ \frac{3}{4}P_r\sum\limits_{r=0}^{s}F[s-r](r+1)(r+2)F[r+2], \end{array} \tag{14}$$

where $F[0]=0$, $F[1]=a_1$, $F[2]=a_2$,

$$\begin{array}{l} (s+1)(s+2)\theta[s+2]=-\frac{3}{4}\sum\limits_{r=0}^{s}F[s-r](r+1)\theta[r+1]- \\ N_b\sum\limits_{r=0}^{s}(s-r+1)\theta[s-r+1](r+1)\theta[r+1]- \\ P_e\sum\limits_{r=0}^{s}(s-r+1)\theta[s-r+1](r+1)(r+2)\theta[r+2], \end{array} \tag{15}$$

where $\theta[0]=1$, $\theta[1]=a_3$,

$$(s+1)(s+2)\phi[s+2] = -\tfrac{3}{4}L_e \sum_{r=0}^{s} F[s-r](r+1)\phi(r+1) - \tfrac{N_t}{N_b}\sum_{r=0}^{s}(s+1)(s+2)\theta[s+2], \tag{16}$$

where $\phi[0] = 1$, $\phi[1] = a_4$,

$$(s+1)(s+2)\chi[s+2] = P_e \sum_{r=0}^{s}(s-r+1)\phi[s-r+1](s+1)\chi[s+1] + P_e \sum_{r=0}^{s}\chi[s-r](s+1)(r+2)\phi[r+2] + P_e\sigma(s+1)(s+2)\phi[s+2] - \tfrac{3}{4}S_c \sum_{r=0}^{s} F[s-r](r+1)\chi[r+1], \tag{17}$$

where $\chi[0] = 1$, $\chi[1] = a_5$.

$F[s]$, $\theta[s]$, $\phi[s]$, and $\chi[s]$ are the differential transforms of $f(\eta)$, $\theta(\eta)$, $\phi(\eta)$, and $\chi(\eta)$, respectively. a_1, a_2, a_3, a_4, and a_5 are the constants, and these can be determined with the aid of Eqs. (14)–(17) and the boundary conditions. For $s = 0, 1, 2, 3\ldots$, we get

$$F[3] = \frac{a_1^2 P_r}{12\left(1+\frac{1}{\beta}\right)} + \frac{a_1\left(M - \frac{1}{Da\sqrt{R_a}}\right)}{6\left(1+\frac{1}{\beta}\right)} + \frac{N_r + R_b - 1}{6\left(1+\frac{1}{\beta}\right)}$$

$$\theta[2] = -\left(\frac{1}{2}a_3 a_4 N_b + \frac{a_3^2 N_t}{2} + \frac{1}{2}a_1^2 E_c P_r\right)$$

$$\phi[2] = \frac{N_t\left(\frac{1}{2}a_3 a_4 N_b + \frac{1}{2}a_3^2 N_t + \frac{1}{2}a_1^2 E_c P_r\right)}{Nb}$$

$$\chi[2] = \frac{1}{2}\left[a_4 a_5 P_e + \frac{N_t P_e\left(a_3 a_4 N_b + a_3^2 N_t + a_1^2 E_c P_r\right)}{N_b} + \frac{N_t P_e\left(a_3 a_4 N_b + a_3^2 N_t + a_1^2 E_c P_r\right)}{N_b}\sigma\right].$$

Similarly, we can find $F[4]$, $\theta[4]$, $\phi[4]$, $\chi[2]$, and $\chi[3]$, and taking $P_r = 6.2$, $\beta = 1$, $M = 5$, $N_r = 0.5$, $R_b = 0.1$, $N_t = 0.1$, $N_b = 0.1$, $L_e = 10$, $S_c = 0.1$, $P_e = 1$, $\sigma = 0.2$, $R_a = 0.5$, and $D_a = 0.5$ and solving for all the five transformed equations in five unknowns using the boundary condition and Pade approximation, we obtain $a_1 = 0.130491$, $a_2 = 0.105687$, $a_3 = 4.053355$, $a_4 = 6.91778$, $a_5 = -0.579615$, and the Taylor's series solutions for Eqs. (8)–(11) are

$$f(\eta) = 0.130491\eta + 0.105687\eta^2 + 0.0254366\eta^3 + 0.00931264\eta^4 + 0.0315593\eta^5 - 0.00705885\eta^6 + 0.0000565214\eta^7 - 0.0010699\eta^8 - 0.000362508\eta^9 - 0.000426626\eta^{10} - 0.000221054\eta^{11} - \ldots.$$

$$\theta(\eta) = 1 + 4.05355\eta - 2.22364\eta^2 + 0.747091\eta^3 - 0.0810705\eta^4 + 0.0117574\eta^5 - 0.0306772\eta^6 + 0.00773456\eta^7 + 0.0069833\eta^8 - 0.00369728\eta^9$$

$$-0.00112102\eta^{10} + 0.000891603\eta^{11} - \ldots.$$

$$\phi[\eta] = 1 + 6.91778\eta + 2.22364\eta^2 - 1.87547\eta^3 - 0.738584\eta^4 + 0.0213219\eta^5$$
$$+0.231329\eta^6 + 0.0247125\eta^7 - 0.0279245\eta^8 + 0.0103749\eta^9$$
$$+0.00405778\eta^{10} + 0.00242809\eta^{11} - \ldots.$$

$$\chi[\eta] = 1 - 0.579615\eta + 0.663541\eta^2 - 1.578781\eta^3 - 2.064390\eta^4 - 4.637221\eta^5$$
$$-5.451773\eta^6 - 6.076239\eta^7 - 4.203908\eta^8 - 1.521497\eta^9$$
$$+1.73707\eta^{10} + 3.626755\eta^{11} - \ldots.$$

4 Results and Discussion

Discussion of bioconvected flow of Casson fluid nanoparticles and gyrotactic microorganisms under the influence of uniformly applied magnetic field normal to the surface through porous media and heat exchange analysis is carried out. In this work, the solution is represented in Taylor's series utilizing DTM for all the governing equations. It is intended to analyze an impact of nondimensional parameters on the velocity of the fluid, temperature, concentration of nanoparticles, and gyrotactic microorganisms through graphs.

Velocity Profile Figures 1, 2, 3, and 4 are the velocity profiles of the Casson fluid for values of a range of parameters. In Fig. 1, the velocity of the Casson fluid flow increases due to applied magnetic field normal to the surface for growing values of Casson fluid parameter. Separation of boundary layers due to magnetic field and bioconvection impact due to the microorganisms and nanofluids together brings increase in the motion of the fluid. From Fig. 2, it is observed that the velocity

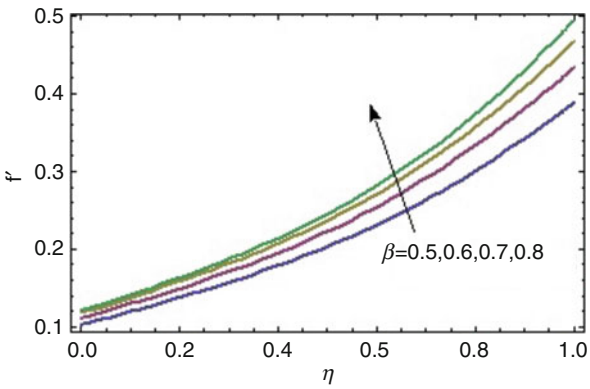


Fig. 1 Velocity profile for $\beta = 0.5, 0.6, 0.7, 0.8$

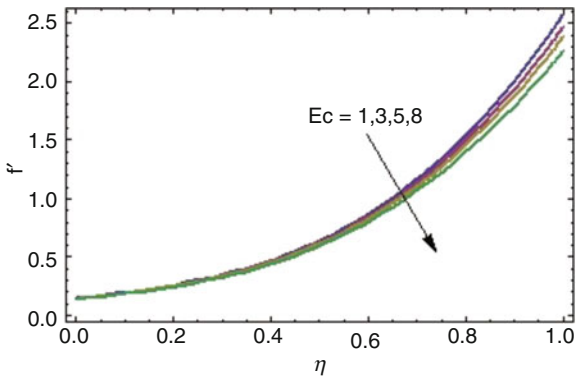


Fig. 2 Velocity profile for $E_c = 1, 3, 5, 8$

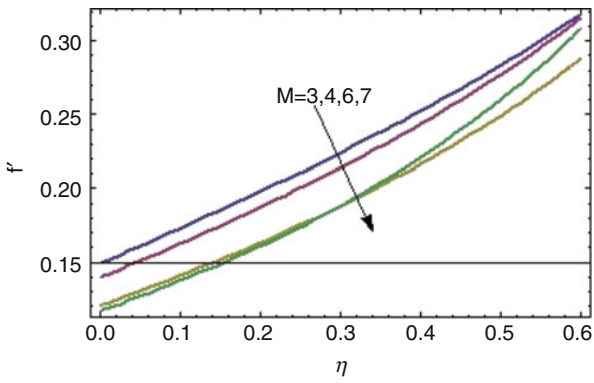


Fig. 3 Velocity profile for $M = 3, 4, 6, 7$

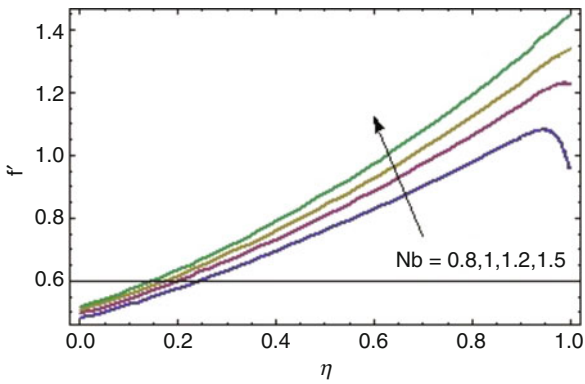


Fig. 4 Velocity profile for $N_b = 0.8, 1, 1.2, 1.5$

of the fluid decreases for the growing values of the viscous dissipation parameter. It is due to the reason that the viscosity of the fluid improves for increasing values of Eckert number, which in turn affects the motion of the fluid. Figure 3 shows the reduction of the velocity with rising values of the magnetic parameter. This is due to the reason that the applied magnetic field is perpendicular to the fluid flow, and hence, the increasing force applied resists the fluid flow. Fig. 4 displays the increase in the velocity for the increase of N_b.

Temperature Profile Figures 5, 6, 7, and 8 are the temperature profiles for different values of nondimensional parameters, and we found that raise of warm for the growing values of Casson fluid parameter, viscous dissipation parameter, Brownian motion parameter, and thermophoresis parameter. In Fig. 5, the growing values of Casson fluid parameter lead to raise in the temperature profile due to the reason that the velocity of the fluid ascends for different values of Casson fluid parameter. Figure 6 shows the enhancement of temperature profile for the increasing values of viscous dissipation parameter (E_c). The influence of the N_b and N_t is shown in Fig. 8. N_b and N_t have effect on fluid flow and heat flow. Temperature profile

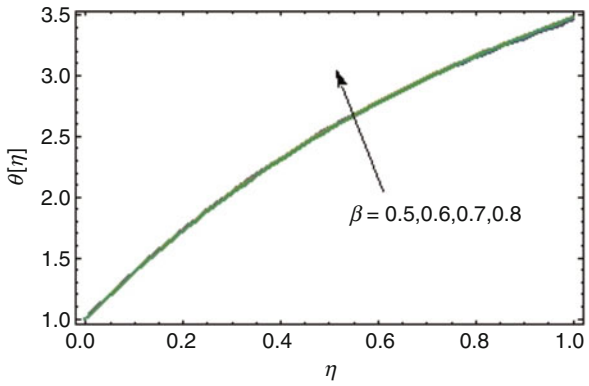


Fig. 5 Temperature profile for $\beta = 0.5, 0.6, 0.7, 0.8$

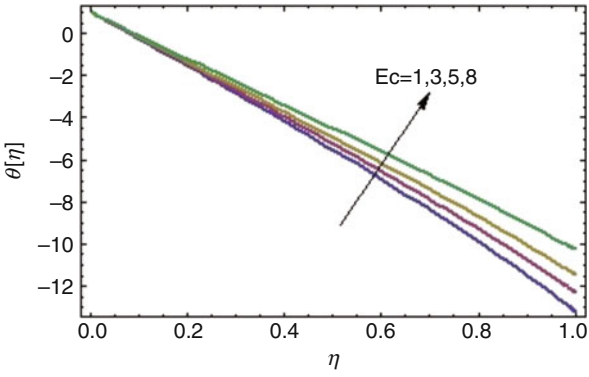


Fig. 6 Temperature profile for $E_c = 1, 3, 5, 8$

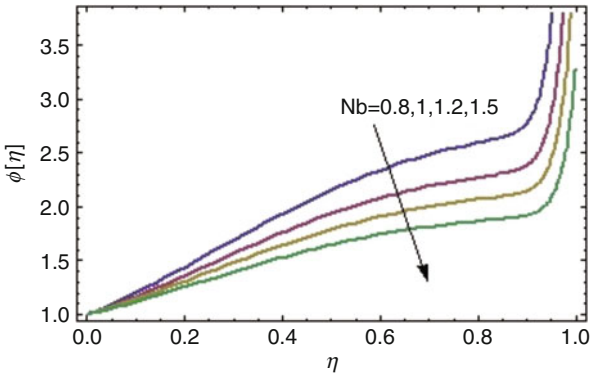


Fig. 7 Concentration of nanoparticles profile for $N_b = 0.8, 1, 1.2, 1.5$ and $N_t = 0.1$

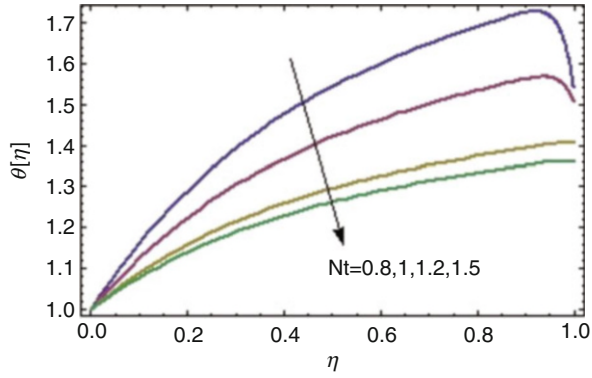


Fig. 8 Temperature profile for $N_t = 0.8, 1, 1.2, 1.5$ and $N_b = 0.1$

decreases with increase in the values of N_b and N_t. The reason is that nanoparticles exhibit different values of N_b and N_t.

Concentration of Nanoparticles and Microorganisms Concentration profiles of Casson nanoparticles and microorganisms are described in Figs. 7, 9, 10, 11, 12, and 13. It can be experiential that concentration profiles of Casson fluid nanoparticles are unlike as compared to the concentration profiles of microorganisms. In Fig. 9, it is found that the concentration of nanoparticles diminishes with growing values of β. Figure 10 describes the concentration of nanofluid particles decreases for raising values of Eckert number. Figure 11 describes the concentration of nanofluid particles increases for raising values of magnetic parameter. Figure 12 give relation between concentration of nanofluid particles and Brownian motion parameter. Figure 13 gives relation between thermophoresis parameter and microorganisms concentration. The graph shows decrease in the concentration of nanofluid particles and microorganisms profiles for the increased values of N_b and N_t. The outcomes of the profiles are due to the dependency of Brownian parameter on reduced thermal enhancement and concentration on temperature field, respectively.

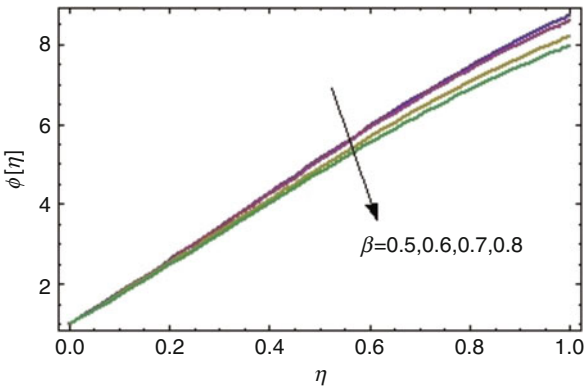


Fig. 9 Concentration of nanoparticles profile for $\beta = 0.5, 0.6, 0.7, 0.8$

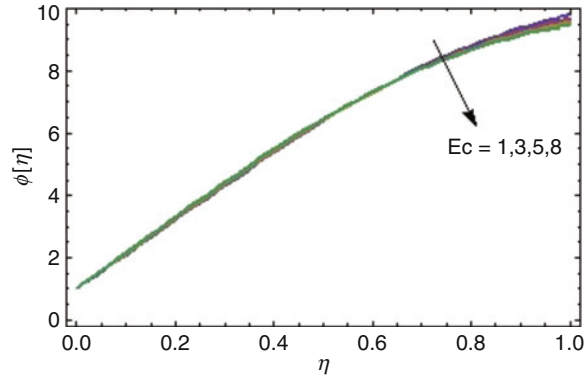


Fig. 10 Concentration of nanoparticles profile for $E_c = 1, 3, 5, 8$

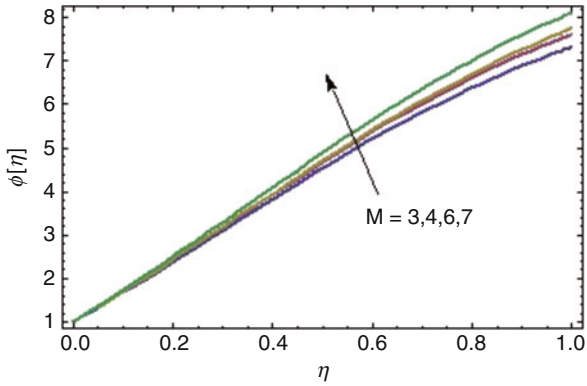


Fig. 11 Concentration of nanoparticles profile for $M = 3, 4, 6, 7$

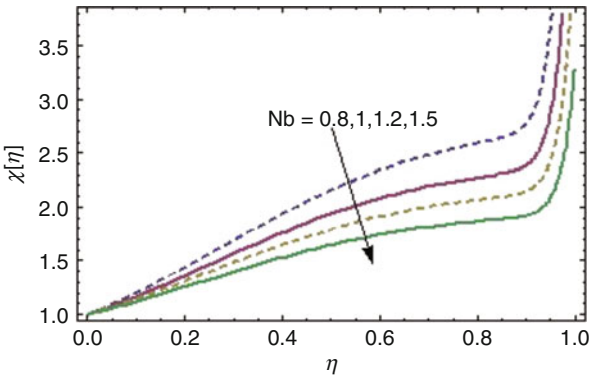


Fig. 12 Concentration of microorganisms for $N_b = 0.8, 1, 1.2, 1.5$ and $N_t = 0.1$

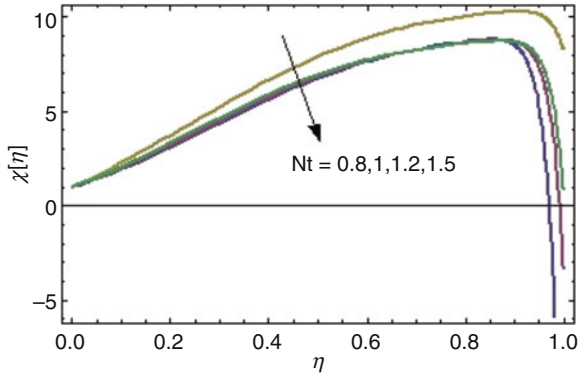


Fig. 13 Concentration of microorganisms for $N_t = 0.8, 1, 1.2, 1.5$ and $N_b = 0.1$

5 Conclusion

Our study addresses an analytic solution of bioconvection of MHD boundary layer flow and heat exchange of Casson nanofluids and gyrotactic microorganisms over a linear stretching sheet through porous media. A Taylor's series solution is obtained for momentum, energy, concentration equation of nanofluids, and density of gyrotactic microorganism's equations using differential transform method. Interesting features of the velocity of the stream and heat exchange are described.

- Velocity of Casson fluid decreases for increasing values of viscous dissipation parameter (E_c).
- Nanoparticles possess unlike values of Brownian motion parameter (N_b) and thermophoresis parameter (N_t). Hence, the concentration of nanofluid and heat flow reduces for growing values of N_b and N_t.
- The concentration of microorganism reduces for an enhanced magnitude of Casson fluid parameter due to sensitivity of microorganisms and thinning of boundary layer thickness.

References

1. B.C. Sakiadis, Boundary layer behavior on continuous solid surfaces: I Boundary layer equations for two dimensional and axisymmetric flow. AICHE. J. **7**, 26-28 (1961
2. S.R. Poo, T. Grosan, I. Pop, Radiation effect on the flow near the stagnation point of stretching sheet. Technische Mechanik, Band 25, Heft 2, 100–106 (2004)
3. S. Mukhopadhyay, G.C. Layek, S.A. Samad, Study of MHD boundary layer flow over a heated stretching sheet with variable viscosity. Int. J. Heat Mass Transfer **48**, 4460–4466 (2205)
4. M.S. Abel, N. Mahesha, Heat transfer in MHD viscoelastic fluid over a stretching sheet with variable thermal conductivity, non-uniform heat source and radiation. Appl. Math. Model. **32**(10), 1965–1983 (2207)
5. T.J. Pedley, N.A. Hill, J.O. Kessler, The growth of bioconvection patterns in a uniform suspension of gyrotactic microorganisms. J. Fluid Mech. **195**, 223–237 (1988)
6. W.A. Khan, O.D. Makinde, MHD boundary layer flow of a nanofluids containing gyrotactic microorganisms past a vertical plate with Navier slip. Int. J. Heat Mass Transfer **74**, 285–291 (2014)
7. S. Pramanik, Casson fluid flow and heat transfer past an exponentially porous stretching surface in presence of thermal radiation. Ain Shams Eng. J. **5**, 205–212 (2014)
8. K. Bhattacharya, M.S. Uddin, G.C. Layek, Exact solution for thermal boundary layer in Casson fluid over permeable shrinking sheet with variable wall temperature and thermal radiation. Alexandria Eng. J. **55**, 1703–1712 (2016)
9. A.A. Pahlavan, V. Aliakbar, F. Farahani, K. Sadeghy, MHD flows of UCM fluids above porous stretching sheets using two-auxiliary-parameter homotopy analysis method. Commun. Nonlinear Sci. Numer. Simul. **14**, 473–488 (2009)
10. M. Hatami, D. Jing, Differential Transformation Method for Newtonian and non-Newtonian nanofluids flow analysis: Compared to numerical solution. Alexandria Eng. J. **55**, 731–739 (2016)
11. S. Sepasgozar, M. Faraji, P. Valipour, Application of differential transformation method (DTM) for heat and mass transfer in a porous channel. Propulsion Power Res. **6**(1), 41–48 (2017)
12. T. Chakraborthy, K. Das, K. Prabhir Kumar, Framing the impact of external magnetic field on bioconvection of nanofluids flow containing gyrotactic microorganisms with convective boundary conditions. Alexandria Eng. J. **57**, 61–71 (2018)
13. D.A. Nield, A.V. Kuznetsov, The onset of bio-thermal convection in a suspension of gyro-tactic microorganisms in a fluid layer: Oscillatory convection. Int. J. Therm. Sci. **45**, 990–997 (2006)
14. R. Cortell, Effect of viscous dissipation and radiation on the thermal boundary layer over a nonlinearly stretching sheet. Phys. Lett. A **372**, 631–636 (2008)

Design of Coupled FIR Filters for Solving the Nuclear Reactor Point Kinetics Equations with Feedback

Dr. M. Mohideen Abdul Razak

Abstract A new method of solving the nuclear reactor point kinetics equations with feedback is presented in this chapter. In small nuclear reactors, the reactor power transients are estimated by solving the stiff point kinetics equations with feedback. Here, a new computational method is developed using finite impulse response (FIR) filters for solving the stiff point kinetics equation with feedback. The point kinetics equations are converted into convolution equation by applying discrete Z transform. The power and precursor concentrations, appearing in the point kinetics equations, are written in terms of convolution equation with different impulse response functions. The impulse response functions characterize the FIR filter. This method is applied to estimate the transients in few benchmark thermal reactors for different types of reactivity perturbations with temperature feedback, i.e., step, ramp, and oscillatory reactivity inputs. This method has high stability, i.e., a small change in the time step of the order of 5 or 10 does not lead to large error in the solution. The transients estimated by this method are compared with other standard methods and they are found to be in good agreement.

Keywords Finite impulse response · Reactor · Transient

1 Introduction

The power transients in nuclear reactors are estimated by solving the time-dependant neutron diffusion equation in three dimensions. For small reactors, the point kinetics equations are sufficient in predicting the power transients caused by reactivity perturbations. The prediction of reactor power under reactivity perturbation is important from the safety point of view. The point kinetics equations describe the space-independent time-evolution of nuclear reactor power and precursor

Dr. M. M. A. Razak (✉)
Bhabha Atomic Research Centre Facilities (BARC-Facilities), Kalpakkam, Tamil Nadu, India
e-mail: mmar@igcar.gov.in

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_5

concentrations under reactivity perturbation. The point kinetics equations are stiff differential equations, and they require a very small time step to solve the equations. There are various methods to solve the point kinetics equations. Aboanber and Nahla [1, 2] developed the analytical inversion method for solving the point reactor kinetics equations with temperature feedback. Nahla [3] applied Taylor's Series Method (TSM) for solving the point kinetics equations. Aboanber [4] and Nahla [5] developed the analytical exponential method and the generalized Runge–Kutta method for solving the point kinetics equations. Li et al. [6] presented the better basis function (BBF) method for solving the point kinetics equations. Recently the modified exponential time differencing method was developed [7] to solve the point kinetics equations using large time step. The major constraint in solving the stiff point kinetics equations is the proper selection of time step. In most of the cases, a small change in the time step may lead to large error in the solution of point kinetics equation.

In the present work, a new computational method is developed using the finite impulse response (FIR) filters for solving the reactor point kinetics equations with feedback. According to this new computational method, the power and precursor concentrations, appearing in the point kinetics equations, are written as convolution integrals. The convolution integrals are solved using discrete Z transform. By applying inverse Z transform, the power and precursor concentrations are written as simple convolution equation with different impulse response functions. The impulse response functions characterize the FIR filters. Here, the impulse response functions are chosen according to the type of reactivity perturbation. By appropriately choosing the impulse response functions, the FIR filters can be designed for solving the point kinetics equations with feedback. The impulse response functions are different for power and precursor concentrations. The impulse response functions are found to be stable and possess finite radius of convergence. This new computational method is applied to estimate the nuclear reactor power transient in few benchmark thermal reactors for different types of reactivity perturbations, i.e., step, ramp, and oscillatory. In all the cases, the estimated power transient is found to be in good agreement with the standard methods. The advantage of this computational method is that the power transient can be estimated using large time step without losing accuracy, and this method has high stability, i.e., a change in the sampling time interval by a factor of 5 or 10 does not alter the solution to a larger extent. In all the cases, the estimated power transient, for various types of reactivity perturbations with feedback, is found to be in good agreement with the reference results. A scheme to choose the sampling time interval is also discussed.

2 Point Kinetics Equations and FIR Filters

Consider the point kinetics equations [8] describing the nuclear reactor power transient:

$$\frac{dp(t)}{dt} = \left(\frac{\rho(t) - \beta}{\Lambda}\right) p(t) + \sum_{i=1}^{6} \lambda_i C_i(t) \tag{1}$$

$$\frac{dC_i}{dt} = \left(\frac{\beta_i}{\Lambda}\right) p(t) - \lambda_i C_i, \quad (i = 1, 2, \ldots 6) \tag{2}$$

In the above Eqs. (1) and (2), p is the power, Λ is the prompt neutron generation time, β_i is the effective fraction of the ith group of delayed neutrons, β is the total effective fraction of delayed neutrons $\left(\beta = \sum_{i=1}^{6} \beta_i\right)$, and λ_i and C_i are the decay constant and precursor concentration of the ith group of the delayed neutron. The initial conditions of the point kinetics equations are given as $p(t = 0) = p_0$, $c_i\ (t = 0) = \frac{\beta_i}{\Lambda \lambda_i} p_0$, where p_0 is the steady state power before the introduction of any external reactivity. In the above equation, $\rho(t) = \rho_{\text{ex}}(t) + \rho_{\text{fb}}(t)$ is the net reactivity acting on the reactor, $\rho_{\text{ex}}(t)$ is the external reactivity, and $\rho_{\text{fb}}(t)$ is the feedback reactivity. In the case of constant reactivity insertion (without feedback), $\rho(t) = \rho_{\text{ex}}(t) = \rho_0$, and the solution of Eqs. (1) and (2) can be written as:

$$p(t) = \sum_{i=1}^{6} \lambda_i \int_{-\infty}^{t} e^{\left(\frac{\rho_0 - \beta}{\Lambda}\right)(t-\tau)} C_i(\tau)\, d\tau \tag{3}$$

$$C_i(t) = \left(\frac{\beta_i}{\Lambda}\right) \int_{-\infty}^{t} e^{-\lambda_i (t-\tau)} p(\tau)\, d\tau \tag{4}$$

Equations (3) and (4) are rewritten as:

$$p(t) = \sum_{i=1}^{6} \lambda_i \int_{-\infty}^{0} e^{\left(\frac{\rho_0 - \beta}{\Lambda}\right)(t-\tau)} C_i(\tau)\, d\tau + \sum_{i=1}^{6} \lambda_i \int_{0}^{t} e^{\left(\frac{\rho_0 - \beta}{\Lambda}\right)(t-\tau)} C_i(\tau)\, d\tau \tag{5a}$$

$$C_i(t) = \left(\frac{\beta_i}{\Lambda}\right) \int_{-\infty}^{0} e^{-\lambda_i (t-\tau)} p(\tau)\, d\tau + \left(\frac{\beta_i}{\Lambda}\right) \int_{0}^{t} e^{-\lambda_i (t-\tau)} p(\tau)\, d\tau \tag{5b}$$

It is assumed that before the application of reactivity perturbation, i.e., $t \leq 0$, the reactor is at constant power, i.e., $p(t) = p_0$, $C_i(t) = C_0$ and net reactivity acting on the reactor is zero. Under this assumption, Eqs. (5a) and (5b) are rewritten as:

$$p(t) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)t} + \sum_{i=1}^{6} \lambda_i \int_0^t e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)(t-\tau)} C_i\,(\tau)\,d\tau \tag{6a}$$

$$C_i(t) = C_0 e^{-\lambda_i t} + \left(\frac{\beta_i}{\Lambda}\right) \int_0^t e^{-\lambda_i (t-\tau)}\, p\,(\tau)\,d\tau \tag{6b}$$

The integrals appearing in Eqs. (6a) and (6b) are convolution integrals. Using Z transform [9], the convolution integrals (Eqs. 6a and 6b) are written as:

$$p(t) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)t} + T_s \sum_{i=1}^{6} \lambda_i g(Z) C_i(Z) \tag{7a}$$

$$C_i(t) = C_0 e^{-\lambda_i t} + T_s \left(\frac{\beta_i}{\Lambda}\right) h(Z) p(Z) \tag{7b}$$

where T_s is the sampling period,

$$\begin{aligned} &g(Z) = \sum_{n=0}^{\infty} g\,[n]\, z^{-n}, \quad h(Z) = \sum_{n=0}^{\infty} h\,[n]\, z^{-n}, \quad p(Z) = \sum_{n=0}^{\infty} p\,[n]\, z^{-n}, \\ &C_i(Z) = \sum_{n=0}^{\infty} C_i\,[n]\, z^{-n}, \quad g\,[n] = e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)n} \text{ and } h_i\,[n] = e^{-\lambda_i n} \end{aligned} \tag{8}$$

Using Eq. (8) and making use of inverse Z transform [9], the power and precursor concentrations (Eqs. 7a and 7b) are written as:

$$p(n) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)n} + T_s \sum_{i=1}^{6} \lambda_i \sum_{m=0}^{n} g\,[n-m]\, C_i\,[n] \tag{9}$$

$$C_i(n) = C_0 e^{-\lambda_i n} + T_s \left(\frac{\beta_i}{\Lambda}\right) \sum_{m=0}^{n} h_i\,[n-m]\, p\,[n] \tag{10}$$

Equations (9) and (10) are the representation of finite impulse response (FIR) filters. In the above equations, $g\,[n] = e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)n}$ is the impulse response function for calculating the power (step reactivity without feedback) and $h_i\,[n] = e^{-\lambda_i n}$ is the impulse response function for calculating the precursor concentration. Here, the FIR filters, (Eqs. 9 and 10), are coupled, i.e., to calculate $p(n)$, the value of $C_i[n]$

is required and to calculate $C_i[n]$, the value of $p(n)$ is required. The power and precursor concentration are obtained from coupled FIR filters as follows. First, an initial guess about $p(n, n > 1)$ is assumed and this is used to get the value of $C_i(n)$. This $C_i(n)$ is again used to get the value of $p(n)$. This process is repeated iteratively till the values of $p(n)$ and $C_i(n)$ are converged. The coupled form of realization of FIR filters for solving the point kinetics equations (Eqs. 9 and 10) with one group of delayed neutron precursor is shown in Fig. 1.

Denoting $\sum_{n=0}^{\infty} g[n-m]\, C_i[n] = g[n] * C_i[n] = C_i[n] * g[n] = y_1[n]$ and $\sum_{n=0}^{\infty} h_i[n-m]\, p[n] = h_i[n] * p[n] = p[n] * h_i[n] = y_{2i}[n]$, the power and precursor concentrations (Eqs. 9 and 10) are rewritten as:

$$p(n) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)n} + T_s \sum_{i=1}^{6} \lambda_i y_1[n] \tag{11}$$

$$C_i(n) = C_0 e^{-\lambda_i n} + T_s \left(\frac{\beta_i}{\Lambda}\right) y_{2i}[n] \tag{12}$$

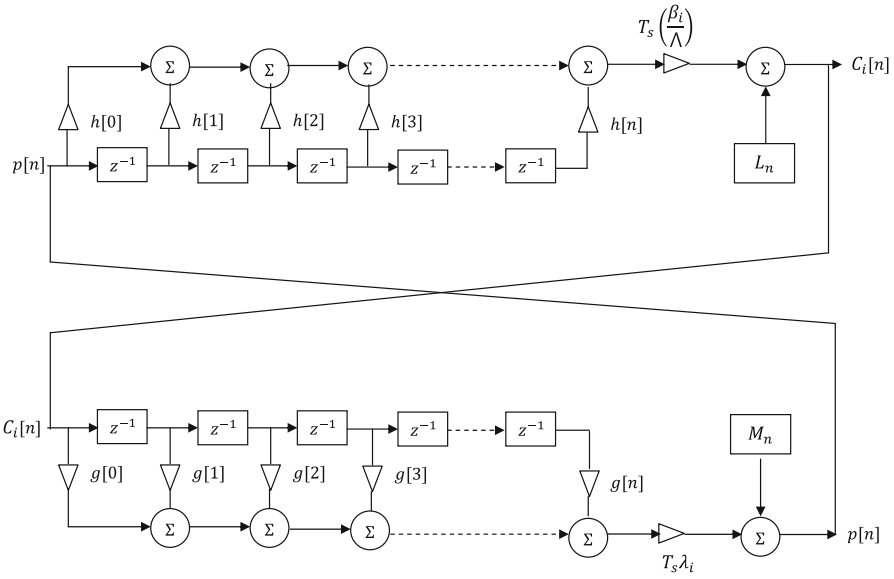


Fig. 1 Realization of coupled FIR filters for solving the point kinetics equations for step reactivity without feedback (assuming one group delayed neutron precursor). $L_n = C_0 e^{-(\lambda_i T_s)n}$ and $M_n = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)T_s n}$

Equations (11) and (12) do not satisfy the initial boundary condition, i.e., to satisfy the initial condition, the impulse response functions, $y_1[n]$ and $y_{2i}[n]$, are improved such that:

$$\tilde{y_1}[n] = y_1[n] - \frac{1}{2}\left[g[n]\,C_i[0] + g[0]\,C_i[n]\right] \tag{13}$$

$$\tilde{y_{2i}}[n] = y_{2i}[n] - \frac{1}{2}\left[h_i[n]\,p[0] + h_i[0]\,p[n]\right] \tag{14}$$

Using the improved impulse response functions, (Eqs. 13 and 14), the FIR filter representations of power and precursor concentrations are given as:

$$p(n) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)n} + T_s \sum_{i=1}^{6} \lambda_i \tilde{y_1}[n] \tag{15}$$

$$C_i(n) = C_0 e^{-\lambda_i n} + T_s \left(\frac{\beta_i}{\Lambda}\right) \tilde{y_{2i}}[n] \tag{16}$$

3 Selection of Sampling Time Interval T_s

For step reactivity (constant input) insertions ($|\rho_0| < \beta$) without feedback, the impulse response functions for power and precursor concentrations are found to be $g[n] = e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)T_s n}$ and $h_i[n] = e^{-\lambda_i T_s n}$, respectively. In this case, the radius of convergence of the impulse response function $g[n]$ is given by $|Z| > e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)T_s}$ and the radius of convergence of $h_i[n]$ is given by $|Z| > e^{-\lambda_i T_s}$ for the precursor concentration "i". For minimum sampling time interval, the radius of convergence is 1 and for maximum sampling time interval, the radius of convergence is 0. In this way, the radius of convergence lies between zero and one, i.e., $0 < |Z| < 1$. This is shown in Figs. 2a, b for power and precursor concentrations. By increasing the number of terms in the summation in Eqs. (9) and (10), the power and precursor concentrations can be accurately estimated. In other words, for a given transient duration, by choosing small sampling time interval, power and precursor concentrations can be estimated accurately. This is equivalent to choosing the radius of convergence nearer to one. Hence by fixing the radius of convergence nearer to one, the sampling time interval, T_s, can be estimated. In the present case, the radius of convergence is fixed as 0.9 and the sampling time interval, for power, is found to be $T_s = \frac{\log_e(0.9)}{\left(\frac{\rho_0-\beta}{\Lambda}\right)}$. In a similar way, the sampling time interval for precursor

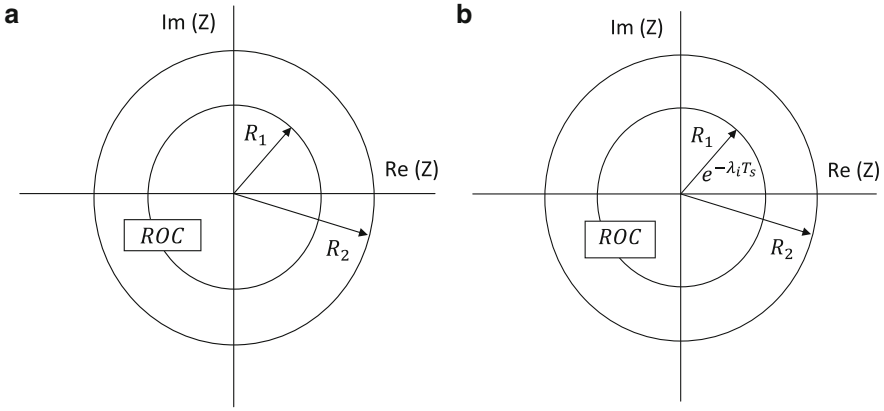


Fig. 2 (**a**) Region of convergence (ROC) of impulse response function $g[n]$ for power under step reactivity of insertion ($|\rho_0| < \beta$) without feedback. The region of convergence is $0 < |Z| < 1$, $R_1 = e^{\left(\frac{\rho_0 - \beta}{\Lambda}\right)T_s}$, and $R_2 = 1$. (**b**) Region of convergence (ROC) of impulse response function $h_i[n]$ for precursor concentration. The region of convergence is $0 < |Z| < 1$, $R_1 = e^{e^{-\lambda_i T_s}}$ and $R_2 = 1$

concentration (using $h_i[n]$) is found to be $T_s(i) = \frac{\log_e(0.9)}{\lambda_i}$. The minimum of T_s and $T_s(i)$ is taken as the sampling time interval.

4 Numerical Results

4.1 Transient from Step Reactivity Without Feedback

Consider the power transients of the thermal reactor described by [3]. The decay constants of the neutron precursors and the delayed neutron fractions of the thermal reactor are taken as $\lambda_1 = 0.0127\,s^{-1}$, $\lambda_2 = 0.0317\,s^{-1}$, $\lambda_3 = 0.115\,s^{-1}$, $\lambda_4 = 0.311\,s^{-1}$, $\lambda_5 = 1.4\,s^{-1}$, $\lambda_6 = 3.87\,s^{-1}$, $\beta_1 = 0.000285$, $\beta_2 = 0.0015975$, $\beta_3 = 0.00141$, $\beta_4 = 0.0030525$, $\beta_5 = 0.00096$, $\beta_6 = 0.000195$, and $\Lambda = 5.0 \times 10^{-4}$ s. Step reactivities $\rho_0 = -1\$$, $\rho_0 = -0.5\$$, $\rho_0 = +0.5\$$ and $\rho_0 = 1.0\$$ are inserted and the resulting power transient is computed using coupled FIR filters. Table 1 shows the values of the power transients obtained from coupled FIR filters along with the exact values given by Nahla [3]. The absolute errors, $|(X_{cal} - X_{exact})|$, are shown in Table 1. From the Table 1, it is observed that the coupled FIR method is capable of estimating the transient to a good accuracy. It is also shown in Table 2 that as the sampling time interval is changed by a factor of 10 or 20, the error in the estimation of power transient is small, indicating that this method has high stability against the change in the sampling time interval. The impulse response functions for power and precursor concentrations are found

Table 1 The power estimated by the coupled FIR filters and the exact values (Nahla [3])

Reactivity	Time	Exact value	Coupled FIR method ($T_s = 1.0$e-3s)	Absolute error
−1.0$	1.00	0.43333	0.43691	0.00358
	10.0	0.23611	0.23687	0.00076
−0.5$	1.00	0.60705	0.61044	0.00339
	10.0	0.39607	0.39701	0.00094
+0.5$	1.00	2.51149	2.46761	0.04388
	10.0	14.2150	14.0498	0.16520
+1.0$	0.50	10.3562	10.3531	0.00310
	1.00	32.1448	32.1356	0.00920

Table 2 The absolute error in the estimation of power transient as the sampling time interval (T_s) is varied

T_s	Reactivity	Time	Exact value	Coupled FIR method	Absolute error
0.001 s	+1.0$	1.0	32.1356	32.1835	0.04790
0.01 s	+1.0$	1.0	32.1356	31.8037	0.37980
0.02 s	+1.0$	1.0	32.1356	31.4398	0.74370

to be $g[n] = e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)n}$ and $h_i[n] = e^{-\lambda_i n}$, respectively. In this case, the radius of convergence of the impulse response function $g[n]$ is given by $|Z| > e^{\left(\frac{\rho_0-\beta}{\Lambda}\right)}$ and the radius of convergence of $h_i[n]$ is given by $|Z| > e^{-\lambda_i}$ for the precursor concentration "i". In general, the radius of convergence of $h_i[n]$ can be taken to be $|Z| > e^{-\lambda_0}$, where λ_0 is the minimum value of decay constant of the precursor group.

4.2 Transient from Step Reactivity with Temperature Feedback

Consider another example of thermal reactor described by Nahla [3] with the following parameters: $\lambda_1 = 0.0124\,\mathrm{s}^{-1}$, $\lambda_2 = 0.0305\,\mathrm{s}^{-1}$, $\lambda_3 = 0.111\,\mathrm{s}^{-1}$, $\lambda_4 = 0.301\,\mathrm{s}^{-1}$, $\lambda_5 = 1.13\,\mathrm{s}^{-1}$, $\lambda_6 = 3.0\,\mathrm{s}^{-1}$, $\beta_1 = 0.00021$, $\beta_2 = 0.00141$, $\beta_3 = 0.00127$, $\beta_4 = 0.00255$, $\beta_5 = 0.00074$, $\beta_6 = 0.00027$, and $\Lambda = 5.0 \times 10^{-5}$ s. A step reactivity $\rho_0 = 0.5$\$ is inserted, and the temperature rise ($T(t)$) with power ($p(t)$) in the reactor is given by:

$$\frac{\partial T(t)}{\partial t} = 0.05\ p(t)^{\circ}\mathrm{C/s}$$

The feedback reactivity is given by [3]:

$$\frac{\partial \rho_{\mathrm{fb}}}{\partial T} = -5.0 \times 10^{-5} \left(\frac{\Delta k}{k}\right)/^{\circ}\mathrm{C}$$

Table 3 The peak power computed using coupled FIR filter ($T_s = 1.0e\text{-}3$) for various step reactivity insertions with temperature feedback

Reactivity	Peak power		Time (s) of occurrence of peak power	
	Coupled FIR filter	TSM	Coupled FIR filter	TSM
+0.5$	44.429	45.754	28.07	28.29
+1.0$	808.0851	807.8765	0.954	0.953
+1.2$	8020.365	8020.848	0.323	0.317
+1.5$	43,023.16	43,021.00	0.174	0.168
+2.0$	167,844.6	167,739.00	0.103	0.098

The estimated peak power is compared with Taylor Series Method (TSM) (Nahla [3])

With temperature feedback, the power and precursor concentration are given by:

$$p(n) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)n} + T_s \sum_{i=1}^{6} \lambda_i \sum_{m=0}^{n} g\,[n-m]\, C_i\,[n] + T_s \sum_{m=0}^{n} g\,[n-m] \left(\frac{\rho_{fb}\,[n]\, p\,[n]}{\Lambda}\right) \tag{17}$$

$$C_i(n) = C_0 e^{-\lambda_i n} + T_s \left(\frac{\beta_i}{\Lambda}\right) \sum_{m=0}^{n} h_i\,[n-m]\, p\,[n] \tag{18}$$

The peak power and the time of occurrence of peak power, under the temperature feedback, are estimated using the coupled FIR filters for various step reactivity insertions. The results are given in Table 3 along with that obtained using Taylor series method (TSM) [3].

4.3 Transient from Ramp Reactivity Without Feedback

4.3.1 Transient from Positive Ramp Reactivity

Consider an example of thermal reactor described by Nahla [5], with the following parameters: $\lambda_1 = 0.0127\,s^{-1}$, $\lambda_2 = 0.0317\,s^{-1}$, $\lambda_3 = 0.115\,s^{-1}$, $\lambda_4 = 0.311\,s^{-1}$, $\lambda_5 = 1.4\,s^{-1}$, $\lambda_6 = 3.87\,s^{-1}$, $\beta_1 = 0.000266$, $\beta_2 = 0.001491$, $\beta_3 = 0.001316$, $\beta_4 = 0.002849$, $\beta_5 = 0.000896$, $\beta_6 = 0.000182$, and $\Lambda = 2.0 \times 10^{-5}$ s. A positive ramp reactivity of the form $\rho(t) = (0.25\$)t/s$ and $\rho(t) = (0.5\$)t/s$ is inserted in the reactor, the transient following this reactivity is estimated by coupled FIR filter, and the result is compared with that of GAEM method [5]. The results are given in Tables 4 and 5. In this case, the power and precursor concentrations are given by:

Table 4 The power transient computed using coupled FIR filter (T_s = 1.0e-3) for ramp reactivity 0.25$/s

Time	Coupled FIR method (T_s = 5.0e-5s) 0.25$/s	GAEM	Absolute error
0.25	1.070897	1.069541	0.001356
0.50	1.159835	1.156694	0.003141
0.75	1.271031	1.265331	0.005700
1.00	1.411403	1.401981	0.009422

The power transient is compared with the GAEM method (Nahla [3]). The absolute error is shown

Table 5 The power is computed using coupled FIR filter (T_s = 1.0e-3) for ramp reactivity 0.50$/s, and it is compared with the GAEM method (Nahla [3])

Time	Coupled FIR method (T_s = 5.0e-5s) 0.5$/s	GAEM	Absolute error
0.25	1.152200	1.149200	0.00300
0.50	1.377465	1.368927	0.00853
0.75	1.727601	1.707600	0.02000
1.00	2.322041	2.275271	0.04677

The absolute error is shown

$$p(n) = p_0 e^{\left(\frac{-\beta}{\Lambda}\right)n} + T_s \sum_{i=1}^{6} \lambda_i \sum_{m=0}^{n} k\,[n-m]\, C_i\,[n] + T_s \sum_{m=0}^{n} k\,[n-m] \left(\frac{\rho_{ex}\,[n]\, p\,[n]}{\Lambda}\right) \tag{19}$$

$$C_i(n) = C_0 e^{-\lambda_i n} + T_s \left(\frac{\beta_i}{\Lambda}\right) \sum_{m=0}^{n} h_i\,[n-m]\, p\,[n] \tag{20}$$

In the above equations (Eqs. 19 and 20), the impulse response function $k\,[n] = e^{\left(\frac{-\beta}{\Lambda}\right)n}$ and $\rho_{ex}(t) = 0.1\beta t$. In this case, the radius of convergence of the impulse response function $k[n]$ is given by $|Z| > e^{\left(\frac{-\beta}{\Lambda}\right)}$, and the radius of convergence of $h_i[n]$ is given by $|Z| > e^{-\lambda_i}$ for the precursor concentration "i".

4.3.2 Transient from Negative Ramp Reactivity

Consider another example of thermal reactor described by Li et al. [6], with the following parameters: $\lambda_1 = 0.0127\,s^{-1}$, $\lambda_2 = 0.0317\,s^{-1}$, $\lambda_3 = 0.115\,s^{-1}$, $\lambda_4 = 0.311\,s^{-1}$, $\lambda_5 = 1.4\,s^{-1}$, $\lambda_6 = 3.87\,s^{-1}$, $\beta_1 = 0.000266$, $\beta_2 = 0.001491$, $\beta_3 = 0.001316$, $\beta_4 = 0.002849$, $\beta_5 = 0.000896$, $\beta_6 = 0.000182$, and $\Lambda = 2.0 \times 10^{-5}$ s. A negative ramp reactivity of the form $\rho(t) = -0.1\$\ t/s$ is inserted in the reactor, the transient following this reactivity is estimated by coupled

Table 6 The power is computed using coupled FIR filter (T_s = 1.0e-3) for negative ramp reactivity, −0.1$/s, and the power transient is compared with the GAEM method (Nahla [3])

Time	Coupled FIR method	TSM	Absolute error
2.0	0.786412	0.791955	0.00554
4.0	0.604639	0.612976	0.00834
6.0	0.464981	0.474027	0.00905
8.0	0.360466	0.369145	0.00868
10.0	0.282778	0.290636	0.00786

The absolute error is shown

Table 7 The power is computed using coupled FIR filter for sinusoidal reactivity insertion $\rho(t) = 0.001 \sin(4\pi t)$ and compared with the modified ETD method (Mohideen Abdul Razak and Devan [7])

Time	Coupled FIR method (T_s = 1.0e-4s)	Modified ETD method	Absolute error
0.0	1.000000	1.0000000	0.000000
0.4	0.876102	0.8828488	0.006747
0.8	0.932204	0.9334387	0.001235
1.2	1.123357	1.1070115	0.016345
1.6	1.182720	1.1615297	0.02119
2.0	1.002299	0.9992939	0.003005
2.4	0.880687	0.8857330	0.005046
2.8	0.937168	0.9366979	0.00047
3.2	1.129228	1.1108737	0.018354
3.6	1.188583	1.1653616	0.023221
4.0	1.007087	1.0024811	0.004606
4.4	0.884874	0.8885748	0.003701
4.8	0.941632	0.9397346	0.001897
5.0	1.009291	1.0039612	0.00533

The absolute error is shown

FIR, and the result is compared with that of Taylor Series Method [3]. The results are shown in Table 6.

4.4 Transient from Oscillatory Reactivity

The power transients caused by a sinusoidal reactivity insertion are analyzed here for the thermal reactor described by Li et al. [6]. The delayed neutron precursor parameters are given as follows: $\lambda_1 = 0.0127\,s^{-1}$, $\lambda_2 = 0.0317\,s^{-1}$, $\lambda_3 = 0.115\,s^{-1}$, $\lambda_4 = 0.311\,s^{-1}$, $\lambda_5 = 1.4\,s^{-1}$, $\lambda_6 = 3.87\,s^{-1}$, $\beta_1 = 0.000266$, $\beta_2 = 0.001491$, $\beta_3 = 0.001316$, $\beta_4 = 0.002849$, $\beta_5 = 0.000896$, $\beta_6 = 0.000182$, and $\Lambda = 2.0 \times 10^{-5}$ s. A sinusoidal reactivity of the form $\rho(t) = 0.001 \sin(4\pi t)$ is inserted in the reactor, the transient following this reactivity is estimated by coupled FIR method, and the result is compared with that of the modified exponential time differencing method [7]. The estimated power transient is given in Table 7.

5 Conclusion

A new computational method for estimating the nuclear reactor power transients using finite impulse response (FIR) filter is developed and presented. The nuclear power transients, in small reactors, are estimated by solving the point kinetics equations. According to this method, the stiff point kinetics equations are written as convolution integrals. The convolution integrals are converted into simple algebraic equations using discrete Z transform. Here, the power and precursor concentrations are written as simple algebraic equations. This method has less computational effort in estimating the transients. The impulse response functions, involved in the convolution, characterize the FIR filters. Here, the reactor power and precursor concentrations are represented by two different FIR filters. The impulse response function is different for different types of reactivity perturbation. The impulse response functions are found to be stable, and they have finite radius of convergence. This method is applied to estimate the power transient of thermal reactor for step (constant) reactivity perturbation with temperature feedback. The power transients estimated with temperature feedback are found to be in good agreement with standard results. In a similar manner, the method is also applied to estimate the power transients for ramp reactivity input. The estimated power transients under ramp reactivity perturbation are found to be in good agreement with reference results. It is also shown that this method has high stability, i.e., any change in the time step by a factor of 10 or 20 will not lead to large error in the estimation of power. From the comparisons of results, it can be concluded that this method is capable of estimating the reactor power transients for various types of reactivity perturbations with feedback. This method can be easily designed and implemented for estimating the power transient with feedback. A scheme to choose the sampling time interval for solving the stiff point kinetics equations is also established.

References

1. A.E. Aboanber, A.A. Nahla, Generalization of the analytical inverse method for the solution of point kinetics equations. J. Phys. A Math. Gen. **35**, 3245–3263 (2002)
2. A.E. Aboanber, A.A. Nahla, Solution of the point kinetics equations in the presence of Newtonian temperature feedback by Pade approximation via the analytical inversion method. J. Phys. A Math. Gen. **35**, 9609–9627 (2002)
3. A.A. Nahla, Taylor series method for solving the nonlinear point kinetics equations. Nucl. Eng. Des. **241**, 1592–1595 (2011)
4. A.E. Aboanber, Stability of generalized RungeKutta methods for stiff kinetics coupled differential equations. J. Phys. A Math. Gen. **30**(2006), 1859–1876 (2006)
5. A.A. Nahla, Generalization of the analytical exponential model to solve the point kinetics equations of Be- and D2O-moderated reactors. Nucl. Eng. Des. **238**, 2648–2653 (2008)
6. H. Li, W. Chen, L. Luo, Q. Zhu, A new integral method for solving the point reactor neutron kinetics equations. Ann. Nucl. Energy **36**, 427–432 (2009)

7. M.M.A. Razak, K. Devan, The modified exponential time differencing method for solving the reactor point kinetics equations. Ann. Nucl. Energy **98**, 1–10 (2015)
8. J.J. Duderstadt, J.J.L.J. Hamilton, *Nuclear Reactor Analysis*, 2nd edn. (Wiley, New York, 1976)
9. J.G. Proakis, D.G. Manolakis, *Digital Signal Processing* (Prentice-Hall of India, New Delhi, 2000)

Convergence of Substructuring Domain Decomposition Methods for Hamilton–Jacobi Equation

Bankim C. Mandal

Abstract I present the convergence behavior of classical Schwarz, Dirichlet–Neumann, and Neumann–Neumann methods for the Hamilton–Jacobi equation. The methods are based on domain decomposition (DD) algorithms, where one semi-discretizes in time or fully discretizes the equation, and use DD method to the resulting equation in phase space. These methods are based on a non-overlapping spatial domain decomposition, and each iteration involves subdomain-solves with Dirichlet or Neumann boundary conditions. However, unlike for elliptic problems, each subdomain solve can also involve a solution in space and time, and the interface conditions may also be time-dependent in case of waveform relaxation version of these methods. I introduce a coarse grid correction to get rid of the convergence dependence on the number of subdomains. Numerical results are shown to illustrate the performance of these algorithms with benchmark examples.

Keywords Dirichlet–Neumann · Neumann–Neumann · Waveform relaxation · Domain decomposition methods · Hamilton–Jacobi equation

Mathematics Subject Classification 65M55, 65Y05, 65M15

1 Introduction

There is a wide variety of plasma simulations where the resolution of key length scales (such as the Debye length and gradients in the self-consistent electromagnetic fields) is crucial, for example, plasma arc formation, virtual cathodes, and plasma opening switches. One common approach to resolve these length scales is to employ a kinetic representation of the plasma; however, such simulations are computational and memory-intensive. Domain decomposition (DD) methods are

B. C. Mandal (✉)
School of Basic Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar, India
e-mail: bmandal@iitbbs.ac.in

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_6

established approach for distributing computational effort and in-memory storage across many computational nodes. However, DD methods have not been formulated for implicit solutions to the Vlasov equations. Furthermore, DD methods have not been formulated for implicit solutions to a wider class of problems, namely the Hamilton–Jacobi (H–J) equations; Pfirsch [22, 23] first showed that the Vlasov equation can be reformulated as an the H–J equation, and for more details, see [24, 28]. Starting from Schwarz [26] through Picard [25], Lindelöf [14], and Bjørstad and Widlund [1], the development of DDM has come a long way. In this chapter, I am particularly interested in DD methods for solving space–time problems. There are mainly two approaches to solve them: firstly, the method of lines, i.e., the classical approach where the problem is discretized in time to obtain a sequence of elliptic problems, and then DDM [3, 4] are applied to the resulting problems. The disadvantage of this approach is that one is forced to include uniform time steps across the whole domain, which is restrictive for multi-scale problems. Classical Schwarz method for PDEs is based on local solutions by using Dirichlet boundary conditions on the artificial interfaces between the subdomains and iteration; see [15, 26]. Dirichlet–Neumann algorithm was first introduced by Bjørstad & Widlund [1] and further studied in [2, 17, 18]. The method is based on a non-overlapping domain decomposition in space, and the iteration requires subdomain solves with Dirichlet boundary conditions followed by subdomain solves with Neumann boundary conditions. The performance of the algorithm for elliptic problems is now well understood; see, for example, the book [27] and the references therein. The second approach is the waveform relaxation (WR) methods; here, one solves the space–time subproblems for the whole time window in one go.

There are multiple variants of the WR-type DDMs, such as Schwarz waveform relaxation (SWR) [7, 11], optimized SWR [8, 9], and more recently the DNWR [10, 19], and NNWR [12, 13, 16]. In these algorithms, one first decomposes the spatial domain into two or several overlapping or non-overlapping subdomains, followed by subdomain solves by adding suitable artificial boundary condition(s). These boundary conditions are usually called *transmission conditions* (TCs), which transmit information between neighboring subdomains via iterations. Depending on the nature of the TCs, one observes difference in convergence behavior.

I consider the following Hamilton–Jacobi equation as our guiding example:

$$\partial_t \phi + H(\nabla \phi) = 0, \quad \phi(x, 0) = \phi_0(x), \quad x \in \Omega, \quad t \in [0, T], \tag{1}$$

where H is a Hamiltonian, and periodic boundary conditions are imposed for $\Omega \subset \mathbb{R}^d$. The domain Ω is a bounded domain of regular or irregular shape; however, I focus on only regular-shaped domain for the discussion. The H–J equations appear in many applications, such as calculus of variations, control theory, and differential games. I consider here three particular types of DD methods, namely, Schwarz method, Dirichlet–Neumann method, and Neumann–Neumann method.

2 Problem Formulation and Domain Decomposition Methods

I start with a correlation between the H–J equation and conservation laws (CLs). Although the solution to (1) is continuous, it has discontinuous derivatives, even if $\phi_0(x)$ is a C^∞ function. By taking a spatial derivative of (1), we obtain

$$(\phi_x)_t + (H(\phi_x))_x = 0.$$

Setting $u := \phi_x$, we can rewrite the above equation as

$$u_t + H(u)_x = 0, \tag{2}$$

which is of the form of a scalar CL. Note that, the solution u to (2) is the derivative of ϕ to (1). Conversely, the solution ϕ to (1) is the integral of a solution u to (2). For more details, see [5, 20, 21]. However, the analogy between the CLs and the H–J equation fails for higher spatial dimensions. The correspondence is somehow effective for some specific the H–J equations, see [21]. So, effective numerical schemes for solving hyperbolic CL can be used to solve the H–J equation.

A standard (mono-domain) approach for the H–J equation is to replace the Hamiltonian with a numerical Hamiltonian, $\hat{H}$, discretize equation (1) in space and time, and solve the resulting non-linear system of equations using a Newton–Raphson method at each time step. For example, one might take the Lax–Friedrich approximation,

$$\hat{H} = H\left(\frac{\phi_x^+ + \phi_x^-}{2}, \frac{\phi_y^+ + \phi_y^-}{2}, \frac{\phi_z^+ + \phi_z^-}{2}\right) - \alpha_x\left(\frac{\phi_x^+ - \phi_x^-}{2}\right) - \alpha_y\left(\frac{\phi_x^+ - \phi_x^-}{2}\right) - \alpha_z\left(\frac{\phi_x^+ - \phi_x^-}{2}\right),$$

where $\alpha_x = \max|\frac{\partial}{\partial u}H(u, v, w)|$, $\alpha_y = \max|\frac{\partial}{\partial v}H(u, v, w)|$, $\alpha_z = \max|\frac{\partial}{\partial w}H(u, v, w)|$, and $\phi_i^\pm$ are the upwind and downwind approximation to the derivatives in the appropriate direction.

Now, I introduce domain decomposition methods for Hamilton–Jacobi equation. Suppose that the domain Ω is decomposed into N overlapping or non-overlapping subdomains, Ω_k, with $\Omega = \cup\Omega_k$ as in Fig. 1. DD methods decompose the original problem into smaller subproblems defined in each subdomain; the subproblems are coupled using transmission conditions $\mathcal{T}$ on the artificial (subdomain) boundaries, i.e.,

$$\partial_t\phi_k + \hat{H}(\nabla\phi_k) = 0, \quad \phi_k(x, 0) = \phi_0(x), \quad x \in \Omega_k, \quad t \in [0, T],$$
$$\mathcal{T}(\phi_k(z, t)) = \mathcal{T}(\phi_j(z, t)), \quad z \in \partial\Omega_k \cap \bar{\Omega}_j, \quad j = 1, \ldots, N.$$

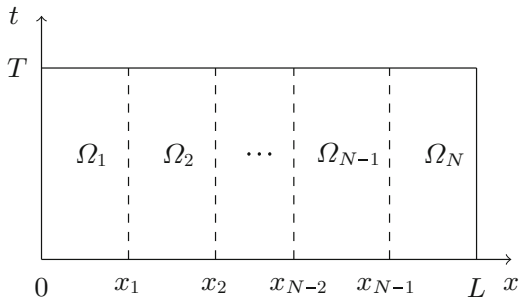


Fig. 1 Decomposition of a space–time domain, $[0, L] \times [0, T]$ domain

These coupled subproblems are decoupled using a Schwarz iteration [26]. Various transmission conditions have previously been analyzed for wide classes of PDEs, for example, for second-order parabolic and hyperbolic equations, see [9] and more recent development in [10], and for Maxwell's equation in [6]. However, no one has addressed DD solutions to Hamilton–Jacobi equations, which are particularly challenging because discontinuities in the derivatives can occur even if smooth initial data is chosen.

2.1 Classical Schwarz Algorithm

To start with, we apply classical Schwarz algorithm to (1) by splitting the spatial domain Ω into Ω_1 and Ω_2, with or without overlap. Suppose we use forward Euler in time and upwind scheme in space to discretize (1):

$$\phi_i^{n+1} = \phi_i^n - \triangle t\, H\left(\frac{\phi_i^n - \phi_{i-1}^n}{\triangle x}\right), \tag{3}$$

with $|H'(\alpha)| \leq \frac{\triangle x}{\triangle t}$. Since the scheme is explicit in nature, the DD method will produce the solution in the first iteration. Two different solution plots are given in Fig. 2, for both with and without overlap. The solutions after first time step are plotted. It is evident that, the DD method with explicit schemes only decouples the big problem into smaller subproblems and solves them to produce complete solution. Hence, it lacks the basic parallel nature of a DD algorithm. One thus needs to focus on implicit schemes. For example, we consider the linear advection equation and use implicit upwind scheme to discretize. Therefore, the classical Schwarz algorithm is given by

$$\begin{cases} \frac{\phi_i^{n+1} - \phi_i^n}{\triangle t} + k\frac{\phi_i^{n+1} - \phi_{i-1}^{n+1}}{\triangle x} = 0, \\ \phi_1^{n+1} = u_{\text{left}}^{n+1}, \\ \phi_J^{n+1} = \varphi_{1+l}^n, \end{cases} \quad \begin{cases} \frac{\varphi_i^{n+1} - \varphi_i^n}{\triangle t} + k\frac{\varphi_i^{n+1} - \varphi_{i-1}^{n+1}}{\triangle x} = 0, \\ \varphi_1^{n+1} = \phi_{J-l}^{n+1}, \\ \varphi_J^{n+1} = u_{\text{right}}^{n+1}, \end{cases} \tag{4}$$

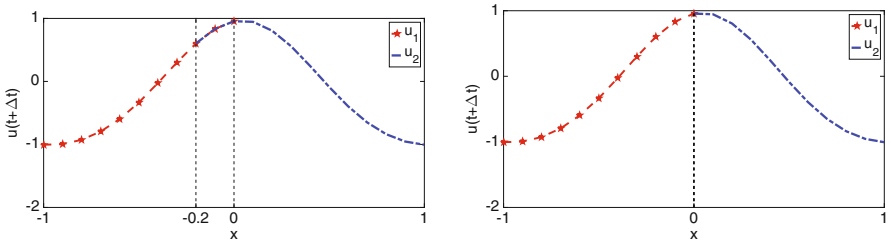


Fig. 2 Convergence of Schwarz algorithm with explicit upwind scheme for Hamilton–Jacobi equation: with overlap on the left and without overlap on the right panel

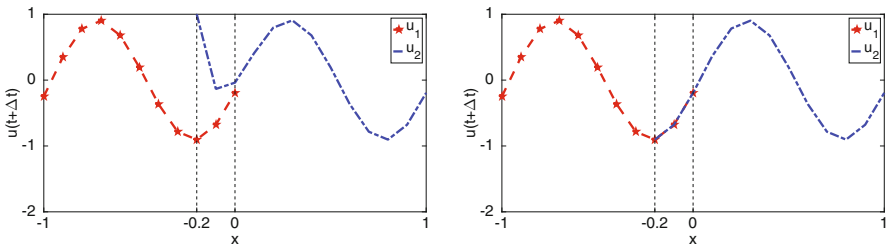


Fig. 3 Convergence of classical Schwarz algorithm with implicit upwind scheme for linear advection equation with overlap: solution after first iteration on the left panel and after second iteration on the right

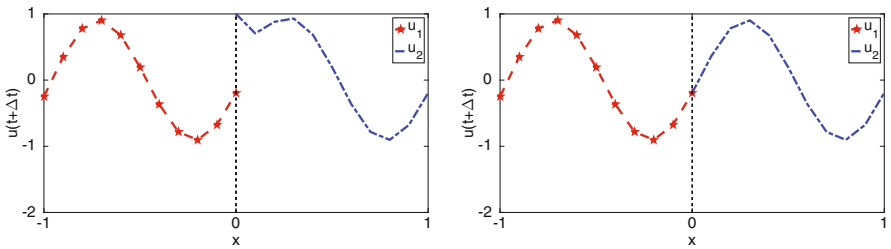


Fig. 4 Convergence of classical Schwarz algorithm with implicit upwind scheme for linear advection equation without overlap: solution after first iteration on the left panel and after second iteration on the right

where $\phi_i^n := \phi(x_i, t_n)$, u_{left} and u_{right} are given physical boundaries, and $\{\phi, \varphi\}$ are subdomain solutions. The algorithm is valid for both with overlap or without overlap ($l = 0$). In Fig. 3, I show two-step convergence of the solution after one time step in overlapping case. The non-overlapping case is plotted in Fig. 4.

2.2 Dirichlet–Neumann and Neumann–Neumann Algorithms

I introduce the DN and NN algorithms for the H–J equation, after semi-discretizing the equation in time. The DN method is given as follows: given initial guesses as interface values g^n,

$$\begin{cases} \frac{\phi_i^{n+1}-\phi_i^n}{\triangle t} + H(\nabla\phi_i^{n+1}) = 0, \\ \phi_1^{n+1} = u_{\text{left}}^{n+1}, \\ \phi_J^{n+1} = g^{n+1}, \end{cases} \quad \begin{cases} \frac{\varphi_i^{n+1}-\varphi_i^n}{\triangle t} + H(\nabla\varphi_i^{n+1}) = 0, \\ \partial_x\varphi_1^{n+1} = \partial_x\phi_J^{n+1}, \\ \varphi_J^{n+1} = u_{\text{right}}^{n+1}, \end{cases} \tag{5}$$

with the update condition $g^{n+1} = \theta g^{n+1} + (1-\theta)\varphi_1^{n+1}$ for $\theta \in (0, 1]$. I iterate at each time step n until convergence before proceeding to the next time step.

The NN method is given as follows: given initial guesses as interface values h^n, I compute the Dirichlet subproblems

$$\begin{cases} \frac{\phi_i^{n+1}-\phi_i^n}{\triangle t} + H(\nabla\phi_i^{n+1}) = 0, \\ \phi_1^{n+1} = u_{\text{left}}^{n+1}, \\ \phi_J^{n+1} = h^{n+1}, \end{cases} \quad \begin{cases} \frac{\varphi_i^{n+1}-\varphi_i^n}{\triangle t} + H(\nabla\varphi_i^{n+1}) = 0, \\ \varphi_1^{n+1} = h^{n+1}, \\ \varphi_J^{n+1} = u_{\text{right}}^{n+1}, \end{cases} \tag{6}$$

followed by the Neumann solves:

$$\begin{cases} \frac{\xi_i^{n+1}-\xi_i^n}{\triangle t} + H(\nabla\xi_i^{n+1}) = 0, \\ \xi_1^{n+1} = 0, \\ \partial_x\xi_J^{n+1} = \partial_x\phi_J^{n+1} - \partial_x\varphi_1^{n+1}, \end{cases} \quad \begin{cases} \frac{\eta_i^{n+1}-\eta_i^n}{\triangle t} + H(\nabla\eta_i^{n+1}) = 0, \\ -\partial_x\eta_J^{n+1} = \partial_x\phi_J^{n+1} - \partial_x\varphi_1^{n+1}, \\ \eta_J^{n+1} = 0, \end{cases} \tag{7}$$

with the updating step: $h^{n+1} = h^{n+1} - \theta\{\phi_J^{n+1} + \varphi_1^{n+1}\}$ for $\theta \in (0, 1]$.

On the contrary, in a waveform relaxation approach, one makes an initial guess of the solution on the artificial boundary, $\phi_k^{[0]}(z, t)$, where $z \in \partial\Omega_k \cap \bar{\Omega}_j$, $j = 1, \ldots, N$, and solves the decoupled problems

$$\partial_t\phi_k^{[\ell]} + \hat{H}(\nabla\phi_k^{[\ell]}) = 0, \quad \phi_k^{[\ell]}(x, 0) = \phi_0(x), \quad x \in \Omega_k, \quad t \in [0, T], \tag{8}$$

$$\mathcal{T}(\phi_k^{[\ell]}(z, t)) = \mathcal{T}(\phi_j^{[\ell-1]}(z, t)), \quad z \in \partial\Omega_k \cap \bar{\Omega}_j, \quad j = 1, \ldots, N. \tag{9}$$

The challenge is to find appropriate (possibly non-linear) transmission conditions (9), which result in a convergent scheme. The choice of transmission conditions is highly problem-dependent, and the convergence behavior differs based on the problem and the artificial conditions.

3 Numerical Illustrations

I implement these DD algorithms numerically for few standard benchmark model problems. I consider three cases: linear flux, convex, and non-convex flux functions. Since these problems are inherently non-linear, one needs to introduce Newton iteration to solve each subdomain problem, and then there is another iterative process for DD methods. I call the first one as inner iteration and the later outer iteration (Tables 1, 2, 3, 4).

(a) **Case 1:** Linear Flux, $H(u) = 3u$, $\phi_0(x) = -\cos(\pi x)$, $T = 1$, and $\Delta x = 1/400$.
(b) **Case 2a:** Convex Flux, $H(u) = (u+1)^2/2$, $\phi_0(x) = -\cos(\pi x)$, $T = 0.5/\pi^2$, and $\Delta x = 1/400$.
(c) **Case 2b:** Convex Flux, $H(u) = (u+1)^2/2$, $\phi_0(x) = \pi x/2 - \tan^{-1}(10^3 x)$, $T = 0.015/\pi^2$, and $\Delta x = 1/400$. For larger final time, mono-domain solution (Newton) does not converge!
(d) **Case 3:** Non-convex Flux, $H(u) = -\cos(u+1)$, $\phi_0(x) = -\cos(\pi x)$, $T = 0.1$, and $\Delta x = 1/400$.

3.1 Comparison with Waveform Relaxation Methods

Now, I compare the performance of classical methods with waveform relaxation-based methods for both convex and non-convex fluxes (Tables 5 and 6).

Table 1 Comparison of inner–outer iterations for linear flux function (Case 1)

Δt	Newton iteration	Schwarz iteration, $\delta = 2\Delta x$
1/20	3	4
1/40	3	3
1/80	3	2
1/160	3	2
1/320	3	2

Table 2 Comparison of inner–outer iterations for convex flux (Case 2a); X denotes the divergence of the method

Δt	Newton iteration	Schwarz iteration, $\delta = 2\Delta x$	Schwarz iteration, $\delta = 20\Delta x$
T/20	X	–	–
T/40	X	–	–
T/80	7	X	2
T/160	6	X	2
T/320	5	3	2

Table 3 Comparison of inner–outer iterations for convex flux and non-smooth initial condition (Case 2b)

Δt	Newton iteration	Schwarz iteration, $\delta = 2\Delta x$
T/20	–	–
T/40	12	X
T/80	10	5
T/160	8	5
T/320	7	3

Table 4 Comparison of inner–outer iterations for non-convex flux (Case 3)

Δt	Newton iteration	Schwarz iteration, $\delta = 2\Delta x$	Schwarz iteration, $\delta = 20\Delta x$
T/20	16	12	3
T/40	11	9	3
T/80	8	8	3
T/160	7	7	2
T/320	6	6	2

Table 5 Comparison in wall time between classical and waveform methods for convex flux (Case 1)

Δt	Methods	Wall time	DD iteration
1/40	CS	1.44	2
	WR	4.38	12
1/80	CS	2.41	2
	WR	7.07	12

Table 6 Comparison in wall time between classical and waveform methods for non-convex flux (Case 2)

Δt	Methods	Wall time	DD iteration
1/40	CS	2.43	2
	WR	14.38	26
1/80	CS	3.94	2
	WR	23.60	25

(a) **Case 1:** Convex Flux, $H(u) = (u+1)^2/2$, $\phi_0(x) = -\cos(\pi x)$, $T = 0.02/\pi^2$, and $\Delta x = 1/400$.
(b) **Case 2:** Non-convex Flux $H(u) = -\cos(u+1)$, $\phi_0(x) = -\cos(\pi x)$, $T = 0.1$, and $\Delta x = 1/400$.

4 Concluding Remarks

I have introduced domain decomposition methods for Hamilton–Jacobi equation. I formulate the Schwarz, the DN, and the NN algorithms to solve the underlying non-linear space–time problem. These algorithms involve Newton iterative method in each of the DD iteration. With numerical experiments, I have shown faster convergence in classical DD methods in comparison to waveform relaxation methods.

Acknowledgments I would like to express my gratitude to Dr. Benjamin W. Ong for his constant support and stimulating suggestions.

References

1. P.E. Bjørstad, O.B. Widlund, Iterative methods for the solution of elliptic problems on regions partitioned into substructures. SIAM J. Numer. Anal. **23**, 1097–1120 (1986)
2. J.H. Bramble, J.E. Pasciak, A.H. Schatz, An iterative method for elliptic problems on regions partitioned into substructures. Math. Comp. **46**, 361–369 (1986)
3. X.-C. Cai, Additive Schwarz algorithms for parabolic convection-diffusion equations. Numer. Math. **60**, 41–61 (1991)
4. X.-C. Cai, Multiplicative Schwarz methods for parabolic problems. SIAM J. Sci. Comput. **15**, 587–603 (1994)
5. M.G. Crandall, P.L. Lions, Two approximations of solutions of Hamilton-Jacobi equations. Math. Comput. **43**, 1–19 (1984)
6. B. Després, P. Joly, J.E. Roberts, *A Domain Decomposition Method for the Harmonic Maxwell Equation* (Amsterdam, 1992), pp. 475–484
7. M.J. Gander, *Overlapping Schwarz for Linear and Nonlinear Parabolic Problems* (1996)
8. M.J. Gander, L. Halpern, Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems. SIAM J. Numer. Anal. **45**, 666–697 (2007)
9. M.J. Gander, L. Halpern, F. Nataf, Optimal Schwarz waveform relaxation for the one dimensional wave equation. SIAM J. Numer. Anal. **41**, 1643–1681 (2003)
10. M.J. Gander, F. Kwok, B.C. Mandal, Dirichlet-Neumann and Neumann-Neumann waveform relaxation algorithms for parabolic problems. Electron. Trans. Numer. Anal. **45**, 424–456 (2016)
11. M.J. Gander, A.M. Stuart, Space-time continuous analysis of waveform relaxation for the heat equation. SIAM J. Sci. Comput. **19**, 2014–2031 (1998)
12. T.T.P. Hoang, Space-time domain decomposition methods for mixed formulations of flow and transport problems in porous media. Ph.D. thesis, University Paris 6, France, 2013
13. F. Kwok, in *Neumann-Neumann Waveform Relaxation for the Time-Dependent Heat Equation*, ed. by J. Erhel, M.J. Gander, L. Halpern, G. Pichot, T. Sassi, O.B. Widlund. Domain Decomposition in Science and Engineering XXI, vol. 98 (Springer, 2014), pp. 189–198
14. E. Lindelöf, Sur l'application des méthodes d'approximations successives à l'étude des intégrales réelles des équations différentielles ordinaires. J. de Math. Pures Appl., 117–128 (1894)
15. P.-L. Lions, *On the Schwarz alternating method I*, in First International Symposium on Domain Decomposition Methods for PDEs, Philadelphia, 1988, pp. 1–42
16. B.C. Mandal, Neumann-Neumann waveform relaxation algorithm in multiple subdomains for hyperbolic problems in 1D and 2D. Numer. Methods Part. Differ. Equ. **33**(2), 514–530 (2016)
17. L.D. Marini, A. Quarteroni, A relaxation procedure for domain decomposition methods using finite elements. Numer. Math. **55**, 575–598 (1989)
18. L. Martini, A. Quarteroni, An iterative procedure for domain decomposition methods: A finite element approach. SIAM Domain Decomposition Methods PDEs I, 129–143 (1988)
19. B.W. Ong, B.C. Mandal, Pipeline Implementations of Neumann–Neumann and Dirichlet–Neumann Waveform Relaxation Methods (to appear). arXiv:1605.08503
20. S. Osher, R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, vol. 153. Applied Mathematical Sciences (Springer, 2003)
21. S. Osher, C.-W. Shu, High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations. SIAM J. Numer. Anal. **28**, 907–922 (1991)
22. D. Pfirsch, Hamilton–Jacobi theory applied to Vlasov's equation. Nucl. Fusion 6 (1966)

23. D. Pfirsch, New variational formulation of Maxwell–Vlasov and guiding-center theories local charge and energy conservation laws. Z. Naturforsch. 39a (1984)
24. D. Pfirsch, P.J. Morrison, The energy-momentum tensor for the linearized Maxwell–Vlasov and kinetic guiding center theories. Phys. Fluids **3B**, 271–283 (1991)
25. E. Picard, Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires. J. Math. Pures Appl., 217–272 (1893)
26. H.A. Schwarz, Über einen Grenzübergang durch alternierendes Verfahren. Vierteljahrsschrift der Naturforschenden Gesellschaft Zürich **15**, 272–286 (1870)
27. A. Toselli, O. Widlund, *Domain Decomposition Methods—Algorithms and Theory*, vol. 34 of Springer Series in Computational Mathematics, (Springer, Berlin, 2005)
28. H. Ye, P.J. Morrison, Action principles for the Vlasov equation. Phys. Fluids **4B**, 771–776 (1992)

Dynamical Behaviour of Dengue: An SIR Epidemic Model

Sudipa Chauhan, Sumit Kaur Bhatia, and Simrat Chaudhary

Abstract In this chapter, we have demonstrated the dynamical behaviour of dengue using an SIR epidemic model, spanning both distributed and discrete time delays. The existence of boundary and interior equilibrium points has been studied. Furthermore, we have discussed the local stability of the equilibrium points. The disease-free equilibrium point is locally asymptotically stable if $R_0 < 1$ and $b > \beta_1 + \beta_2 e^{-b\tau}$ and unstable for $R_0 > 1$. The endemic equilibrium point is locally asymptotically stable for $[0,\tau^{'})$, and it undergoes Hopf bifurcation at $\tau = \tau^{'}$. The direction and stability of Hopf bifurcation have been established using the normal form theory and the centre manifold theorem, and lastly the analytical results are verified numerically, and further, sensitivity analysis is conducted to show how the periodic solution of the system is dependent upon delay, rate of infection, and birth and death rate.

Keywords Discrete delay · Distributed delay · Basic reproduction number · Local stability · Hopf bifurcation

Mathematics Subject Classification 34A38, 92D30

1 Introduction

From an endemic to a pandemic, dengue has stretched its length and breadth with an estimated 50 million infections per year across 100 countries approximately. This arbovirus is primarily attributed to the distribution of able mosquito vectors across tropical and sub-tropical areas. Prima facie here is the urban-adapted "Aedes aegypti." It breeds in dense man-made environments. The accelerated urbanization in Asia and Latin America propelled an increased population, leading to ample

S. Chauhan (✉) · S. K. Bhatia · S. Chaudhary
Department of Mathematics, Amity Institute of Applied Science, Amity University, Noida, UP, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_7

vector-breeding sites within densely packed urban communities and surrounding areas instigating its endemic nature. According to the World Health Organization (WHO), about 390 million cases of dengue fever occur worldwide each year, with around 96 million requiring medical treatment. Thailand has reached 10,446 recorded cases of dengue fever in 2018, as of mid-May, with 15 related deaths. Sri Lanka reported 80,732 cases of dengue fever, with 215 deaths from January to July 2017. New Delhi, India, reported an outbreak of dengue fever, with 1872 testing positive for the illness in September 2015. Dengue fever has been a recurrent problem in West Bengal with a major outbreak in 2012, which involved several districts of West Bengal [1]. And, the list is endless. Dengue virus infection is inapparent sometimes, but it can trigger versatile clinical manifestations starting from mild fever to life-threatening dengue shock syndrome [2]. The incubation period is approximately of 10–15 days. The patient starts showing symptom after incubation period. Till now, many mathematical models have been discussed by the researchers to study the qualitative and quantitative analysis of dengue or other epidemic disease [3–5]. One of the simplest compartmental models in epidemiology to formulate any disease dynamics is an SIR model. This consists of three compartments, S stands for the number of susceptible, I for the number of infected, and R for the recovered population. This was proposed by Kermack and McKendrick [6], and it is as follows:

$$\frac{dS}{dt} = -\beta SI \tag{1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I. \tag{3}$$

Here, β is the contact rate and γ is the recovery rate from the infected compartment. This model takes the population size to be fixed (i.e. there are no births, no deaths due to disease, or deaths by natural causes). Also, the incubation period (i.e. the period between exposure to an infection and the appearance of the first symptoms) of the infectious agent is instantaneous. But, simple mathematical models cannot be used to understand the rich dynamics of the disease like dengue. Since we know that time lag is present in the transmission phase of dengue, it is required to incorporate delay to study the rich dynamics of such models. Delay differential equations are widely used in epidemiology, and problems related to delay have been investigated by a number of authors [7–11]. Several authors have investigated this disease and presented their work regarding the same [12–16]. Recently, in 2019, the authors [17] developed a dengue transmission mathematical model with discrete time delays and estimated the reproduction number. However, they did not incorporate distributive delay in the model which is a more generalized case. The distributive delay is considered in reference to [18]. Furthermore, the existence of periodic solution and

the direction of Hopf bifurcation have not been discussed in their paper, and the effect of rest of the parameters on the stability of the system was also untouched by them.

Hence, motivated by the above literature, in this chapter, we have constructed the SIR model with distributed and discrete time delays. The chapter is organized as follows: the mathematical model is proposed in Sect. 2, followed by the existence of equilibrium points in Sect. 3. The local stability, the existence of Hopf bifurcation, and its direction of stability are discussed in Sects. 4, 5, and 6. Finally, the analytic results are validated numerically in the last section with conclusion and supporting graphs.

2 Mathematical Model Formulation

In this section, we will propose our new model from the basic SIR epidemic model, by incorporating distributed and discrete time delays:-

$$\begin{aligned} \frac{dS}{dt} &= -\beta_1 I \int_{-\infty}^{t} F(t-\tau)S(\tau)d\tau - bS + b(S+I+R) \\ \frac{dI}{dt} &= \beta_1 I \int_{-\infty}^{t} F(t-\tau)S(\tau)d\tau - \beta_2 e^{-b\tau} S(t-\tau)I(t-\tau) - bI \\ \frac{dR}{dt} &= \beta_2 e^{-b\tau} S(t-\tau)I(t-\tau) - bR. \end{aligned} \tag{4}$$

Here, $F(t)$, called the delay kernel, is a weighting factor that indicates how much emphasis should be given to the size of the population at earlier times to determine the present effect on resource availability, and we are normalizing it to $\int_0^{+\infty} F(\tau)d\tau = 1$. It is done so that distributive delay must not affect the equilibrium values. Furthermore, we have considered $F(t) = ae^{-at}, a > 0$, which signifies weak delay kernel, which indicates that the maximum weighted response of the growth rate of population is due to current population density, while past densities have (exponentially) decreasing influence. In addition, a few standard assumptions are taken which are as follows:

- All newborns are considered to be susceptible as soon as they are born.
- The population considered has a constant size N, and the variables are normalized to $N = 1$, that is, $S(t) + I(t) + R(t) = 1$ for all t.
- Births and deaths occur at equal rates b in N, and all the newborns are susceptible.
- Infected individuals after recovering are transferred to the "removed" class R through the infections period τ that is given by $\beta_2 e^{-b\tau} S(t-\tau)I(t-\tau)$.

Thus, we only need to consider the following system:

$$\begin{aligned} \frac{dS}{dt} &= -\beta_1 I \int_{-\infty}^{t} F(t-\tau)S(\tau)d\tau - bS + b \\ \frac{dI}{dt} &= \beta_1 I \int_{-\infty}^{t} F(t-\tau)S(\tau)d\tau - \beta_2 e^{-b\tau} S(t-\tau)I(t-\tau) - bI \\ \frac{dR}{dt} &= \beta_2 e^{-b\tau} S(t-\tau)I(t-\tau) - bR. \end{aligned} \tag{5}$$

Table 1 Meaning of variables and parameters

Variable/Parameter	Meaning
S	Number of susceptible
I	Number of infective
R	Number of recovered
β_1	Interaction rate of S and Z
β_2	Contact rate of susceptible and infective
τ	Discrete time delay
$e^{-b\tau}$	Survival rate of individuals
b	Daily death removal rate

The initial conditions of the system take the form $S(\theta) = \phi_1(\theta), I(\theta) = \phi_2(\theta), R(\theta) = \phi_3(\theta)$ and $\phi_1(\theta) > 0, \phi_2(\theta) > 0$, and $\phi_3(\theta) > 0$ for $\theta = [-\tau, 0]$ and $\phi_1(0) > 0, \phi_2(0) > 0$, and $\phi_3(0) > 0$, where $\phi = (\phi_1(\theta), \phi_2(\theta), \phi_3(\theta)) \in C^+ \times C^+$. Here, C is the Banach space $C = C([-\tau, 0], \mathbb{R})$ of continuous functions mapping the interval $[-\tau, 0]$ into $\mathbb{R}$, equipped with the supremum norm. The non-negative cone is defined as $C^+ = C([-\tau, 0], \mathbb{R}^+)$.

Furthermore, we reduce the system using linear chain trick [19] by defining (proof given in the Appendix)

$$Z(t) = \int_{-\infty}^{t} F(t-\tau)S(\tau)d\tau. \tag{6}$$

In the above system (5), the first two equations are independent of R, and hence, the final system becomes

$$\begin{aligned} \frac{dS}{dt} &= -\beta_1 I Z - bS + b \\ \frac{dI}{dt} &= \beta_1 I Z - \beta_2 e^{-b\tau} S(t-\tau) I(t-\tau) - bI \\ \frac{dZ}{dt} &= a(S - Z), \end{aligned} \tag{7}$$

and the parameters and the variables have already been defined in Table 1.

3 Existence of Equilibrium Points

In this section, the disease-free equilibrium and the endemic equilibrium points would be discussed.

- The disease-free equilibrium is $E_0(1, 0, 1)$.

- The endemic equilibrium point is $E^*(\frac{b}{(\beta_1-\beta_2 e^{-b\tau})}, (1-\frac{1}{R_0}), \frac{b}{(\beta_1-\beta_2 e^{-b\tau})})$, which exists if $R_0 > 1$ and $\beta_1 > \beta_2 e^{-b\tau}$, where $R_0 = \frac{\beta_1}{b+\beta_2 e^{-b\tau}}$ is the basic reproduction number. Which reduces to $R_0^{''} = \frac{\beta_1}{b+\beta_2}$ in the absence of delay.

4 Local Stability of Equilibrium Points

In this section, we will be discussing the local stability of the disease-free and endemic equilibrium points.

Theorem 4.1

1. *The disease-free equilibrium point is locally asymptotically stable for* $R_0 < 1$ *if* $b > \beta_1 + \beta_2 e^{-b\tau}$.
2. *The endemic equilibrium point is locally stable for* $R_0 > 1$ *in* $[0, \tau^*]$ *and possesses Hopf bifurcation for* $\tau > \tau^*$.

Proof The general Jacobian matrix for the given system of equations is

$$J(S, I, Z) = \begin{bmatrix} -b & -\beta_1 Z & -\beta_1 I \\ 0 & \beta_1 Z - b & \beta_1 I \\ a & 0 & -a \end{bmatrix} + e^{-\lambda\tau} \begin{bmatrix} 0 & 0 & 0 \\ -\beta_2 e^{-b\tau} I & -\beta_2 e^{-b\tau} S & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The characteristic equation corresponding to the disease-free equilibrium $E_0(1, 0, 1)$ is given by

$$(-b-\lambda)(\beta_1 - b - \beta_2 e^{-(b+\lambda)\tau} - \lambda)(-a-\lambda) = 0. \tag{8}$$

Clearly, (8) always has two negative roots, which are $\lambda = -a, -b$. All other roots of (8) are determined by the following equation:

$$\lambda + b + \beta_2 e^{-(b+\lambda)\tau} - \beta_1 = 0. \tag{9}$$

Let $H(\lambda) = \lambda + b + \beta_2 e^{-(b+\lambda)\tau} - \beta_1$. We note that $H(0) = b + \beta_2 e^{-b\tau} - \beta_1 < 0$ if $R_0 = \frac{\beta_1}{b+\beta_2 e^{-b\tau}} > 1$, and $\lim_{\lambda\to+\infty} H(\lambda) = +\infty$. It follows from the continuity of the function $H(\lambda)$ on $(-\infty, +\infty)$ that the equation $H(\lambda) = 0$ has at least one positive root. Hence, (9) has at least one positive root. Therefore, the disease-free equilibrium E_0 is unstable for $R_0 > 1$.

Now, we prove that E_0 is locally stable for $R_0 < 1$. Let us suppose that $\lambda = \alpha + i\beta$ be the root of Eq. (9); then, separating the real and imaginary parts, we get

$\beta_1 - b = \beta_2 e^{-b\tau} e^{-\alpha\tau} cos\beta\tau$
$\omega = \beta_2 e^{-b\tau} e^{-\alpha\tau} sin\beta\tau.$

On squaring and adding the above equations, we get $(\beta_1 - b)^2 + \omega^2 = (\beta_2 e^{-b\tau} e^{-\alpha\tau})^2$
$(\beta_1 - b)^2 + \omega^2 \leq (\beta_2 e^{-b\tau})^2$
$\omega^2 \leq (b + \beta_2 e^{-b\tau})(\beta_2 e^{-b\tau} + \beta_1 - b)(1 - R_0)$.
ω^2 has no positive solution for $R_0 < 1$ if $b > \beta_1 + \beta_2 e^{-b\tau}$. Thus, it is proved that disease-free equilibrium point is locally asymptotically stable for $R_0 < 1$ if $b > \beta_1 + \beta_2 e^{-b\tau}$. □

The characteristic equation corresponding to the endemic equilibrium point $E^*(\frac{b}{(\beta_1 - \beta_2 e^{-b\tau})}, 1 - \frac{1}{R_0}, \frac{b}{(\beta_1 - \beta_2 e^{-b\tau})})$ is given by

$$\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0 + e^{-\lambda\tau}\left(D_2\lambda^2 + D_1\lambda + D_0\right) = 0, \tag{10}$$

where

$C_2 = -a - \beta_1 S^* + b + bS^*$,
$C_1 = -b^2 S^* - \beta_1(1 - \frac{1}{R_0}) - 2ab + a\beta_1 + bS^*\beta_1)$,
$C_0 = -a\beta_1\beta_2 Z^*(1 - \frac{1}{R_0}) + ab(\beta_1 - b) + a\beta_1 S^*(1 - \frac{1}{R_0})(\beta_1 - b)$,
$D_2 = \beta_2 e^{-b\tau} S^*$,
$D_1 = \beta_2 e^{-b\tau}(a + bS^*)$, and
$D_0 = -a\beta_2 e^{-b\tau}(\beta_1(1 - \frac{1}{R_0})S^* - b) + \beta_2 I^* Z^* e^{-b\tau}(-a\beta_1 + 1)$

In this section, two cases would be discussed, i.e. when $\tau = 0$ and $\tau \neq 0$. We begin with the case of $\tau = 0$ as it is necessary that the nontrivial equilibrium point should be locally stable for $\tau = 0$, so that we can obtain the local stability for all non-negative values of delay and further can find the critical value that may destabilize the system.

Case 1: $\tau = 0$
The endemic equilibrium point in this case is given by $E^*(\frac{b}{\beta_1 - \beta_2}, 1 - \frac{1}{R_0''}, \frac{b}{\beta_1 - \beta_2})$, which exists when $R_0'' > 1$ and $\beta_1 > \beta_2$, where $R_0'' = \frac{\beta_1}{b + \beta_2}$. The characteristic equation reduces to

$$\lambda^3 + (C_2 + D_2)\lambda^2 + (C_1 + D_1)\lambda + (C_0 + D_0) = 0. \tag{11}$$

By Routh–Hurwitz criteria, E^* is locally asymptotically stable, if $(C_2 + D_2) > 0$ and $(C_2 + D_2)(C_1 + D_1) - (C_0 + D_0) > 0$. We assume in the sequel that this condition is true and now discuss the case for $\tau > 0$.

Case 2: $\tau > 0$
As we know that the characteristic equation in this case is given by (10). Now, let $\lambda = \iota\zeta$. Hence, (10) can be written as

$(\iota\zeta)^3 + C_2(\iota\zeta)^2 + C_1(\iota\zeta) + C_0 + [D_2(\iota\zeta)^2 + D_1(\iota\zeta) + D_0]e^{-\iota\zeta\tau} = 0$
$\Rightarrow -\iota\zeta^3 - C_2\zeta^2 + C_1\iota\zeta + C_0 + [-D_2\zeta^2 + D_1\iota\zeta + D_0][\cos(\zeta\tau) - \iota\sin(\zeta\tau)] = 0.$

Separating real and imaginary parts, we get

$-\zeta^3 + C_1\zeta + D_1\zeta\cos(\zeta\tau) + \sin(\zeta\tau)(D_2\zeta^2 - D_0) = 0$
and, $-C_2\zeta^2 + C_0 - D_2\zeta^2\cos(\zeta\tau) + D_0\cos(\zeta\tau) + D_1\zeta\sin(\zeta\tau) = 0,$

or

$D_1\zeta\cos(\zeta\tau) + \sin(\zeta\tau)(D_2\zeta^2 - D_0) = \zeta^3 - C_1\zeta,$

and $D_1\zeta\sin(\zeta\tau) - \cos(\zeta\tau)(D_2\zeta^2 - D_0) = C_2\zeta^2 - C_0.$

And, after squaring both sides and adding both the equations, we get

$$\zeta^6 + G_2\zeta^4 + G_1\zeta^2 + G_0 = 0, \tag{12}$$

where

$G_2 = C_2^2 - 2C_1 - D_2^2,$
$G_1 = C_1^2 - 2C_0C_2 + 2D_0D_2 - D_1^2,$ and
$G_0 = C_0^2 - D_0^2.$

Next, let $H = \zeta^2$. Therefore, (12) reduces to

$$H^3 + G_2H^2 + G_1H + G_0 = 0. \tag{13}$$

According to the Routh–Hurwitz stability criterion, (13) has roots with negative real parts if $G_0 \geq 0$, $G_2 \geq 0$, and $G_2G_1 \geq G_0$. But, $H = \zeta^2 \geq 0$ clearly indicates that our assumption $\lambda = \iota\zeta$ is wrong. Hence, the characteristic equation has no positive roots, and the real part of all the eigenvalues is negative for all $\tau \geq 0$.

Therefore, the system of equations is stable when $\tau \geq 0$.

5 Existence of Hopf Bifurcation

In this section, we discuss the criteria for the existence of Hopf bifurcation. The characteristic equation for the system of equations at the endemic equilibrium point E^* is

$$\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0 + e^{-\lambda\tau}\left(D_2\lambda^2 + D_1\lambda + D_0\right) = 0, \tag{14}$$

where

$C_2 = -a - \beta_1S^* + b + bS^*,$
$C_1 = -b^2S^* - \beta_1(1 - \frac{1}{R_0}) - 2ab + a\beta_1 + bS^*\beta_1),$
$C_0 = -a\beta_1\beta_2Z(1 - \frac{1}{R_0}) + ab(\beta_1 - b) + a\beta_1S^*(1 - \frac{1}{R_0})(\beta_1 - b),$
$D_2 = \beta_2e^{-b\tau}S^*,$

$D_1 = \beta_2 e^{-b\tau}(a + bS^*)$, and
$D_0 = -a\beta_2 e^{-b\tau}(\beta_1(1 - \frac{1}{R_0})S^* - b) + \beta_2 I^* Z^* e^{-b\tau}(-a\beta_1 + 1)$.

On multiplying (14) by $e^{\lambda\tau}$ on both sides, we get

$$(\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0)e^{\lambda\tau} + D_2\lambda^2 + D_1\lambda + D_0 = 0. \tag{15}$$

Let $\lambda = \iota\sigma$. And hence, (15) can be written as

$[(\iota\sigma)^3 + C_2(\iota\sigma)^2 + C_1(\iota\sigma) + C_0]e^{\iota\sigma\tau} + D_2(\iota\sigma)^2 + D_1(\iota\sigma) + D_0 = 0$
$\Rightarrow [-\iota\sigma^3 - C_2\sigma^2 + C_1(\iota\sigma) + C_0][\cos(\sigma\tau) + \iota\sin(\sigma\tau)] - D_2\sigma^2 + D_1\iota\sigma + D_0 = 0.$

Separating real and imaginary parts, we get

$[-\sigma^3 + C_1\sigma]\cos(\sigma\tau) + \sin(\sigma\tau)[-C_2\sigma^2 + C_0] = -D_1\sigma,$
and $[-C_2\sigma^2 + C_0]\cos(\sigma\tau) - \sin(\sigma\tau)[-\sigma^3 + C_1\sigma] = D_2\sigma^2 - D_0.$

On solving the above two equations, we get

$$\sin(\sigma\tau) = \frac{g_4\sigma^5 + g_5\sigma^3 + g_6\sigma}{\sigma^6 + g_1\sigma^4 + g_2\sigma^2 + g_3} \tag{16}$$

$$\cos(\sigma\tau) = \frac{g_7\sigma^4 + g_8\sigma^2 + g_9}{\sigma^6 + g_1\sigma^4 + g_2\sigma^2 + g_3}, \tag{17}$$

where

$g_1 = C_2^2 - 2C_1,$
$g_2 = C_1^2 - 2C_0C_2,$
$g_3 = C_0^2,$
$g_4 = D_2,$
$g_5 = C_2D_1 - C_1D_2 - D_0,$
$g_6 = C_1D_0 - C_0D_1,$
$g_7 = D_1 - C_2D_2,$
$g_8 = C_2D_0 + C_0D_2 - C_1D_1,$ and
$g_9 = -C_0D_0.$

Furthermore, adding the square of Eqs. (16) and (17), we have

$$\sigma^{12} + z_1\sigma^{10} + z_2\sigma^8 + z_3\sigma^6 + z_4\sigma^4 + z_5\sigma^2 + z_6 = 0, \tag{18}$$

where

$z_1 = 2g_1 - g_4^2,$
$z_2 = 2g_2 + g_1^2 - 2g_4g_5 - g_7^2,$
$z_3 = 2g_3 + 2g_1g_2 - g_5^2 - 2g_4g_6 - 2g_7g_8,$
$z_4 = 2g_1g_3 + g_2^2 - g_8^2 - 2g_7g_9 - 2g_5g_6,$

$z_5 = 2g_2g_3 - 2g_8g_9 - g_6^2$, and
$z_6 = g_3^2 - g_9^2$.

Next, we take $a = \sigma^2$. Then, Eq. (18) gets reduced to

$$a^6 + z_1a^5 + z_2a^4 + z_3a^3 + z_4a^2 + z_5a + z_6 = 0. \tag{19}$$

Now, let

$$\tilde{L}(a) = a^6 + z_1a^5 + z_2a^4 + z_3a^3 + z_4a^2 + z_5a + z_6. \tag{20}$$

Since $\tilde{L}(a) \to \infty$ as $a \to \infty$, and $z_6 < 0$ if $g_3^2 < g_9^2$, then by Descartes' rule of signs, Eq. (20) has at least one positive real root.

Let us assume that we have six positive roots for Eq. (20), denoted by a_1, a_2, a_3, a_4, a_5 and a_6. Then,

$\sigma_1 = \sqrt{a}_1, \sigma_2 = \sqrt{a}_2, \sigma_3 = \sqrt{a}_3, \sigma_4 = \sqrt{a}_4, \sigma_5 = \sqrt{a}_5, \sigma_6 = \sqrt{a}_6$.

From (17), we have

$$\cos(\sigma_j\tau) = \frac{g_7\sigma_j^4 + g_8\sigma_j^2 + g_9}{\sigma_j^6 + g_1\sigma_j^4 + g_2\sigma_j^2 + g_3}, \tag{21}$$

where $j = 1, 2, 3, 4, 5, 6$. Hence, we get

$$\tau_j^{(k)} = \frac{1}{\sigma_j}\left[\arccos\left(\frac{g_7\sigma_j^4 + g_8\sigma_j^2 + g_9}{\sigma_j^6 + g_1\sigma_j^4 + g_2\sigma_j^2 + g_3}\right) + 2k\pi\right], \tag{22}$$

where $j = 1, 2, 3, 4, 5, 6$ and $k = 0, 1, 2, 3, \ldots$

Then, the pair of imaginary roots is $\pm\iota\sigma_j$. Next, we define $\tau' = \min \tau_j^{(0)}$ and $\sigma' = \sigma(\tau')$. To establish Hopf bifurcation at $\tau = \tau'$, we need to prove that $Re(\frac{d\lambda}{d\tau})_{\tau=\tau'} \neq 0$.

Taking the derivative of (14) with respect to τ, we get

$$\frac{d\lambda}{d\tau} = -\left[\frac{(\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0)\lambda e^{\lambda\tau}}{e^{\lambda\tau}((3\lambda^2 + 2C_2\lambda + C_1) + \tau(\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0)) + (2D_2\lambda + D_1)}\right]. \tag{23}$$

And hence, it follows that

$$\left(\frac{d\lambda}{d\tau}\right)^{-1} = -\left[\frac{(3\lambda^2 + 2C_2\lambda + C_1)e^{\lambda\tau} + (2D_2\lambda + D_1)}{\lambda e^{\lambda\tau}(\lambda^3 + C_2\lambda^2 + C_1\lambda + C_0)} + \frac{\tau}{\lambda}\right]. \tag{24}$$

Furthermore, substituting $\lambda = \iota\sigma^{'}$, we get

$$\left(\frac{d\lambda}{d\tau}\right)^{-1}|_{\tau=\tau'} = -\left[\frac{\tilde{d_1} + \iota\tilde{d_2}}{\tilde{d_3} + \iota\tilde{d_4}}\right] + \iota\frac{\tau^{'}}{\sigma^{'}}, \tag{25}$$

where

$\tilde{d_1} = -3(\sigma^{'})^2\cos(\sigma^{'}\tau^{'}) + C_1\cos(\sigma^{'}\tau^{'}) - 2C_2\sigma^{'}\sin(\sigma^{'}\tau^{'}) + D_1$,
$\tilde{d_2} = 2C_2\sigma^{'}\cos(\sigma^{'}\tau^{'}) - 3(\sigma^{'})^2\sin(\sigma^{'}\tau^{'}) + C_1\sin(\sigma^{'}\tau^{'}) + 2D_2\sigma^{'}$,
$\tilde{d_3} = (\sigma^{'})^4\cos(\sigma^{'}\tau') + C_2(\sigma^{'})^3\sin(\sigma^{'}\tau^{'}) - C_1(\sigma^{'})^2\cos(\sigma^{'}\tau^{'}) - C_0\sigma^{'}\sin(\sigma^{'}\tau^{'})$, and
$\tilde{d_4} = (\sigma^{'})^4\sin(\sigma^{'}\tau') - C_2(\sigma^{'})^3\cos(\sigma^{'}\tau^{'}) - C_1(\sigma^{'})^2\sin(\sigma^{'}\tau^{'}) + C_0\sigma^{'}\cos(\sigma^{'}\tau^{'})$.

Thus,

$$Re\left(\frac{d\lambda}{d\tau}\right)^{-1}|_{\tau=\tau'} = -\left[\frac{\tilde{d_1}\tilde{d_3} + \tilde{d_2}\tilde{d_4}}{\tilde{d_3}^2 + \tilde{d_4}^2}\right]. \tag{26}$$

We notice that

$$\text{sign}\left\{Re\left(\frac{d\lambda}{d\tau}\right)|_{\tau=\tau'}\right\} = \text{sign}\left\{Re(\frac{d\lambda}{d\tau})^{-1}|_{\tau=\tau'}\right\}. \tag{27}$$

And hence, we can conclude that the endemic equilibrium point of the given system of equations is asymptotically stable for $[0, \tau^{'})$, and it undergoes Hopf bifurcation at $\tau = \tau^{'}$.

6 Direction and Stability of Hopf Bifurcation

In the previous section, we obtained certain conditions under which the given system of equations undergoes Hopf bifurcation, with time delay $\tau = \tau^{'}$ being the critical parameter. In this section, by taking into account the normal form theory and the centre manifold theorem, which were introduced by Hassard et al. [20], we will be presenting the formula determining the direction of Hopf bifurcation and will be obtaining conditions for the stability of bifurcating periodic solutions, as well. Since Hopf bifurcation occurs at the critical value $\tau^{'}$ of τ, there exists a pair of pure imaginary roots $\pm\iota\sigma(\tau^{'})$ of the characteristic equation (14).
Let, $x_1 = S - S^*$, $x_2 = I - I^*$, and $x_3 = Z - Z^*$ (where S^*, I^*, and Z^* are the values of S, I, and Z in the case of endemic equilibrium point).

Thus, the given system of equations gets transformed into the following system:

$$\frac{dx_1}{dt} = -\beta_1 x_2 x_3 - \beta_1 I^* Z^* - \beta_1 Z^* x_2 - \beta_1 I^* x_3 - bx_1 - bS^* + b$$

$$\frac{dx_2}{dt} = \beta_1 x_2 x_3 + \beta_1 I^* Z^* + \beta_1 Z^* x_2 + \beta_1 I^* x_3 - \beta_2 e^{-b\tau} x_1(t-\tau) x_2(t-\tau)$$
$$- \beta_2 e^{-b\tau} S^* x_2(t-\tau)$$
$$- \beta_2 e^{-b\tau} I^* x_1(t-\tau) - \beta_2 e^{-b\tau} S^* I^* - b x_2 - b I^*$$
$$\frac{dx_3}{dt} = a x_1 + a S^* - a x_3 - a Z^*.$$

We also let $t \to \tau t$ and $\tau = \tau^{'} + \mu$. Then, the system finally takes the form of an FDE in $C = C([-1, 0], R^3$ as

$$\dot{x}(t) = L_\mu(x_t) + F(\mu, x_t), \tag{28}$$

where $x(t) = (x_1(t), x_2(t), x_3(t))^T \in R^3$ and $L_\mu : C \to R^3$, $F : C \times R \to R^3$ are given, respectively, by

$L_\mu(\psi) = (\tau^{'} + \mu) L_1 \psi(0) + (\tau^{'} + \mu) L_2 \psi(-1)$, and $F(\mu, \psi) = (\tau^{'} + \mu) F_1$ where,

$$L_1 = \begin{bmatrix} -b & -\beta_1 Z^* & -\beta_1 I^* \\ 0 & \beta_1 Z^* - b & \beta_1 I^* \\ a & 0 & -a \end{bmatrix},$$
$$L_2 = \begin{bmatrix} 0 & 0 & 0 \\ -\beta_2 e^{-b(\tau^{'}+\mu)} I^* & -\beta_2 e^{-b(\tau^{'}+\mu)} S^* & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and}$$
$$F_1 = \begin{bmatrix} -\beta_1 \psi_2(0) \psi_3(0) \\ \beta_1 \psi_2(0) \psi_3(0) - \beta_2 e^{-b(\tau^{'}+\mu)} \psi_1(-1) \psi_2(-1) \\ 0 \end{bmatrix}.$$

We also have that, $\psi = (\psi_1, \psi_2, \psi_3)^T \in C$, and $x_t(\theta) = x(t + \theta)$ for $\theta \in [-1, 0]$.

By the Riesz representation theorem, there exists a function $\eta(\theta, \mu)$ of bounded variation for $\theta \in [-1, 0]$, such that

$$L_\mu(\psi) = \int_{-1}^{0} d\eta(\theta, \mu) \psi(\theta). \tag{29}$$

This equation holds for $\psi \in C$.

In fact, we can take

$$\eta(\theta, \mu) = \left(\tau^{'} + \mu\right) L_1 \delta(\theta) + \left(\tau^{'} + \mu\right) L_2 \delta(\theta + 1), \tag{30}$$

where, L_1 and L_2 have already been given above, and $\delta(\theta)$ is Dirac delta function.

Next, for $\psi \in C^1([-1, 0], R^3)$, we define the following:

$$A(\mu)\psi = \begin{cases} \frac{d\psi(\theta)}{d\theta}, \ \theta \in [-1, 0) \\ \int_{-1}^{0} d\eta(s, \mu)\psi(s), \ \theta = 0, \end{cases}$$

and

$$R(\mu)\psi = \begin{cases} 0, \ \theta \in [-1, 0) \\ F(\mu, \psi), \ \theta = 0 \end{cases}$$

Then, the system (28) is equivalent to

$$\dot{x}_t = A(\mu)x_t + R(\mu)x_t, \tag{31}$$

where $x_t(\theta) = x(t + \theta)$ for $\theta \in [-1, 0]$.

Next, for $\varphi \in C^1([0, 1], R^3)$, the adjoint operator A^* of A can be defined as

$$A^*\varphi(s) = \begin{cases} \frac{-d\varphi(s)}{ds}, \ s \in (0, 1] \\ \int_{-1}^{0} d\eta^T(t, 0)\varphi(-t), \ s = 0, \end{cases}$$

and hence for $\psi \in ([-1, 0], R^3)$, $\varphi \in ([0, 1], R^3)$, a bilinear inner product, in order to normalize the eigenvalues of A and A^*, can be defined as follows:

$$< \varphi(s), \psi(\theta) >= \bar{\varphi}(0)\psi(0) - \int_{-1}^{0}\int_{\gamma=0}^{\theta} \bar{\varphi}(\gamma - \theta)d\eta(\theta)\psi(\gamma)d\gamma, \tag{32}$$

where $\eta(\theta) = \eta(\theta, 0)$, and $\bar{\varphi}$ is the complex conjugate of φ. It can be verified that the operators A and A^* are adjoint operators with respect to this bilinear form. Thus, since $\pm\iota\sigma^{'}\tau'$ are eigenvalues of $A(0)$, they are the eigenvalues of A^* as well.

We need to compute the eigenvectors of $A(0)$ and A^* corresponding to the eigenvalues $\iota\sigma^{'}\tau^{'}$ and $-\iota\sigma^{'}\tau^{'}$, respectively.

Let us suppose that $q(\theta) = (1, \alpha^{'}, \beta^{'})^T e^{\iota\sigma^{'}\tau^{'}\theta}$ is the eigenvector of $A(0)$ corresponding to $\iota\sigma^{'}\tau'$.

Then, $A(0)q(\theta) = \lambda q(\theta)$, that is, $A(0)q(\theta) = \iota\sigma^{'}\tau' q(\theta)$,

or

$[\lambda I - A(0)]q(0) = 0$,

which gives the following:

$$\tau^{'}\begin{bmatrix} \iota\sigma^{'} + b & \beta_1 Z^* & \beta_1 I^* \\ \beta_2 e^{-b\tau^{'}} I^* e^{-\iota\sigma^{'}\tau^{'}} & \iota\sigma^{'} + \beta_2 e^{-b\tau^{'}} S^* e^{-\iota\tau^{'}\sigma^{'}} - \beta_1 Z^* + b & -\beta_1 I^* \\ -a & 0 & \iota\sigma^{'} + a \end{bmatrix}\begin{bmatrix} 1 \\ \alpha^{'} \\ \beta^{'} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} \iota\sigma^{'}+b & \beta_1 Z^* & \beta_1 I^* \\ \beta_2 e^{-b\tau^{'}} I^* e^{-\iota\sigma^{'}\tau^{'}} & \iota\sigma^{'}+\beta_2 e^{-b\tau^{'}} S^* e^{-\iota\tau^{'}\sigma^{'}}-\beta_1 Z^*+b & -\beta_1 I^* \\ -a & 0 & \iota\sigma^{'}+a \end{bmatrix} \begin{bmatrix} 1 \\ \alpha^{'} \\ \beta^{'} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(since $\tau^{'} \neq 0$).

And, on solving this, we get $q(0) = (1, \alpha^{'}, \beta^{'})^T$, where

$\alpha^{'} = \frac{\beta_1 I^* \beta^{'} - \beta_2 e^{-b\tau^{'}} e^{-\iota\sigma^{'}\tau^{'}} I^*}{\iota\sigma^{'}+\beta_2 S^* e^{-b\tau^{'}} e^{-\iota\sigma^{'}\tau^{'}}-\beta_1 Z^*+b}$, and $\beta^{'} = \frac{a}{a+\iota\sigma^{'}}$.

Next, let us suppose that $q^*(\theta) = D(1, (\alpha^{'})^*, (\beta^{'})^*)e^{\iota\sigma^{'}\tau^{'}\theta}$ be the eigenvector of A^* corresponding to the eigenvalue $-\iota\sigma^{'}\tau^{'}$, and hence in a similar manner, we can obtain

$(\alpha^{'})^* = \frac{-\beta_1 Z^*}{-\iota\sigma^{'}-\beta_1 Z^*+b+\beta_2 S^* e^{-b\tau^{'}} e^{\iota\sigma^{'}\tau^{'}}}$, and $(\beta^{'})^* = \frac{(\beta I^* \alpha^{'}-\beta_1 I^*)^*}{a-\iota\sigma^{'}}$.

From (32), we get

$$\begin{aligned} < q^*(s), q(\theta) > &= \bar{D}\left(1, (\bar{\alpha^{'}})^*, (\bar{\beta^{'}})^*\right)(1, \alpha^{'}, \beta^{'})^T \\ &\quad - \int_{-1}^{0}\int_{\gamma=0}^{\theta} \bar{D}(1, (\bar{\alpha^{'}})^*, (\bar{\beta^{'}})^*) e^{-\iota\sigma^{'}\tau^{'}(\gamma-\theta)} d\eta(\theta) \\ &\quad \times (1, \alpha^{'}, \beta^{'})^T e^{\iota\sigma^{'}\tau^{'}\gamma} d\gamma \\ &= \bar{D}[1 + (\alpha^{'})^*\alpha^{'} + (\beta^{'})^*\beta^{'} - (1, (\bar{\alpha^{'}})^*, (\bar{\beta^{'}})^*) \\ &\quad \times \int_{-1}^{0} \theta e^{\iota\sigma^{'}\tau^{'}\theta} d\eta(\theta)(1, \alpha^{'}, \beta^{'})^T. \end{aligned}$$

Now, let $\psi(\theta) = \theta e^{\iota\sigma^{'}\tau^{'}\theta}$

$\Rightarrow \psi(0) = 0$, and $\Rightarrow \psi(-1) = -e^{-\iota\sigma^{'}\tau^{'}}$.

Thus, from (32) and the definition of ψ as taken above, we finally get that

$< q^*(s), q(\theta) >= \bar{D}[1 + (\alpha^{'})^*\alpha^{'} + (\beta^{'})^*\beta^{'} - \tau'(I^* + \alpha^{'} S^*)(\bar{\alpha^{'}})^*\beta e^{-b\tau^{'}} e^{-\iota\sigma^{'}\tau^{'}}]$.

Hence, $\bar{D} = \frac{1}{[1+(\alpha^{'})^*\alpha^{'}+(\beta^{'})^*\beta^{'}-\tau'(I^*+\alpha^{'}S^*)(\bar{\alpha^{'}})^*\beta e^{-b\tau^{'}} e^{-\iota\sigma^{'}\tau^{'}}]}$,

such that $< q^*(s), q(\theta) >= 1$ and $< q^*(s), \bar{q}(\theta) >= 0$.

In the remaining part of this section, using the same ideas as in [20], we now compute the coordinates in order to describe the centre manifold C_0 at $\mu = 0$. Let x_t be the solution of (28) when $\mu = 0$.

Next, define

$$\tilde{z}(t) =< q^*, x_t >, W(t, \theta) = x_t - 2Re[\tilde{z}(t)q(\theta)]. \tag{33}$$

Now, on the centre manifold C_0, we have

$$W(t,\theta) = W(\tilde{z}(t),\bar{\tilde{z}}(t),\theta) = W_{20}(\theta)\frac{\tilde{z}^2}{2} + W_{11}(\theta)\tilde{z}\bar{\tilde{z}} + W_{02}(\theta)\frac{\bar{\tilde{z}}^2}{2} + \ldots, \quad (34)$$

where $\tilde{z}$ and $\bar{\tilde{z}}$ are local coordinates for the centre manifold C_0 in the direction of q^* and $\bar{q^*}$. We note that W is real if x_t is real, and we will be considering the real solutions only.

From (33), we have $\dot{\tilde{z}}(t) =< q^*, \dot{x_t} >$

$=< q^*, A(\mu)x_t + R(\mu)x_t >$
$=< A^*(\mu)q^*, x_t > + < q^*, R(\mu)x_t >$
$= \iota\sigma'\tau'\tilde{z}(t) + < q^*, R(\mu)x_t >$ (since $A^*q^* = \bar{\lambda}q^*$)
$= \iota\sigma'\tau'\tilde{z}(t) + \bar{q^*}(0)F(0, x_t)$ (from the definition of bilinear product and taking $\theta = 0$)
$= \iota\sigma'\tau'\tilde{z}(t) + \bar{q^*}(0)F(0, W(\tilde{z},\bar{\tilde{z}},0) + 2Re[\tilde{z}q(0)])$
$= \iota\sigma'\tau'\tilde{z}(t) + \bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})$
$= \iota\sigma'\tau'\tilde{z}(t) + g(\tilde{z},\bar{\tilde{z}})$,

where

$$g(\tilde{z},\bar{\tilde{z}}) = \bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}}) = g_{20}\frac{\tilde{z}^2}{2} + g_{11}\tilde{z}\bar{\tilde{z}} + g_{02}\frac{\bar{\tilde{z}}^2}{2} + g_{21}\frac{\tilde{z}^2\bar{\tilde{z}}}{2} + \ldots \quad (35)$$

From (35), we have

$$x_t(\theta) = (x_{1t}(\theta), x_{2t}(\theta), x_{3t}(\theta)) = W(t,\theta) + \tilde{z}q(\theta) + \bar{\tilde{z}}\bar{q}(\theta),$$

and $q(\theta) = (1, \alpha', \beta')^T e^{\iota\theta\sigma'\tau'}$.

And thus, we can obtain that

$$\begin{aligned}
x_{1t}(0) &= W_{20}^{(1)}(0)\tfrac{\tilde{z}^2}{2} + W_{11}^{(1)}(0)\tilde{z}\bar{\tilde{z}} + W_{02}^{(1)}(0)\tfrac{\bar{\tilde{z}}^2}{2} + \tilde{z} + \bar{\tilde{z}} + O(|\tilde{z},\bar{\tilde{z}}|^3),\\
x_{2t}(0) &= W_{20}^{(2)}(0)\tfrac{\tilde{z}^2}{2} + W_{11}^{(2)}(0)\tilde{z}\bar{\tilde{z}} + W_{02}^{(2)}(0)\tfrac{\bar{\tilde{z}}^2}{2} + \alpha'\tilde{z} + \bar{\alpha'}\bar{\tilde{z}} + O(|\tilde{z},\bar{\tilde{z}}|^3),\\
x_{3t}(0) &= W_{20}^{(3)}(0)\tfrac{\tilde{z}^2}{2} + W_{11}^{(3)}(0)\tilde{z}\bar{\tilde{z}} + W_{02}^{(3)}(0)\tfrac{\bar{\tilde{z}}^2}{2} + \beta'\tilde{z} + \bar{\beta'}\bar{\tilde{z}} + O(|\tilde{z},\bar{\tilde{z}}|^3),\\
x_{1t}(-1) &= W_{20}^{(1)}(-1)\tfrac{\tilde{z}^2}{2} + W_{11}^{(1)}(-1)\tilde{z}\bar{\tilde{z}} + W_{02}^{(1)}(-1)\tfrac{\bar{\tilde{z}}^2}{2} + \tilde{z} + \bar{\tilde{z}} + O(|\tilde{z},\bar{\tilde{z}}|^3),\\
x_{2t}(-1) &= W_{20}^{(2)}(-1)\tfrac{\tilde{z}^2}{2} + W_{11}^{(2)}(-1)\tilde{z}\bar{\tilde{z}} + W_{02}^{(2)}(-1)\tfrac{\bar{\tilde{z}}^2}{2} + \alpha'\tilde{z}e^{-\iota\sigma'\tau'} + \bar{\alpha'}\bar{\tilde{z}}e^{\iota\sigma'\tau'}\\
&\quad + O(|\tilde{z},\bar{\tilde{z}}|^3),\\
x_{3t}(-1) &= W_{20}^{(3)}(-1)\tfrac{\tilde{z}^2}{2} + W_{11}^{(3)}(-1)\tilde{z}\bar{\tilde{z}} + W_{02}^{(3)}(-1)\tfrac{\bar{\tilde{z}}^2}{2} + \beta'\tilde{z}e^{-\iota\sigma'\tau'} + \bar{\beta'}\bar{\tilde{z}}e^{\iota\sigma'\tau'}\\
&\quad + O(|\tilde{z},\bar{\tilde{z}}|^3).
\end{aligned}$$

From the definition of $F(\mu, x_t)$, we get

$$g(\tilde{z},\bar{\tilde{z}}) = \tau'\bar{D}\left(1, \left(\bar{\alpha'}\right)^*, \left(\bar{\beta'}\right)^*\right)\begin{bmatrix} -\beta_1 x_{2t}(0)x_{3t}(0) \\ \beta_1 x_{2t}(0)x_{3t}(0) - \beta_2 e^{-b\tau'}x_{1t}(-1)x_{2t}(-1) \\ 0 \end{bmatrix}$$

$$= \tau'\bar{D}\left\{\tilde{z}^2\left[\alpha'\beta'\left(-\beta_1 + \left(\bar{\alpha'}\right)^*\beta\right) + \alpha' e^{-\iota\sigma'\tau'}\left(-\left(\bar{\alpha'}\right)^*\beta_2 e^{-b\tau'}\right)\right]\right.$$

$$
\begin{aligned}
&+ 2\tilde{z}\bar{\tilde{z}}\Big[\Big(-\beta_1 + \Big(\bar{\alpha'}\Big)^*\beta\Big)Re\Big\{\alpha'\bar{\beta'}\Big\} + Re\Big\{\alpha' e^{-\iota\sigma'\tau'}\Big\}\Big(-\Big(\bar{\alpha'}\Big)^*\beta_2 e^{-b\tau'}\Big)\Big]\\
&+ \bar{\tilde{z}}^2\Big[\Big(-\beta_1 + \Big(\bar{\alpha'}\Big)^*\beta\Big)\bar{\alpha'}\bar{\beta'} + \bar{\alpha'}e^{\iota\sigma'\tau'}\Big(-\Big(\bar{\alpha'}\Big)^*\beta_2 e^{-b\tau'}\Big)\Big]\\
&+ \frac{z^2\bar{z}}{2}\Big[\Big(-\beta_1 + \Big(\bar{\alpha'}\Big)^*\beta\Big)\Big(\bar{\beta'}W_{20}^{(2)}(0) + 2\beta' W_{11}^{(2)}(0) + 2\alpha' W_{11}^{(2)}(0)\\
&+ \bar{\alpha'}W_{20}^{(3)}(0)\Big)\\
&- \Big(\bar{\alpha'}\Big)^*\beta_2 e^{-b\tau'}\Big(W_{20}^{(2)}(-1) + 2W_{11}^{(2)}(-1) + 2\alpha' e^{-\iota\sigma'\tau'}W_{11}^{(1)}(-1)\\
&+ \bar{\alpha'}W_{20}^{(1)}(-1)e^{\iota\sigma'\tau'}\Big)\Big]\Big\}.
\end{aligned}
$$

Now, comparing the coefficients, we get

$$
\begin{aligned}
g_{20} &= 2\tau'\bar{D}[\alpha'\beta'(-\beta_1 + (\bar{\alpha'})^*\beta) + \alpha' e^{-\iota\sigma'\tau'}(-(\bar{\alpha'})^*\beta_2 e^{-b\tau'})]\\
g_{11} &= 2\tau'\bar{D}[(-\beta_1 + (\bar{\alpha'})^*\beta)Re\{\alpha'\bar{\beta'}\} + Re\{\alpha' e^{-\iota\sigma'\tau'}\}(-(\bar{\alpha'})^*\beta_2 e^{-b\tau'})]\\
g_{02} &= 2\tau'\bar{D}[(-\beta_1 + (\bar{\alpha'})^*\beta)\bar{\alpha'}\bar{\beta'} + \bar{\alpha'}e^{\iota\sigma'\tau'}(-(\bar{\alpha'})^*\beta_2 e^{-b\tau'})]\\
g_{21} &= \tau'\bar{D}[(-\beta_1 + (\bar{\alpha'})^*\beta)(\bar{\beta'}W_{20}^{(2)}(0) + 2\beta' W_{11}^{(2)}(0) + 2\alpha' W_{11}^{(2)}(0) + \bar{\alpha'}W_{20}^{(3)}(0))\\
&\quad -(\bar{\alpha'})^*\beta_2 e^{-b\tau'}(W_{20}^{(2)}(-1) + 2W_{11}^{(2)}(-1) + 2\alpha' e^{-\iota\sigma'\tau'}W_{11}^{(1)}(-1)\\
&\quad +\bar{\alpha'}W_{20}^{(1)}(-1)e^{\iota\sigma'\tau'})].
\end{aligned}
$$

We can clearly see that in order to determine g_{21}, we will have to compute $W_{20}(\theta)$ and $W_{11}(\theta)$.

From (33) and (35), we have

$$
\begin{aligned}
\dot{W} &= \dot{x}_t - 2Re\big[\dot{\tilde{z}}(t)q(\theta)\big]\\
&= A(\mu)x_t + R(\mu)x_t - 2Re\Big[\Big(\iota\sigma'\tau'\tilde{z}(t) + \bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})\Big)q(\theta)\Big]\\
&= A(\mu)x_t + R(\mu)x_t - 2Re\Big[\iota\sigma'\tau'\tilde{z}(t)q(\theta)\Big] - 2Re\Big[\bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})q(\theta)\Big].
\end{aligned}
$$

Therefore,

$$
\dot{W} = \begin{cases} AW - 2Re[\bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})q(\theta)],\ \theta \in [-1,0)\\ AW - 2Re[\bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})q(\theta)] + F_0,\ \theta = 0 \end{cases}
$$

(using the definition of AW and $R(\mu)x_t$).

Therefore, let

$$
\dot{W} = AW + \tilde{H}\big(\tilde{z},\bar{\tilde{z}},\theta\big), \tag{36}
$$

where

$$\tilde{H}\left(\tilde{z},\bar{\tilde{z}},\theta\right)=\tilde{H}_{20}(\theta)\frac{\tilde{z}^2}{2}+\tilde{H}_{11}(\theta)\tilde{z}\bar{\tilde{z}}+\tilde{H}_{02}(\theta)\frac{\bar{\tilde{z}}^2}{2}+\ldots \tag{37}$$

On the other hand, on the centre manifold C_0 near the origin, $\dot{W}=W_{\tilde{z}}\dot{\tilde{z}}+W_{\bar{\tilde{z}}}\dot{\bar{\tilde{z}}}$.

Using (37) to compare the coefficients, we finally deduce

$$\left(A-2\iota\sigma^{'}\tau^{'}\right)W_{20}(\theta)=-\tilde{H}_{20}(\theta),\ AW_{11}(\theta)=-\tilde{H}_{11}(\theta). \tag{38}$$

From (37), we also have $\tilde{H}(\tilde{z},\bar{\tilde{z}},\theta) = -2Re[\bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})q(\theta)]$, for $\theta\in[-1,0)$;that is,

$$\begin{aligned}\tilde{H}(\tilde{z},\bar{\tilde{z}},\theta)&=-\bar{q^*}(0)F_0(\tilde{z},\bar{\tilde{z}})q(\theta)-q^*(0)\bar{F}_0(\tilde{z},\bar{\tilde{z}})\bar{q}(\theta)\\&=-(g_{20}\tfrac{\tilde{z}^2}{2}+g_{11}\tilde{z}\bar{\tilde{z}}+g_{02}\tfrac{\bar{\tilde{z}}^2}{2}+g_{21}\tfrac{\tilde{z}^2\bar{\tilde{z}}}{2}+\ldots..)q(\theta)-(\bar{g}_{20}\tfrac{\bar{\tilde{z}}^2}{2}\\&\quad+\bar{g}_{11}\tilde{z}\bar{\tilde{z}}+\bar{g}_{02}\tfrac{\tilde{z}^2}{2}+\bar{g}_{21}\tfrac{\bar{\tilde{z}}^2\tilde{z}}{2})q(\theta).\end{aligned}$$

Now, equating this with (37), and comparing the coefficients, we have

$$\tilde{H}_{20}(\theta)=-g_{20}q(\theta)-\bar{g_{02}}\bar{q}(\theta),\ \tilde{H}_{11}(\theta)=-g_{11}q(\theta)-\bar{g_{11}}\bar{q}(\theta). \tag{39}$$

From (39), (38), and the definition of A for $\theta\in[-1,0)$, we get

$$\dot{W}_{20}(\theta)=2\iota\sigma^{'}\tau^{'}W_{20}(\theta)+g_{20}q(\theta)+\bar{g_{02}}\bar{q}(\theta). \tag{40}$$

Note that, $q(\theta)=q(0)e^{\iota\sigma^{'}\tau^{'}\theta}$. Hence, putting this value in (40), and solving it being a linear differential equation, we get

$$W_{20}(\theta)=\frac{\iota g_{20}}{\sigma^{'}\tau^{'}}q(0)e^{\iota\sigma^{'}\tau^{'}\theta}+\frac{\iota\bar{g}_{02}}{3\sigma^{'}\tau^{'}}\bar{q}(0)e^{-\iota\sigma^{'}\tau^{'}\theta}+\tilde{E}_1e^{2\iota\sigma^{'}\tau^{'}\theta}, \tag{41}$$

where $\tilde{E}_1=(\tilde{E}_1^{(1)},\tilde{E}_1^{(2)},\tilde{E}_1^{(3)})\in R^3$ is a constant vector. Similarly, we can get

$$W_{11}(\theta)=-\frac{\iota g_{11}}{\sigma^{'}\tau^{'}}q(0)e^{\iota\sigma^{'}\tau^{'}\theta}+\frac{\iota\bar{g}_{11}}{\sigma^{'}\tau^{'}}\bar{q}(0)e^{-\iota\sigma^{'}\tau^{'}\theta}+\tilde{E}_2, \tag{42}$$

where $\tilde{E}_2=(\tilde{E}_2^{(1)},\tilde{E}_2^{(2)},\tilde{E}_2^{(3)})\in R^3$ is a constant vector.

Furthermore, we will be finding $\tilde{E}_1$ and $\tilde{E}_2$.

From the definition of A at $\theta=0$ and (40), we have

$$\int_{-1}^{0}d\eta(\theta)W_{20}(\theta)=2\iota\sigma^{'}\tau^{'}W_{20}(0)-\tilde{H}_{20}(0) \tag{43}$$

$$\int_{-1}^{0} d\eta(\theta) W_{11}(\theta) = -\tilde{H}_{11}(0), \tag{44}$$

where $\eta(\theta) = \eta(0, \theta)$ (since $\mu = 0$).

Also, for $\theta = 0$, we have $\tilde{H}(\tilde{z}, \bar{\tilde{z}}, \theta) = -2Re[\bar{q}^*(0) F_0(\tilde{z}, \bar{\tilde{z}}) q(\theta)] + F_0$.That is,

$$\begin{aligned}\tilde{H}(\tilde{z}, \bar{\tilde{z}}, \theta) &= -\bar{q}^*(0) F_0(\tilde{z}, \bar{\tilde{z}}) q(\theta) - q^*(0) \bar{F}_0(\tilde{z}, \bar{\tilde{z}}) \bar{q}(\theta) + F_0, \\ &= -(g_{20} \frac{\tilde{z}^2}{2} + g_{11} \tilde{z}\bar{\tilde{z}} + g_{02} \frac{\bar{\tilde{z}}^2}{2} + g_{21} \frac{\tilde{z}^2 \bar{\tilde{z}}}{2} + \ldots..) q(\theta) - (\bar{g}_{20} \frac{\bar{\tilde{z}}^2}{2} + \bar{g}_{11} \tilde{z}\bar{\tilde{z}} \\ &\quad + \bar{g}_{02} \frac{\tilde{z}^2}{2} + \bar{g}_{21} \frac{\bar{\tilde{z}}^2 \tilde{z}}{2}) q(\theta) + F_0\end{aligned}$$

$$\text{where } F_0 = \tau' \begin{bmatrix} -\beta_1 x_{2t}(0) x_{3t}(0) \\ \beta_1 x_{2t}(0) x_{3t}(0) - \beta_2 e^{-b\tau'} x_{1t}(-1) x_{2t}(-1) \\ 0 \end{bmatrix}$$

$$= \tau' \begin{bmatrix} -\beta_1 \alpha' \beta' \\ \beta_1 \alpha' \beta' - \beta_2 e^{-b\tau'} \alpha' e^{-\iota\sigma'\tau'} \\ 0 \end{bmatrix} \tilde{z}^2$$

$$+ \begin{bmatrix} -\beta_1 2Re\{\alpha' \bar{\beta}'\} \\ \beta_1 2Re\{\alpha' \bar{\beta}'\} - \beta_2 e^{-b\tau'} 2Re\{\alpha' e^{-\iota\sigma'\tau'}\} \\ 0 \end{bmatrix} \tilde{z}\bar{\tilde{z}} + \ldots$$

And thus, after comparing the coefficients, we get

$$\tilde{H}_{20}(0) = -g_{20} q(0) - g_{\bar{0}2} \bar{q}(0) + 2\tau' \begin{bmatrix} -\beta_1 \alpha' \beta' \\ \beta_1 \alpha' \beta' - \beta_2 e^{-b\tau'} \alpha' e^{-\iota\sigma'\tau'} \\ 0 \end{bmatrix}, \tag{45}$$

and

$$\tilde{H}_{11}(0) = -g_{11} q(0) - g_{\bar{1}1} \bar{q}(0) + 2\tau' \begin{bmatrix} -\beta_1 Re\{\alpha' \bar{\beta}'\} \\ \beta_1 Re\{\alpha' \bar{\beta}'\} - \beta_2 e^{-b\tau'} Re\{\alpha' e^{-\iota\sigma'\tau'}\} \\ 0 \end{bmatrix}. \tag{46}$$

Substituting the above values in (37), and noticing that $(\iota\sigma'\tau' I - \int_{-1}^{0} d\eta(\theta) e^{\iota\sigma'\tau'\theta})$ $q(0) = 0$, and $(-\iota\sigma'\tau' I - \int_{-1}^{0} d\eta(\theta) e^{-\iota\sigma'\tau'\theta}) \bar{q}(0) = 0$ (since $\iota\sigma'\tau'$ is the eigenvalue of $A(0)$ and $q(0)$ is the corresponding eigenvector), we obtain

$$\tilde{E}_1 \left(2\iota\sigma'\tau' I - \int_{-1}^{0} d\eta(\theta) e^{2\iota\sigma'\tau'} \right) = 2\tau' \begin{bmatrix} -\beta_1 \alpha' \beta' \\ \beta_1 \alpha' \beta' - \beta_2 e^{-b\tau'} \alpha' e^{-\iota\sigma'\tau'} \\ 0 \end{bmatrix},$$

which leads to

$$\tilde{E}_1 \begin{bmatrix} 2\iota\sigma^{'}+b & \beta_1 Z^* & \beta_1 I^* \\ \beta_2 e^{-b\tau^{'}} I^* e^{-2\iota\sigma^{'}\tau^{'}} & 2\iota\sigma^{'}-\beta_1 Z^*+b+\beta_2 e^{-b\tau^{'}} S^* e^{-2\iota\sigma^{'}\tau^{'}} & -\beta I^* \\ -a & 0 & 2\iota\sigma'+a \end{bmatrix}$$
$$= 2 \begin{bmatrix} -\beta_1\alpha^{'}\beta^{'} \\ \beta_1\alpha^{'}\beta^{'}-\beta_2 e^{-b\tau^{'}}\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}} \\ 0 \end{bmatrix}.$$

And, from Crammer's rule for solving system of linear equations, we get

$$\tilde{E}_1^{(1)}$$
$$= \frac{2}{\tilde{M}_1} \begin{vmatrix} -\beta_1\alpha^{'}\beta^{'} & \beta_1 Z^* & \beta_1 I^* \\ \beta_1\alpha^{'}\beta^{'}-\beta_2 e^{-b\tau^{'}}\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}} & 2\iota\sigma^{'}-\beta_1 Z^*+b+\beta_2 e^{-b\tau^{'}} S^* e^{-2\iota\sigma^{'}\tau^{'}} & -\beta_1 I^* \\ 0 & 0 & 2\iota\sigma'+a \end{vmatrix}$$

$$\tilde{E}_1^{(2)}$$
$$= \frac{2}{\tilde{M}_1} \begin{vmatrix} 2\iota\sigma^{'}+b & -\beta_1\alpha^{'}\beta^{'} & \beta_1 I^* \\ \beta_2 e^{-b\tau^{'}} I^* e^{-2\iota\sigma^{'}\tau^{'}} & \beta_1\alpha^{'}\beta^{'}-\beta_2 e^{-b\tau^{'}}\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}} & -\beta_1 I^* \\ -a & 0 & 2\iota\sigma'+a \end{vmatrix}$$

$$\tilde{E}_1^{(3)}$$
$$= \frac{2}{\tilde{M}_1} \begin{vmatrix} 2\iota\sigma^{'}+b & \beta_1 Z^* & -\beta_1\alpha^{'}\beta^{'} \\ \beta_2 e^{-b\tau^{'}} I^* e^{-2\iota\sigma^{'}\tau^{'}} & 2\iota\sigma^{'}-\beta_1 Z^*+b+\beta_2 e^{-b\tau^{'}} S^* e^{-2\iota\sigma^{'}\tau^{'}} & \beta_1\alpha^{'}\beta^{'}-\beta_2 e^{-b\tau^{'}}\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}} \\ -a & 0 & 0 \end{vmatrix},$$

where $\tilde{M}_1 = \begin{vmatrix} 2\iota\sigma^{'}+b & \beta_1 Z^* & \beta_1 I^* \\ \beta_2 e^{-b\tau^{'}} I^* e^{-2\iota\sigma^{'}\tau^{'}} & 2\iota\sigma^{'}-\beta_1 Z^*+b+\beta_2 e^{-b\tau^{'}} S^* e^{-2\iota\sigma^{'}\tau^{'}} & -\beta_1 I^* \\ -a & 0 & 2\iota\sigma'+a. \end{vmatrix}$

Next, working in a similar pattern as above, we get

$$\tilde{E}_2^{(1)} = \frac{2}{\tilde{M}_2} \begin{vmatrix} \beta_1 Re\{\alpha^{'}\bar{\beta^{'}}\} & -\beta_1 Z^* & -\beta_1 I^* \\ -\beta_1 Re\{\alpha^{'}\bar{\beta^{'}}\}+\beta_2 e^{-b\tau^{'}} Re\{\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}}\} & \beta_1 Z^*-b-\beta_2 e^{-b\tau^{'}} S^* & \beta_1 I^* \\ 0 & 0 & -a \end{vmatrix}$$

$$\tilde{E}_2^{(2)} = \frac{2}{\tilde{M}_2} \begin{vmatrix} -b & \beta_1 Re\{\alpha^{'}\bar{\beta^{'}}\} & -\beta_1 I^* \\ -\beta_2 e^{-b\tau^{'}} I^* & -\beta_1 Re\{\alpha^{'}\bar{\beta^{'}}\}+\beta_2 e^{-b\tau^{'}} Re\{\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}}\} & \beta_1 I^* \\ a & 0 & -a \end{vmatrix}$$

$$\tilde{E}_2^{(3)} = \frac{2}{\tilde{M}_2} \begin{vmatrix} -b & -\beta_1 Z^* & \beta Re\{\alpha^{'}\bar{\beta^{'}}\} \\ -\beta_2 e^{-b\tau^{'}} I^* & \beta_1 Z^*-b-\beta_2 e^{-b\tau^{'}} S^* & -\beta_1 Re\{\alpha^{'}\bar{\beta^{'}}\}+\beta_2 e^{-b\tau^{'}} Re\{\alpha^{'}e^{-\iota\sigma^{'}\tau^{'}}\} \\ a & 0 & 0 \end{vmatrix},$$

where $\tilde{M}_2 = \begin{vmatrix} -b & -\beta_1 Z^* & -\beta_1 I^* \\ -\beta_2 e^{-b\tau'} I^* & \beta_1 Z^* - b - \beta_2 e^{-b\tau'} S^* & \beta_1 I^* \\ a & 0 & -a \end{vmatrix}$.

Thus, we can determine $W_{20}(\theta)$ and $W_{11}(\theta)$, and hence, we can compute g_{21}.

Therefore, the behaviour of bifurcating periodic solutions in the centre manifold at the critical value $\tau = \tau'$ is computed by the following values:

- $\tilde{C}_1(0) = \frac{\iota}{2\sigma'\tau'}(g_{20}g_{11} - 2|g_{11}|^2 - \frac{|g_{02}|^2}{3}) + \frac{g_{21}}{2}$,
- $\tilde{\mu}_2 = -\frac{Re\{\tilde{C}_1(0)\}}{Re\{\frac{d\lambda(\tau')}{d\tau}\}}$,
- $\tilde{\beta''} = 2Re\{\tilde{C}_1(0)\}$,
- $\tilde{T}_2 = -\frac{Im\{\tilde{C}_1(0)\} + \tilde{\mu}_2 Im\{\frac{d\lambda(\tau')}{d\tau}\}}{\sigma'\tau'}$,

where

- $\tilde{\mu}_2$ determines the direction of Hopf bifurcation, for if $\tilde{\mu}_2 > 0$, the Hopf bifurcation will be supercritical, and if $\tilde{\mu}_2 < 0$, the Hopf bifurcation will be subcritical, and the bifurcating periodic solutions exist for $\tau > \tau'$ or $\tau < \tau'$.
- $\tilde{\beta''}$ determines the stability of the bifurcating periodic solutions, for if $\tilde{\beta''} < 0$, the bifurcating periodic solutions will be stable, and if $\tilde{\beta''} > 0$, the bifurcating periodic solutions will be unstable.
- $\tilde{T}_2$ determines the period of the bifurcating periodic solutions, for if $\tilde{T}_2 > 0$, the period increases, and if $\tilde{T}_2 < 0$, the period decreases.

7 Numerical Simulation

In this section, we have plotted graphs in support of our analytical results. We have shown that how the parameters β_1, β_2, a, and b along with τ shape the dynamics of the system. Initially, the trajectories approach to disease-free equilibrium point E_0 at $b = 0.5, \beta_1 = 0.1, \beta_2 = 2, \tau = 3.22$, and $a = 0.2$ for initial conditions $(0.5, 0.5, 0.2)$, and the value of $R_0 = 0.1111 < 1$ along with $b > \beta_1 + \beta_2 e^{-b\tau}$ is satisfied as per Theorem 4.1 of local stability of DFE.

Modulating the parameters to $b = 0.2, \beta_1 = 10.6, \beta_2 = 5.7$, and $a = 0.25$ pertaining to same initial conditions $(0.5, 0.5, 0.2)$, endemic equilibrium point exists, and it also violates the condition $b > \beta_1 + \beta_2 e^{-b\tau}$ of disease-free equilibrium point. The equilibrium point obtained is $E^*(0.0465, 0.1613, 0.0256)$ with $R_0 = 3.4310 > 1$. However, interestingly, the system remains stable for lower value of τ, i.e. $\tau \leq \tau^* = 1.9$ (Fig. 1a) and undergoes Hopf bifurcation as $\tau > 1.9$. This shows that as we increase the rate of infection, endemicity exits in the system for even lower value of τ. Furthermore, as we increase the value of τ, the system undergoes bifurcation (Fig. 1d).

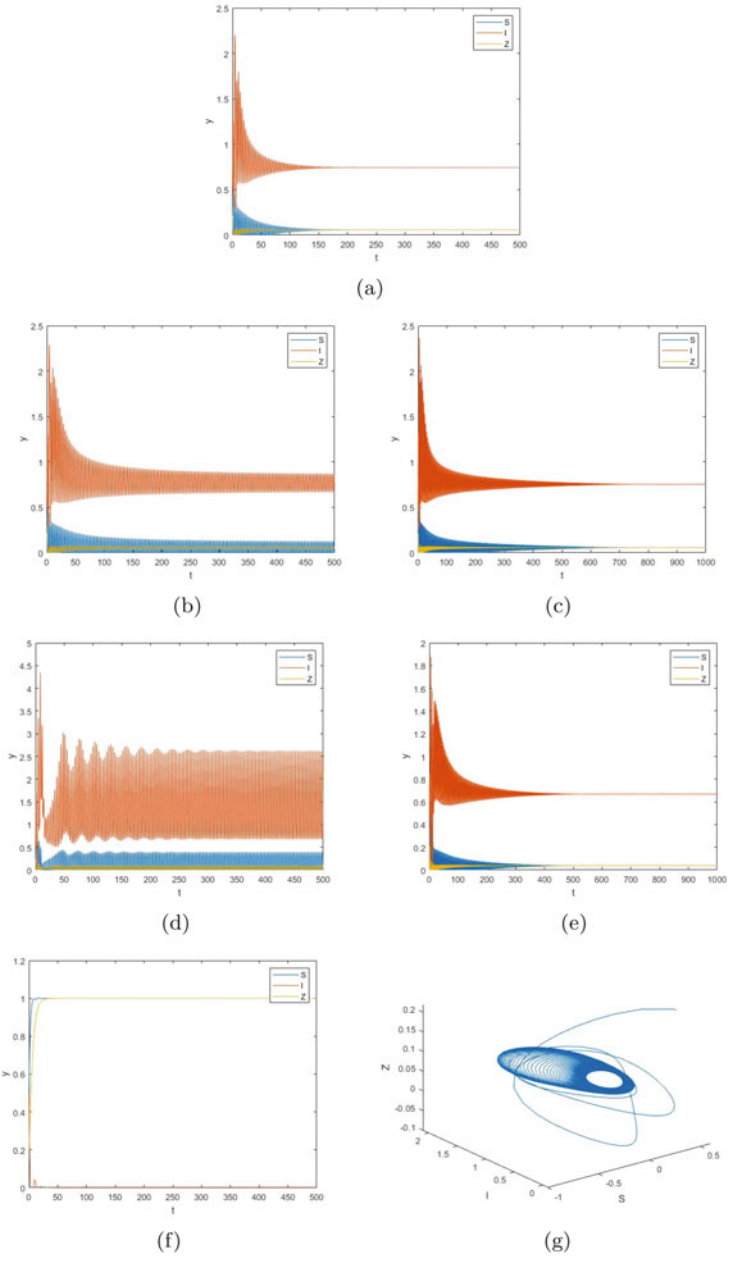


Fig. 1 Dynamics of the system. (**a**) Behaviour of system at $\tau = 1.9$. (**b**) Behaviour of system at $\tau = 2$. (**c**) Behaviour of system at $a = 0.25$. (**d**) Periodic solution of system at $\tau = 3$. (**e**) Stable behaviour of system at $b = 0.29$. (**f**) 2D graph for disease-free equilibrium. (**g**) Phase plane behaviour of Endemic equilibrium point

Contributing to the sensitivity of the model parameters, a, b, β_1, and β_2 elevate system behaviour prominently. The system at $b = 0.2, \beta_1 = 10.6, \beta_2 = 5.7$, and $\tau = 1.9$ shows Hopf bifurcation for $a > 0.25$ (Fig. 1b), but as soon as the value reduces, i.e. $a \leq 0.25$ (Fig. 1c), the system stabilizes. This change in the dynamics is also visible for b where the system at $a = 0.25, \beta_1 = 10.6, \beta_2 = 5.7$, and $\tau = 1.9$ is stable only till $b < 0.3$ (Fig. 1e), which concludes that the birth rate and death rate should be controlled to avoid periodic solution in the system. β_1 and β_2 also mark their presence. Increasing them beyond a certain limit, i.e. $\beta_1 > 10.6$ (Fig. 1b, $\beta_1 = 11$) and $\beta_2 > 5.7$ (Fig. 1b, $\beta_2 = 6$), leads to periodic solution in the system at $b = 0.2, a = 0.25$, and $\tau = 1.9$. Hence, infection rate plays a vital role in destabilizing the system, and its control should be the utmost priority for any government to fight against this dreadful disease.

8 Discussion

To sum up, in this chapter, we have studied the dynamical behaviour of a Dengue-SIR epidemic model that involves both discrete and distributed delays. We have studied the existence of disease-free and endemic equilibrium points. Furthermore, we have studied the local stability of disease-free equilibrium point, and it has been proved that DFE is stable for $R_0 < 1$ if $b > \beta_1 + \beta_2 e^{-b\tau}$. The endemic equilibrium point exists if the basic reproduction number, $R_0 > 1$ and is locally asymptotically stable when $\tau \in [0, \tau')$ and possesses periodic solution for $\tau > \tau'$. Moreover, using the normal form theory and centre manifold theorem, we have derived explicit formulae in order to determine the stability and direction of the bifurcating periodic solutions and obtained sensitivity analysis for all the parameters, i.e. β_1, β_2, a, b, and τ involved in the system depicting their influence on system stabilization. Therefore, in order to control dengue, government should consider all these parameters collectively for effective eradication of dengue.

Appendix

The reduction of $Z = \int_{-\infty}^{t} F(t-\tau)S(\tau)d\tau$ to ordinary equation is done by the Leibnitz rule, which states that

$$\frac{d}{dx}\Big(\int_{a(x)}^{b(x)} f(x,t)dt\Big) = f(x,b(x)).\frac{d}{dx}b(x) - f(x,a(x)).\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)dt,$$

which gives

$$\frac{dZ}{dt} = aS - aZ.$$

References

1. B. Bandyopadhyay, I. Bhattacharyya, et al., A comprehensive study on the 2012 Dengue fever outbreak in Kolkata, India. Int. Sch. Res. Notices **2013**, 1–5 (2013)
2. C.P. Simmons, J.J. Farrar, N. van Vinh Chau, W.B. Dengue. New England J. Med. **366**, 1423–1432 (2012)
3. M. Andraud, et al. A simple periodic-forced model for dengue fitted to incidence data in Singapore. Math. Biosci. **244**, 22–28 (2013)
4. C. Favier, et al., Early determination of the reproductive number for vector-borne diseases: the case of dengue in Brazil. Trop. Med. Int. Health **11**, 332–340 (2006)
5. S.B. Halstead, Dengue. Lancet **370**, 1644–1652 (2007)
6. W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics. Bull. Math. Biol. **53**, 33–55 (1991)
7. X. Meng, L. Chen, Global dynamical behaviors for an SIR epidemic model with time delay and pulse vaccination. Taiwan. J. Math. **12**, 1107–1122 (2008)
8. W. Zhao, T. Zhang, Z. Chang, X. Meng, Y. Liu, Dynamical analysis of SIR epidemic models with distributed delay. J. Appl. Math. **2013**, 1–15 (2013)
9. S. Chauhan, S.K. Bhatia, S. Sharma, Effect of delay on single population with infection in polluted environment. Int. J. Math. Comput. **29**, 132–150 (2018)
10. J. Ma, Q. Gao, Stability and Hopf bifurcations in a business cycle model with delay. Appl. Math. Comput. **215**, 829–834 (2009)
11. D. Lv, W. Zhang, Y. Tang, Bifurcation analysis for a ratio-dependent predator-prey system with multiple delays. J. Nonlinear Sci. Appl. **9**, 3479–3490 (2016)
12. N. Gupta, S. Srivastava, et al., Dengue in India. Indian J. Med. Res. **136**, 373–390 (2012)
13. V. Racloz, R. Ramsey, et al., Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. PLoS Neglected Trop. Disease **6**, 1–9 (2012)
14. A. Asmaidi, P. Sianturi, et al., A SIR mathematical model of dengue transmission and its simulation. IOSR J. Math. **10**, 56–65 (2014)
15. M. Derouich, A. Boutayeb, Dengue fever: mathematical modelling and computer simulation. Appl. Math. Comput. **177**, 528–544 (2006)
16. M.R. Calsavara, et al., An analysis of a mathematical model describing the geographic spread of dengue disease. J. Math. Anal. Appl. **444**, 298–325 (2016)
17. C. Wu, P.J.Y. Wong, Dengue transmission: mathematical model with discrete time delays and estimation of the reproduction number. J. Biol. Dyn. **13**, 1–25 (2019)
18. W. Zhao, T. Zhang, Z. Chang, X. Meng, Y. Liu, Dynamical analysis of SIR epidemic models with distributed delay. J. Appl. Math. **2013**, 1–15 (2013)
19. N. MacDonald, *Time Lags in Biological Models*. Lecture Notes in Biomathematics, vol. 27 (Springer, Heidelberg, 1978)
20. B. Hassard, D. Kazarinoff, Y. Wan, *Theory and Applications of Hopf Bifurcation*. Contributions to Nonlinear Functional Analysis (Cambridge University Press, Cambridge, 1981)

Deformable Derivative of Fibonacci Polynomials

Krishna Kumar Sharma

Abstract The Fibonacci sequence is the most spectacular subject in mathematics, and the Fibonacci polynomials are generalizations of Fibonacci numbers made by various authors. The main objective of this research paper is to construct the relation between deformable derivative and Fibonacci polynomials. Using this relationship, the basic properties of the Fibonacci polynomial are proposed and discussed. In this article, the generating function of the Fibonacci polynomial for the deformable derivative is also explained.

Keywords k-Fibonacci sequence · Fractional derivative

2010 Mathematics Subject Classification 11B39, 26A33

1 Introduction

In the present time, there are innumerable applications of Fibonacci numbers [1–3]. It has resulted in a variety of competing conceptual and mathematical models that have been conceptualized to describe the applications of Fibonacci numbers. Fibonacci numbers have originated from the well known Fibonacci series that was innovated during the study of the population growth of Rabbits. Lovers of art, nature, mathematics etc. have continuously been awe-struck by famous Fibonacci numbers. For centuries, researchers have been working on this concept specially those associated with Fibonacci Association. Their efforts have opened new doors for research in connected areas. It cannot be denied that almost every field of

K. K. Sharma (✉)
School of Vocational Studies and Applied Sciences, Gautam Buddha University, Greater Noida, UP, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_8

science and arts has utilized Fibonacci numbers and their generalizations to suit their purpose. Mathematicians of different realms have worked on Fibonacci numbers in geometry, algebra, number theory, and many other branches of mathematics.

1.1 Fibonacci Sequence

The Fibonacci sequence is defined by the recurrence relation

$$F_n = F_{n-1} + F_{n-2}, F_0 = 0, F_1 = 1. \tag{1}$$

1.2 Binet's Formula for Fibonacci Sequence

The general term of Fibonacci sequence can be defined by Binet's formula

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta} = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right], \tag{2}$$

where α and β are the roots of polynomial $t^2 - t - 1 = 0$.

1.3 k-Fibonacci Sequence

Falcon et al. [4, 5] defined k-Fibonacci sequence in this way.

For any real number k,

$$F_{k,n+1} = kF_{k,n} + F_{k,n-1}, F_{k,0} = 0, F_{k,1} = 1. \tag{3}$$

The original Fibonacci sequence can be obtained by putting $k = 1$.

$$F_{n+1} = F_n + F_{n-1}, F_0 = 0, F_1 = 1; \tag{4}$$

if ζ denotes the positive roots of the equation $t^2 = kt + 1$, then the general term can be expressed as

$$F_{k,n} = \frac{\zeta^n - (-\zeta)^{-n}}{\zeta + \zeta^{-1}}, \tag{5}$$

where $\zeta = \frac{k+\sqrt{k^2+4}}{2}$.

Falcon [4] obtained two formulae for the general term of the k-Fibonacci sequence

$$F_{k,n} = \frac{1}{2^{n-1}} \sum_{i=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n}{2i+1} k^{(n-2i-1)} (k^2+4)^i \tag{6}$$

$$F_{k,n} = \sum_{i=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1-i}{2i+1} k^{(n-2i-1)}. \tag{7}$$

1.4 The Fibonacci Polynomials

Firstly, Fibonacci polynomials were studied by Catalan, and these Fibonacci polynomials are defined as recurrence relation

$$F_{n+1} = tF_n(t) + F_{n-1}(t), n \geq 2, F_1(t) = 1, F_2(t) = t, \tag{8}$$

from where the first Fibonacci polynomials are

$F_1(t) = 1$
$F_2(t) = t$
$F_3(t) = t^2 + 1$
$F_4(t) = t^3 + 2t$
$F_5(t) = t^4 + 3t^2 + 1$
$F_6(t) = t^5 + 4t^3 + 3t$
$F_7(t) = t^6 + 5t^4 + 6t^2 + 1$
$F_8(t) = t^7 + 6t^5 + 10t^3 + 4t$

From these Fibonacci polynomials, Falcon [4] mentioned the following result:

$$F_{n+1} = \sum_{i=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-i}{i} t^{n-2i}, n \geq 0. \tag{9}$$

1.5 Derivative of the Fibonacci Polynomials

$F_1'(t) = 0$
$F_2'(t) = 1$
$F_3'(t) = 2t$
$F_4'(t) = 3t^2 + 2$

$F_5'(t) = 4t^3 + 6t$
$F_6'(t) = 5t^4 + 12t^2 + 3$
$F_7'(t) = 6t^5 + 20t^3 + 12t$
$F_8'(t) = 7t^6 + 30t^4 + 30t^2 + 4$

2 Deformable Fractional Derivative

Fractional calculus [6] is an effective tool that has been commonly used in many aspects of electronics engineering and computer science applications. Although it has a complex mathematical background, fractional calculus was discovered due to very simple problems associated with the concept of derivation. In the case, the first order derivative represents the slope of a function; What does a half-order derivative of a function represent? The results of such questions have resulted in many new unexplored diversions in the field of mathematical research. Fractional calculus has been playing an important role in the fields of signal and image processing, mechanics, control theory, biology, chemistry, economics, etc. In the current era, fractional differentiation has been investigated simultaneously by several authors and researchers. The main thrust of the research article is on deformable derivation. In the present study, we introduce the deformable derivative of the Fibonacci polynomial, which is an extension of the Fibonacci numbers. We construct many interesting relationships with its deformable derivatives. These derivatives give us a new set of integer sequences.

According to Zulfequarr et al. [7], fractional derivative can be stated as follows.

For a given number α, $0 \leq \alpha \leq 1$.

$$D^{\alpha} f(t) = \lim_{\epsilon \to 0} \frac{(1 + \epsilon\beta) f(t + \epsilon\alpha) - f(t)}{\epsilon}, \alpha + \beta = 1. \tag{10}$$

They also defined the connection between α-derivative and ordinary derivative.

$$D^{\alpha} f(t) = \beta f(t) + \alpha D f(t), \alpha + \beta = 1. \tag{11}$$

2.1 Basic Properties of the Deformable Fractional Derivative

The operator D^{α} possesses the following properties:

(1) Linearity: $D^{\alpha}(af + bg) = aD^{\alpha}(f) + bD^{\alpha}(g)$.
(2) Commutativity: $D^{\alpha_1} D^{\alpha_2} = D^{\alpha_2} D^{\alpha_1}$.
(3) For a constant function K, $D^{\alpha}(K) = \beta K$.

(4) $D^{\alpha}(f.g) = D^{\alpha}(f).g + \alpha f.Dg, \implies D^{\alpha}$ does not obey product rule.
(5) $D^{\alpha}(t^r) = \beta t^r + r\alpha t^{r-1}, r \in \mathbb{R}$.
(6) $D^{\alpha}(e^t) = e^t$.
(7) $D^{\alpha}(\sin t) = \beta \sin t + \alpha \cos t$.
(8) $D^{\alpha}(\log t) = \beta \log t + \frac{\alpha}{t}, t > 0$.

2.2 Deformable Derivative of the Fibonacci Polynomial

$D^{\alpha}[F_1(t)] = \beta$
$D^{\alpha}[F_2(t)] = \beta t + \alpha$
$D^{\alpha}[F_3(t)] = \beta(t^2 + 1) + 2\alpha t$
$D^{\alpha}[F_4(t)] = \beta(t^3 + 2t) + \alpha(3t^2 + 2)$
$D^{\alpha}[F_5(t)] = \beta(t^4 + 3t^2 + 1) + \alpha(4t^3 + 6t)$
$D^{\alpha}[F_6(t)] = \beta(t^5 + 4t^3 + 3t) + \alpha(5t^4 + 12t^2 + 3)$
$D^{\alpha}[F_7(t)] = \beta(t^6 + 5t^4 + 6t^2 + 1) + \alpha(6t^5 + 20t^3 + 12t)$
$D^{\alpha}[F_8(t)] = \beta(t^7 + 6t^5 + 10t^3 + 4t) + \alpha(7t^6 + 30t^4 + 30t^2 + 4)$

2.3 Relation Between the Deformable Derivative Sequence and Fibonacci Sequence

$$D^{\alpha}[F_n(t)] = \frac{\left\{\beta(t^2+4) - \alpha t\right\} F_n(t) + n\alpha \left\{F_{n+1}(t) + F_{n-1}(t)\right\}}{t^2+4}, \alpha + \beta = 1. \tag{12}$$

Proof We know that

$D^{\alpha} F_n(t) = \beta F_n(t) + \alpha D F_n(t)$;
$= \beta F_n(t) + \alpha F_n'(t)$.

□

S. Falcon [8] constructed the following relation between derivative sequence and the Fibonacci sequence:

$$F_n'(t) = \frac{n F_{n+1}(t) - t F_n(t) + n F_{n-1}(t)}{t^2+4}. \tag{13}$$

Then, $D^{\alpha} F_n(t) = \beta F_n(t) + \alpha \left[\frac{n F_{n+1}(t) - t F_n(t) + n F_{n-1}(t)}{t^2+4}\right]$
$= \frac{\beta(t^2+4) F_n(t) + \alpha[n F_{n+1}(t) - t F_n(t) + n F_{n-1}(t)]}{t^2+4}$.

Hence,

$$D^{\alpha}F_n(t) = \frac{\{\beta(t^2+4)-\alpha t\}F_n(t)+n\alpha\{F_{n+1}(t)+F_{n-1}(t)\}}{t^2+4}, \alpha+\beta=1.$$

This result can also be proved by Binet's Formula

$$F_n(t) = \frac{\zeta^n-(-\zeta)^{-n}}{\zeta+\zeta^{-1}}, \text{ where } \zeta=\frac{t+\sqrt{t^2+4}}{2}.$$

Using Binet's form, $D^{\alpha}F_n(t) = \beta\left[\frac{\zeta^n-(-\zeta)^{-n}}{\zeta+\zeta^{-1}}\right] + \alpha D[\frac{\zeta^n-(-\zeta)^{-n}}{\zeta+\zeta^{-1}}\Big]$

$D^{\alpha}F_n(t) = \beta F_n(t) + \alpha\left[n\left\{\frac{\zeta^{n-1}-(-\zeta)^{n-1}}{(\zeta+\zeta^{-1})}\zeta'\right\} - \left\{\frac{\zeta^n-(-\zeta)^{-n}}{(\zeta+\zeta^{-1})^2}(1-\zeta'^2)\zeta'\right\}\right]$. Here

$\zeta' = \frac{\zeta}{\zeta+\zeta^{-1}}$ and$(1-\zeta'^2) = \frac{t}{\zeta}$;

$D^{\alpha}F_n(t) = \beta F_n(t) + \alpha\left[n\left\{\frac{\zeta^n+(-\zeta)^{-n}}{(\zeta+\zeta^{-1})^2}\right\} - \left\{\frac{\zeta^n-(-\zeta)^{-n}}{(\zeta+\zeta^{-1})}\frac{t}{(\zeta+\zeta^{-1})^2}\right\}\right]$.

$= \beta F_n(t) + \alpha[n\left\{\frac{\zeta^n+(-\zeta)^{-n}}{(\zeta+\zeta^{-1})^2}\right\} - \left\{\frac{tF_n(t)}{(\zeta+\zeta^{-1})^2}\right\}]$.

We have

$$F_{n+1}(t)+F_{n-1}(t) = \frac{\zeta^{n+1}-(-\zeta)^{-n-1}}{\zeta+\zeta^{-1}} + \frac{\zeta^{n-1}-(-\zeta)^{-n+1}}{\zeta+\zeta^{-1}} = \zeta^n+(-\zeta)^{-n}.$$

$$\Longrightarrow\ D^{\alpha}[F_n(t) = \beta F_n(t) + \alpha\left[\frac{n\{F_{n+1}(t)+F_{n-1}(t)\}-tF_n(t)}{(\zeta+\zeta^{-1})^2}\right].$$

Hence,

$$D^{\alpha}[F_n(t)] = \frac{\{\beta(t^2+4)-\alpha t\}F_n(t)+n\alpha\{F_{n+1}(t)+F_{n-1}(t)\}}{t^2+4}, \text{ where } \alpha+\beta=1.$$

2.4 Expression of α-Deformable Derivative of Fibonacci Polynomials

From the result (8), we have

$$F_{n+1}(t) = \sum_{i=0}^{\left\lfloor\frac{n-1}{2}\right\rfloor}\binom{n-i}{i}t^{n-2i}, n\geq 0,$$

and by definition of α-deformable derivative,

$D^{\alpha}[F_{n+1}(t)] = \beta F_{n+1}(t) + \alpha D[F_{n+1}(t)]$;

$D^{\alpha}[F_{n+1}(t)] = \beta\sum_{i=0}^{\left\lfloor\frac{n-1}{2}\right\rfloor}\binom{n-i}{i}t^{n-2i} + \alpha\sum_{i=0}^{\left\lfloor\frac{n-1}{2}\right\rfloor}(n-2i)\binom{n-i}{i}t^{n-2i-1}$, where $\alpha+\beta=1$ and $D^{\alpha}F_1(t)=\beta$.

2.5 α-Deformable Derivative of Fibonacci Polynomials and Convolved Fibonacci Polynomials

$$D^{\alpha}[F_n(t)] = \beta F_n(t) + \alpha \sum_{i=1}^{n-1} F_i(t) F_{n-i}(t),\ n > 1,\ D^{\alpha}[F_1(t)] = \beta. \tag{14}$$

Proof By the induction method, for $n = 2$,

$D^{\alpha}[F_2(t)] = \beta F_2(t) + \alpha \sum_{i=1}^{1} F_1(t) F_1(t)$;
$\implies D^{\alpha}[F_2(t)] = \beta t + \alpha.$

□

Let us suppose that the result is true for every polynomial $D^{\alpha}[F_m(t)], m \leq n$

$D^{\alpha}[F_{n-1}(t)] = \beta F_{n-1}(t) + \alpha \sum_{i=1}^{n-1} F_i(t) F_{n-1-i}(t)$.

By the definition of Fibonacci polynomial, we have $F_{n+1} = tF_n(t) + F_{n-1}(t)$. Then,

$$\begin{aligned}
D^{\alpha}[F_{n+1}(t)] &= D^{\alpha}[tF_n(t)] + D^{\alpha}[F_{n-1}(t)] \\
&= D^{\alpha}[tF_n(t)] + D^{\alpha}[F_{n-1}(t)] \\
&= (\beta t + \alpha) F_n(t) + \alpha t F_n'(t) + \beta F_{n-1}(t) + \alpha F_{n-1}'(t) \\
&= (\beta t + \alpha) F_n(t) + \alpha t \sum_{i=1}^{n-1} F_i(t) F_{n-i}(t) \\
&\quad + \beta F_{n-1}(t) + \alpha \sum_{i=1}^{n-2} F_i(t) F_{n-1-i}(t) \\
&= (\beta t + \alpha) F_n(t) + \beta F_{n-1}(t) + \alpha t F_{n-1}(t) F_1(t) \\
&\quad + \alpha \sum_{i=1}^{n-2} F_i(t) F_{n-i}(t) + \alpha \sum_{i=1}^{n-2} F_i(t) F_{n-1-i}(t) \\
&= \beta [tF_n(t) + F_{n-1}(t)] + \alpha F_n(t) \\
&\quad + \alpha t F_{n-1}(t) F_1(t) + \alpha \sum_{i=1}^{n-2} F_i(t) [tF_{n-i}(t) + F_{n-1-i}(t)] \\
&= \beta F_{n+1}(t) + \alpha F_n(t) F_1(t) + \alpha F_{n-1}(t) F_2(t) \\
&\quad + \alpha \sum_{i=1}^{n-2} F_i(t) [tF_{n-i}(t) + F_{n-1-i}(t)].
\end{aligned}$$

Hence,

$$D^{\alpha}[F_{n+1}(t)] = \beta F_{n+1}(t) + \alpha \sum_{i=1}^{n-2} F_i(t) F_{n+1-i}(t), \tag{15}$$

and it implies that the result is true for $n = n + 1$.

We can also conclude by Eqs. (12) and (14)

$$\begin{aligned}
&\beta F_n(t) + \alpha \sum_{i=1}^{n-1} F_i(t) F_{n-i}(t) \\
&\quad = \left[\frac{\left\{\beta(t^2+4) - \alpha t\right\} F_n(t) + n\alpha \left\{F_{n+1}(t) + F_{n-1}(t)\right\}}{t^2+4} \right] \\
&\alpha \sum_{i=1}^{n-1} F_i(t) F_{n-i}(t) \\
&\quad = \left[\frac{\left\{\beta(t^2+4) - \alpha t\right\} F_n(t) + n\alpha \left\{F_{n+1}(t) + F_{n-1}(t)\right\}}{t^2+4} \right] \\
&\quad - \beta F_n(t)
\end{aligned}$$

$$\sum_{i=1}^{n-1} F_i(t) F_{n-i}(t) = \frac{[(n-1)t F_n(t) + 2n F_{n-1}(t)]}{t^2+4}. \tag{16}$$

Proposition 2.1 *Let $F_n(t)$ be the Fibonacci polynomial; then,*

$$D^{\alpha}[F_{n+1}(t) + F_{n-1}(t)] = \beta[F_{n+1}(t) + F_{n-1}(t)] + n\alpha F_n(t). \tag{17}$$

Proof We know that

$$F_{n+1}(t) = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n-i}{i} t^{n-2i} \text{ and } F_{n-1}(t) = \sum_{i=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-2-i}{i} t^{n-2-2i}$$

$$\begin{aligned}
F_{n+1}(t) + F_{n-1}(t) &= t^n + \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n-i}{i} t^{n-2i} + \sum_{i=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-2-i}{i} t^{n-2-2i} \\
&= t^n + \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \left[\binom{n-i}{i} + \binom{n-1-i}{i} \right] t^{n-2i}.
\end{aligned}$$

We also know $[\binom{n-i}{i} + \binom{n-1-i}{i}] = \binom{n-1-i}{i-1}(\frac{n-i}{i} + 1) = \binom{n-1-i}{i}\frac{n}{i}$.

Then,

$$F_{n+1}(t) + F_{n-1}(t) = t^n + \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n-1-i}{i-1} \frac{n}{i} t^{n-2i}.$$

Taking α-deformable derivative both the sides,

$$\begin{aligned} D^{\alpha}[F_{n+1}(t) + F_{n-1}(t)] &= \beta t^n + \alpha n t^{n-1} + n\beta \left[\sum_{i=1}^{\frac{n}{2}} \binom{n-1-i}{i-1} \frac{1}{i} t^{n-2i} \right] \\ &\quad + \alpha n \sum_{i=1}^{\frac{n}{2}} \binom{n-1-i}{i} \frac{n-2i}{i} t^{n-1-2i} \\ &= \beta t^n + n\beta \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n-1-i}{i-1} \frac{1}{i} t^{n-2i} \\ &\quad + \alpha n \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n-1-i}{i} \frac{n-2i}{i} t^{n-1-2i}. \end{aligned}$$

Hence,

$$D^{\alpha}[F_{n+1}(t) + F_{n-1}(t)] = \beta[F_{n+1}(t) + F_{n-1}(t)] + n\alpha F_n(t).$$

□

2.6 Generating Function for the α-Deformable Derivative of Fibonacci Polynomials

In 2007, Falcon [5] obtained the generating function of the k-Fibonacci polynomial

$$F_k(t) = \frac{t}{1 - kt - t^2}.$$

Now, we find the deformable derivative of $F_k(t)$

$$D^{\alpha}[F_k(t)] = D^{\alpha}\left[\frac{t}{1 - kt - t^2}\right]$$

$$= t\left[\left\{\beta\left(\frac{1}{1-kt-t^2}\right)\right\} + \alpha\left\{\frac{1}{(1-kt-t^2)^2}\right\}\right]$$
$$= \frac{\beta t}{1-kt-t^2} + \alpha\left\{\frac{t}{1-kt-t^2}\right\}^2.$$

3 Conclusion

It cannot be denied that the fractional differentiation is a generalization of classical calculus and the k-Fibonacci numbers are the generalization of the Fibonacci numbers. In the current study, deformable derivative of Fibonacci polynomials has been obtained. During this study, many identities have been constructed and the deformable derivative of these polynomials has been expressed in the form of convolution of Fibonacci Polynomials. The results of this paper will hopefully act as a stimulus for researchers and mathematicians to work in the field of fractional calculus in a prolific manner.

References

1. N.N. Vorobyou, *The Fibonacci Numbers* (D.C. Health Company, Boston, 1963)
2. S. Vajda, Fibonacci and Lucas numbers and the golden section, *Theory and Applications* (Ellis Horwood, Chichester, 1989)
3. T. Koshy, *Fibonacci and Lucas Numbers with Applications* (Wiley-Interscience Publication, New York, 2001)
4. S. Falcon, A. Plaza, The k-Fibonacci sequence and the Pascal 2-triangles Chaos, Solitons Fractals **33**(1), 38–49 (2007)
5. S. Falcon, A. Plaza, On the Fibonacci k-numbers, Chaos, Solitons Fractals **32**, 1615–24 (2007)
6. K.S. Miller, B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations* (Willey, New York, 1993)
7. F. Zulfequarr, A. Ujlayan, P. Ahuja, A New Fractional Derivative and its Fractional Integral with Some Applications (2017). arXiv:1705.00962v1
8. S. Falcon, A. Plaza, On k-Fibonacci sequence and polynomials and their derivatives. Chaos, Solitons Fractals **39**, 1005–1019 (2009)

Part II
Optimization and Optimal Control

Simulation and Analysis of 5G Wireless mm-Wave Modulation Technique for High Capacity Communication System

M. Vinothkumar and Vinod Kumar

Abstract Millimeter-wave (mm-wave) technology has been observed as an active part in 5th generation (5G) systems because of its potential for low-latency, multi-gigabit wireless links. The challenging fact in mm-wave technology is higher propagation losses at advanced carrier frequencies and also the increased complexity of hardware required. Multiple-input multiple-output (MIMO) is a key technology for increasing the capacity of 5G networks and the capability of supporting a large number of users. mm-Wave MIMO is considered to be the significant enabler for 5G wireless networks, which causes the maximum growth in the network capacity. The massive antennas connection and multiple radio frequency (RF) chains in wireless communication system cause excessive power consumption. In MIMO system, spatial modulation (SM) technique enables less complexity and low power consumption by reducing RF chain counts. This chapter elaborates simulation and analyses the result of SM technique that can be effectively implemented in mm-wave MIMO system to reduce power consumption.

Keywords Multi-input multi-output · Millimetre-wave · Spatial modulation · Spectral efficiency

1 Introduction

In wireless communication systems, significant capacity boosting can be done by MIMO systems that gained massive research attention recently. Achieving a high throughput and cost-effective deployment are the requirements for new transmission technologies to overcome today's rapid proliferation of mobile data traffic. High capacity, low latency, and huge connectivity over the scarce wireless resources are the scope for the future wireless communication systems where tremendous efforts

M. Vinothkumar · Vinod Kumar (✉)
SRMIST - NCR Campus, Modinagar, UP, India
e-mail: vinothkm@srmist.edu.in; vinodkur1@srmist.edu.in

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_9

have been put through to have a better knowledge about the benefits of mm-wave MIMO systems under different considerations. A MIMO system enhances spectral efficiency (SE) by multiple antennas used concurrently to transmit information bits to the receiver. Large bandwidth (30–300 GHz) provided by mm-wave bands [2] leads it to become the great source of future wireless communication systems.

Three main types of MIMO techniques are (1) improving power efficiency by maximizing spatial diversity, (2) exploiting knowledge of the channel at the transmitter, and (3) layered space–time method of transmitting many independent data streams through the antenna, which increases capacity [6]. The small wavelength in mm-wave also inspired by deploying a large number of antennas in MIMO technology results in considerable gain improvement in terms of spectral efficiency. The critical challenge in implementing mm-wave MIMO and massive MIMO is the large number of RF chains required to process signals, which increases hardware complexity and also costs more power consumption [9].

As an emerging technique, spatial modulation (SM) [5] has become the solution to the above challenge to reduce the number of RF chains used in traditional mm-wave and massive MIMO systems. The SM technique uses antenna indexes in a multiple antenna system and can be a new hopeful transmission technique for means of data transmissions. Random antennas switching method of SM is setting one antenna active at a moment and other antennas silent. This reduces RF chain usage in MIMO communication [1], a good trade-off between achieving spectral efficiency and the number of RF chains required to implement MIMO system by incorporating SM. Section II elaborates spatial modulation (SM) technique and further the simulation results showing bit error rate (BER) performance of spatial modulation (SM) and signal-to-noise ratio (SNR).

2 Spatial Modulation

Hardware complexity in MIMO system implementation is effectively reduced by incorporating SM techniques that makes one antenna active to transmit bits of information at a time and the other antenna is kept silence so that SM is using only one RF chain. Selection of active antenna is made using $m = \log_2 N_t$ bits. Modulation techniques [3] such as binary phase shift keying (BPSK), quadrature phase shift keying (QPSK) and quadrature amplitude modulation (QAM) are basically mapping a group of information bits into a symbol that represents a constellation point in complex two-dimensional diagrams. The approach of extending two dimensions into three dimensions [8] known as spatial dimension demonstrated a flexible mechanism that achieved high spectral efficiency and low complexity. The information bits to be transmitted depend on constellation diagram and the number of transmit antennas. On the selected active antenna, a symbol from M-ary modulation such as M-PSK and M-QAM is sent. The remaining antennas $(N_t - 1)$ are silent except active antenna. Therefore, bits to be transmitted per channel use (bpcu) are $\log_2 N_t + \log_2 M$.

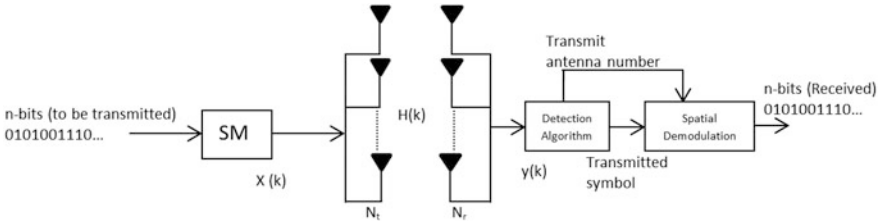


Fig. 1 Spatial modulation

Figure 1 shows SM system model where n bits to be transmitted are spatially modulated resulting in the vector $x(k) = [0xl \ldots .0]$ of size N_t. xl is the symbol transmitted from antenna number l over channel $H(k)$. $H(k)$ (where, $H = [h_1 h_2 h_3 \ldots hN_t]$) is the vector corresponding to channel path gain between transmitting and receiving antennas. The received vector is $y(k)$ ($y = Hx_1 +$ w), w-additive white Gaussian noise vector(AWGN). The total number of bits transmitted using SM is $n = \log_2(N_t) + m$. Estimating the transmitting antenna number is an important key in SM. The iterative maximum ratio combining (i-MRC) algorithm can be used to find antenna number of the active antenna to transmit bits at a time [3]. The channel vector $H(k)$ is considered between transmitting antenna and receiving antenna.

Transmitting antenna number and points in constellation complex diagrams are used in SM to transmit information bits [4]. So the information bits include transmitted symbol that is chosen from complex signal constellation diagram and actual location of active antenna chosen from antenna array shown in Fig. 2 with spatial constellation line 00, 01, 10, 11. A simple example is shown in figure 2 [7], a linear antenna array with 4 number of antennas and quadrature phase shift keying (QPSK). SM technique is reduced to space shift keying (SSK) when information bits carry only transmitting antenna index [3]. SM will perform coding and decoding process while transmitting information bits that carry both antenna index and symbol from digital modulation. Simulated SM technique with $N_t = 4$, $N_r = 4$ and $m = 2$ is discussed in next section.

3 Simulation Result

Transmitting antenna and receiving antenna are $N_t = 4$ and $N_r = 4$. Change in SM compared to SSK is $m = 2$, which is BPSK modulation (coding) that is two-symbol constellation. Bits per channel use (bpcu) is $\log_2 N_t + \log_2 M$. Transmit bits are 3 bits that are antenna bits along with message bit, where $\log_2 N_t$ being the number of bits identifying transmitting antenna from array and $\log_2 M$ is a symbol in BPSK. Each block is processed in SM mapper and divided into two sub-blocks: $\log_2 N_t$, which selects active antenna while keeping the other antenna silent, and $\log_2(m)$, which chooses symbol in signal constellation diagram.

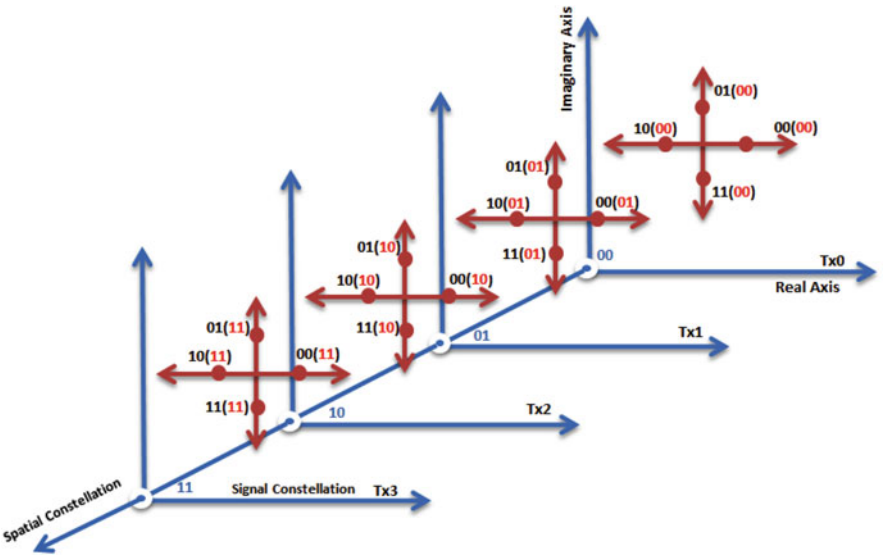


Fig. 2 Spatial modulation: Tri-dimensional constellation diagram

Rayleigh fading channel matrix $H(k)$ and Gaussian noise are calculated to depict the transmission of generated data over wireless medium. The information bits -1 to 0 and $+1$ to 1 are formulated, which represent symbols as coded in BPSK. Multiple antennas at the receiver are exploited, under the assumption of ML optimum detection, which is to attain receiver diversity gains through MRC.

Simulation results of SM analysed in this chapter are shown in Fig. 3, where a flat Rayleigh fading channel is considered with AWGN and receiver is having knowledge of channel. The bit error ratio (BER) for 4*4 BPSK SM is plotted.

4 Conclusion

This chapter has reviewed and analysed spatial modulation and its recent research achievements. One RF chain usage in SM is effectively reducing hardware complexity and its cost. SM has been known as useful physical layer transmission technique by the combination of digital modulation and multiple antenna transmission in MIMO wireless communication system. From the technique, it is clearly studied that antenna number and symbol are conveying information bits, which can be a hopeful method for low complexity MIMO implementations. SM technique avoids inter-channel interference at the receiver input and produces no correlation among transmitting antennas and also there is no requirement of synchronization between antenna.

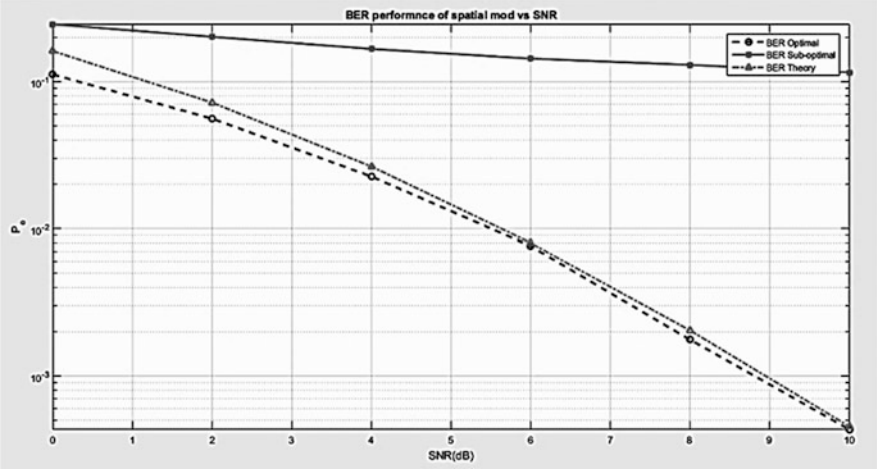


Fig. 3 Bit error performance (BER) of SM vs. signal-to-noise ratio (SNR)

References

1. A. Alsanie, K.M. Humadi, A.I. Sulyman, Spatial modulation concept for massive multiuser MIMO systems. Int. J. Antennas Propag. **2014**, 9 (2014)
2. M.V. Kumar, V. Kumar, Relative investigation of methods to generate millimeter wave in radio-over-fiber communication. Micro-Electro. Telecommun. Eng. Lecture Notes Netw. Syst. **106**, 567–574 (2020)
3. Y. Li, I.A. Hemadeh, M. El-Hajjar, L. Hanzo, Radio over fiber downlink design for spatial modulation and multi-set space-time shift-keying. IEEE Access **6**, 21812–21827 (2018)
4. Y. Li, Q. Yang, I.A. Hemadeh, M. El-Hajjar, C.-K. Chan, L. Hanzo, Experimental characterization of the radio over fiber aided twin-antenna spatial modulation downlink. Opt. Express **26**(10), 12432–12440 (2018)
5. R. Mesleh, H. Haas, C.W. Ahn, S. Yun, Spatial modulation - a new low complexity spectral efficiency enhancing technique, in *2006 First International Conference on Communications and Networking in China* (2006), pp. 1–5
6. R.Y. Mesleh, H. Haas, S. Sinanovic, C.W. Ahn, S. Yun, Spatial modulation. IEEE Trans. Vehicular Technol. **57**(4), 2228–2241 (2008)
7. M.D. Renzo, H. Haas, P.M. Grant, Spatial modulation for multiple-antenna wireless systems: a survey. IEEE Commun. Mag. **49**(12), 182–191 (2011)
8. J. Wang, L. He, J. Song, An overview of spatial modulation techniques for millimeter wave MIMO systems, in *2017 IVth International Conference on Engineering and Telecommunication (EnT)* (2017), pp. 51–56
9. M. Wen, B. Zheng, K.J. Kim, M. Di Renzo, T.A. Tsiftsis, K. Chen, N. Al-Dhahir, A survey on spatial modulation in emerging wireless systems: Research progresses and applications. IEEE J. Sel. Areas Commun. **37**(9), 1949–1972 (2019)

Controllability of Fractional Stochastic Delayed System with Nonlocal Conditions

Surendra Kumar

Abstract This chapter concerns with approximate controllability for a class of fractional stochastic control systems with nonlocal conditions and fixed delay. The existence of a solution is shown via the contraction mapping principle by assuming Lipschitz continuity of nonlinear terms. A set of sufficient conditions is also constructed which ensure that the fractional stochastic control system is approximately controllable. The main results are verified through an example.

Keywords Fractional calculus · Stochastic analysis · Mild solution · Approximate controllability · Fixed point theory

Mathematics Subject Classification (2000) 93B05, 93E, 60G

1 Introduction

Fractional calculus is about to differentiation and integration of non-integer order. The potential applications of fractional calculus are in diffusion process, electrical science, electrochemistry, viscoelasticity, control science, electromagnetic theory, and many more [1–8]. In real-life problems, such as population dynamics, finance, physical systems subject to thermal fluctuations involve some randomness. Therefore, it seems reasonable to modify deterministic systems to stochastic ones. Moreover, it is observed that the nonlocal initial conditions, introduced by Byszewski [9, 10], provide better effect in applications than the classical ones.

Controllability is a qualitative property of a dynamical control system and is of particular importance in both deterministic and stochastic control theories. In infinite-dimensional spaces, exact controllability of fractional semilinear deterministic and stochastic systems has been investigated by many researchers [11–16]

S. Kumar (✉)
Faculty of Mathematical Sciences, Department of Mathematics, University of Delhi, Delhi, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_10

and the references therein. However, the concept of exact controllability is usually too strong in infinite-dimensional setting. Therefore, it seems necessary to study a weaker concept, namely approximate controllability for semilinear control systems. For integer order, several researchers have studied approximate controllability for various types of deterministic and stochastic systems with nonlocal conditions [17–24]. The issue of approximate controllability for deterministic fractional control systems has been raised by several authors (see [25–33]).

On the other hand, some researchers also paid their attention towards the theory of fractional stochastic systems and developed some interesting results. Utilizing the Krasnoselskii fixed point theorem and stochastic analysis, Sakthivel et al. [34] proved approximate controllability of neutral stochastic fractional integro-differential system with infinite delay. In [35], Sakthivel et al. established sufficient conditions for approximate controllability of fractional stochastic differential equations. Using the Sadovskii fixed point theorem, Muthukumar and Rajivganthi [36] discussed approximate controllability for fractional neutral stochastic integro-differential system with nonlocal conditions and infinite delay. Kerboua et al. [37] obtained some sufficient conditions for approximate controllability of fractional stochastic control systems. Using fixed point theorem for multivalued operators, approximate controllability for fractional stochastic differential inclusions with nonlocal conditions and delay has been studied in [38]. Balasubramaniam et al. [39] used Bohnenblust–Karlin's fixed point theorem and discussed some sufficient conditions for approximate controllability of fractional neutral stochastic integro-differential inclusions with infinite delay. Boudaoui et al. [40] obtained approximate controllability for fractional impulsive stochastic system with nonlocal conditions and infinite delay. Shukla et al. [41] studied the concept of approximate controllability of fractional stochastic differential system under simple sufficient conditions. Chadha et al. [42] studied the approximate controllability of an impulsive fractional neutral stochastic system with nonlocal conditions in a Hilbert space.

Motivated by the above cited work, the main objective of this article is to investigate the approximate controllability of the following nonlocal fractional stochastic delayed differential equation:

$$ {}^{C}D_t^{\alpha}\xi(t) = \mathcal{A}\xi(t) + \mathcal{B}v(t) + f(t, \xi(t-\gamma)) + g(t, \xi(t-\gamma))\frac{d\omega(t)}{dt},\ t \in (0, \tau]; \tag{1} $$

$$ \xi(0) = \xi_0 + h(\xi);\ \xi(t) = \varphi(t),\ \text{for } t \in [-\gamma, 0), \tag{2} $$

where ${}^{C}D_t^{\alpha}$ is the Caputo fractional derivative operator of order $\alpha \in (0, 1)$, $\mathcal{A}$ is a sectorial operator densely defined on the separable Hilbert space $\mathcal{X}$, the state $\xi(\cdot)$ is $\mathcal{X}$-valued stochastic processes, the control function $v(\cdot)$ takes values in $L_2([0, \tau], \mathcal{F}, \mathcal{V})$, $\mathcal{B}$ is a bounded linear operator from $\mathcal{V}$ to $\mathcal{X}$, $f : [0, \tau] \times \mathcal{X} \to \mathcal{X}$ and $g : [0, \tau] \times \mathcal{X} \to L_Q(\mathcal{K}, \mathcal{X})$ are appropriate functions to be defined latter, and $h : C([0, \tau], \mathcal{X}) \to \mathcal{X}$ is a continuous function. The random variable

$\xi_0 \in \mathcal{X}$ satisfies $\mathbb{E}\|\xi_0\|^2 < \infty$; the initial data $\varphi = \{\varphi(t) : t \in [-\gamma, 0)\}$ is an $\mathcal{F}_0$-measurable, $\mathcal{X}$-valued random variable independent of ω with finite second moments.

The chapter is organized as follows: in Sect. 2, we present basic definitions and results as preliminaries. In Sect. 3, first, the existence and uniqueness of mild solution are discussed, and thereafter approximate controllability is studied. In Sect. 4, an example is given to illustrate the developed theory.

2 Preliminaries

This section contains basic definitions and preliminary results, which help us to develop further results. Throughout this chapter, we use the following notations: let $\mathcal{X}$, $\mathcal{V}$, and $\mathcal{K}$ be the separable Hilbert spaces. For convenience, we denote the inner products and norms in all spaces by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a complete probability space furnished with complete family of right-continuous increasing sub-σ-algebras $\{\mathcal{F}_t : 0 \le t \le \tau\}$ satisfying $\mathcal{F}_t \subset \mathcal{F}$. Let $\{e_n\}_{n=1}^{\infty} \subset \mathcal{K}$ be a complete orthonormal set and $\{\beta_n\}_{n=1}^{\infty}$ a sequence of independent Brownian motions such that

$$\omega(t) := \sum_{n=1}^{\infty} \sqrt{\lambda_n} e_n \beta_n(t), \ t \in [0, \tau],$$

where the sequence $\{\lambda_n \ge 0 : n \in \mathbb{N}\}$ is bounded, and $Qe_n = \lambda_n e_n$, $n \in \mathbb{N}$ with trace $tr(Q) = \sum_{n=1}^{\infty} \lambda_n < \infty$. The $\mathcal{K}$-valued stochastic process $\omega(\cdot)$ is called the Winner process. The normal filtration $\mathcal{F}_t$ is the σ-algebra generated by $\{\omega(s) : 0 \le s \le t\}$ and $\mathcal{F}_\tau = \mathcal{F}$.

Denoted by $\mathcal{L}(\mathcal{K}, \mathcal{X})$ the space of all bounded continuous operators from $\mathcal{K}$ to $\mathcal{X}$ equipped with the usual operator norm. For $\psi \in \mathcal{L}(\mathcal{K}, \mathcal{X})$, define

$$\|\psi\|_Q^2 = tr(\psi Q \psi^*) = \sum_{n=1}^{\infty} \|\sqrt{\lambda_n} \psi e_n\|^2.$$

If $\|\psi\|_Q^2 < \infty$, then ψ is called a Q-Hilbert–Schmidt operator. Let $L_Q(\mathcal{K}, \mathcal{X})$ be the space of all Q-Hilbert–Schmidt operators $\psi : \mathcal{K} \to \mathcal{X}$. The completion $L_Q(\mathcal{K}, \mathcal{X})$ of $\mathcal{L}(\mathcal{K}, \mathcal{X})$ with respect to the topology induced by the norm $\| \cdot \|_Q$ is a Hilbert space.

The space of strongly measurable, $\mathcal{X}$-valued, square integrable random variables, denoted by $L_2(\Omega, \mathcal{X})$, is a Banach space equipped with the norm topology $\|\xi(\cdot)\| = (\mathbb{E}\|\xi(\cdot, w)\|^2)^{1/2}$, where $w \in \Omega$ and the expectation $\mathbb{E}(\cdot)$ is defined by $\mathbb{E}(\ell) = \int_\Omega \ell(w) d\mathbf{P}$. Let $C([-\gamma, \tau], L_2(\Omega, \mathcal{X}))$ be the Banach space of continuous maps from $[-\gamma, \tau]$ to $L_2(\Omega, \mathcal{X})$ satisfying $\sup_{-\gamma \le t \le \tau} \mathbb{E}\|\xi(t)\|^2 < \infty$. Let $\mathcal{X}_2$ be the

closed subspace of $C([-\gamma,\tau],L_2(\Omega,\mathcal{X}))$ consisting of measurable, $\mathcal{F}_t$-adapted, $\mathcal{X}$-valued processes $x\in C([0,\tau],L_2(\Omega,\mathcal{X}))$ equipped with the norm

$$\|\xi\|_{\mathcal{X}_2}:=\left(\sup_{0\le t\le\tau}\mathbb{E}\|\xi(t)\|^2\right)^{1/2}.$$

Let us recall the following well-known definitions. For more details on fractional calculus, one can see [1, 4].

Definition 1 The Riemann–Liouville fractional integral operator of order $\alpha>0$ of a function $f:[0,\infty)\to\mathbb{R}$ with the lower limit 0 is defined as

$$I^{\alpha}f(t)=\frac{1}{\Gamma(\alpha)}\int_0^t(t-s)^{\alpha-1}f(s)ds,$$

where Γ is the Euler gamma function.

Definition 2 The Caputo fractional derivative of order $\alpha>0$ for the function $f\in C^m([0,\tau],\mathbb{R})$ is defined by

$${}^{C}D_t^{\alpha}f(t)=\frac{1}{\Gamma(m-\alpha)}\int_0^t(t-s)^{m-\alpha-1}f^{(m)}(s)ds,\quad m-1\le\alpha<m\in\mathbb{N}.$$

If f is $\mathcal{X}$-valued, the integrals in Definitions 1 and 2 are taken in Bochner's sense.

Definition 3 ([38, 43]) A closed linear operator $\mathcal{A}$ is called sectorial of type $\mu\in\mathbb{R}$ if there are $\pi/2\le\theta\le\pi$ and $\tilde{M}>0$ such that the following holds: $\rho(\mathcal{A})\subset\sum_{(\theta,\mu)}:=\{\lambda\in\mathbb{C}:\lambda\ne\mu,\ |\arg(\lambda-\mu)|<\theta\}$, and $\|R(\lambda,\mathcal{A})\|:=\|(\lambda-\mathcal{A})^{-1}\|\le\frac{\tilde{M}}{|\lambda-\mu|},\ \lambda\in\sum_{(\theta,\mu)}$.

Lemma 1 ([43]) *For $0<\alpha<2$, a linear, closed, and densely defined operator $\mathcal{A}$ is in $\mathcal{A}^{\alpha}(\theta_0,\mu_0)$ if and only if $\lambda^{\alpha}\in\rho(\mathcal{A})$ for each $\lambda\in\Sigma_{(\theta_0+\pi/2,\mu)}$ and for $\mu>\mu_0$, $\theta<\theta_0$ there is a constant k_0 depending on θ and μ such that*

$$\|\lambda^{\alpha-1}R(\lambda^{\alpha},\mathcal{A})\|\le\frac{k_0}{|\lambda-\mu|}.$$

Lemma 2 ([43]) *If f satisfies the uniform Hölder condition with the exponent $0<\delta\le1$ and $\mathcal{A}$ is a sectorial operator, then*

$$\left.\begin{aligned}&{}^{C}D_t^{\alpha}\xi(t)=A\xi(t)+f(t),\ \ 0<\alpha<1,\quad t\in(0,\tau],\\&\xi(0)=\xi_0,\end{aligned}\right\}\tag{3}$$

has a unique solution that is given by

$$\xi(t)=S_{\alpha}(t)\xi_0+\int_0^t T_{\alpha}(t-s)f(s)ds,$$

where

$$S_\alpha(t) = E_{\alpha,1}(At^\alpha) = \frac{1}{2\pi i}\int_{\hat{B}_\varrho} e^{\lambda t}\frac{\lambda^{\alpha-1}}{\lambda^\alpha - \mathcal{A}}d\lambda,$$

$$T_\alpha(t) = t^{\alpha-1}E_{\alpha,\alpha}(At^\alpha) = \frac{1}{2\pi i}\int_{\hat{B}_\varrho} e^{\lambda t}\frac{1}{\lambda^\alpha - \mathcal{A}}d\lambda,$$

where $\hat{B}_\varrho$ denotes the Bromwich path.

The operator $\mathcal{A}$ belongs to $\mathfrak{C}^\alpha(\tilde{M},\mu)$ if problem (3) with $f = 0$ has a solution operator $S_\alpha(t)$ satisfying $\|S_\alpha(t)\| \le \tilde{M}e^{\mu t}$. Denote $\mathfrak{C}^\alpha(\mu) := \cup\{\mathfrak{C}^\alpha(\tilde{M},\mu) : \tilde{M} \ge 1\}$, $\mathfrak{C}^\alpha := \{\mathfrak{C}^\alpha(\mu) : \mu \ge 0\}$, and $\mathcal{A}^\alpha(\theta_0,\mu_0) = \{\mathcal{A} \in \mathfrak{C}^\alpha : \mathcal{A}$ generates analytic solution operators $S_\alpha(t)$ of type $(\theta_0,\mu_0)\}$.

For $0 < \alpha < 1$ and $\mathcal{A} \in \mathcal{A}^\alpha(\theta_0,\mu_0)$, $\|S_\alpha(t)\| \le \tilde{M}e^{\mu t}$ and $\|T_\alpha(t)\| \le Ce^{\mu t}(1 + t^{\alpha-1})$, $t > 0$, $\mu > \mu_0$. Set

$$M_S := \sup_{0\le t\le\tau}\|S_\alpha(t)\|, \quad M_T := \sup_{0\le t\le\tau} Ce^{\mu t}\left(1 + t^{1-\alpha}\right).$$

Then, $\|S_\alpha(t)\| \le M_S$, $\|T_\alpha(t)\| \le t^{\alpha-1}M_T$.

By virtue of Lemma 2, we define the solution of system (1)–(2) as follows:

Definition 4 A continuous $\mathcal{F}_t$-adapted stochastic process $x : [-\gamma,\tau] \to \mathcal{X}$ is called a mild solution of system (1)–(2) if for every $v(\cdot) \in L_2([0,\tau],\mathcal{F},\mathcal{V})$, $\xi(t)$ is measurable and satisfying

$$\xi(t) = \begin{cases} S_\alpha(t)[\xi_0 + h(\xi)] + \int_0^t T_\alpha(t-s)[\mathcal{B}v(s) + f(s,\xi(s-\gamma))]ds \\ + \int_0^t T_\alpha(t-s)g(s,\xi(s-\gamma))d\omega(s), & t \in [0,\tau]; \\ \varphi(t), & t \in [-\gamma,0). \end{cases}$$

Let $\xi(t,\xi_0,v)$ be the state value of system (1)–(2) at time t corresponding to the control $v(\cdot) \in L_2([0,\tau],\mathcal{F},\mathcal{V})$. Then,

$$\mathfrak{R}(\tau,\xi_0,v) = \{\xi(\tau,\xi_0,v) : v \in L_2([0,\tau],\mathcal{F},\mathcal{V})\}$$

is called the reachable set of system (1)–(2) at terminal time τ, and its closure in $L_2(\Omega,\mathcal{X})$ is denoted by $\overline{\mathfrak{R}(\tau,\xi_0,v)}$.

Definition 5 The system (1)–(2) is said to be approximately controllable on $[0,\tau]$ if and only if $\overline{\mathfrak{R}(\tau,\xi_0,v)} = L_2(\Omega,\mathcal{X})$.

To discuss approximate controllability of system (1)–(2), introduce the operator $\mathcal{L}_\tau : L_2([0,\tau],\mathcal{F},\mathcal{V}) \to L_2(\Omega,\mathcal{X})$ by

$$\mathcal{L}_\tau v := \int_0^\tau T_\alpha(\tau - s)\mathcal{B}v(s)ds,$$

and the adjoin operator $\mathcal{L}_\tau^* : L_2(\Omega, \mathcal{X}) \to L_2([0, \tau], \mathcal{F}, \mathcal{V})$ is given by

$$\mathcal{L}_\tau^* z = \mathcal{B}^* T_\alpha^*(\tau - s)\mathbb{E}\{z|\mathcal{F}_t\},$$

where $\mathcal{B}^*$ and T_α^* denote the adjoint operators of $\mathcal{B}$ and T_α, respectively.

Define the controllability operator $\Pi_0^\tau : L_2(\Omega, \mathcal{X}) \to L_2(\Omega, \mathcal{X})$ associated with the linear part of (1)–(2) by

$$\Pi_0^\tau\{\cdot\} := \mathcal{L}_\tau \mathcal{L}_\tau^*\{\cdot\} = \int_0^\tau T_\alpha(\tau - t)\mathcal{B}\mathcal{B}^* T_\alpha^*(\tau - t)\mathbb{E}\{\cdot|\mathcal{F}_t\}dt,$$

and the controllability operator associated with the linear part of fractional deterministic system

$$\left.\begin{array}{l} {}^C D_t^\alpha \xi(t) = \mathcal{A}\xi(t) + \mathcal{B}v(t), \ \ t \in (0, \tau]; \\ \xi(0) = \xi_0, \end{array}\right\} \tag{4}$$

is given by

$$\Psi_t^\tau := \int_t^\tau T_\alpha(\tau - s)\mathcal{B}\mathcal{B}^* T_\alpha^*(\tau - s)ds.$$

It is easy to see that system (4) is approximately controllable on $[0, \tau]$ if and only if $\beta(\beta I + \Psi_0^\tau)^{-1} \to 0$ strongly as $\beta \to 0^+$ [44]. For more details on approximate controllability of linear fractional deterministic control system, one can also see [38, 45]. Moreover, for each $0 \le t \le \tau$, Ψ_t^τ is a bounded linear operator and $\|(\beta I + \Psi_0^\tau)^{-1}\| \le \frac{1}{\beta}$.

3 Main Results

In this section, we formulate and prove a set of sufficient conditions for approximate controllability of system (1)–(2). For this purpose, we first examine the existence of mild solution of system (1)–(2) by using Banach's fixed point theorem. In particular, we convert the controllability issue into a fixed point problem. Next, we show that under some natural assumptions, the system (1)–(2) is approximately controllable. To obtain desired results, we need the following hypotheses:

(H1) The function h satisfies linear growth and Lipschitz conditions. That is, there are positive constants l_1 and l_2 such that

$$\mathbb{E}\|h(\xi) - h(\zeta)\|^2 \le l_1 \mathbb{E}\|\xi - \zeta\|_{\mathcal{X}_2}^2, \quad \mathbb{E}\|h(\xi)\|^2 \le l_2\left(1 + \mathbb{E}\|\xi\|_{\mathcal{X}_2}^2\right).$$

(H2) (i) The functions $f(\cdot) : \mathcal{X} \to \mathcal{X}$ and $g(\cdot) : \mathcal{X} \to L_Q(\mathcal{K}, \mathcal{X})$ are continuous for almost all $0 \le t \le \tau$, and $f(\cdot, \xi) : [0, \tau] \to \mathcal{X}$ and $g(\cdot, \xi) : [0, \tau] \to L_Q(\mathcal{K}, \mathcal{X})$ are strongly measurable for each $\xi \in \mathcal{X}$.

(ii) There are positive constants l_3, l_4, l_5, and l_6 such that

$$\mathbb{E}\|f(t,\xi) - f(t,\zeta)\|^2 \le l_3\mathbb{E}\|\xi - \zeta\|^2, \ \ \mathbb{E}\|f(t,\xi)\|^2 \le l_4\left(1 + \mathbb{E}\|\xi\|^2\right),$$

$$\mathbb{E}\|g(t,\xi) - g(t,\zeta)\|_Q^2 \le l_5\mathbb{E}\|\xi - \zeta\|^2, \ \ \mathbb{E}\|g(t,\xi)\|_Q^2 \le l_6\left(1 + \mathbb{E}\|\xi\|^2\right).$$

(H3) For $t > 0$, $S_\alpha(t)$ and $T_\alpha(t)$ are compact.

To define the control function, we need the following result. For more details, one can see [46, 47].

Lemma 3 *For any $\xi_\tau \in L_2(\Omega, \mathcal{X})$, there exists $\tilde{\phi} \in L_2(\Omega, \mathcal{F}, L_2([0,\tau], \mathcal{L}(\mathcal{K}, \mathcal{X})))$ such that*

$$\xi_\tau = \mathbb{E}\xi_\tau + \int_0^\tau \tilde{\phi}(s)d\omega(s).$$

Thus, for any $\beta > 0$ and $\xi_\tau \in L_2(\Omega, \mathcal{X})$, define the control function

$$\begin{aligned} v^\beta(t,\xi) := {} & \mathcal{B}^*T_\alpha^*(\tau - t)\Big[\left(\beta I + \Psi_0^\tau\right)^{-1}[\mathbb{E}\tilde{x}_\tau - S_\alpha(\tau)(\xi_0 + h(\xi))] \\ & + \int_0^\tau \left(\beta I + \Psi_s^\tau\right)^{-1}\tilde{\phi}(s)d\omega(s)\Big] \\ & - \mathcal{B}^*T_\alpha^*(\tau - t)\int_0^\tau \left(\beta I + \Psi_s^\tau\right)^{-1}T_\alpha(\tau - s)f(s, \xi(s-\gamma))ds \\ & - \mathcal{B}^*T_\alpha^*(\tau - t)\int_0^\tau \left(\beta I + \Psi_s^\tau\right)^{-1}T_\alpha(\tau - s)g(s, \xi(s-\gamma))d\omega(s). \end{aligned}$$

Lemma 4 *There is a constant $\hat{M} > 0$ such that for all $\xi, \zeta \in \mathcal{X}_2$, the following hold:*

$$\mathbb{E}\|v^\beta(t,\xi) - v^\beta(t,\zeta)\|^2 \le \frac{\hat{M}}{\beta^2}\|\xi - \zeta\|_{\mathcal{X}_2}^2,$$

$$\mathbb{E}\|v^\beta(t,\xi)\|^2 \le \frac{\hat{M}}{\beta^2}\left(1 + \|\xi\|_{\mathcal{X}_2}^2\right).$$

Proof We prove the first inequality only, since the second inequality can be established in a similar way. Let $\xi, \zeta \in \mathcal{X}_2$ be arbitrary. Then, the Cauchy–Schwartz inequality and assumptions (H1) and (H2) yield that

$$
\begin{aligned}
&\mathbb{E}\|v^{\beta}(t,\xi)-v^{\beta}(t,\zeta)\|^2\\
&\quad\le 3\mathbb{E}\left\|\mathcal{B}^*T_{\alpha}^*(\tau-t)(\beta I+\Psi_0^{\tau})^{-1}S_{\alpha}(\tau)[h(\xi)-h(\zeta)]\right\|^2\\
&\qquad+3\mathbb{E}\left\|\mathcal{B}^*T_{\alpha}^*(\tau-t)\int_0^{\tau}(\beta I+\Psi_s^{\tau})^{-1}T_{\alpha}(\tau-s)\right.\\
&\qquad\left.\times[f(s,\xi(s-\gamma))-f(s,\zeta(s-\gamma))]ds\right\|^2\\
&\qquad\times+3\mathbb{E}\left\|\mathcal{B}^*T_{\alpha}^*(\tau-t)\int_0^{\tau}(\beta I+\Psi_s^{\tau})^{-1}T_{\alpha}(\tau-s)\right.\\
&\qquad\left.\times\ [g(s,\xi(s-\gamma))-g(s,\zeta(s-\gamma))]d\omega(s)\right\|^2.\\
&\quad\le\frac{3}{\beta^2}\|\mathcal{B}\|^2\tau^{2(\alpha-1)}M_T^2\Big[M_S^2l_1\|\xi-\zeta\|_{\mathcal{X}_2}^2\\
&\qquad+\frac{M_T^2\tau^{\alpha}}{\alpha}[l_3+tr(Q)l_5]\int_0^{\tau}(\tau-s)^{\alpha-1}\mathbb{E}\|\xi(s-\gamma)-\zeta(s-\gamma)\|^2ds\\
&\quad\le\frac{3}{\beta^2}\|\mathcal{B}\|^2\tau^{2(\alpha-1)}M_T^2\Big[M_S^2l_1+\frac{M_T^2\tau^{\alpha}(\tau-\gamma)^{\alpha}}{\alpha^2}[l_3+tr(Q)l_5]\Big]\|\xi-\zeta\|_{\mathcal{X}_2}^2\\
&\quad\le\frac{3}{\alpha^2\beta^2}\|\mathcal{B}\|^2\tau^{2(\alpha-1)}M_T^2\Big[M_S^2\alpha^2l_1+M_T^2\tau^{\alpha}(\tau-\gamma)^{\alpha}[l_3+tr(Q)l_5]\Big]\|\xi-\zeta\|_{\mathcal{X}_2}^2.
\end{aligned}
$$

Now, compute

$$
\begin{aligned}
&\int_0^{\tau}(\tau-s)^{\alpha-1}\mathbb{E}\|\xi(s-\gamma)-\zeta(s-\gamma)\|^2ds\\
&=\int_{-\gamma}^{\tau-\gamma}(\tau-\gamma-\sigma)^{\alpha-1}\mathbb{E}\|\xi(\sigma)-\zeta(\sigma)\|^2d\sigma\\
&=\int_{-\gamma}^{0}(\tau-\gamma-\sigma)^{\alpha-1}\mathbb{E}\|\xi(\sigma)-\zeta(\sigma)\|^2d\sigma\\
&\quad+\int_0^{\tau-\gamma}(\tau-\gamma-\sigma)^{\alpha-1}\mathbb{E}\|\xi(\sigma)-\zeta(\sigma)\|^2d\sigma.
\end{aligned}
$$

But, for $-\gamma\le t<0$, $\xi(t)=\zeta(t)=\varphi(t)$, and hence

$$
\begin{aligned}
&\int_0^{\tau}(\tau-s)^{\alpha-1}\mathbb{E}\|\xi(s-\gamma)-\zeta(s-\gamma)\|^2ds\\
&\quad=\int_0^{\tau-\gamma}(\tau-\gamma-\sigma)^{\alpha-1}\mathbb{E}\|\xi(\sigma)-\zeta(\sigma)\|^2d\sigma
\end{aligned}
$$

$$\leq \left(\int_0^{\tau-\gamma} (\tau - \gamma - \sigma)^{\alpha-1} d\sigma \right) \|\xi - \zeta\|^2_{\mathcal{X}_2}$$
$$= \frac{(\tau-\gamma)^\alpha}{\alpha} \|\xi - \zeta\|^2_{\mathcal{X}_2}.$$

Therefore, we have

$$\mathbb{E}\|v^\beta(t,\xi) - v^\beta(t,\zeta)\|^2 \leq \frac{\hat{M}}{\beta^2} \|\xi - \zeta\|^2_{\mathcal{X}_2},$$

where $\hat{M}$ is a suitable constant. This completes the proof. □

Theorem 1 *Suppose assumptions (H1)–(H3) hold. Then, the system* (1)–(2) *has a mild solution on* $[-\gamma, \tau]$ *provided*

$$4\left[M_S^2 l_1 + \frac{M_T^2 \tau^\alpha}{\alpha^2} \left(\frac{\hat{M} \|\mathcal{B}\|^2 \tau^\alpha}{\beta^2} + [l_3 + tr(Q) l_5](\tau - \gamma)^\alpha \right) \right] < 1.$$

Proof For $\beta > 0$, consider a map $\Phi_\beta : \mathcal{X}_2 \to \mathcal{X}_2$ defined by

$$(\Phi_\beta \xi)(t)$$
$$:= \begin{cases} S_\alpha(t)[\xi_0 + h(\xi)] + \int_0^t T_\alpha(t-s)[\mathcal{B}v^\beta(s,\xi) + f(s,\xi(s-\gamma))]ds \\ + \int_0^t T_\alpha(t-s) g(s,\xi(s-\gamma)) d\omega(s), & t \in [0,\tau]; \\ \varphi(t), & t \in [-\gamma, 0). \end{cases}$$

Now, we show that the operator Φ_β has a fixed point in $\mathcal{X}_2$. The proof is divided into three steps. □

Step 1 For any $\xi \in \mathcal{X}_2$, $(\Phi_\beta \xi)(t)$ is continuous on $[-\gamma, \tau]$. If $t \in [-\gamma, 0)$, then $(\Phi_\beta \xi)(t) = \varphi(t)$ is clearly continuous. So, let $0 \leq t_1 < t_2 \leq \tau$, then

$$\mathbb{E}\|(\Phi_\beta \xi)(t_2) - (\Phi_\beta \xi)(t_1)\|^2$$
$$\leq 7\mathbb{E}\,\|[S_\alpha(t_2) - S_\alpha(t_1)]h(\xi)\|^2$$
$$+ 7\mathbb{E}\left\| \int_0^{t_1} [T_\alpha(t_2 - s) - T_\alpha(t_1 - s)]\mathcal{B}v^\beta(s,\xi) ds \right\|^2$$
$$+ 7\mathbb{E}\left\| \int_{t_1}^{t_2} T_\alpha(t_2 - s)\mathcal{B}v^\beta(s,\xi) ds \right\|^2$$
$$+ 7\mathbb{E}\left\| \int_0^{t_1} [T_\alpha(t_2 - s) - T_\alpha(t_1 - s)] f(s,\xi(s-\gamma)) ds \right\|^2$$

$$
\begin{aligned}
&+ 7\mathbb{E}\left\|\int_{t_1}^{t_2} T_\alpha(t_2-s) f(s,\xi(s-\gamma))ds\right\|^2 \\
&+ 7\mathbb{E}\left\|\int_0^{t_1} [T_\alpha(t_2-s)-T_\alpha(t_1-s)] g(s,\xi(s-\gamma))d\omega(s)\right\|^2 \\
&+ 7\mathbb{E}\left\|\int_{t_1}^{t_2} T_\alpha(t_2-s) g(s,\xi(s-\gamma))d\omega(s)\right\|^2 \\
\leq\ & 7\mathbb{E}\,\|[S_\alpha(t_2)-S_\alpha(t_1)]h(\xi)\|^2 \\
&+ 7t_1\|\mathcal{B}\|^2 \int_0^{t_1} \mathbb{E}\left\|[T_\alpha(t_2-s)-T_\alpha(t_1-s)]v^\beta(s,\xi)\right\|^2 ds \\
&+ 7M_T^2\|\mathcal{B}\|^2 \frac{(t_2-t_1)^\alpha}{\alpha} \int_{t_1}^{t_2} (t_2-s)^{\alpha-1}\mathbb{E}\left\|v^\beta(s,\xi)\right\|^2 ds \\
&+ 7t_1 \int_0^{t_1} \mathbb{E}\,\|[T_\alpha(t_2-s)-T_\alpha(t_1-s)] f(s,\xi(s-\gamma))\|^2 ds \\
&+ 7M_T^2 \frac{(t_2-t_1)^\alpha}{\alpha} \int_{t_1}^{t_2} (t_2-s)^{\alpha-1}\mathbb{E}\,\|f(s,\xi(s-\gamma))\|^2 ds \\
&+ 7t_1 tr(Q) \int_0^{t_1} \mathbb{E}\,\|[T_\alpha(t_2-s)-T_\alpha(t_1-s)] g(s,\xi(s-\gamma))\|^2 ds \\
&+ 7tr(Q)M_T^2 \frac{(t_2-t_1)^\alpha}{\alpha} \int_{t_1}^{t_2} (t_2-s)^{\alpha-1}\mathbb{E}\,\|T_\alpha(t_2-s) g(s,\xi(s-\gamma))\|^2 ds.
\end{aligned}
$$

The operators $S_\alpha(t)$ and $T_\alpha(t)$ are continuous in the uniform operator topology due to (H3). Thus, using Lebesgue's dominated convergence theorem and the continuity of $S_\alpha(t)$ and $T_\alpha(t)$ in the uniform operator topology, we conclude that the right-hand side of the above inequality tends to zero as $t_2 - t_1 \to 0$. Therefore, the operator $(\Phi_\beta\xi)(t)$ is continuous from the right in $[0,\tau)$. A similar argument gives the continuity of the operator $(\Phi_\beta\xi)(t)$ from the left in $(0,\tau]$. Hence, $(\Phi_\beta\xi)(t)$ is continuous on $[-\gamma,\tau]$.

Step 2 We now show that Φ_β maps $\mathcal{X}_2$ into itself. For $t \in [-\gamma, 0)$, the proof is trivial. So, let $t \in [0,\tau]$, and using Lemma 4 and assumptions (H1) and (H2), we get

$$
\begin{aligned}
&\mathbb{E}\|(\Phi_\beta\xi)(t)\|^2 \\
&\quad \leq 5\|S_\alpha(t)\|^2[\mathbb{E}\|\xi_0\|^2 + \mathbb{E}\|h(\xi)\|^2] \\
&\qquad + 5\mathbb{E}\left\|\int_0^t T_\alpha(t-s)\mathcal{B}v^\beta(s,\xi)ds\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&+ 5\mathbb{E}\left\| \int_0^t T_\alpha(t-s) f(s, \xi(s-\gamma)) ds \right\|^2 \\
&+ 5\mathbb{E}\left\| \int_0^t T_\alpha(t-s) g(s, \xi(s-\gamma)) d\omega(s) \right\|^2 \\
&\leq 5M_S^2 \mathbb{E}\|\xi_0\|^2 + 5M_S^2 l_2 (1 + \|\xi\|_{\mathcal{X}_2}^2) + 5\frac{\hat{M}\|\mathcal{B}\|^2 \tau^{2\alpha}}{\alpha^2 \beta^2}\left(1 + \|\xi\|_{\mathcal{X}_2}^2\right) \\
&+ 5\frac{M_T^2 \tau^{2\alpha}}{\alpha^2}[l_4 + tr(Q) l_6]\left[1 + \|\varphi\|^2 + \|\xi\|_{\mathcal{X}_2}^2\right].
\end{aligned}
$$

This yields that $\mathbb{E}\|\Phi_\beta \xi\|_{\mathcal{X}_2}^2 < \infty$, and hence $\Phi_\beta \xi \in \mathcal{X}_2$. Thus, for each $\beta > 0$, we have $\Phi_\beta(\mathcal{X}_2) \subseteq \mathcal{X}_2$.

Step 3 Finally, we use the contraction mapping principle to show that Φ_β has a fixed point in $\mathcal{X}_2$, which is the mild solution of fractional control system (1)–(2). Let $\beta > 0$ and $\xi, \zeta \in \mathcal{X}_2$. Then, for $t \in [-\gamma, 0)$, we have $\xi(t) = \zeta(t) = \varphi(t)$. Next, let $t \in [0, \tau]$, then

$$
\begin{aligned}
&\mathbb{E}\|(\Phi_\beta \xi)(t) - (\Phi_\beta \zeta)(t)\|^2 \\
&\quad \leq \mathbb{E}\Big\| S_\alpha(t)[h(\xi) - h(\zeta)] \\
&\qquad + \int_0^t T_\alpha(t-s)\mathcal{B}[v^\beta(s, \xi) - v^\beta(s, \zeta)] ds \\
&\qquad + \int_0^t T_\alpha(t-s)[f(s, \xi(s-\gamma)) - f(s, \zeta(s-\gamma))] ds \\
&\qquad + \int_0^t T_\alpha(t-s)[g(s, \xi(s-\gamma)) - g(s, \zeta(s-\gamma))] d\omega(s) \Big\|^2 \\
&\quad \leq 4\left[M_S^2 l_1 + \frac{M_T^2 \tau^\alpha}{\alpha^2}\left(\frac{\hat{M}\|\mathcal{B}\|^2 \tau^\alpha}{\beta^2} + [l_3 + tr(Q) l_5](\tau - \gamma)^\alpha \right) \right] \|\xi - \zeta\|_{\mathcal{X}_2}^2.
\end{aligned}
$$

This implies that

$$
\|\Phi_\beta \xi - \Phi_\beta \zeta\|_{\mathcal{X}_2}^2 \leq l(\alpha, \beta)\|\xi - \zeta\|_{\mathcal{X}_2}^2,
$$

where $l(\alpha, \beta) := 4\left[M_S^2 l_1 + \frac{M_T^2 \tau^\alpha}{\alpha^2}\left(\frac{\hat{M}\|\mathcal{B}\|^2 \tau^\alpha}{\beta^2} + [l_3 + tr(Q) l_5](\tau - \gamma)^\alpha \right) \right] < 1$.

Hence, by contraction mapping principle, we conclude that Φ_β admits a unique fixed point in $\mathcal{X}_2$.

Theorem 2 *Let f and g be uniformly bounded and suppose that (H1)–(H3) hold. If the linear system (4) is approximately controllable, then system (1)–(2) is approximately controllable on $[0, \tau]$.*

Proof Let ξ_β be the fixed point of the operator Φ_β in $\mathcal{X}_2$. By the stochastic Fubini theorem, it is easy to see that

$$\begin{aligned}\xi_\beta(\tau) = \xi_\tau &- \beta(\beta I + \Psi_0^\tau)^{-1}\left[\mathbb{E}\xi_\tau - S_\alpha(\tau)(\xi_0 + h(\xi_\beta))\right] \\ &+\beta\int_0^\tau (\beta I + \Psi_s^\tau)^{-1} T_\alpha(\tau - s) f(s, \xi_\beta(s-\gamma))ds \\ &+\beta\int_0^\tau (\beta I + \Psi_s^\tau)^{-1}[T_\alpha(\tau - s) g(s, \xi_\beta(s-\gamma)) - \tilde{\phi}(s)]d\omega(s).\end{aligned}$$

□

The uniform boundedness of f and g yields that there is a subsequence denoted by $\{f(\cdot, \xi_\beta(\cdot)),\ g(\cdot, \xi_\beta(\cdot))\}$ converging to, say, $\{f(\cdot, w),\ g(\cdot, w)\}$ weakly in $L_2([0,\tau];\mathcal{X}) \times L_2(L_Q(\mathcal{K},\mathcal{X}))$. Then, in view of (H3), we have $T_\alpha(\tau - s) f(s, \xi_\beta(s-\gamma)) \to T_\alpha(\tau - s) f(s, \omega)$ and $T_\alpha(\tau - s) g(s, \xi_\beta(s-\gamma)) \to T_\alpha(\tau - s) g(s, \omega)$ in $[0,\tau] \times \Omega$. Thus, we obtain

$$\begin{aligned}&\mathbb{E}\|\xi_\beta(\tau) - \xi_\tau\|^2 \\ &\le 7\left\|\beta(\beta I + \Psi_0^\tau)^{-1}\left[\mathbb{E}\xi_\tau - S_\alpha(\tau)(\xi_0 + h(\xi_\beta)\right]\right\|^2 \\ &\quad + 7\mathbb{E}\left(\int_0^\tau \left\|\beta(\beta I + \Psi_s^\tau)^{-1} T_\alpha(\tau - s)[f(s, \xi_\beta(s-\gamma)) - f(s)]\right\| ds\right)^2 \\ &\quad + 7\mathbb{E}\left(\int_0^\tau \left\|\beta(\beta I + \Psi_s^\tau)^{-1} T_\alpha(\tau - s) f(s)\right\| ds\right)^2 \\ &\quad + 7\mathbb{E}\left(tr(Q)\int_0^\tau \left\|\beta(\beta I + \Psi_s^\tau)^{-1} T_\alpha(\tau - s)\right.\right. \\ &\quad \left.\left.\times\ [g(s, \xi_\beta(s-\gamma)) - g(s)]\right\|^2_{L_2(L_Q(\mathcal{K},\mathcal{X}))} ds\right) \\ &\quad + 7\mathbb{E}\left(tr(Q)\int_0^\tau \left\|\beta(\beta I + \Psi_s^\tau)^{-1} T_\alpha(\tau - s) g(s)\right\|^2_{L_2(L_Q(\mathcal{K},\mathcal{X}))} ds\right) \\ &\quad + 7\mathbb{E}\left(tr(Q)\int_0^\tau \left\|\beta(\beta I + \Psi_s^\tau)^{-1}\tilde{\phi}(s)\right\|^2_{L_2(L_Q(\mathcal{K},\mathcal{X}))} ds\right).\end{aligned}$$

On the other hand, the approximate controllability of linear system (4) implies that for all $0 \le s \le \tau$, $\beta(\beta I + \Psi_s^\tau)^{-1} \to 0$ strongly as $\beta \to 0^+$, and moreover

$\|\beta(\beta I + \Psi_s^\tau)^{-1}\| < 1$. Thus, by the Lebesgue dominated convergence theorem, it follows that

$$\mathbb{E}\|\xi_\beta(\tau) - \xi_\tau\|^2 \to 0 \text{ as } \beta \to 0^+.$$

This shows that the system (1)–(2) is approximately controllable.

4 Example

Example 1 Let $\mathcal{X} = L_2([0,\pi],\mathbb{R})$. Define the operator $\mathcal{A} : \mathcal{X} \to \mathcal{X}$ by $A\varsigma = \varsigma''$ with domain

$$D(\mathcal{A}) = \{\varsigma \in \mathcal{X};\ \varsigma, \varsigma' \text{ are absolutely continuous}, \varsigma'' \in \mathcal{X}, \varsigma(0) = \varsigma(\pi) = 0\}.$$

It is well known that the set $\{\varphi_n,\ n = 1, 2, \ldots\}$ forms an orthonormal basis for $\mathcal{X}$, where $\varphi_n(\varsigma) = \sqrt{(\frac{2}{\pi})}\sin(n\varsigma)$, $n \in \mathbb{N}$, are the normalized eigenfunctions corresponding to the eigenvalues $\lambda_n = -n^2$ of the operator $\mathcal{A}$. Then, the C_0-semigroup generated by $\mathcal{A}$ is defined by

$$T(t)\varsigma := \sum_{n=1}^{\infty} \exp(\lambda_n t)\langle \varsigma, \varphi_n\rangle \varphi_n,\quad \varsigma \in \mathcal{X}.$$

Thus, it follows that $\{T(t) :\ t > 0\}$ is uniformly bounded compact semigroup, so that $R(\lambda, \mathcal{A}) := (\lambda - \mathcal{A})^{-1}$ is a compact operator for $\lambda \in \rho(\mathcal{A})$ [38].

Consider the following fractional stochastic delayed control system:

$$\begin{cases} {}^{C}D_t^{3/4}\varsigma(t,z) = \frac{\partial^2}{\partial z^2}\varsigma(t,z) + \vartheta(t,z) + f(t,\varsigma(t-\gamma,z)) + g(t,\varsigma(t-\gamma,z))\frac{d\omega(t)}{dt} \\ \varsigma(0,z) = \varsigma_0(z) + \sum_{j=1}^{k}\gamma_j\varsigma(t_j,z),\ z \in [0,\pi] \\ \varsigma(t,z) = \varphi(t,z),\ t \in [-\gamma, 0) \\ \varsigma(t,0) = \varsigma(t,\pi) = 0,\ t \in [0,\tau] \end{cases} \tag{5}$$

where $0 < t_j < \tau$, $\gamma_j \in \mathbb{R}$, and $\vartheta : [0,\tau] \times [0,\pi] \to \mathbb{R}$ is continuous.

Define $\varphi(t)(z) = \varphi(t,z)$, $\varsigma(t)(z) = \varsigma(t,z)$, and $(\mathcal{B}v)(t)(z) = \vartheta(t,z)$. Set $\alpha = 3/4$ and $h(t)(z) = \sum_{j=1}^{k}\gamma_j\varsigma(t_j,z)$. Then, with the choices of $\mathcal{A}$, f, g, and h, system (5) can be written in the abstract form given by (1)–(2). Therefore, by virtue of Theorem 2, if the hypotheses (H1)–(H3) are fulfilled and the linear deterministic system associated with (5) is approximately controllable, then the system (5) is approximately controllable on $[0,\tau]$.

References

1. K.B. Oldham, J. Spanier, *The Fractional Calculus, Theory and Applications of Differentiation and Integration to Arbitrary Order* (Academic, New York, 1974)
2. W.G. Glockle, T.F. Nonnenmacher, A fractional calculus approach of self-similar protein dynamics. Biophys. J. **68**, 46–53 (1995)
3. R. Hilfer, *Applications of Fractional Calculus in Physics* (World Scientific, Singapore, 2000)
4. A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, *Theory and Applications of Fractional Differential Equations* (Elsevier, Amsterdam, 2006)
5. A.D. Fitt, A.R.H. Goodwin, W.A. Wakeham, A fractional differential equation for a MEMS viscometer used in the oil industry. J. Comput. Appl. Math. **229**, 373–381 (2009)
6. C.A. Monje, Y.Q. Chen, B.M. Vinagre, D. Xue, V. Feliu, *Fractional-Order Systems and Controls, Fundamentals and Applications* (Springer, New York, 2010)
7. K. Diethelm, *The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type* (Springer, New York, 2010)
8. M. Rahimy, Applications of fractional differential equations. Appl. Math. Sci. **4**, 2453–2461 (2010)
9. L. Byszewski, Existence and uniqueness of solutions of nonlocal problems for hyperbolic equation $u_{xt} = F(x, t, u, u_\xi)$. J. Appl. Math. Stochastic Anal. **3**, 163–168 (1990)
10. L. Byszewski, Existence and uniqueness of classical solution to Darboux problem together with nonlocal conditions. Ann. Math. Sil. **27**, 67–74 (2013)
11. Z. Tai, Controllability of fractional impulsive neutral integrodifferential systems with a nonlocal Cauchy condition in Banach spaces. Appl. Math. Lett. **24**, 2158–2161 (2011)
12. H.M. Ahmed, Controllability of fractional stochastic delay equations. Lobachevskii J. Math. **30**, 195–202 (2009)
13. L. Kexue, P. Jigen, Controllability of fractional neutral stochastic functional differential systems. Z. Angew. Math. Phys. **65**, 941–959 (2014)
14. X. Yang, H. Gu, Complete controllability for fractional evolution equations. Abstr. Appl. Anal. **2014**, Article ID 730695, 8 (2014)
15. S. Kailasavalli, S. Suganya, M.M. Arjunan, Exact controllability of fractional neutral integro-differential systems with state-dependent delay. Nonlinear Stud. **22**, 687–704 (2015)
16. J. Liang, H. Yang, Controllability of fractional integro-differential evolution equations with nonlocal conditions. Appl. Math. Comput. **254**, 20–29 (2015)
17. L. Chen, G. Li, Approximate controllability of impulsive differential equations with nonlocal conditions. Int. J. Nonlinear Sci. **10**(4), 438–446 (2010)
18. A. Shukla, N. Sukavanam, D.N. Pandey, Approximate controllability of semilinear stochastic control system with nonlocal conditions. Nonlinear Dyn. Syst. Theory **15**(3), 321–333 (2015)
19. D. Ahluwalia, N. Sukavanam, U. Arora, Approximate controllability of abstract semilinear stochastic control systems with nonlocal conditions. Cogent Math. **3**, 1191409 (2016)
20. A. Babiarz,, J. Klamka, M. Niezabitowski, Schauder's fixed-point theorem in approximate controllability problems. Int. J. Appl. Math. Comput. Sci. **26**(2), 263–275 (2016)
21. X. Fu, H. Rong, Approximate controllability of semilinear non-autonomous evolutionary systems with nonlocal conditions. Autom. Remote Control **77**(3), 428–442 (2016)
22. U. Arora, N. Sukavanam, Approximate controllability of second order semilinear stochastic system with variable delay in control and with nonlocal conditions. Rend. Circ. Mat. Palermo **65**, 307–322 (2016)
23. A. Shukla, U. Arora, N. Sukavanam, Approximate controllability of retarded semilinear stochastic system with non local conditions. J. Appl. Math. Comput. **49**, 513–527 (2015)
24. F.Z. Mokkedem, X. Fu, Approximate controllability for a retarded semilinear stochastic evolution system. IMA J. Math. Control Inf. **36**, 285–315 (2019)
25. R. Sakthivel, R. Yong, N.I. Mahmudov, On the approximate controllability of semilinear fractional differential systems. Comput. Math. Appl. **62**, 1451–1459 (2011)

26. N. Sukavanam, S. Kumar, Approximate controllability of fractional order semilinear delay systems. J. Optim. Theory Appl. **151**(2), 373–384 (2011)
27. S. Kumar, N. Sukavanam, Approximate controllability of fractional order semilinear systems with bounded delay. J. Differ. Equ. **252**, 6163–6174 (2012)
28. S. Kumar, N. Sukavanam, Approximate controllability of fractional order semilinear delayed control systems. Nonlinear Stud. **20**(1), 73–83 (2013)
29. S. Kumar, N. Sukavanam, Controllability of fractional order system with nonlinear term having integral contractor. Fract. Calc. Appl. Anal. **16**(4), 791–801 (2013)
30. A. Debbouche, D.F.M. Torres, Approximate controllability of fractional nonlocal delay semilinear systems in Hilbert spaces. Int. J. Control **86**, 1577–1585 (2013)
31. R. Sakthivel, Y. Ren, Approximate controllability of fractional differential equations with state-dependent delay. Results Math. **63**, 949–963 (2013)
32. Z. Liu, X. Li, Approximate controllability of fractional evolution systems with Riemann–Liouville fractional derivatives. SIAM J. Control Optim. **53**, 1920–1933 (2015)
33. T. Lian, Z. Fan, G. Li, Approximate controllability of semilinear fractional differential systems of order $1 < q < 2$ via resolvent operators. Filomat **31**, 5769–5781 (2017)
34. R. Sakthivel, R. Ganesh, S. Suganya, Approximate controllability of fractional neutral stochastic system with infinite delay? Rep. Math. Phys. **70**, 291–311 (2012)
35. R. Sakthivel, S. Suganya, S.M. Anthoni, Approximate controllability of fractional stochastic evolution equations. Comput. Math. Appl. **63**, 660–668 (2012)
36. P. Muthukumar, C. Rajivganthi, Approximate controllability of fractional order neutral stochastic integro-differential system with nonlocal conditions and infinite delay. Taiwanese J. Math. **17**(5), 1693–1713 (2013)
37. M. Kerboua, A. Debbouche, D. Baleanu, Approximate controllability of Sobolev type fractional stochastic nonlocal nonlinear differential equations in Hilbert spaces. Electron. J. Qual. Theory Differ. Equ. Paper No. **58**, 1–16 (2014)
38. R. Sakthivel, Y. Ren, A. Debbouche, N.I. Mahmudov, Approximate controllability of fractional stochastic differential inclusions with nonlocal conditions. Appl. Anal. **95**(11), 2361–2382 (2016)
39. P. Balasubramaniam, P. Tamilalagan, Approximate controllability of a class of fractional neutral stochastic integro-differential inclusions with infinite delay by using Mainardi's function. Appl. Math. Comput. **256**, 232–246 (2015)
40. A. Boudaoui, A. Slama, Approximate controllability of nonlinear fractional impulsive stochastic differential equations with nonlocal conditions and infinite delay. Nonlinear Dyn. Syst. Theor. **16**, 35–48 (2016)
41. A. Shukla, N. Sukavanam, D.N. Pandey, Approximate controllability of fractional semilinear stochastic system of order $\alpha \in (1, 2]$. J. Dyn. Control Syst. **23**, 679–691 (2017)
42. A. Chadha, S.N. Bora, R. Sakthivel, Approximate controllability of impulsive stochastic fractional differential equations with nonlocal conditions. Dyn. Syst. Appl. **27**(1), 1–29 (2018)
43. X.B. Shu, Y. Lai, Y. Chen, The existence of mild solutions for impulsive fractional partial differential equations. Nonlinear Anal. **74**, 2003–2011 (2011)
44. K. Jeet, D. Bahuguna, Approximate controllability of nonlocal neutral fractional integro-differential equations with finite delay. J. Dyn. Control Syst. **22**(3), 485–504 (2016)
45. N.I. Mahmudov, S. Zorlu, Approximate controllability of fractional integro-differential equations involving nonlocal initial conditions. Bound. Value Probl. **2013**, 118 (2013). https://doi.org/10.1186/1687-2770-2013-118
46. N.I. Mahmudov, Approximate controllability of semilinear deterministic and stochastic evolution equations in abstract spaces. SIAM J. Control Optim. **42**, 1604–1622 (2003)
47. J.P. Dauer, N.I. Mahmudov, Controllability of stochastic semilinear functional differential equations in Hilbert spaces. J. Math. Anal. Appl. **290**, 373–394 (2004)

On Noncritical Solutions of Complementarity Systems

Andreas Fischer and Mario Jelitte

1 Introduction

The present paper is devoted to characterize local Lipschitzian error bounds for a system of nonsmooth equations

$$F(u) = 0, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a (at least) locally Lipschitz continuous function whose further properties will be specified later on. The solution set of equation (1) is denoted by

$$F^{-1}(0) := \{u \in \mathbb{R}^n \mid F(u) = 0\}.$$

It is said that F *provides a local Lipschitzian error bound* at $\hat{u} \in F^{-1}(0)$, if there are constants $c, \varepsilon > 0$ such that

$$c\,\mathrm{dist}[u, F^{-1}(0)] \leq \|F(u)\| \quad \text{for all} \quad u \in \hat{u} + \varepsilon\mathbb{B}, \tag{2}$$

where $\|\cdot\|$ stands for the Euclidean norm, $\mathbb{B} := \{u \in \mathbb{R}^n \mid \|u\| \leq 1\}$ is the unit ball, and $\mathrm{dist}[u, U] := \inf\{\|u - y\| \mid y \in U\}$ denotes the point-to-set distance of u to a nonempty set $U \subset \mathbb{R}^n$.

Different types of error bounds are fundamental for the design and analysis of numerical methods in the field of mathematical programming [29]. Let us, in particular, consider local Lipschitzian error bounds for systems of equations. This

A. Fischer (✉) · M. Jelitte
Faculty of Mathematics, Institute of Numerical Mathematics, Technische Universität Dresden, Dresden, Germany
e-mail: Andreas.Fischer@tu-dresden.de; Mario.Jelitte@tu-dresden.de

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_11

type of error bounds turned out to be of high relevance for the construction of Newton-type methods for solving these systems even if a system has nonisolated solutions, see [1, 9, 11, 13, 25, 34, 35], for example. Later on, more difficult cases were dealt with, which allow that a solution is not only nonisolated but that the function F is also not differentiable there, for instance, see [2, 7, 12, 16]. Again, local Lipschitzian error bounds played an important role.

Several problems in mathematical programming, such as generalized Nash equilibrium problems [15], possess nonisolated solutions. It is well known that necessary optimality conditions of those problems can be written as complementarity systems (for a definition, see Sect. 3 or [16]). To apply Newton-type methods for the solution of complementarity systems, they are often reformulated as systems of equations (1). However, the resulting systems might be nonsmooth at possibly nonisolated solutions. Therefore, Newton-type methods described in the previous paragraph can be helpful but require the knowledge whether appropriate local Lipschitzian error bounds hold.

The present paper contributes to this question. More in detail, based on recent results in [17], we first extend the notion of a *noncritical solution* developed in [23] to a more general case with reduced smoothness. Moreover, we will see that the relation between the existence of a local Lipschitzian error bound and the noncriticality carries over to our more general case. Then, in Sect. 3, we introduce the *Switching Index Condition (SIC)*. This new condition will allow us to prove that a reformulation of the complementarity system as nonsmooth system of equations (1) is strictly semidifferentiable in a certain sense. Finally, in Sect. 4, we will consider the Karush–Kuhn–Tucker (KKT) conditions arising from smooth inequality constrained optimization problems and, related to this, a reformulation of the KKT conditions as a nonsmooth system of equations (1). Then, based on results in Sects. 2 and 3, we will characterize the existence of a local Lipschitzian error bound under SIC but without conditions on the local isolatedness of primal or dual variables.

Notation A nonempty set $C \subset \mathbb{R}^n$ is called *cone* if $v \in C$ implies $\lambda v \in C$ for all $\lambda \in [0, \infty)$. For $C \subset \mathbb{R}^n$ and $v \in C$, we write $v' \overset{C}{\to} v$ to say that all sequences $(v^k) \subset C$ with $v^k \to v$ are meant. Instead of $t \overset{(0,\infty)}{\to} 0$, we write $t \downarrow 0$. Moreover, $o : (0, \infty) \to \mathbb{R}$ is used to denote any function satisfying $o(t)/t \to 0$ as $t \downarrow 0$.

2 Definitions and Preliminary Results

In this section, we exploit some results from [17], where the concept of *noncritical solutions* of differentiable systems of equations, introduced in [23], is extended, in particular to cases of nonsmooth systems.

Let $C \subset \mathbb{R}^n$ be a given set. Then, the *regular tangent cone to C at $u \in C$* is given by

$$\widehat{T}^{C}(u) := \left\{ v \in \mathbb{R}^n \mid \forall (u^k) \overset{C}{\to} u \forall (t_k) \downarrow 0 \exists (v^k) \to v : (u^k + t_k v^k) \subset C \right\}, \tag{3}$$

for example, see [32, Definition 11.1.1].

Throughout the paper, $\hat{u} \in F^{-1}(0)$ denotes an arbitrary but fixed solution of equation (1). In what follows, we want to approximate $\widehat{T}^{F^{-1}(0)}(\hat{u})$ by the cone

$$\mathfrak{D}F(\hat{u}) := \left\{ v \in \mathbb{R}^n \mid \exists (t_k) \downarrow 0 \exists (v^k) \to v : \|F(\hat{u} + t_k v^k)\| = o(t_k) \right\}. \tag{4}$$

To this end, we will make use of two notions of directional differentiability. The function F is called *semidifferentiable at* $\hat{u}$ if there exists a continuous and positively homogeneous function $F'(\hat{u}) : \mathbb{R}^n \to \mathbb{R}^m$ so that, for all $v \in \mathbb{R}^n$,

$$\|F(\hat{u} + tv') - F(\hat{u}) - tF'(\hat{u})(v')\| = o(t) \quad \text{as } t \downarrow 0 \text{ and } v' \to v \tag{5}$$

is valid. Notice that, due to the local Lipschitz continuity of F, semidifferentiability is the same as *B-differentiability* or directional differentiability *in the sense of Hadamard*, see the discussions in [26] for instance. According to [17, Definition 3], we call F *strictly semidifferentiable at* $\hat{u}$ *with respect to* $F^{-1}(0)$, if there exists a continuous and positively homogeneous function $F'(\hat{u}) : \mathbb{R}^n \to \mathbb{R}^m$ so that, for all $v \in \mathbb{R}^n$,

$$\|F(u + tv') - F(u) - tF'(\hat{u})(v')\| = o(t) \quad \text{as } u \overset{F^{-1}(0)}{\to} \hat{u},\ t \downarrow 0,\ \text{and } v' \to v \tag{6}$$

holds. For equivalent and more extended definitions of semidifferentiability and strict semidifferentiability and for related discussions, we refer to [17]. Finally, note that in (5) and (6), $F(\hat{u})$ and $F(u)$ are equal to 0 since $\hat{u}$ and u used in these formulas are solutions of (1).

To proceed, we need two results from [17]. Firstly, the inclusion

$$\widehat{T}^{F^{-1}(0)}(\hat{u}) \subset \mathfrak{D}F(\hat{u}) \tag{7}$$

is valid according to [17, Lemma 8]. Secondly, if F is semidifferentiable at $\hat{u}$, then [17, Lemma 5 a)] yields

$$\mathfrak{D}F(\hat{u}) = \left\{ v \in \mathbb{R}^n \mid F'(\hat{u})(v) = 0 \right\}. \tag{8}$$

Definition 1 A solution $\hat{u}$ of (1) is called *noncritical*, if $\widehat{T}^{F^{-1}(0)}(\hat{u}) = \mathfrak{D}F(\hat{u})$ holds. Otherwise, $\hat{u}$ is called *critical*.

This definition extends the one given in [23] to some nonsmooth setting. For more discussion, see Remark 2 below.

Remark 1 If F is strictly differentiable at $\hat{u}$ [32, Definition 3.2.2] with Jacobian $F'(\hat{u})$ and $\ker F'(\hat{u})$ denoting the nullspace of $F'(\hat{u})$, then (8) yields

$$\mathfrak{D}F(\hat{u}) = \ker F'(\hat{u}).$$

In this case, a condition ensuring that $\hat{u}$ is noncritical is $\operatorname{rank} F'(\hat{u}) = m$, or equivalently, $F'(\hat{u})\mathbb{R}^n = \mathbb{R}^m$, see [31, 6.32 Exercise] for instance. The latter condition is also known as *Lyusternik's regularity condition*, cf. [4, 19, 20]. □

Theorem 1 *Suppose that F is strictly semidifferentiable at $\hat{u}$ with respect to $F^{-1}(0)$. Then, $\hat{u}$ is noncritical if and only if F provides a local Lipschitzian error bound at $\hat{u}$.*

Proof This theorem is a special case of [17, Corollary 1]. □

Remark 2 If $\hat{u}$ is noncritical, then [17, Lemma 8] yields that the set $F^{-1}(0)$ is *Clarke regular* at $\hat{u}$, see [3, Definition 2.4.6] for the notion of Clarke regularity. Thus, a crucial difference between Definition 1 and [23, Definition 1] is that $\widehat{T}^{F^{-1}(0)}(\hat{u})$ is now not restricted to be a linear subspace of $\mathbb{R}^n$, which can be seen in Example 1 below. Somehow surprisingly, [17, Lemma 11] shows that noncriticality of $\hat{u}$ together with strict semidifferentiability of F at $\hat{u}$ with respect to $F^{-1}(0)$ implies that $\widehat{T}^{F^{-1}(0)}(\hat{u})$ (and thus too $\mathfrak{D}F(\hat{u})$) is a linear subspace. □

3 The Switching Index Condition

For continuously differentiable functions $\varphi, \psi : \mathbb{R}^n \to \mathbb{R}^m$, let us consider the *complementarity system*

$$\varphi(u) \geq 0, \quad \psi(u) \geq 0, \quad \varphi(u)^\top \psi(u) = 0. \tag{9}$$

There are various ways to rewrite this system as an equation [5]. Here, we use the min-function approach and define $F : \mathbb{R}^n \to \mathbb{R}^m$ by

$$F(u) := \min\{\varphi(u), \psi(u)\} \quad \text{for } u \in \mathbb{R}^n, \tag{10}$$

where min is taken componentwise. Since

$$\min\{a, b\} = 0 \quad \Leftrightarrow \quad a \geq 0,\ b \geq 0,\ ab = 0$$

holds for all $a, b \in \mathbb{R}$, we see that $u \in \mathbb{R}^n$ solves (9) if and only if $F(u) = 0$. However, note that the min-function is not differentiable if $a = b$. Hence, F is in general not everywhere differentiable. To deal with this problem, let us define the index set

$$\mathcal{V}(u) := \{i \mid \varphi_i(u) = \psi_i(u)\} \tag{11}$$

for any $u \in \mathbb{R}^n$. Our aim is to apply Theorem 1 to Eq. (1) with F defined by (10). To this end, we introduce the following condition that, as we will show later on, ensures strict semidifferentiability with respect to the solution set of (1).

Switching Index Condition (SIC) at $\hat{u}$.

$$\exists \varepsilon > 0 \forall u \in F^{-1}(0) \cap (\hat{u} + \varepsilon \mathbb{B}) : \mathcal{V}(\hat{u}) = \mathcal{V}(u).$$

As the continuity of φ and ψ yields $\mathcal{V}(u) \subset \mathcal{V}(\hat{u})$ for all $u \in F^{-1}(0)$ sufficiently close to $\hat{u}$, SIC guarantees that the index set $\mathcal{V}(u)$ does not change for all solutions u in a sufficiently small neighborhood of $\hat{u}$, i.e., $\mathcal{V}(\hat{u})$ does not contain an index i that switches to $\{1, \ldots, m\} \setminus \mathcal{V}(u)$ for $u \in F^{-1}(0)$ sufficiently close to $\hat{u}$.

Notice that SIC is weaker than the *strict complementarity condition* at $\hat{u}$. The latter can be written as $\mathcal{V}(\hat{u}) = \varnothing$. A trivial example, where SIC is fulfilled but strict complementarity is violated for any solution of (9), is given by $\varphi(u) := \psi(u) := 0$ for $u \in \mathbb{R}$.

The next lemma is the main result of this section. On the one hand, it is shown that SIC ensures strict semidifferentiability of F at $\hat{u}$ with respect to $F^{-1}(0)$. On the other hand, we provide a formula to compute $\mathfrak{D}F(\hat{u})$.

Lemma 1 *F is semidifferentiable at $\hat{u}$, where $F'(\hat{u}) : \mathbb{R}^n \to \mathbb{R}^m$ is given by*

$$F'(\hat{u})_i(v) = \begin{cases} \psi_i'(\hat{u})v, & \text{if } \varphi_i(\hat{u}) > 0, \\ \varphi_i'(\hat{u})v, & \text{if } \psi_i(\hat{u}) > 0, \\ \min\{\varphi_i'(\hat{u})v,\ \psi_i'(\hat{u})v\}, & \text{if } i \in \mathcal{V}(\hat{u}) \end{cases} \tag{12}$$

for $i \in \{1, \ldots, m\}$ and $v \in \mathbb{R}^n$. Moreover, $v \in \mathfrak{D}F(\hat{u})$ holds if and only if there is some constant $\tau > 0$ with

$$\min\left\{\varphi(\hat{u}) + t\,\varphi'(\hat{u})v,\ \psi(\hat{u}) + t\,\psi'(\hat{u})v\right\} = 0 \quad \text{for all } t \in [0, \tau]. \tag{13}$$

If SIC is valid at $\hat{u}$, then F is strictly semidifferentiable at $\hat{u}$ with respect to $F^{-1}(0)$.

Proof Because φ and ψ are both continuous near $\hat{u}$, there exists $\delta > 0$ such that for all $u \in \hat{u} + \delta\mathbb{B}$ the inclusions

$$\{i \mid 0 < \varphi_i(\hat{u})\} \subset \{i \mid \psi_i(u) \leq \varphi_i(u)\} \subset \{i \mid \psi_i(\hat{u}) = 0\} \tag{14}$$

are valid. Pick $\hat{v} \in \mathbb{R}^n$ arbitrarily. For any $i \notin \mathcal{V}(\hat{u})$, we can assume without loss of generality that $\varphi_i(\hat{u}) > 0$. Then, (14) implies that for all $u \in F^{-1}(0) \cap (\hat{u} + \delta\mathbb{B})$, $t \geq 0$ and $v \in \hat{v} + \mathbb{B}$ with $u + tv \in \hat{u} + \delta\mathbb{B}$,

$$F_i(u + tv) = \min\{\varphi_i(u + tv), \psi_i(u + tv)\} = \psi_i(u + tv) \tag{15}$$

holds true. Therefore, and because ψ is continuously differentiable, [32, Proposition 3.4.2] yields

$$F_i(u+tv) - F_i(u) = \psi_i(u+tv) - \psi_i(u) = t\,\psi_i'(\hat{u})v + o(t), \tag{16}$$

as $u \to \hat{u}, t \downarrow 0$, and $v \to \hat{v}$, implying that F_i with $i \notin \mathcal{V}(\hat{u})$ is strictly differentiable at $\hat{u}$. For $i \in \mathcal{V}(\hat{u})$, we obtain that $\varphi_i(\hat{u}) = \psi_i(\hat{u}) = 0$ and

$$\begin{aligned} F_i(\hat{u}+tv) - F_i(\hat{u}) &= \min\{\varphi_i(\hat{u}+tv), \psi_i(\hat{u}+tv)\} \\ &= \min\left\{\varphi_i(\hat{u}) + t\varphi_i'(\hat{u})v + o(t), \psi_i(\hat{u}) + t\psi_i'(\hat{u})v + o(t)\right\} \\ &= t\,\min\left\{\varphi_i'(\hat{u})v,\ \psi_i'(\hat{u})v\right\} + o(t), \end{aligned}$$

as $t \downarrow 0$ and $v \to \hat{v}$, implying that F_i is semidifferentiable at $\hat{u}$. Thus, F is semidifferentiable at $\hat{u}$. Moreover, the latter calculation together with (16) gives the formula (12). The representation of $\mathfrak{D}F(\hat{u})$ by (13) is an immediate consequence of (8), (12), and the continuity of φ and ψ.

Let us finally assume that SIC is satisfied at $\hat{u}$ and consider $i \in \mathcal{V}(\hat{u})$. Then, there exists $\varepsilon \in (0, \delta]$ so that $\varphi_i(u) = \psi_i(u) = 0$ holds for all $u \in F^{-1}(0) \cap (\hat{u} + \varepsilon\mathbb{B})$. Therefore, we obtain

$$F_i(u+tv) - F_i(u) = t\,\min\left\{\varphi_i'(\hat{u})v,\ \psi_i'(\hat{u})v\right\} + o(t),$$

as $F^{-1}(0) \ni u \to \hat{u}$, $t \downarrow 0$, and $v \to \hat{v}$, implying that F_i is strictly semidifferentiable at $\hat{u}$ with respect to $F^{-1}(0)$. Taking into account (16) for $i \notin \mathcal{V}(\hat{u})$, we have shown that F is strictly semidifferentiable at $\hat{u}$ with respect to $F^{-1}(0)$. □

Remark 3 According to Lemma 1, we obtain for some $v \in \mathbb{R}^n$ that $v \in \mathfrak{D}F(\hat{u})$ if and only if v solves the linear complementarity system

$$\left.\begin{aligned} \psi_i'(\hat{u})v = 0, &\quad \text{if } \varphi_i(\hat{u}) > 0, \\ \varphi_i'(\hat{u})v = 0, &\quad \text{if } \psi_i(\hat{u}) > 0, \\ \varphi_i'(\hat{u})v \geq 0,\ \psi_i'(\hat{u})v \geq 0,\ (\varphi_i'(\hat{u})v)(\psi_i'(\hat{u})v) = 0, &\quad \text{if } i \in \mathcal{V}(\hat{u}). \end{aligned}\right\} \tag{17}$$

Therefore, if $\mathcal{V}(\hat{u}) \neq \varnothing$, we cannot expect $\mathfrak{D}F(\hat{u})$ to be convex in general. In contrast to this, $\widehat{T}^{F^{-1}(0)}(\hat{u})$ is always convex, see [32, Proposition 11.1.2] for instance. Thus, the equality $\widehat{T}^{F^{-1}(0)}(\hat{u}) = \mathfrak{D}F(\hat{u})$ (as requested in Definition 1 for the noncriticality of $\hat{u}$) can be violated particularly if $\mathcal{V}(\hat{u}) \neq \varnothing$. □

Let us close the section with examples. The first example demonstrates that noncriticality of $\hat{u}$ does not imply that the regular tangent cone to $F^{-1}(0)$ at $\hat{u}$ is a linear subspace. The second example shows that in Theorem 1, strict semidifferentiability with respect to $F^{-1}(0)$ cannot be replaced by semidifferentiability in general.

Example 1 Consider $\varphi(u) := (u_1, u_2)$ and $\psi(u) := (1, u_1)$ for $u = (u_1, u_2) \in \mathbb{R}^2$. Then, we find $F^{-1}(0) = \{0\} \times \mathbb{R}_+$. Moreover, let $\hat{u} := 0$. Now, it is easily seen that $\widehat{T}^{F^{-1}(0)}(\hat{u}) = \{0\} \times \mathbb{R}_+$. According to (17) in Remark 3, we further observe that $\mathfrak{D}F(\hat{u}) = \{0\} \times \mathbb{R}_+$. Thus, we conclude that $\hat{u}$ is noncritical, whereas $\widehat{T}^{F^{-1}(0)}(\hat{u})$ is clearly not a linear subspace but just a half-space. □

Example 2 Consider $\varphi(u) := u_1$ and $\psi(u) := u_2$ for $u = (u_1, u_2) \in \mathbb{R}^2$. Then, Lemma 1 yields that F is semidifferentiable at $\hat{u} := 0$ but not strictly semidifferentiable. Now, on the one hand, it is known [30, Proposition 1] that F provides a local Lipschitzian error bound at each $u \in F^{-1}(0)$. On the other hand, $F^{-1}(0)$ is clearly not Clarke regular at $\hat{u}$. With Remark 2 in mind, $\hat{u}$ is thus referred to as critical. This shows that in general, strict semidifferentiability with respect to $F^{-1}(0)$ cannot be replaced by semidifferentiability in Theorem 1. □

4 An Application to KKT Systems

Let us now consider the inequality constrained nonlinear optimization problem

$$\theta(x) \to \min \quad \text{s.t.} \quad g(x) \leq 0, \tag{18}$$

where $\theta : \mathbb{R}^l \to \mathbb{R}$ and $g : \mathbb{R}^l \to \mathbb{R}^m$ are assumed to be twice continuously differentiable. The KKT system for (18) reads as

$$L(x, \mu) := \theta'(x)^\top + g'(x)^\top \mu = 0, \quad g(x) \leq 0, \quad \mu \geq 0, \quad \mu^\top g(x) = 0. \tag{19}$$

With

$$\varphi(x, \mu) := \begin{pmatrix} -L(x, \mu) \\ -g(x) \end{pmatrix} \quad \text{and} \quad \psi(x, \mu) := \begin{pmatrix} L(x, \mu) \\ \mu \end{pmatrix}, \tag{20}$$

we obtain that any solution of (19) is a solution of the complementarity system (9) with $u = (x, \mu)$ and vice versa. Thus, in the following, we consider F as in (10), that is,

$$F(x, \mu) = \min\{\varphi(x, \mu), \psi(x, \mu)\} \quad \text{for } (x, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$$

and fix any $(\hat{x}, \hat{\mu}) \in F^{-1}(0)$.

Remark 4 Let the multifunction $M : \mathbb{R}^l \rightrightarrows \mathbb{R}^m$ be defined by

$$M(x) := \left\{ \mu \in \mathbb{R}^m \mid (x, \mu) \in F^{-1}(0) \right\} \quad \text{for } x \in \mathbb{R}^l. \tag{21}$$

Because $F^{-1}(0) = \left\{(x,\mu) \in \mathbb{R}^l \times \mathbb{R}^m \mid \mu \in M(x)\right\}$, we note that the computation of $\widehat{T}^{F^{-1}(0)}(\hat{x},\hat{\mu})$ can be challenging [31, 8.33 Definition and formula 8(16)]. However, if $\hat{x}$ is an isolated primal solution of (19), that is, there exists $\varepsilon > 0$ such that

$$F^{-1}(0) \cap ((\hat{x},\hat{\mu}) + \varepsilon\mathbb{B}) = \{\hat{x}\} \times \left(M(\hat{x}) \cap (\hat{\mu} + \varepsilon\mathbb{B})\right),$$

then we get from [31, 6.41 Proposition] that $\widehat{T}^{F^{-1}(0)}(\hat{x},\hat{\mu}) = \{0\} \times \widehat{T}^{M(\hat{x})}(\hat{\mu})$. □

In the remainder, the index sets

$$\begin{aligned} I_g &:= \{i \mid g_i(\hat{x}) = 0\}, \quad & I_< &:= \{1,\dots,m\} \setminus I_g, \\ I_\mu &:= \{i \mid \hat{\mu}_i = 0\}, & I_> &:= \{1,\dots,m\} \setminus I_\mu \end{aligned}$$

are used. Moreover, by L'_x, the Jacobian of L with respect to x is denoted.

Lemma 2 *Let $(v,w) \in \mathbb{R}^l \times \mathbb{R}^m$. Then, $(v,w) \in \mathfrak{D}F(\hat{x},\hat{\mu})$ if and only if (v,w) solves the linear complementarity system*

$$\begin{aligned} L'_x(\hat{x},\hat{\mu})v + g'(\hat{x})^\top w &= 0, \\ g'_i(\hat{x})v &= 0 \ \text{for all } i \in I_>, \\ g'_i(\hat{x})v &\le 0 \ \text{for all } i \in I_g, \\ w_i &= 0 \ \text{for all } i \in I_<, \\ w_i &\ge 0 \ \text{for all } i \in I_\mu, \\ w_i g'_i(\hat{x})v &= 0 \ \text{for all } i \in I_g \cap I_\mu. \end{aligned} \tag{22}$$

Proof Lemma 1 yields $(v,w) \in \mathfrak{D}F(\hat{x},\hat{\mu})$ if and only if there is $\tau > 0$ such that

$$\begin{aligned} 0 &= \min\left\{\varphi(\hat{x},\hat{\mu}) + t\varphi'(\hat{x},\hat{\mu})\begin{pmatrix} v \\ w \end{pmatrix}, \psi(\hat{x},\hat{\mu}) + t\psi'(\hat{x},\hat{\mu})\begin{pmatrix} v \\ w \end{pmatrix}\right\} \\ &= \min\left\{-\begin{pmatrix} t\left(L'_x(\hat{x},\hat{\mu})v + g'(\hat{x})^\top w\right) \\ g(\hat{x}) + tg'(\hat{x})v \end{pmatrix}, \begin{pmatrix} t\left(L'_x(\hat{x},\hat{\mu})v + g'(\hat{x})^\top w\right) \\ \hat{\mu} + tw \end{pmatrix}\right\} \end{aligned}$$

for all $t \in [0,\tau]$. This, in turn, holds if and only if

$$\left.\begin{aligned} L'_x(\hat{x},\hat{\mu})v + g'(\hat{x})^\top w &= 0 \\ g(\hat{x}) + tg'(\hat{x})v &\le 0 \\ \hat{\mu} + tw &\ge 0 \\ g(\hat{x})^\top w + \hat{\mu}^\top g'(\hat{x})v + t\, w^\top g'(\hat{x})v &= 0 \end{aligned}\right\} \quad \text{for all } t \in [0,\tau]. \tag{23}$$

Finally, since each solution of (22) is a solution of (23) and vice versa, the assertion of the lemma is true. □

Lemma 3 (Sufficient and Necessary Conditions for SIC)

(a) SIC holds at $(\hat{x}, \hat{\mu})$ if and only if there exists $\varepsilon > 0$ such that, for each $i \in I_g \cap I_\mu$,

$$g_i(x) = \mu_i \quad \textit{for all } (x, \mu) \in F^{-1}(0) \cap \left((\hat{x}, \hat{\mu}) + \varepsilon \mathbb{B}\right)$$

is valid.

(b) If there is some $\varepsilon > 0$ so that, for each $i \in I_g \cap I_\mu$,

$$g_i(x) = \mu_i \quad \textit{for all } (x, \mu) \in (\hat{x}, \hat{\mu}) + \varepsilon \mathbb{B} \textit{ with } \min\{-g(x), \mu\} = 0,$$

then SIC is fulfilled at $(\hat{x}, \hat{\mu})$.

(c) Let SIC be satisfied at $(\hat{x}, \hat{\mu})$. Then, there exists $\delta > 0$ such that, for each $(x, \mu) \in (\hat{x}, \hat{\mu}) + \delta \mathbb{B}$, we have $(x, \mu) \in F^{-1}(0)$ if and only if

$$L(x, \mu) = 0, \quad \mu_i = 0 \textit{ for } i \in I_\mu \quad \textit{and} \quad g_i(x) = 0 \textit{ for } i \in I_g. \tag{24}$$

Proof

(a) It suffices to notice that each $(x, \mu) \in F^{-1}(0)$ solves $L(x, \mu) = 0$.

(b) This assertion follows from (a) because

$$F^{-1}(0) \subset \left\{(x, \mu) \in \mathbb{R}^l \times \mathbb{R}^m \mid \min\{-g(x), \mu\} = 0\right\}.$$

(c) If SIC holds at $(\hat{x}, \hat{\mu})$, then assertion (a) yields $\varepsilon > 0$ such that, for any $(x, \mu) \in (\hat{x}, \hat{\mu}) + \varepsilon \mathbb{B}$, we have $(x, \mu) \in F^{-1}(0)$ if and only if

$$L(x, \mu) = 0, \quad \mu_i \begin{cases} = 0 \text{ for } i \in I_< \cup (I_g \cap I_\mu) \\ > 0 \text{ for } i \in I_g \setminus I_\mu \end{cases},$$

$$g_i(x) \begin{cases} = 0 \text{ for } i \in I_> \cup (I_g \cap I_\mu) \\ < 0 \text{ for } i \in I_\mu \setminus I_g \end{cases}.$$

Since $I_\mu = I_< \cup (I_g \cap I_\mu)$ and $I_g = I_> \cup (I_g \cap I_\mu)$, the latter system equals

$$L(x, \mu) = 0, \quad \mu_i \begin{cases} = 0 \text{ for } i \in I_\mu \\ > 0 \text{ for } i \in I_g \setminus I_\mu \end{cases}, \quad g_i(x) \begin{cases} = 0 \text{ for } i \in I_g \\ < 0 \text{ for } i \in I_\mu \setminus I_g \end{cases}.$$

Hence, the assertion is true for some $\delta \in (0, \varepsilon]$ because g is continuous.

□

Notice that the property stated in item (a) of the previous lemma corresponds to the property given in item (d) of [8, Theorem 5], which is used (in combination with a local Lipschitzian error bound) to ensure quadratic convergence of the LP-Newton method for the solution of KKT systems.

We can now formulate the main result of this section.

Theorem 2 *Suppose that SIC is satisfied at* $(\hat{x}, \hat{\mu})$. *Then,* $(\hat{x}, \hat{\mu})$ *is noncritical if and only if* F *provides a local Lipschitzian error bound at* $(\hat{x}, \hat{\mu})$.

Proof According to Lemma 1, SIC at $(\hat{x}, \hat{\mu})$ implies that F is strictly semidifferentiable at $(\hat{x}, \hat{\mu})$ with respect to $F^{-1}(0)$. Therefore, Theorem 1 yields the equivalence stated. □

Remark 5 Notice that, although SIC is a strong assumption, it does neither imply isolatedness of the primal solution $\hat{x}$ of the KKT system (19) nor uniqueness of the multiplier $\hat{\mu}$. Hence, Theorem 2 extends the existing results as those given in [6, 8, 10, 18, 22]. □

Recall [21, Definition 2], where $\hat{\mu}$ is called *noncritical multiplier*, if there exists no $(v, w) \in \mathbb{R}^l \times \mathbb{R}^m$, with $v \neq 0$, that solves the complementarity system (22). Otherwise, $\hat{\mu}$ is called critical multiplier. Equivalently, one can say that $\hat{\mu}$ is a noncritical multiplier if and only if

$$(v, w) \in \mathfrak{D}F(\hat{x}, \hat{\mu}) \quad \Longrightarrow \quad v = 0. \tag{25}$$

In the following, we show that noncriticality of $(\hat{x}, \hat{\mu})$ corresponds to noncriticality of the multiplier $\hat{\mu}$ provided $\hat{x}$ is an isolated primal solution of (19).

Lemma 4 *Suppose that* $\hat{x}$ *is an isolated primal solution of* (19). *Then,* $(\hat{x}, \hat{\mu})$ *is noncritical if and only if* $\hat{\mu}$ *is a noncritical multiplier.*

Proof At first, we obtain from (21) that

$$M(\hat{x}) = \left\{ \mu \in \mathbb{R}^m \mid \sum_{i=1}^{m} \nabla g_i(\hat{x}) \mu_i = -\nabla \theta(\hat{x}), \quad \mu_i \begin{cases} = 0 \text{ for } i \in I_< \\ \geq 0 \text{ for } i \in I_g \end{cases} \right\}. \tag{26}$$

Evidently, $M(\hat{x})$ is a (closed and convex) polyhedron. Therefore, a combination of [31, 6.9 Theorem] and [31, 6.29 Corollary (e)] allows to apply [31, 6.46 Theorem] to (26), and we obtain

$$\widehat{T}^{M(\hat{x})}(\hat{\mu}) = \left\{ w \in \mathbb{R}^m \mid \sum_{i \in I_g} \nabla g_i(\hat{x}) w_i = 0, \quad w_i \begin{cases} = 0 \text{ for } i \in I_< \\ \geq 0 \text{ for } i \in I_\mu \end{cases} \right\}.$$

Taking into account Remark 4, we thus have

$$\widehat{T}^{F^{-1}(0)}(\hat{x},\hat{\mu}) = \{0\} \times \left\{ w \in \mathbb{R}^m \mid \sum_{i \in I_g} \nabla g_i(\hat{x}) w_i = 0,\ w_i \begin{cases} = 0 \text{ for } i \in I_< \\ \geq 0 \text{ for } i \in I_\mu \end{cases} \right\}.$$

Therefore, with Lemma 2 in mind, we observe that $\widehat{T}^{F^{-1}(0)}(\hat{x},\hat{\mu}) = \mathfrak{D}F(\hat{x},\hat{\mu})$ (i.e., $(\hat{x},\hat{\mu})$ is noncritical) is satisfied if and only if the implication in (25) holds, and thus, if and only if $\hat{\mu}$ is a noncritical multiplier. □

5 Conclusions

In the present article, the concept of noncritical solutions of nonlinear equations, introduced in [23] and recently extended in [17], is studied. We showed how the newly introduced Switching Index Condition (SIC) allows us to employ the latter concept for a nonsmooth reformulation of a complementarity system. Finally, an application to KKT systems, arising from smooth nonlinear programs with inequality constraints, is given. As a consequence, we achieved new conditions ensuring a local Lipschitzian error bound for such KKT systems under SIC, but without assuming isolatedness of the primal solution or the uniqueness of a multiplier. If, instead, the primal solution is isolated, we showed that noncritical solutions of the nonsmooth min-reformulation of the KKT system correspond to noncritical multipliers of this system and vice versa.

In the present paper, the problem functions φ and ψ appearing in the complementarity system were assumed to be continuously differentiable. If this assumption is violated, particular difficulties may arise and it might be helpful to exploit weaker differentiability notions for the problem functions, for example, see [14, 24, 27, 28, 33].

Acknowledgments We gratefully acknowledge the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—409756759. The authors are thankful to A. F. Izmailov for various comments and discussions.

References

1. R. Behling, A. Fischer, A unified local convergence analysis of inexact constrained Levenberg-Marquardt methods. Optim. Lett. **6**, 927–940 (2012)
2. R. Behling, A. Fischer, K. Schönefeld, N. Strasdat, A special complementarity function revisited. Optimization **68**, 65–79 (2019)
3. F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983)
4. A.V. Dmitruk, A.A. Milyutin, N.P. Osmolovskii, Lyusternik's theorem and the theory of extrema. Russ. Math. Surv. **35 (6)**, 11–51 (1980)
5. F. Facchinei, J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems* (Springer, New York, 2003)

6. F. Facchinei, A. Fischer, V. Piccialli, Generalized Nash equilibrium problems and Newton methods. Math. Program. **117**, 163–194 (2009)
7. F. Facchinei, A. Fischer, M. Herrich, A family of Newton methods for nonsmooth constrained systems with nonisolated solution. Math. Methods Oper. Res. **77**, 433–443 (2013)
8. F. Facchinei, A. Fischer, M. Herrich, An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. **146**, 1–36 (2014)
9. J. Fan, Y. Yuan, On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. Computing **74**, 23–39 (2005)
10. D. Fernández, M. Solodov, Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**, 47–73 (2010)
11. A. Fischer, Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
12. A. Fischer, M. Herrich, Newton-type methods for Fritz John systems of generalized Nash equilibrium problems. Pure Appl. Funct. Anal. **3**, 587–602 (2018)
13. A. Fischer, P.K. Shukla, A Levenberg-Marquardt algorithm for unconstrained multicriteria optimization. Oper. Res. Lett. **36**, 643–646 (2008)
14. A. Fischer, V. Jeyakumar, D.T. Luc, Solution point characterizations and convergence analysis of a descent algorithm for nonsmooth continuous complementarity problems. J. Optim. Theory Appl. **110**, 493–513 (2001)
15. A. Fischer, M. Herrich, K. Schönefeld, Generalized Nash equilibrium problems – Recent advances and challenges. Pesqui. Oper. **34**, 521–558 (2014)
16. A. Fischer, M. Herrich, A.F. Izmailov, M.V. Solodov, Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. Comput. Optim. Appl. **63**, 425–459 (2016)
17. A. Fischer, A.F. Izmailov, M. Jelitte, Constrained Lipschitzian error bounds and noncritical solutions of constrained equations. Set-Valued Var. Anal. (2020). https://doi.org/10.1007/s11228-020-00568-8
18. W. Hager, M. Gowda, Stability in the presence of degeneracy and error estimation. Math. Program. **85**, 181–192 (1999)
19. A.D. Ioffe, *Variational Analysis of Regular Mappings, Theory and Applications* (Springer, Cham, 2017)
20. A.D. Ioffe, V.M. Tihomirov, *Theory of Extremal Problems* (North Holland, Amsterdam, 1979)
21. A.F. Izmailov, M.V. Solodov, Stabilized SQP revisited. Math. Program. **133**, 93–120 (2012)
22. A.F. Izmailov, A.S. Kurennoy, M.V. Solodov, A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. Math. Program. **142**, 591–604 (2013)
23. A.F. Izmailov, A.S. Kurennoy, M.V. Solodov, Critical solutions of nonlinear equations: stability issues. Math. Program. **168**, 475–507 (2018)
24. V. Jeyakumar, D.T. Luc, Approximate Jacobian matrices for nonsmooth continuous maps and C^1-optimization. SIAM J. Control Optim. **36**, 1815–1832 (1998)
25. C. Kanzow, N. Yamashita, M. Fukushima, Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J. Comput. Appl. Math. **172**, 375–397 (2004)
26. A.J. King, R.T. Rockafellar, Sensitivity analysis for nonsmooth generalized equations. Math. Program. **55**, 193–212 (1992)
27. D. Klatte, B. Kummer, *Nonsmooth Equations in Optimization* (Kluwer, Dordrecht, 2002)
28. B.S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings. J. Math. Anal. Appl. **183**, 250–288 (1994)
29. J.S. Pang, Error bounds in Mathematical Programming. Math. Program. **79**, 299–332 (1997)
30. S.M. Robinson, Some continuity properties of polyhedral multifunctions. Math. Program. Study **14**, 206–214 (1981)
31. R.T. Rockafellar, R.J.B. Wets, *Variational Analysis* (Springer, Berlin, 1998)
32. W. Schirotzek, *Nonsmooth Analysis* (Springer, Berlin, 2007)

33. M.A. Tawhid, M.S. Gowda, On two applications of H-Differentiability to optimization and complementarity problems. Comput. Optim. Appl. **17**, 279–299 (2000)
34. S.J. Wright, Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**, 253–275 (1998)
35. N. Yamashita, M. Fukushima, On the rate of convergence of the Levenberg-Marquardt method, in *Topics in Numerical Analysis, Computing Supplementa*, ed. by G. Alefeld, X. Chen, vol. 15 (Springer, Vienna, 2001), pp. 239–249

Testing the Performance of Some New Hybrid Metaheuristic Algorithms for High-Dimensional Optimization Problems

Souvik Ganguli

Abstract This chapter tests the performance of five new firefly-based hybrid algorithms to solve unconstrained high-dimensional as well as fixed-dimensional optimization problems. Firefly algorithm has been successfully combined with bacterial foraging, flower pollination, pattern search, and grey wolf optimizer to present these high performing computing algorithms. Three types of benchmark functions are taken up to justify each of the hybrid propositions. The first two sets of test functions, namely, the unimodal and the multimodal functions, are employed to validate the exploitation and exploration features, respectively, of the suggested techniques. The convergence plots show good promise in terms of convergence speed and accuracy than the existing algorithms. Even the nonparametric test, viz., the Kruskal–Wallis diagram, was used to validate the test results. The third type of test functions constituting some fixed-dimensional multimodal functions is also evaluated using the integrated methods. The statistical measures of the error function are also performed considering 50 independent test runs. The rank-sum test of Wilcoxon is carried out to validate the test outcomes. The proposed methods can also be suitably utilized to solve constrained as well as multi-objective optimization problems.

1 Introduction

Metaheuristic algorithms have now become important tools for different applications. The word "meta" means "beyond" or "higher." They outnumber ordinary heuristics. With the aid of randomization, the variety of solutions obtained using metaheuristics is always achieved. While metaheuristic algorithms are widely popular, the literature still offers no clear concept of heuristics and metaheuristics. They

S. Ganguli (✉)
Department of Electrical and Instrumentation Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
e-mail: souvik.ganguli@thapar.edu

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_12

are also used almost interchangeably by many researchers. But the general trend is aimed at marking all stochastic algorithms as metaheuristic by randomization and global exploration. Randomization offers a positive contribution to moving away from a local search to a global one. Nearly all metaheuristic algorithms are thus strongly suited to nonlinear modeling and control. Metaheuristic algorithms provide an effective means to provide, in relatively good time, appropriate solutions to a complex problem by trial and error. These algorithms are aimed not at finding any possible solution in the search space but at finding a feasible solution within a reasonable time limit. But there is no guarantee that we can get the right solutions. Any metaheuristic algorithm has two main components, namely exploration (diversification) and exploitation (intensification). Exploration produces a range of solutions for using the entire search space, while exploitation focuses on searching in a particular area by exercising the knowledge that a successful current solution is located in that area. A good balance between those two would ensure a global solution [29].

Though human beings' problem-solving abilities have always been heuristic or metaheuristic since the early periods of human history, yet its scientific study is relatively a budding venture. It was Alan Turing, who was perhaps the first person to coin the heuristic search technique during World War II. The 1960s and 1970s witnessed the development of Genetic Algorithms (GAs). Another breakthrough contribution is the proposition of the Simulated Annealing (SA) method in 1982. In 1992 and 1995, significant progress took place through the developments of Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO), respectively. In around 1996 and later in 1997, a vector-based evolutionary algorithm was coined as Differential Evolution (DE) came into existence. With the advent of the twenty-first century, things became even more fascinating. Many new algorithms like Bacterial Foraging Algorithm (BFA), Harmony Search (HS), Artificial Bee Colony (ABC) optimization, Firefly Algorithm (FA), Cuckoo Search (CS), Bat Algorithm (BA), and Flower Pollination Algorithm (FPA) also evolved [29].

Few metaheuristic algorithms directly associated with this research work are described as follows. Bacterial Foraging Algorithm (BFA), coined by Passino, was developed on the theme of the foraging strategy of *E. coli* bacteria that reside in the intestine of human beings[20]. However, investigation with complicated problems discloses that the BFA possesses poor convergence and its performance highly decreases with dimensionality and problem complexity. Another newcomer in the list of metaheuristic algorithms, viz. Firefly Algorithm (FA), is motivated by the communication and the flashing patterns of fireflies found in the tropical climatic conditions [27]. This algorithm has shown promising superiority over several algorithms in the recent past. An additional entrant in the tally of metaheuristic algorithms is the Flower Pollination Algorithm (FPA) that depicts the process of pollination in the plants having flowers. There are two fundamental mechanisms in FPA: global and local pollination. The switching probability in the algorithm is utilized to shift between the common global pollination and the intensive local pollination [28]. Even dynamism in switch probability often leads to improved solutions for different optimization problems. Grey wolf optimizer (GWO) is

another metaheuristic technique that is based on the leadership hierarchy and hunting behavior of grey wolves available mostly in the northern part of America [19]. The GWO method is tremendously popular among the researchers and thus has wide acceptance in diverse fields. Though several new algorithms [15–18, 24] have also evolved, yet pure algorithms cannot always deliver an optimal solution and are almost inferior to hybridizations. Moreover, Pattern Search (PS) algorithm [5] acts as a potential candidate to offer good local search capabilities and has been widely employed to constitute a hybrid combination with other metaheuristic algorithms. Further, it is also found that the one-dimensional chaotic maps play a dominant role to refine the capabilities of any metaheuristic algorithm [6].

The hybrid methods usually consist of two or more algorithms that work in cohesion to achieve a successful integration synergy. Hybridization of the algorithms is popular, due partly to improved noise handling efficiency, uncertainty, vagueness, and inaccuracy. The hybrid topologies perform a crucial role in the search power of the algorithms. The integration targets, at the same time seeking to mitigate any significant drawback, to incorporate the advantages of a single algorithm into the integrated algorithm. Overall, some improvements in both computational speed and precision are typical of the hybridization produced [26].

Hybridization with other algorithms is one of the standard variants of any metaheuristic algorithm. In this aspect, FA is no exception. ACO [21], GA [22], PSO [1], DE [14], and other algorithms have successfully combined FA. FA's strength in hybridizing with different algorithms is found suitable in both global and local search. While the hybrid variations of the firefly algorithm are available in the existing literature, there is still room for the development of new hybrid algorithms with FA. The methodology inspired by nature, coined as the Bacterial Foraging Algorithm (BFA) [20] depending on the foraging pattern of *E. coli* bacteria, suffers from the drawback of premature convergence for which a hybrid combination of FA and BFA can be called for. To develop a new hybrid algorithm, the flower pollination algorithm [28] with dynamic switch probability can be combined with FA. Also, the pattern search (PS) algorithm [5] is an ideal candidate to provide excellent local search capabilities [13, 23] and can be used as a hybrid combination with FA. In the literature, the hybridization of the grey wolf optimizer [19] with FA has been lacking and can be checked out. It is also found that chaos plays an essential role in improving any metaheuristic algorithm's performance [6]. It is also possible to hybridize chaotic firefly algorithms (CFAs) with GWO to provide improved convergence and accuracy compared to the parent algorithms. Making use of fewer fixed parameters in the algorithm has been the main motive of these hybrid topologies discussed so far. Another specialty of these algorithms is that higher-dimensional optimization problems have been checked. Also, these algorithms were evaluated with a competitive number of function evaluations and contrasted with some metaheuristic algorithms reported recently.

The remainder of the chapter is constructed in the manner discussed. In Sect. 2, the various hybrid topologies with firefly algorithms are detailed. Section 3 contains the experiments and their results. Eventually, in Sect. 4, the key findings are discussed with some directions for future scope.

2 A Brief Overview of Hybrid Methodologies

The main goal of hybridizing different algorithms is to build improved performance structures that incorporate the strengths of the individual strategies of algorithms. Talbi proposed terminology for hybrid metaheuristic algorithms in which two high-level or low-level algorithms can be hybridized as homogeneous or heterogeneous with a relay or co-evolutionary method [25]. The hybrid algorithms mentioned in this chapter are having heterogeneous architecture and are usually designated as the low-level relay type. The hybrid topologies are low level with the impression that both parent algorithms maintain functionalities within the hybrid system. The hybrid approaches are relay type as the parent algorithms are used one after the other. Two different algorithms are connected in all the hybrid propositions to return the desired outcomes. The hybrid methods are thus heterogeneous. Figure 1 generalizes the flowchart of the hybrid architecture discussed in this chapter.

The main reason behind this amalgamation is to resolve the drawbacks of the single optimization algorithm and to achieve an enhanced rendition. Furthermore, to reach the best solution possible in the time defined and finally harmonize the diversification and intensification, it is also important to define the strength of the proposed process. Those two concepts are utilized to investigate the new likely outcomes and strengthen the existing method to make it more praiseworthy.

To identify the various identification models in the delta domain, Ganguli et al. [10] have formed a merger of BFA and FA. In the literature, it was found that FA would divide the entire population automatically into subgroups with light-intensity

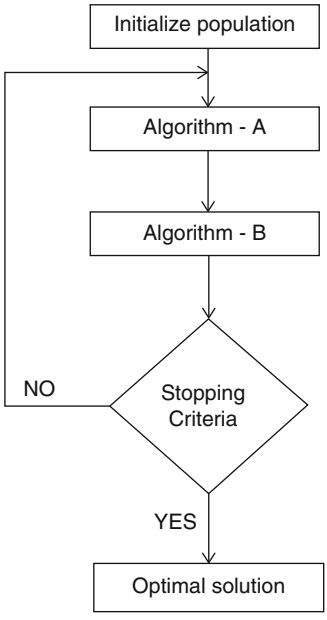


Fig. 1 Flowchart of proposed hybrid methods

variation as regards the attraction mechanism. Moreover, the firefly technique can also escape from local minimum conditions owing to the long-distance mobility using the Lévy flight mechanism. These merits of FA help it to explore. This results in two aspects of the hybridization of the FA with BFA. In the first step, FA undertook the heuristic research to explore the whole search area, and BFA was utilized to change the solution consistency to the optimization problem. The technique suggested was coined as a hybrid firefly algorithm (HFA).

A brand-new hybrid topology called the FAdFPA algorithm was devised by combining FA with FPA to solve system identification problem [11]. By employing FA for exploration and FPA for exploitation, the balance between diversification and intensification was achieved in this approach. This algorithm used both the merits of FA and FPA approaches successfully and avoided their drawbacks. In FAdFPA, a collection of random operators initialized the search process with the FA. For a certain count of iterations, the calculation proceeded to look for the overall best position in the entire search domain with the help of the firefly method. As the initial starting point for FPA, the best solution obtained via FA was taken, and then the search process was shifted to FPA to step up the confluence toward the optimal solution. In the hybrid algorithm, the switching probability of FPA was made adaptive by the formula:

$$p = p_{max} - (p_{max} - p_{min}) \times \frac{t}{T}. \tag{1}$$

In Eq. 1, p_{max} and p_{min} are the two fixed parameters of the algorithm having a standard choice of 0.9 and 0.4, respectively, as per literature, "T" represents the maximum iterations, while "t" denotes the present iteration [3].

Another hybrid algorithm called the FAPS algorithm has been developed by the authors that combine FA with PS for reduced-order modeling in the continuous-time domain [7]. In this method, the parity between exploration and exploitation was attained by employing FA as a global optimizer to perform exploration, while PS performed a local search to deliver exploitation functionality. The algorithm used the benefits of FA and PS approaches correctly and also minimized their limitations. In FAPS, a bunch of random agents initialized the search with FA. For some iterations, the calculation proceeded with FA to look for the best global location in the search space. FA's approach was taken as the starting point for PS. The quest method was then transferred to PS to speed up the optimal global convergence. Therefore, the integrated technique found an optimal solution faster and yet reliably.

The fusion of GWO and FA presented a new hybrid algorithm, called FAGWO to diminish the order of single-input single-output systems in the delta domain [8]. The balance between exploration and exploitation was achieved by applying FA globally, while GWO conducted a quest in the local search space to reveal the intensification functionality. The search method commenced with the initialization process from FA with the help of a class of agents randomly. The evaluation continued for a fixed count of iterations to attain the finest location within the global search domain. The

outcome determined with the help of the firefly technique was taken up as the initial point of the grey wolf optimizer. The search mechanism finally switched to the GWO method for the convergence process to reach the optimal solution quickly. In this way, the fusion method was able to obtain a global optimum correctly.

Another novel hybrid algorithm was also introduced to assess the Hammerstein and Wiener model's parameters in a unified domain incorporating the merits of GWO with CFA [9]. The hybridization took place in two stages. First of all, GWO was used to achieve the diversification of the algorithm to discover the optimal solution in the complete search space. Also, using the swarm behavior of the firefly algorithm powered by the iterative chaotic map, this algorithm's dominance in searching for the optimal solution has been improved. FA has a demerit to be stuck in several local optimums. However, since the parameters used are set and do not change with iterations, FA cannot come out of the local search. Therefore, an effort was made to use the iterative chaotic map to modify the algorithm parameters.

The GWO method, the other constituent algorithm, also suffered from the drawbacks of untimely convergence and sometimes get stuck at the local minimum. Thus, a hybrid algorithm known as GWOCFA was developed to solve their demerits. The balance between exploration and exploitation was achieved by using GWO as a global optimizer, while the chaotic firefly technique supported local search to deliver exploitation functionality. With a group of random agents as an initializer, the search process began with GWO. The computation continued to find the global best position in the complete search domain for a fixed number of iterations. GWO's findings were then well-selected as the source of CFA. The search technique then shifted to the chaotic firefly algorithm for the convergence to the global optimum within a quick time. Therefore, the combined approach could found optimum meticulously. The algorithm parameters of FA, viz., α and γ, were varied adaptively by applying iterative chaotic map defined by 2:

$$x_{k+1} = \sin\left(\frac{\pi a}{x_k}\right), \tag{2}$$

where $a \in (0, 1)$ is a suitable parameter [6].

3 Results and Discussions

The global optimization techniques discussed in Sect. 2 are now used to test some of the standard benchmark functions available in the literature. Three types of test functions are thus considered for the study. The first category belongs to the unimodal test functions and is characterized by a single global optimum. But there are no local minima in these types of functions. Hence these functions test the exploitation capability of the algorithms. The other set of functions is the multimodal functions having quite a several local minima. The third and the last

Table 1 Unimodal problems and their descriptions

Test functions	Dimension	Search domain	
Sphere (F1)	100	[−100, 100]	0
Schwefel 2.22 (F2)	100	[−10, 10]	0
Schwefel 2.21 (F3)	100	[−100, 100]	0
Rosenbrock (F4)	100	[−30, 30]	0
Step (F5)	100	[−100, 100]	0

category of the testbed is the fixed-dimensional multimodal functions that have only multiple optimum peaks. The second and third categories of test functions are normally employed to test the exploration capability of any algorithm. Further, they also test how the algorithm can avoid being getting trapped in local searches. The test functions considered in this chapter are taken up to minimize. A representative set of unimodal test functions (F1–F5) are thus shown in Table 1. More information about these test functions is available in [12].

From Table 1, it is seen that a hundred decision variables are taken up to evaluate the unimodal test functions. The population size considered to solve the different categories of optimization problems is set as 30, while the total number of iterations has been taken up as 500. The choice is made in such a way that the number of function evaluations (NFEs) turns out to be 15,000, quite competitive enough for a hundred decision variables. The results of the test functions are normalized between 0 and 1, zero being the best value and 1 being the worst value of the data set. The plot between the normalized value of the functional value and the number of iterations for the test function F1 is carried out in Fig. 2 to test both the convergence speed and accuracy of the HFA method over the parent and standard heuristic techniques reported in the literature.

The graphical representation of the convergence plot in Fig. 2 shows clearly that HFA outperforms the originator techniques FA and BFA. Moreover, our proposed method at the same time supersedes a few classical techniques like PSO, DE, and HS. To test further, the convergence characteristics for the test function F5 are also plotted corresponding to the hybrid topology FAPS. The firefly technique and some latest techniques like grasshopper optimization algorithm (GOA), whale optimization algorithm (WOA), sine cosine algorithm (SCA), and salp swarm algorithm (SSA) are used for comparison.

The FAPS techniques outperform the WOA, SCA, and SSA methods, while the FA and GOA approaches produce convergence plots that are quite close to the convergence of the proposed topology as seen in Fig. 3. As metaheuristic algorithms are stochastic processes, they do not yield unique results every time they are run on a PC. Hence some statistical assessments must take place to validate the results. Since multiple algorithms including the parent methods are used to compare with the proposed methods, hence Kruskal–Wallis test [2] proves to be a good measure to validate the outcomes obtained. The test function F4 is considered as a sample to verify the results of the FAGWO method as significant in terms of about ten heuristic algorithms, which is provided in Fig. 4.

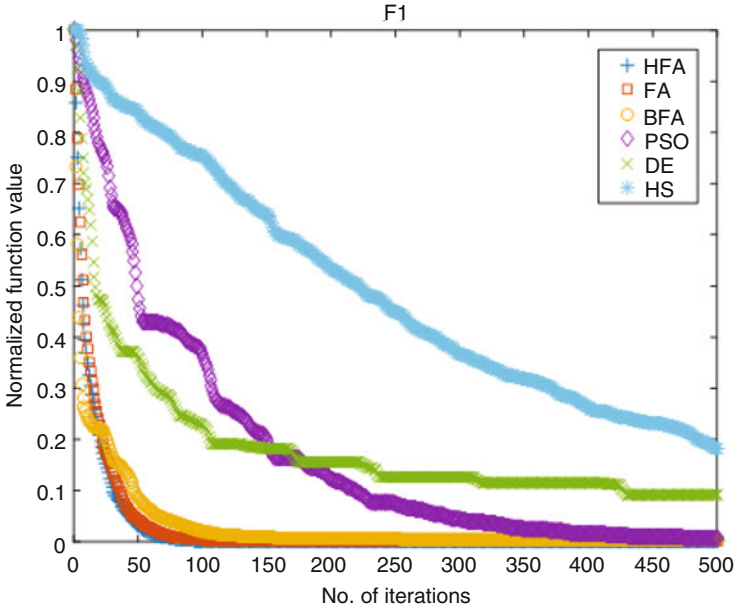


Fig. 2 Convergence plot of the test function F1 with HFA

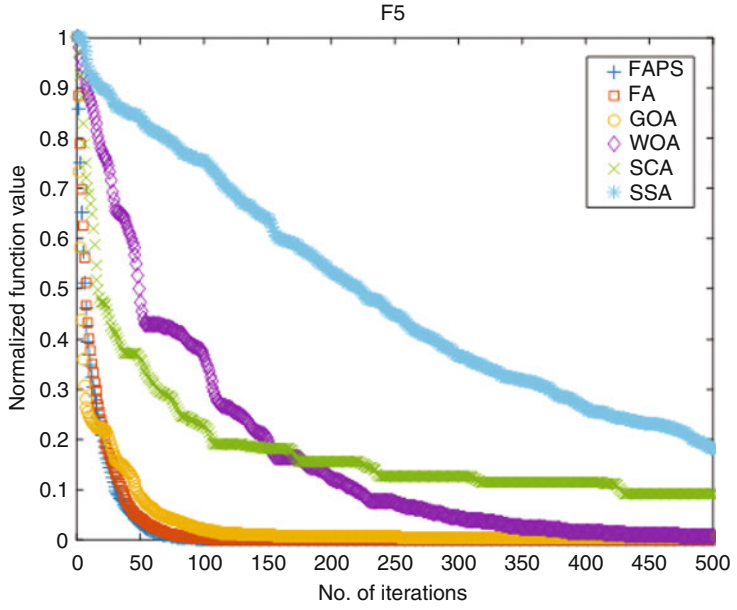


Fig. 3 Convergence characteristics of the test function F5 applying the FAPS technique

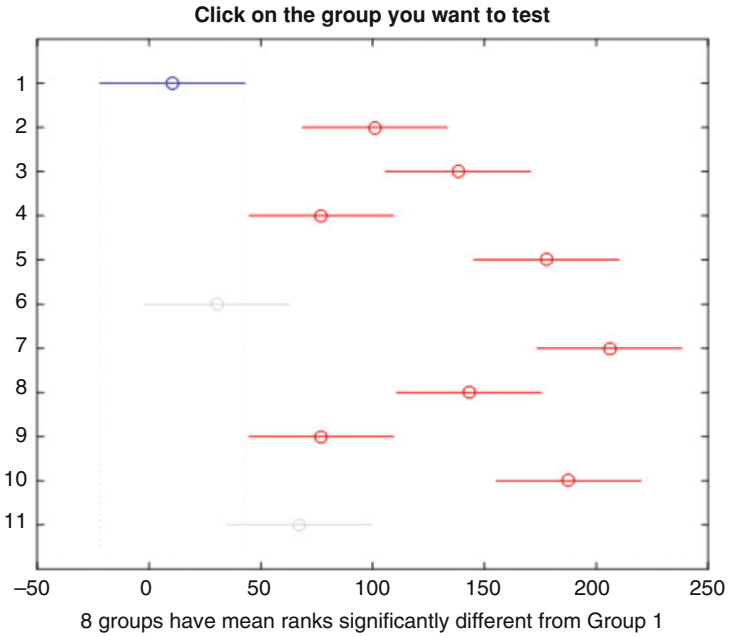


Fig. 4 Kruskal–Wallis test diagram for the benchmark function F4

Table 2 List of high-dimensional multimodal test functions

Test functions	Dimension	Search domain	*fmin*
Rastrigin (F6)	100	[−5.12, 5.12]	0
Ackley (F7)	100	[−32, 32]	0
Griewank (F8)	100	[−600, 600]	0
Penalized-1 (F9)	100	[−50, 50]	0
Penalized-2 (F10)	100	[−50, 50]	0

From the Kruskal–Wallis test results of the test function F4, it is observed that the mean ranks of the FAGWO algorithm differ significantly from the mean ranks of the eight algorithms out of the ten algorithms considered for this work. Some multimodal functions (F6–F10) are likewise chosen as found popular in the literature of benchmark functions [12]. The descriptions of these mathematical functions are presented in Table 2.

On a similar note, the convergence curves are drawn for the multimodal functions and compared with the parent methods as well as some metaheuristic techniques widely cited in the literature. As a sample, the test function F7 is chosen whose convergence plot is shown in Fig. 5. The GWOCFA method is expected to supersede the parent algorithms FA and GWO. The techniques like DE, HS, and the bird swarm algorithm (BSA) is used for comparison.

Only the GWO algorithm is somewhat close to the convergence plot of the GWOCFA technique. The rest of the algorithms compared are outperformed by the

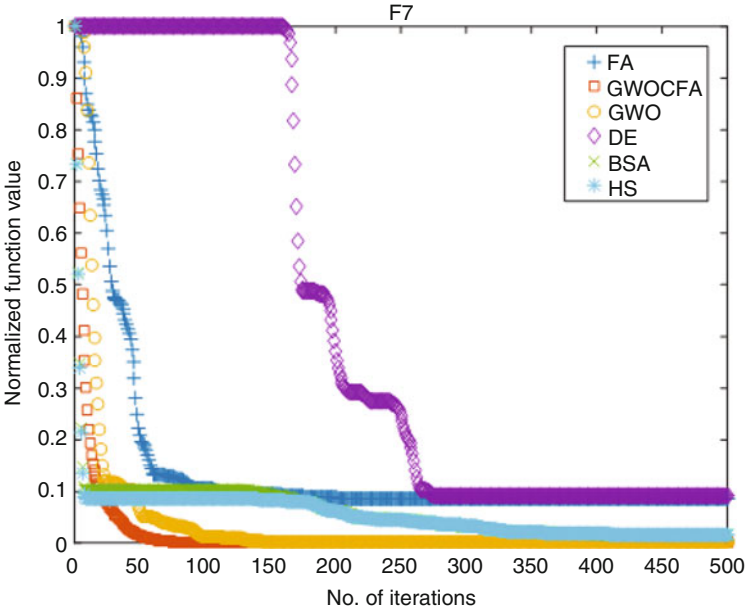


Fig. 5 Convergence curve of the test function F7 for the GWOCFA approach

suggested method. The Kruskal–Wallis test is carried out for the test function F9 as shown in Fig. 6.

The results as shown in Fig. 6 suggest that the mean ranks of the proposed technique outperform those of the other algorithms on most of the occasions. Thus, the results obtained are significant as well. Quite a handful number of fixed-dimensional multimodal test functions are also considered for the acid test of the proposed methods. Their descriptions are appended in Table 3.

Fifty test runs are taken up for obtaining meaningful statistical measures. The best, worst, mean, and standard deviation of each of the test functions are then calculated. The statistical analysis of $fmin$ is showcased in Table 4.

From Table 4, it is evident that the hybrid methods are either competitive or better than the parent and the standard heuristic methods used for comparison. The results ought to be followed by some nonparametric tests. Therefore, the rank-sum test of Wilcoxon [4] is taken up to prove that the results obtained are meaningful with respect to the other algorithms. Some selected p-values of the Wilcoxon test are provided in Table 5. Any p-value that is greater than 0.05 will be considered to be insignificant as per 95% confidence interval. They are underlined in Table 5 to make it noticeable. Only selected test functions, viz., F12, F18, F19, and F20, are taken up for the discussion.

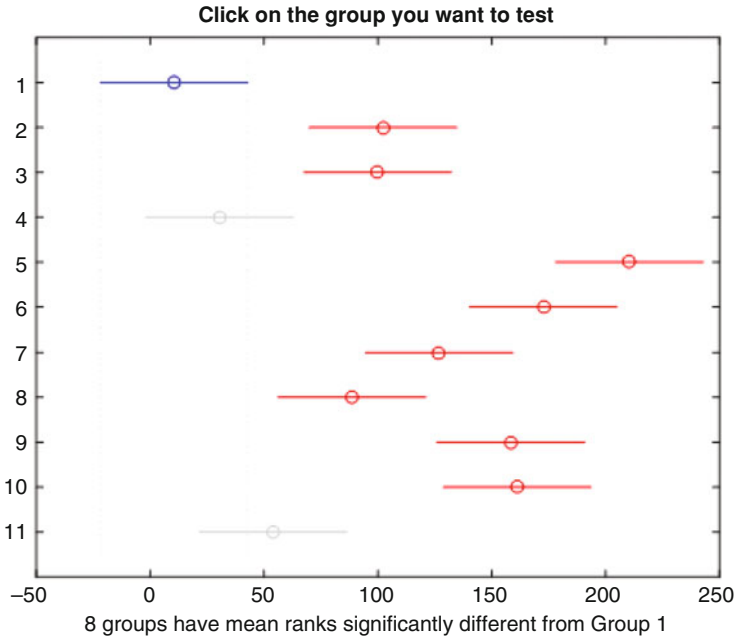


Fig. 6 Kruskal–Wallis test outcomes for the benchmark function F9

Table 3 Fixed-dimensional multimodal benchmark functions and their descriptions

Test functions	Dimension	Search domain	*fmin*
Foxholes (F11)	2	[−65.536, 65.536]	1
Kowalik (F12)	4	[−5, 5]	0
Six-hump camel (F13)	2	[−5, 5]	−1.0316
Branin (F14)	2	[−5, 15]	0.3979
Goldstein-price (F15)	2	[−2, 2]	3
Hartman 3 (F16)	3	[0, 1]	−3.8626
Hartman 6 (F17)	6	[0, 1]	−3.3220
Shekel 5 (F18)	4	[0, 10]	−10.1532
Shekel 7 (F19)	4	[0, 10]	−10.4029
Shekel 10 (F20)	4	[0, 10]	−10.5364

Most of the p-values in Table 5 are found to be quite less than 0.05, except on one occasion that is underlined in the table. Hence, the proposed algorithms generated valid results. The methods proposed can also be extended to solve both inequality and equality constraints-based design problems. There can also be proposed multi-objective variants of these algorithms. For these computational algorithms, parametric study and analysis can be performed. Also, higher-dimensional issues

Table 4 Experimental test results of fixed-dimensional multimodal benchmark functions (F11–F20)

Test functions	Algorithms	Best	Worst	Average	Std. dev.
F11	HFA	0.998	0.998	0.998	5.61E–16
	FAdFPA	0.998	0.998	0.998	5.61E–16
	FAPS	0.998	0.998	0.998	5.61E–16
	FAGWO	0.998	0.998	0.998	5.61E–16
	GWOCFA	0.998	0.998	0.998	5.61E–16
	PSOGSA	0.998	21.9884	6.2402	5.8503
	FA	0.998	0.998	0.998	5.61E–16
	BFA	0.998	1.0837	1.9921	0.1389
	FPA	0.998	1.0709	1.0059	0.022
	GWO	0.998	12.6705	3.8989	3.8756
	PSO	0.998	11.7187	5.7043	3.9947
	DE	0.998	0.998	0.998	5.61E–16
	HS	0.998	0.998	0.998	5.61E–16
	CSO	0.998	0.998	0.998	5.61E–16
	BSA	0.998	12.6705	7.4437	4.6237
	MFO	0.998	10.7632	2.4187	2.1847
	ALO	0.998	6.9033	1.9871	1.5817
	DA	0.998	3.9683	1.1964	0.6009
	MVO	0.998	0.998	0.998	5.61E–16
	SCA	0.998	2.9821	1.6786	0.9454
	GOA	0.998	0.998	0.998	5.61E–16
	SSA	0.998	2.9821	1.2959	0.6099
	WOA	0.998	10.7632	2.9515	2.9054
F12	HFA	3.07E–04	3.07E–04	3.07E–04	2.19E–19
	FAdFPA	3.07E–04	3.07E–04	3.07E–04	2.19E–19
	FAPS	3.07E–04	3.07E–04	3.07E–04	2.19E–19
	FAGWO	3.07E–04	3.07E–04	3.07E–04	2.19E–19
	GWOCFA	3.07E–04	3.07E–04	3.07E–04	2.19E–19
	PSOGSA	3.83E–04	0.0565	0.0107	0.006
	FA	3.08E–04	4.20E–04	3.25E–04	3.40E–05
	BFA	3.08E–04	7.63E–04	6.00E–04	1.32E–04
	FPA	3.58E–04	7.66E–04	5.97E–04	1.23E–04
	GWO	3.07E–04	0.0204	0.0085	0.0053
	PSO	3.07E–04	0.0204	0.0055	0.0021
	DE	4.32E–04	0.0012	7.46E–04	1.63E–04
	HS	5.51E–04	0.0204	0.0079	0.0052
	CSO	3.10E–04	0.0016	7.84E–04	3.79E–04
	BSA	5.34E–04	0.0226	0.0065	0.0054
	MFO	6.29E–04	0.0204	0.0038	0.0018
	ALO	5.85E–04	0.021	0.0047	0.0022
	DA	5.02E–04	0.0226	0.0069	0.0049

(continued)

Table 4 (continued)

Test functions	Algorithms	Best	Worst	Average	Std. dev.
F12	MVO	4.77E–04	0.0204	0.0076	0.0045
	SCA	3.36E–04	0.0016	0.001	3.75E–04
	GOA	3.13E–04	0.0633	0.0116	0.0079
	SSA	3.08E–04	0.0633	0.0098	0.0033
	WOA	3.13E–04	0.0078	0.0011	8.67E–04
F13	HFA	−1.0316	−1.0316	−1.0316	4.5563E–16
	FAdFPA	−1.0316	−1.0316	−1.0316	4.5563E–16
	FAPS	−1.0316	−1.0316	−1.0316	4.5563E–16
	FAGWO	−1.0316	−1.0316	−1.0316	4.5563E–16
	GWOCFA	−1.0316	−1.0316	−1.0316	4.5563E–16
	PSOGSA	−1.0316	−1.0316	−1.0316	4.5563E–16
	FA	−1.0316	−1.0316	−1.0316	4.5563E–16
	BFA	−1.0316	−1.0316	−1.0316	4.5563E–16
	FPA	−1.0316	−1.0316	−1.0316	4.5563E–16
	GWO	−1.0316	−1.0316	−1.0316	4.5563E–16
	PSO	−1.0316	−1.0316	−1.0316	4.5563E–16
	DE	−1.0316	−1.0316	−1.0316	4.5563E–16
	HS	−1.0316	−1.0316	−1.0316	4.5563E–16
	CSO	−1.0316	−1.0316	−1.0316	4.5563E–16
	BSA	−1.0316	−1.0316	−1.0316	4.5563E–16
	MFO	−1.0316	−1.0316	−1.0316	4.5563E–16
	ALO	−1.0316	−1.0316	−1.0316	4.5563E–16
	DA	−1.0316	−1.0316	−1.0316	4.5563E–16
	MVO	−1.0316	−1.0316	−1.0316	4.5563E–16
	SCA	−1.0316	−1.0316	−1.0316	4.5563E–16
	GOA	−1.0316	−1.0316	−1.0316	4.5563E–16
	SSA	−1.0316	−1.0316	−1.0316	4.5563E–16
	WOA	−1.0316	−1.0316	−1.0316	4.5563E–16
F14	HFA	0.3979	0.3979	0.3979	0
	FAdFPA	0.3979	0.3979	0.3979	0
	FAPS	0.3979	0.3979	0.3979	0
	FAGWO	0.3979	0.3979	0.3979	0
	GWOCFA	0.3979	0.3979	0.3979	0
	PSOGSA	0.3979	0.3979	0.3979	0
	FA	0.3979	0.3979	0.3979	0
	BFA	0.3979	0.3979	0.3979	0
	FPA	0.3979	0.3979	0.3979	0
	GWO	0.3979	0.3979	0.3979	0
	PSO	0.3979	0.3979	0.3979	0
	DE	0.3979	0.3979	0.3979	0
	HS	0.3979	0.3979	0.3979	0
	CSO	0.3979	0.3979	0.3979	0

(continued)

Table 4 (continued)

Test functions	Algorithms	Best	Worst	Average	Std. dev.
F14	BSA	0.3979	0.3979	0.3979	0
	MFO	0.3979	0.3979	0.3979	0
	ALO	0.3979	0.3979	0.3979	0
	DA	0.3979	0.3979	0.3979	0
	MVO	0.3979	0.3979	0.3979	0
	SCA	0.3979	0.3979	0.3979	0
	GOA	0.3979	0.3979	0.3979	0
	SSA	0.3979	0.3979	0.3979	0
	WOA	0.3979	0.3979	0.3979	0
F15	HFA	3	3	3	0
	FAdFPA	3	3	3	0
	FAPS	3	3	3	0
	FAGWO	3	3	3	0
	GWOCFA	3	3	3	0
	PSOGSA	3	3	3	0
	FA	3	3	3	0
	BFA	3	3	3	0
	FPA	3	3	3	0
	GWO	3	3.0002	3	5.03E–05
	PSO	3	3	3	0
	DE	3	3	3	0
	HS	3	3	3	0
	CSO	3	3	3	0
	BSA	3	3	3	0
	MFO	3	3	3	0
	ALO	3	3	3	0
	DA	3	3.0001	3	1.20E–05
	MVO	3	3	3	0
	SCA	3	3.0007	3.0001	1.36E–04
	GOA	3	3	3	0
	SSA	3	3	3	0
	WOA	3	3.0007	3.0001	1.10E–04
F16	HFA	−3.8628	−3.8628	−3.8628	1.3669E–15
	FAdFPA	−3.8628	−3.8628	−3.8628	1.3669E–15
	FAPS	−3.8628	−3.8628	−3.8628	1.3669E–15
	FAGWO	−3.8628	−3.8628	−3.8628	1.3669E–15
	GWOCFA	−3.8628	−3.8628	−3.8628	1.3669E–15
	PSOGSA	−3.8628	−3.8628	−3.8628	1.3669E–15
	FA	−3.8628	−3.8628	−3.8628	1.3669E–15
	BFA	−3.8628	−3.8628	−3.8628	1.3669E–15
	FPA	−3.8628	−3.8628	−3.8628	1.3669E–15
	GWO	−3.8628	−3.8628	−3.8628	1.3669E–15

(continued)

Table 4 (continued)

Test functions	Algorithms	Best	Worst	Average	Std. dev.
F16	PSO	−3.8628	−3.8628	−3.8628	1.3669E–15
	DE	−3.8628	−3.8628	−3.8628	1.3669E–15
	HS	−3.8628	−3.8628	−3.8628	1.3669E–15
	CSO	−3.8628	−3.8628	−3.8628	1.3669E–15
	BSA	−3.8628	−3.8628	−3.8628	1.3669E–15
	MFO	−3.8628	−3.8628	−3.8628	1.3669E–15
	ALO	−3.8628	−3.8628	−3.8628	1.3669E–15
	DA	−3.8628	−3.8628	−3.8628	1.3669E–15
	MVO	−3.8628	−3.8628	−3.8628	1.3669E–15
	SCA	−3.8628	−3.8628	−3.8628	1.3669E–15
	GOA	−3.8628	−3.8628	−3.8628	1.3669E–15
	SSA	−3.8628	−3.8628	−3.8628	1.3669E–15
	WOA	−3.8628	−3.8628	−3.8628	1.3669E–15
F17	HFA	−3.322	−3.322	−3.322	2.24E–15
	FAdFPA	−3.322	−3.322	−3.322	2.24E–15
	FAPS	−3.322	−3.322	−3.322	2.24E–15
	FAGWO	−3.322	−3.322	−3.322	2.24E–15
	GWOCFA	−3.322	−3.322	−3.322	2.24E–15
	PSOGSA	−3.322	−3.1376	−3.2755	0.0607
	FA	−3.322	−3.2031	−3.2507	0.0588
	BFA	−3.322	−3.2872	−3.3021	0.0159
	FPA	−3.322	−3.266	−3.3091	0.0145
	GWO	−3.322	−3.038	−3.2494	0.0912
	PSO	−3.322	−3.2031	−3.2816	0.0569
	DE	−3.322	−3.2473	−3.3199	0.0108
	HS	−3.322	−3.2031	−3.2839	0.056
	CSO	−3.3215	−3.2012	−3.2902	0.0383
	BSA	−3.2998	−2.8517	−3.0735	0.1444
	MFO	−3.322	−3.0867	−3.2194	0.0641
	ALO	−3.322	−3.2003	−3.279	0.058
	DA	−3.322	−2.9201	−3.233	0.106
	MVO	−3.322	−3.1933	−3.242	0.058
	SCA	−3.1746	−1.4568	−2.8877	0.3633
	GOA	−3.322	−3.1796	−3.2744	0.0616
	SSA	−3.322	−3.1555	−3.22	0.0558
	WOA	−3.3211	−2.9849	−3.2352	0.0983
F18	HFA	−10.1532	−10.1532	−10.1532	1.26E–14
	FAdFPA	−10.1532	−10.1532	−10.1532	1.26E–14
	FAPS	−10.1532	−10.1532	−10.1532	1.26E–14
	FAGWO	−10.1532	−10.1532	−10.1532	1.26E–14
	GWOCFA	−10.1532	−10.1532	−10.1532	1.26E–14
	PSOGSA	−10.1532	−2.6305	−5.1565	3.3812

(continued)

Table 4 (continued)

Test functions	Algorithms	Best	Worst	Average	Std. dev.
F18	FA	−10.1532	−2.6829	−8.8964	2.5951
	BFA	−10.1531	−10.1445	−10.1532	0.0014
	FPA	−10.1531	−10.1495	−10.1516	0.0012
	GWO	−10.1531	−2.6303	−9.0447	2.4426
	PSO	−10.1532	−2.6305	−6.1962	3.5019
	DE	−10.1532	−4.6636	−9.9631	0.8438
	HS	−10.1532	−2.6305	−4.6641	3.3046
	CSO	−10.148	−2.6783	−8.1096	2.6988
	BSA	−10.1532	−2.5537	−6.2203	3.4874
	MFO	−10.1532	−2.6305	−6.1342	3.1982
	ALO	−10.1532	−2.6305	−6.1293	3.2014
	DA	−10.1532	−2.6201	−7.0129	2.7744
	MVO	−10.1532	−2.6304	−7.6806	3.0299
	SCA	−5.0337	−0.4965	−1.7471	1.5185
	GOA	−10.1532	−2.6305	−4.141	2.3026
	SSA	−10.1532	−2.6305	−7.7015	3.3676
	WOA	−10.1511	−2.6238	−8.0544	2.7745
F19	HFA	−10.4029	−10.4029	−10.4029	7.18E–15
	FAdFPA	−10.4029	−10.4029	−10.4029	7.18E–15
	FAPS	−10.4029	−10.4029	−10.4029	7.18E–15
	FAGWO	−10.4029	−10.4029	−10.4029	7.18E–15
	GWOCFA	−10.4029	−10.4029	−10.4029	7.18E–15
	FAGWO	−10.4029	−1.8376	−5.4272	3.3512
	GWOCFA	−10.4029	−3.7243	−9.735	2.0239
	PSOGSA	−10.4028	−10.4051	−10.3754	0.0312
	FA	−10.4023	−10.3085	−10.375	0.0304
	BFA	−10.4024	−10.3984	−10.4014	7.36E–04
	FPA	−10.4029	−2.7519	−7.5045	3.499
	GWO	−10.4029	−9.545	−10.3645	0.1431
	PSO	−10.4029	−2.7519	−5.3703	3.3616
	DE	−10.402	−2.7638	−8.3964	2.8284
	HS	−10.4027	−2.7388	−6.516	3.1794
	CSO	−10.4029	−1.8376	−6.8382	3.6494
	BSA	−10.4029	−1.8376	−6.246	3.142
	MFO	−10.4029	−1.8369	−7.3726	3.1526
	ALO	−10.4029	−1.8376	−8.1977	3.0567
	DA	−7.1794	−0.5239	−3.2448	1.7055
	MVO	−10.4029	−1.8376	−5.929	3.4966
	SCA	−10.4029	−1.8376	−8.1932	3.0845
	GOA	−10.4029	−1.8361	−6.4511	3.2374

(continued)

Table 4 (continued)

Test functions	Algorithms	Best	Worst	Average	Std. dev.
F20	HFA	−10.5364	−10.5364	−10.5364	8.97E–15
	FAdFPA	−10.5364	−10.5364	−10.5364	8.97E–15
	FAPS	−10.5364	−10.5364	−10.5364	8.97E–15
	FAGWO	−10.5364	−10.5364	−10.5364	8.97E–15
	GWOCFA	−10.5364	−10.5364	−10.5364	8.97E–15
	PSOGSA	−10.5364	−1.8595	−5.6667	3.7404
	FA	−10.5364	−2.8711	−9.7699	2.3229
	BFA	−10.5364	−10.4534	−10.5329	0.0234
	FPA	−10.5293	−10.4575	−10.5055	0.0229
	GWO	−10.5362	−2.4217	−10.102	1.7584
	PSO	−10.5364	−1.6766	−6.4895	3.8488
	DE	−10.5364	−10.4027	−10.5316	0.0213
	HS	−10.5364	−2.4217	−6.0312	3.7453
	CSO	−10.3956	−2.7923	−7.2906	3.3233
	BSA	−10.5336	−1.6753	−6.7317	3.9575
	MFO	−10.5364	−2.4217	−8.5059	3.3297
	ALO	−10.5364	−1.6766	−6.2825	3.615
	DA	−10.5364	−2.4158	−6.6537	3.1252
	MVO	−10.5364	−2.4273	−8.4828	3.0974
	SCA	−9.4397	−0.9403	−3.4949	1.7251
	GOA	−10.5364	−1.6766	−4.969	3.5748
	SSA	−10.5364	−2.4217	−8.2053	3.3666
	WOA	−10.5351	−1.6741	−6.905	3.1312

can be tackled to present strategies with improved challenges. Fractional chaos is nowadays also a common field of study. GWOCFA can therefore include operators of fractional chaos to improve upon the solution. It is also possible to think of different hybrid combinations with the firefly technique. In recent times, several new algorithms have been devised, such as Equilibrium Optimizer (EO), Political Optimizer (PO), or Marine Predator Algorithm (MPA). Researchers are anticipated to establish their obvious variant hybridizing with FA in the coming years.

However, there are few major disadvantages to the relay hybridization scheme suggested here. The parameter tuning is too vulnerable to these hybrid algorithms. To minimize or optimize objective functionality, the correct choice of parameters in these hybrid schemes is essential. In addition, the algorithms depend primarily on the requirements for termination. Therefore, it is also necessary to choose the appropriate NFEs to achieve the desired outcomes. The functionality of any algorithm together with its demerits can deteriorate, as low-level hybrid topologies have been created.

Table 5 Selected p-values for fixed-dimensional multimodal test functions

Test functions	Algorithms						
F12		PSOGSA	FA	BFA	FPA	GWO	PSO
	HFA	2.8858e–20	2.9029e–20	8.3029e–19	8.0547e–19	1.3469e–20	1.7620e–20
		DE	CSO	BSA	HS	MFO	ALO
		3.2055e–11	8.6170e–19	9.3344e–06	1.8753e–20	5.3066e–06	2.7497e–20
		DA	MVO	SCA	GOA	SSA	WOA
		2.5711e–20	8.2570e–21	3.1494e–20	3.1217e–20	3.1494e–20	3.1494e–20
F18		PSOGSA	FA	BFA	FPA	GWO	PSO
	FAdFPA	2.6202e–11	4.6326e–13	3.2917e–20	3.6404e–06	2.5911e–11	4.5419e–19
		DE	CSO	BSA	HS	MFO	ALO
		3.2994e–20	1.3285e–17	3.2714e–20	2.5398e–04	4.1929e–10	8.5605e–14
		DA	MVO	SCA	GOA	SSA	WOA
		9.7841e–13	4.9074e–04	3.0086e–20	3.1217e–20	1.6576e–18	3.0086e–20
F19		PSOGSA	FA	BFA	FPA	GWO	PSO
	FAGWO	1.0022e–11	1.0289e–11	3.3091e–20	1.5787e–05	1.4535e–08	1.2075e–19
		DE	CSO	BSA	HS	MFO	ALO
		1.2279e–19	2.5813e–11	3.2379e–20	2.5398e–04	3.7461e–07	1.2717e–12
		DA	MVO	SCA	GOA	SSA	WOA
		1.2804e–12	0.0230	3.1217e–20	3.1217e–20	3.2122e–20	3.1171e–20
F20		PSOGSA	FA	BFA	FPA	GWO	PSO
	GWOCFA	1.8009e–10	6.1260e–14	3.3111e–20	7.7031e–06	6.4867e–05	1.1892e–19
		DE	CSO	BSA	HS	MFO	ALO
		3.2810e–20	4.6690e–13	3.2666e–20	0.0018	2.6641e–09	1.7395e–10
		DA	MVO	SCA	GOA	SSA	WOA
		2.6932e–11	0.0955	3.1217e–20	3.1217e–20	3.2379e–20	3.1217e–20

4 Conclusions

This chapter examines five new, high-dimensional as well as fixed-dimensional optimization problems with firefly-based hybrid algorithms. To present these high-performance computing algorithms, the firefly technique was integrated successfully with bacterial foraging, flower pollination, pattern search, and grey wolf optimizer algorithms. To support each hybrid proposition, three types of benchmark functions are employed. The two first sets of test functions, the unimodal and the multimodal functions, are used to verify the exploitation and exploration capabilities of the techniques recommended. In terms of speed and precision, the convergence characteristics show greater promise than the current algorithms. To test the validity of the results, the Kruskal–Wallis analysis was also used. Besides, the integrated methods are used to analyze the third category of test functions consisting of several fixed-dimensional multimodal functions. Even in 50 separate test trials, statistical tests for the error function are taken. The findings are confirmed by a Wilcoxon ranking assessment. The methods proposed can also be used properly to handle both constrained and multi-objective problems in optimization techniques.

References

1. I.B. Aydilek, A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. Appl. Soft Comput. **66**, 232–249 (2018)
2. N. Breslow, A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. Biometrika **57**(3), 579–594 (1970)
3. D. Chakraborty, S. Saha, Dutta, O.: DE-FPA: a hybrid differential evolution-flower pollination algorithm for function minimization, in *2014 International Conference on High Performance Computing and Applications (ICHPCA)* (IEEE, Piscataway, 2014), pp. 1–6
4. J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evolut. Comput. **1**(1), 3–18 (2011)
5. E.D. Dolan, R.M. Lewis, V. Torczon, On the local convergence of pattern search. SIAM J. Optim. **14**(2), 567–583 (2003)
6. A.H. Gandomi, X.S. Yang, S. Talatahari, A.H. Alavi, Firefly algorithm with chaos. Commun. Nonlinear Sci. Num. Simulation **18**(1), 89–98 (2013)
7. S. Ganguli, G. Kaur, P. Sarkar, Model order reduction of continuous time system using hybrid metaheuristic algorithm, in *2016 7th India International Conference on Power Electronics (IICPE)* (IEEE, Piscataway, 2016), pp. 1–5
8. S. Ganguli, G. Kaur, P. Sarkar, A novel hybrid metaheuristic algorithm for model order reduction in the delta domain: a unified approach. Neural Comput. Appl. **31**(10), 6207–6221 (2019)
9. S. Ganguli, G. Kaur, P. Sarkar, Identification in the delta domain: a unified approach via GWOCFA. Soft Comput. **24**(7), 4791–4808 (2020)
10. S. Ganguli, G. Kaur, P. Sarkar, A new hybrid algorithm for identification in the unified delta framework, in *AIP Conference Proceedings*, vol. 2207 (AIP Publishing LLC, Melville, 2020), p. 040002

11. S. Ganguli, G. Kaur, P. Sarkar, S.S. Rajest, An algorithmic approach to system identification in the delta domain using FAdFPA algorithm, in *Business Intelligence for Enterprise Internet of Things* (Springer, Berlin, 2020), pp. 203–211
12. M. Jamil, X.S. Yang, A literature survey of benchmark functions for global optimisation problems. Int. J. Math. Modell. Num. Optim. **4**(2), 150–194 (2013)
13. F. Kang, J. Li, H. Li, Artificial bee colony algorithm and pattern search hybridized for global optimization. Appl. Soft Comput. **13**(4), 1781–1791 (2013)
14. Q.X. Lieu, D.T. Do, J. Lee, An adaptive hybrid evolutionary firefly algorithm for shape and size optimization of truss structures with frequency constraints. Comput. Struct. **195**, 99–112 (2018)
15. S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput. Appl. **27**(4), 1053–1073 (2016)
16. S. Mirjalili, SCA: a sine cosine algorithm for solving optimization problems. Knowledge-Based Syst. **96**, 120–133 (2016)
17. S. Mirjalili, A. Lewis, The whale optimization algorithm. Adv. Eng. Softw. **95**, 51–67 (2016)
18. S. Mirjalili, S.M. Mirjalili, A. Hatamlou, Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Comput. Appl. **27**(2), 495–513 (2016)
19. S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)
20. K.M. Passino, Biomimicry of bacterial foraging for distributed optimization and control. IEEE Control Syst. Mag. **22**(3), 52–67 (2002)
21. R.M. Rizk-Allah, E.M. Zaki, A.A. El-Sawy, Hybridizing ant colony optimization with firefly algorithm for unconstrained optimization problems. Appl. Math. Comput. **224**, 473–483 (2013)
22. O. Roeva, Genetic algorithm and firefly algorithm hybrid schemes for cultivation processes modelling, in *Transactions on Computational Collective Intelligence XVII* (Springer, Berlin, 2014), pp. 196–211
23. R.K. Sahu, S. Panda, S. Padhan, A novel hybrid gravitational search and pattern search algorithm for load frequency control of nonlinear power system. Appl. Soft Comput. **29**, 310–327 (2015)
24. S. Saremi, S. Mirjalili, A. Lewis, Grasshopper optimisation algorithm: theory and application. Adv. Eng. Softw. **105**, 30–47 (2017)
25. E.G. Talbi, A taxonomy of hybrid metaheuristics. J. Heurist. **8**(5), 541–564 (2002)
26. T. Ting, X.S. Yang, S. Cheng, K. Huang, Hybrid metaheuristic algorithms: past, present, and future, in *Recent Advances in Swarm Intelligence and Evolutionary Computation* (Springer, Berlin, 2015), pp. 71–83
27. X.S. Yang, Firefly algorithm, stochastic test functions and design optimisation. Int. J. Bio-Inspired Comput. **2**(2), 78–84 (2010)
28. X.S. Yang, Flower pollination algorithm for global optimization, in *International Conference on Unconventional Computing and Natural Computation* (Springer, Berlin, 2012), pp. 240–249
29. X.S. Yang, *Nature-Inspired Optimization Algorithms* (Elsevier, Amsterdam, 2014)

Consumer Decisions in the Age of the Internet: Filtering Information When Searching for Valuable Goods

David M. Ramsey

Abstract The Internet allows people to access information about a large number of offers almost at the click of a button. For consumers, this has a number of advantages. However, humans have a limited capacity for processing the available information. Hence, consumers often use simple rules of thumb (heuristics) to process this information. Such heuristics allow consumers to choose a good offer, while keeping the search costs low. One such heuristic is the concept of a shortlist, which is useful when searching for a unique valuable good, e.g., a second-hand car or flat. A consumer can find basic information about an offer from the Internet. This information is used to choose a shortlist of offers to inspect more closely, before a final decision is made. This paper gives an overview of recent research on mathematical models of such search processes. These models can be split into three categories: (a) optimization models considering a single decision maker, (b) game theoretic models, and (c) models of group decision procedures. Directions for future research are also considered.

1 Introduction

People are increasingly using the Internet when making important consumer decisions, even when the final purchase is not made online. This is due to the fact that basic information regarding a large number of offers is available at very little cost. For example, suppose an individual wishes to buy a new flat in a large city. Fundamental information, e.g., price, floor space, and location, on a large number of flats can be found very easily via the Internet. Purchasing a flat simply on the basis of such information from the Internet is highly risky. Unless a flat is observed in real life, it is impossible to accurately assess how appropriate an offer is. However,

D. M. Ramsey (✉)
Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wrocław, Poland
e-mail: david.ramsey@pwr.edu.pl

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_13

viewing each flat that meets defined conditions according to size, price, and location may lead to prohibitively large search costs. Hence, in such scenarios, consumers often apply strategies based on constructing a shortlist of offers, i.e., a relatively small set of offers that appear to be attractive from initial information. This article gives an overview of recent research on mathematical models of procedures using shortlists to choose a unique, valuable good from a large set of offers.

Since decision makers (DMs) have limited cognitive abilities, heuristic rules, such as the formation of a shortlist, can be very useful in choosing an appropriate offer while controlling search costs. Heuristics should be adapted to both the process of acquiring information and the cognitive abilities of DMs (see Simon [18, 19], Todd and Gigerenzer [20], as well Bobadilla-Suarez and Love [3]). Shortlists can be successful when basic information can be gained at little cost, while the costs of exhaustive search in terms of time and/or cognitive effort are very high (see Masatlioglu et al. [11] and Lleras et al. [8]). For example, suppose someone is choosing a holiday destination. He/she may select a shortlist of propositions using information from friends and colleagues (see Bora and Kops [4]). Shortlists are also practical when offers can be categorized (Armouti-Hansen and Kops [2]). When a DM is searching for offers described by multiple characteristics on the Internet, filters may be applied by ordering offers according to the traits judged to be the most crucial (see Rubinstein and Salant [17] and Mandler et al. [9]). Kimya [5] describes a similar model where offers assessed the least positively on the basis of a given trait are successively eliminated, in decreasing order of the importance of traits. Such an approach may be interpreted as a procedure that constructs shortlists of ever decreasing size until a final decision is taken.

The models described here differ from Kimya's [5] model, as the search process considered here is split into two stages that have clearly different natures. Search costs are not explicitly considered in Kimya's model, and thus, this model is more appropriate when search costs are low or at least uniform, e.g., search is carried out purely on the Internet. The models presented here assume that search costs in stage one are low but are high in stage two. Such a strategy is often used by an employer looking for a specialist employee. The employer invites written applications via the Internet. The costs of such an invitation and assessing the written applications are assumed to be small. Written applications commonly only give a rough estimate of the abilities of the applicants. Thus, the employer invites the most promising candidates for interview. These interviews are generally costly, as they involve using a set of experts for a relatively long period and the employers pay the travel costs of the interviewees. An important aspect in such procedures is determining an appropriate length for the shortlist (the number of candidates to be interviewed) according to the nature of the information gained at each stage of the process and the costs incurred. Ramsey [13] describes a model of an individual DM searching for a valuable good based on constructing a shortlist. Analytis [1] considers a similar model with two stages of inspection. This approach involves what might be called prioritizing, rather than construction of a shortlist. In the first stage (parallel search), offers are ranked according to an initial signal. In the second stage (sequential search), the DM observes offers sequentially from the most highly ranked to the

least highly ranked and stops when the value of an offer is greater than the reward expected from future search.

Section 2 presents a model of a single DM using a shortlist to choose an offer from a large set of goods. Some extensions of the basic model involving a single DM are considered in Sect. 3. Extensions to problems involving multiple DMs are considered in Sect. 4. Such extensions can be broadly classified into game theoretic models and models of group decision making. Section 5 gives a final summary and directions for future research.

2 The Basic Model

The following model is based on the one described in Ramsey [13]. A DM must choose one of n offers. Firstly, the DM observes in parallel an initial signal of each offer's value. The DM cannot measure these signals precisely but is able to rank these signals according to their attractiveness. Such a ranking will be called the initial ranking. According to this ranking, the DM selects k offers for further inspection, where $1 \leq k \leq n$. The strategy of the DM is defined by the value of k (the length of this shortlist). Secondly, the DM obtains another signal of the value of each offer on the shortlist. The DM then makes his/her final selection of an offer. By assumption, if the DM observes all of the offers in both rounds, then he/she is able to rank these offers on the basis of the two signals combined. This ranking will be called the overall ranking. However, in the second round of inspection, the DM can only compare the k offers on the shortlist with each other. Such a ranking will be called the DM's partial ranking. By assumption, this partial ranking is completely consistent with the overall ranking, i.e., offer i is ranked above offer j in a partial ranking if and only if offer i is ranked above offer j in the overall ranking.

By assumption, the two signals of an offer's value to the DM are realizations from a continuous joint distribution. The pair of signals of an offer's value may be correlated, but the pair of signals associated with offer i is independent of the pair of signals associated with offer j, $i \neq j$. An offer's value is a function of the two signals. Let X_m denote the value of the m-th signal ($m = 1, 2$) and W an offer's value. Given $x > y$, it is assumed that the random variable $W|X_1 = x$ stochastically dominates the random variable $W|X_1 = y$. In addition, an offer's value is increasing in the value of the second signal. Hence, high values of the signals correspond to valuable offers.

The DM's goal is to maximize his/her expected reward from search, defined to be the value of the offer accepted minus the costs of searching. Search costs are split into the costs of initial inspection (round one) and the costs of close inspection of the items on the shortlist (round two). The costs of search in round one, denoted as $c_1(k, n)$, are strictly increasing in both the length of the shortlist and the total number of offers, k and n, respectively. These costs reflect the effort involved in the initial inspection of the offers and forming the shortlist. In addition, by assumption, c_1 is assumed to be convex in k, i.e., $c_1(k, n) - c_1(k-1, n)$ is non-decreasing in k. A cost

function of this form reflects the large cognitive effort necessary to form shortlists of long length. Note that this is a simplification, as when $k = n$ the DM automatically inspects every offer closely. Thus, in this case, the search costs should not consider the costs of controlling the shortlist. By assumption, the costs of searching in round two, denoted as $c_2(k)$, are increasing and convex in k. Note that it may be natural to suppose that these costs are linear in k (when $k \geq 2$, then each successive offer on the shortlist is inspected and only needs to be compared with the most highly ranked of the previously inspected offers). Let $c(k, n) = c_1(k, n) + c_2(k)$ be the overall search costs and $C_k = c(k, n) - c(k-1, n)$ denote the marginal cost of increasing the length of the shortlist from $k-1$ to k.

The form of the function describing search costs is not based on a procedure for constructing the initial shortlist. A procedure for forming a shortlist based on pairwise comparisons is described in Sect. 3. The structure of this procedure can be used to define the search costs incurred. Also, note that a shortlist of length k should simply include the k highest ranked offers on the basis of the initial round of observations. This comes directly from the fact that the reward obtained by selecting from such a set of offers stochastically dominates the reward obtained by choosing from another set of k offers.

2.1 Some Theoretical Results

These results are taken from Ramsey [13]. Let W_i denote the value of the i-th ranked offer based on the initial ranking and V_k denote the value of the offer accepted when the shortlist is of length k. Hence, $V_k = \max_{1 \leq i \leq k}\{W_k\}$. Thus, when $i > j$, then V_i stochastically dominates V_j. Let M_k denote the marginal increase in the expected value of the offer accepted when the length of the shortlist increases from $k-1$ to k, i.e., $M_k = E[V_k - V_{k-1}]$.

Theorem 1 *The marginal increase in the expected value of the offer accepted, M_k, is non-increasing in k.*

Proof By definition,

$$M_k = E[\max\{0, W_k - V_{k-1}\}]; \qquad M_{k+1} = E[\max\{0, W_{k+1} - V_k\}].$$

The fact that $M_k \geq M_{k+1}$ follows directly from the fact that W_k stochastically dominates W_{k+1} and V_k stochastically dominates V_{k-1}. □

The following theorem gives a criterion that the optimal length of the shortlist must satisfy.

Theorem 2 *Suppose that $M_2 > C_2$. The optimal length of the shortlist, k^*, is the largest integer k, such that $k \leq n$ and $M_k > C_k$.*

This theorem follows directly from the fact that C_k is non-decreasing in k and M_k is non-increasing in k. The condition $M_2 > C_2$ ensures that it is better to create a shortlist of length two than automatically accept the highest ranked offer according to the initial ranking. When $k \leq k^*$, it follows that $M_k > C_k$, and when $k > k^*$, then $M_k \leq C_k$. Thus, when $k < k^*$, the DM expects to gain overall by increasing the length of the shortlist, but when $k \geq k^*$ the gains expected from increasing the length of the shortlist are not expected to outweigh the costs. Thus, k^* is the optimal length of the shortlist.

It should be noted that if $M_k = C_k$, then the DM is indifferent between forming a shortlist of length $k-1$ and forming a shortlist of length k. The condition described above assumes that when there is not a unique optimal length of shortlist, then the smallest length from the set of optimal lengths is chosen.

2.2 Some Empirical Results from Simulations

Simulations of the search procedure were carried out using a program written in R under the following model. The pair of signals describing an offer (X_1, X_2) is assumed to come from a bivariate normal distribution. The marginal distribution of X_1 is assumed to be standard normal (i.e., of mean zero and variance one). The coefficient of correlation between X_1 and X_2 is denoted by ρ, and the residual variance of X_2, i.e., the variance in X_2 that is not explained by X_1, is defined to be σ^2. Thus, given X_1, X_2 has a normal distribution with mean ρX_1 and variance σ^2. It follows that the overall variance of the signal X_2 is $\frac{\sigma^2}{1-\rho^2}$. The value of an offer is defined to be $W = X_1 + X_2$. From these assumptions, $E(W) = 0$ and

$$\begin{aligned} Var(W) &= Var(X_1) + Var(X_2) + 2\rho\sqrt{Var(X_1)Var(X_2)} \\ &= 1 + \frac{\sigma^2}{1-\rho^2} + \frac{2\rho\sigma}{\sqrt{1-\rho^2}}. \end{aligned} \tag{1}$$

Simple differentiation indicates that this variance is increasing in ρ. The costs of closer inspection are assumed to be proportional to the residual variance of X_2. On one hand, when ρ increases, the increase in the overall variance of the offer favors more intense search (a longer shortlist). On the other hand, the amount of information about the overall value of an offer given by X_1 also increases. This effect favors shortlists with fewer items. The search costs incurred in round one are $c_1(k, n) = 0.0001(n + k^2)$, and the costs incurred in round two are $c_2(k) = c_0\sigma$, where c_0 is a constant, $c_0 \in \{0.02, 0.05, 0.1\}$. These cost functions reflect the logic that strategies based on shortlists should be successful when the costs of initial observation are low relative to the costs of closer inspection. The costs of close inspection are assumed to be proportional to the standard deviation of the second signal, since under sequential search based purely on the second signal the expected number of offers that are seen when c_0 is fixed is independent of σ (see Ramsey [12].

Table 1 Optimal lengths of shortlists for relatively high costs of close inspection ($c = 0.1$) and correlated signals. The components of the vector in each cell give the optimal thresholds for $\rho = 0, 0.2, 0.4, 0.6$ and 0.8, sequentially

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
$\sigma = 1/5$	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma = 1/3$	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma = 1$	(3, 3, 3, 2, 2)	(3, 3, 3, 3, 2)	(4, 3, 3, 3, 3)	(4, 3, 3, 3, 3)
$\sigma = 3$	(4, 4, 3, 3, 3)	(5, 4, 4, 3, 3)	(5, 4, 4, 3, 3)	(5, 5, 4, 4, 3)
$\sigma = 5$	(5, 4, 4, 3, 3)	(5, 5, 4, 3, 3)	(5, 4, 4, 4, 3)	(5, 5, 4, 4, 3)

These assumptions are made so that changes in the optimal length of the shortlist when σ increases and c_0 is fixed reflect the amount of information contained in the second signal relative to the information contained in the first signal (as σ increases, the importance of the second signal compared to the first signal increases).

The optimal lengths of shortlists described in Table 1 were derived empirically on the basis of 100,000 simulations, based on a program written in R, of the search procedure for each possible length of shortlist ($2 \leq k \leq n - 1$) for each combination of parameters: $n \in \{20, 50, 100, 200\}$, $\sigma \in \{1/5, 1/3, 1, 3, 5\}$ and $\rho \in \{0, 0.2, 0.4, 0.6, 0.8\}$. The relative costs of closer inspection are $c_0 = 0.1$.

The results from these simulations lead to the following conclusions:

1. The optimal length of the shortlist is positively associated with the amount of information given by the second signal, and the residual variance is σ^2.
2. The optimal length of the shortlist is negatively associated with the relative costs of search in the second round.
3. Fixing the residual variance of the second signal, the optimal length of the shortlist is non-increasing in the level of correlation between the two signals, ρ.
4. The optimal length of the shortlist is almost unaffected by changes in the total number of offers, n.
5. When the two signals contain a similar amount of information, $\sigma \approx 1$, shortlists of moderate size (4 or 5) are optimal or close to optimal over a wide range of parameters describing the search costs.

The final two comments above indicate that strategies based on shortlists that are used in practice (e.g., when looking for an employee) are very robust.

3 Extensions to the Basic Model

In this section, we consider three extensions to the basic model. The first extension describes a model for constructing a shortlist based on pairwise comparisons. By assuming that pairwise comparisons in round i have cost c_i, $i \in \{1, 2\}$, this model can be used to derive the overall search costs. The second extension concerns

another aspect of limits on the cognitive abilities of DMs. It is assumed that DMs cannot perfectly compare signals of the value of an offer. These two extensions are described more fully in Ramsey [14]. The third extension considers the formation of a shortlist based on quantitative, multivariate data. This extension is described in detail in [10].

3.1 A Model of Shortlist Formation

Note that the DM does not need to construct a full ranking of the offers on the basis of the first signal, in order to construct a shortlist. Assume that the costs of constructing a shortlist are proportional to the number of pairwise comparisons carried out and the DM applies a two-step heuristic procedure that ensures that the number of pairwise comparisons implemented is close to the minimum required. By assumption, the offers appear in random order in the first round of observations. A complete ranking of the first k offers is created using the optimal procedure for ordering a set of values (as described below). This forms an initial shortlist. From the $k+1$-th offer onwards, the DM first decides whether the current offer should be placed on the present shortlist. If not, then the DM proceeds to the next offer. Otherwise, the current offer replaces the offer ranked k on the present shortlist and is then ranked with respect to the remaining $k-1$ offers on the present shortlist. Once the initial signals have been observed for each of the offers, the present shortlist becomes the official shortlist.

First, consider the procedure for ordering the first k offers. This ordering is formed iteratively by ranking the i-th offer to appear relative to the previous $i-1$ offers, for $i = 2, 3, \ldots, k$. Let T_i denote the expected number of pairwise comparisons necessary to create a full ranking of i offers and E_i the expected number of pairwise comparisons necessary to rank the i-th offer with respect to the previous $i-1$ offers. Thus, $T_k = \sum_{i=2}^{k} E_k$. When $k = 2$, just one pairwise comparison is needed to form the initial shortlist, thus $E_2 = T_2 = 1$. Using the optimal procedure for ordering offers, the current offer is first compared with a median ranked item from the previous $i-1$ offers. After this comparison, the current offer is successively compared to a median offer from the subset of offers it should be compared with, until its position in the ordering has been uniquely defined (see Knuth [6]).

When i is odd, the i-th offer may be initially compared to the presently $\frac{i-1}{2}$-th ranked offer (a median from the previous $i-1$ offers). When comparison is perfect, the i-th offer is ranked more highly than this median offer with probability $\frac{i-1}{2i}$, and it now suffices to compare the current offer with $\frac{i-3}{2}$ others. Otherwise, it suffices to compare the current offer with $\frac{i-1}{2}$ others. Thus, for odd i,

$$E_i = 1 + \frac{i-1}{2i} E_{(i-1)/2} + \frac{i+1}{2i} E_{(i+1)/2}. \tag{2}$$

When i is even, independently of whether the i-th offer is better or worse than the median from the previous $i-1$ offers (ranked $\frac{i}{2}$), after the initial comparison, it suffices to compare the current offer with $\frac{i}{2}-1$ previous offers. Hence, for even i,

$$E_i = 1 + E_{i/2}. \tag{3}$$

After forming the initial shortlist, each new offer is firstly compared with the offer currently ranked k (this offer is labelled D_k). If the new offer is ranked more highly than D_k, then D_k is replaced by the new offer, which is then ranked with respect to the remaining $k-1$ offers currently on the shortlist (using the approach adopted when forming the initial shortlist). Given that comparisons are perfect, for $i = k+1, k+2, \ldots, N$, the initial comparison is always carried out, and with probability k/i, the mean number of additional comparisons made is E_k. It follows that the expected number of comparisons from offer $k+1$ onwards is $U_{k,n}$, where

$$U_{k,n} = n - k + \sum_{i=k+1}^{n} \frac{kE_k}{i}. \tag{4}$$

The expected number of pairwise comparisons overall is $W_{k,n} = T_k + U_{k,n}$. Suppose that each pairwise comparison during the initial inspection costs c_1. It follows that the expected search costs during the first round of inspection are $c_1 W_{k,n}$.

It should be noted that for i slightly greater than k, it may be more efficient to apply a similar procedure to the one used for the first k offers. That is to say, the DM compares the present offer with the median ranked offer from the appropriate set of offers until either it is decided that the current offer should not be placed on the present shortlist or the current offer occupies the appropriate position on the present shortlist. Such a procedure could reduce the expected number of pairwise comparisons implemented. However, this comes at the cost of making the procedure much less intuitive (or difficult to formulate/program).

When the shortlist has been finalized, the offers placed on it are then inspected more closely. By assumption, after this second round of inspection, the DM accepts the offer ranked most highly on the basis of both signals. Hence, after observing the first offer on the shortlist, it suffices to compare each new offer with the presently highest ranked offer. Hence, $k-1$ pairwise comparisons are necessary in the second round of inspection. It follows that the search costs incurred in the second round are $(k-1)c_2$.

Inspection of the form of the overall search costs, $c(k,n) = c_1 W_{k,n} + (k-1)c_2$ indicates that the function c is not always convex in c. For example, numerical calculations give $W_{2,100} \approx 106.4142$, $W_{3,100} \approx 116.5840$, and $W_{4,100} \approx 125.8113$. Hence, $c(3,100) - c(2,100) > c(4,100) - c(3,100)$. This is due to the fact that the procedure for forming a shortlist is relatively efficient when k is an integer power of 2. However, for a large, fixed value of n, numerical calculations indicate that the inequality $c(k,n) - c(k-1,n) \leq c(k+1,n) - c(k,n)$ is satisfied for a large majority of the possible values of k. This indicates that any length of shortlist satisfying

the condition given in Theorem 2 (the smallest k such that marginal gain from increasing k does not exceed the marginal increase in search costs from increasing k) will be at least close to optimal.

Note that in practical problems of this type, the costs of gathering the information required to compare two offers may be much greater than the effort required then to decide which is the better of two offers. For example, if someone is searching for a new flat, the time required to travel to a flat and then observe it will be much greater than the time required to then mentally compare two flats. Hence, one might instead assume that, in addition to the costs of comparison, a cost is incurred for observing each signal. Suppose the costs of observing a signal in round i are b_i, $i \in \{1, 2\}$. We may thus define the overall search costs by $b(k, n)$, where $b(k, n) = c(k, n)+b_1n+b_2k$. Since the additional costs of search under this model are linear in both n and k, this does not affect the convexity (or lack of convexity) of the function determining the overall search costs. Hence, again any length of shortlist satisfying the condition given in Theorem 2 will be close to optimal.

The optimal lengths of the shortlists based on this model show a very similar pattern to those derived under the basic model.

3.2 Errors in Pairwise Comparisons

Since the basic model already assumes that the cognitive abilities of DMs are limited, one natural way of extending the model is to assume that the DM cannot perfectly compare options. This section briefly describes the approach taken in Ramsey [14].

It is assumed that ranks are assigned to offers based on imperfect pairwise comparisons according to the information available. Let $p(x, y)$, where $x \geq y$, be the probability that the DM assesses x to be greater than y. It is assumed that

$$p(x, y) = 1 - \frac{\exp(-r[x - y]/\sigma)}{2}, \tag{5}$$

where $r > 0$ and σ is the standard deviation of the distribution of the signal (or sum of signals, as appropriate). Thus, when x and y have the same value, then the result of the comparison is completely random. When $x - y \to \infty$, the probability of correct comparison tends to 1. The parameter r is a measure of the accuracy of perception. One might alternatively give the probability, p, of correct comparison when the difference between the realizations x and y is equal to the standard deviation of the distribution they come from. Hence, $p = 1 - 0.5e^{-r}$, or equivalently, $r = -\ln(2 - 2p)$. It is assumed that the probability of an error in a pairwise comparison does not depend on the results of other comparisons. Apart from the accuracy of perception, the parameters in this model are the same as those used in the original model. It is assumed that the two signals of the value of an offer are independent, i.e., $\rho = 0$.

In the original model, the order in which the offers on the shortlist are observed has no effect on the expected reward from a given strategy. This is due to the fact that the number of pairwise comparisons in the second round is always k and the offer accepted does not depend on the order in which the offers on the shortlist are observed. However, this is not true when pairwise comparisons are not perfect. Intuitively, given the initial rank of an offer on the shortlist, the later it is observed in the second round of observation, the greater the probability that it is accepted. This implies that the DM should observe the offers on the shortlist in reverse order from the k-th ranked to the highest ranked. This is confirmed by simulations using two protocols. According to one protocol, the offers on the shortlist are observed in reverse order. Based on the other protocol, the i-th ranked offer according to the initial ranking is the i-th to be observed. Table 2 presents the empirically derived optimal lengths of the shortlist based on the reverse order protocol. The components in each vector give the optimal length of the shortlist in order of increasing accuracy of perception (the final entry corresponds to the original model).

The results from these simulations lead to the following conclusions:

1. When search costs in the second round are relatively small and the first signal is at least as important as the second signal, i.e., $\sigma \leq 1$, then the optimal length of the shortlist tends to increase as the probability of error increases. Shortlists of greater length ensure that the probability of potentially attractive offers being omitted from the official shortlist by mistake is significantly reduced.
2. Apart from the cases described immediately above, the optimal length of the shortlist is robust to changes in the accuracy of perception.
3. Low error rates do not have a large impact on the expected reward from search.

It should be noted that when the optimal length of the shortlist is relatively small compared to the number of offers available, then the expected number of pairwise comparisons required in the first round is increasing in the error rate. This results from the fact that the likelihood of placing the current offer on the present shortlist tends to increase with the error rate (this is associated with having to make additional comparisons).

Table 2 Empirically derived optimal lengths of the shortlist for the "in reverse order" protocol. The five results given in each cell correspond to increasing levels of accuracy of perception $p = 0.9, 0.99, 0.999, 0.9999, 1$, respectively

	$c = 0.02$	$c = 0.05$	$c = 0.10$
$\sigma = 1/5$	(5, 4, 4, 4, 2)	(4, 2, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma = 1/3$	(5, 4, 4, 4, 3)	(4, 4, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma = 1$	(8, 7, 7, 7, 6)	(4, 4, 4, 5, 4)	(3, 3, 3, 3, 3)
$\sigma = 3$	(10, 12, 13, 12, 11)	(6, 7, 7, 7, 7)	(4, 4, 4, 4, 4)
$\sigma = 5$	(12, 14, 14, 14, 14)	(7, 8, 8, 8, 8)	(4, 5, 5, 5, 5)

3.3 Forming a Shortlist Based on Multiple Criteria

According to the basic model, the information about the initial offers is reduced to a single variable that indicates the potential attractiveness of an offer. In reality, the initial signal will generally include a number of variables. For example, when searching for a flat, information on the price, size, and location of a flat can be obtained from the Internet. This information can be used to derive a shortlist of offers to be physically viewed according to the preferences of the DM. Instead of forming a shortlist of length k simply by choosing the k offers assessed to be the most attractive according to procedure for multiple criteria decision making, we form an optimal shortlist based on the following two criteria: A shortlist should contain offers that (a) are potentially very attractive to the DM and (b) show diversity in their characteristics. The second criterion can be useful when a DM has little knowledge about a given market. Suppose somebody is looking for a flat when moving to a new city. The benefits obtained from viewing flats with various locations/characteristics may be often greater than those from viewing offers that are all very close to the assumed ideal in terms of location and size. Giving some weight to variation in the offers viewed will probably lead to a more informed final decision.

Here, we briefly describe an algorithm for constructing a shortlist of length k from N offers based on n numeric traits, denoted as $x_1, x_2, \ldots, x_n$. The goal is to construct a list that maximizes a weighted sum of the mean attractiveness scores of the offers on the list and the mean distance between them. Hence, we should define (1) a measure of an offer's attractiveness, (2) a measure of the variety of offers on a shortlist, and (3) a measure of a shortlist's attractiveness [based on (1) and (2)]. For example, suppose a DM wishes to construct a shortlist of flats to view based on: price, floor space, and distance from the city center. It should be noted that distance from the city center is a specific variable, since, e.g., two flats that are the same distance from the city center might be a large distance away from each other. In order to define both the distance of an offer to the city center and the physical distance between offers, we need to have two coordinates specifying the physical location of an offer. For example, we may specify the distance of a flat both to the north and the east of the city center. Negative values of these coordinates indicate that the flat lies to the south and west, respectively, of the city center. An example of such data is given in Table 3.

Denote price, size, location north, and location east by x_1, x_2, x_3, and x_4, respectively. The three traits required to define the attractiveness of an offer are denoted by y_1, y_2, and y_3, where $y_1 = x_1$ is the price, $y_2 = x_2$ is the size, and $y_3 = \sqrt{x_3^2 + x_4^2}$ is the distance from the city center. The variables y_1, y_2, y_3 may be standardized using a linear transformation that maps the minimum value of a variable to zero and the maximum value of a variable to one (see Table 3). Various methods are available for measuring the attractiveness of an offer based on the vector (y_1, y_2, y_3). For example, TOPSIS (see Yoon and Hwang [21]), which is based on the relative distance of an offer from an ideal offer and an "anti-ideal" offer based on

Table 3 Raw data and standardized data (in brackets) describing flats for sale

Number	Price (Euro)	Size (m^2)	Location north (km)	Location east (km)
1	450,000 (0.60)	84 (0.84)	2.5 (0.70)	−1.4(0.00)
2	390,000 (0.00)	68 (0.20)	1.4 (0.48)	−0.9(0.10)
3	440,000 (0.50)	63 (0.00)	−0.8(0.04)	0.6 (0.40)
4	420,000 (0.30)	76 (0.52)	3.1 (0.82)	1.4 (0.56)
5	410,000 (0.20)	88 (1.00)	4.0 (1.00)	3.6 (1.00)
6	490,000 (1.00)	72 (0.36)	−1.00(0.00)	1.2 (0.52)

standardized data. Mariański et al. (2020) adopt simple additive weighting (SAW). Using this approach, each trait describing an offer is assigned an attractiveness score between zero and one. The overall attractiveness score for an offer is given by a weighted average of these individual attractiveness scores, where the weights correspond to the importance ascribed to a particular trait. Using either TOPSIS or SAW, we can obtain an overall attractiveness score for each offer that can vary between zero and one.

Due to the differences in the scales of the variables observed, the distance between offers should be based on standardized values of the observations. Let $\tilde{x}_{i,j}$ be the standardized observation of x_i for offer j. The measure of the variety of offers on shortlist S, $w(S)$, is given by the mean of the distances, d, between the offers on the list. The distance between offer j and m is given by

$$d(j,m) = \left[\sum_{i=1}^{n}(\tilde{x}_{i,j} - \tilde{x}_{i,m})^2\right]^{0.5}. \tag{6}$$

This is the standard Euclidean measure of distance in an n-dimensional space.

The overall attractiveness of a shortlist is assumed to be a weighted average of the mean score of the measures of overall attractiveness (given a weight $1 - v$) and the mean distance between the offers (given a weight v). It should be noted that the maximum distance between offers is $\sqrt{n}$, while the attractiveness score is defined to be in the interval [0, 1]. Hence, when defining an algorithm that is robust to changes in the number of variables observed, one might, e.g., scale the attractiveness scores (multiply them by $\sqrt{n}$).

When no weight is ascribed to the variety of the offers on the shortlist, the optimal shortlist simply includes the k offers with the highest attractiveness scores. The number of possible shortlists is given by

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}. \tag{7}$$

When $v > 0$ and k is small, it is reasonable to find the optimal shortlist by exhaustive calculation. However, since such an algorithm should be able to work online, for larger values of k, one might use the following greedy algorithm:

1. Let $i = 1$. Place the most attractive offer on the current shortlist.
2. Add the offer that maximizes the weighted average of the mean attractiveness and the mean distance between offers.
3. Let $i = i + 1$. If $i = k$, then STOP, otherwise return to 2.

This algorithm is based on the concept of dynamic programming. However, when $v > 0$, the objective function cannot be written as a sum of the values of component parts. Hence, in this case, the shortlist formed is not necessarily optimal according to the given criterion.

This algorithm was tested using a database of nearly 10,000 properties on offer in the city of Wrocław, Poland, to form a shortlist of six offers (data accessed on 14/4/2020 from otodom.pl). Based on SAW, the DM enters minimum and maximum values for each trait. Properties that do not satisfy these criteria are eliminated from consideration. By also entering the ideal value of each trait and the weights of the traits, the algorithm calculates a measure of the attractiveness of each offer. When the weight ascribed to variety, v, was less than 0.5, the algorithm constructed shortlists of properties that were all very highly ranked (in the top 20, where 195 properties satisfied the basic criteria). Future research will concentrate on what values of v should be used over a range of problems and given the DM's level of knowledge regarding a market.

4 Models Involving Multiple Decision Makers

In practice, the search for a valuable resource may involve a number of DMs, e.g., a family looking for a flat. In such situations, two approaches to decision making are often used: game theoretic and procedures for group decision making. As in the basic model, it is assumed that there are two decision points: (a) selecting a shortlist based on initial information and (b) choosing an offer from the shortlist.

Using a game theoretic approach, at each decision point, one or more DMs choose an action and the outcome out of each stage of the game (interpreted as a decision making process) depends on the set of decisions made. In classical game theory, it is assumed that the players choose their actions independently. On the other hand, when a married couple is looking for a flat, they will consult with each other before making a final decision. In a game theoretic framework, this could, however, be interpreted as the couple agreeing on the rules of the game before the search process begins. A Nash equilibrium of such a game is defined by a set of strategies for each DM, such that no DM expects to gain by changing their strategy given that the other DMs do not change their strategy.

Using a procedure for group decision making, at each decision point, each DM either gives an attractiveness score to each offer or ranks the offers. These scores or rankings are then used to define an overall attractiveness score to each offer according to a procedure that is agreed upon before search begins. The k offers with the highest overall attractiveness scores based on the initial round are placed on a

shortlist. After the second round of observation, the offer with the highest overall attractiveness score based on all the information available is chosen.

The game theoretic approach described below is described more fully in Ramsey [15], while a group decision procedure is considered in Ramsey [16]. It should be noted that DMs may have different preferences (modelled by allowing the ratios between the variances of the signals as observed by the DMs to differ) and different search costs. For simplicity, here we consider a symmetric model in which the relative importance of the signals to two DMs is the same. For a more general model of the DM's preferences and their search costs, see Ramsey [15].

4.1 A Game Theoretic Model

Assume that two DMs, DM1 and DM2, must choose one of the n offers to be used as a commonly held good. They agree to use a procedure based on constructing a shortlist. For convenience, DM1 will be referred to as "he" and DM2 as "she." We consider the following approach to such a problem: DM1 observes the initial signal for each of the offers. Based on these signals, he chooses a shortlist of offers. After closer inspection of the offers on the shortlist, DM2 then selects an offer. Thus, DM1 may be treated as a Stackelberg leader (see Leitmann [7]). The strategy of Player 1 is his choice of the length of the shortlist.

In such problems, the common interest of the DMs needs to be taken into account. This might come from two sources: (a) the two DMs might have correlated (i.e., similar) preferences and (b) one DM might show altruism toward the other DM. Assume that DM1 may show altruism toward DM2, while DM2 only considers her own payoff. It follows that the optimal response of DM2 is to choose the offer on the shortlist that she ranks most highly. By assumption, in the first round of inspection, only DM1 incurs search costs (e.g., DM1 searches via the Internet). Both players incur the same search costs in the second round of inspection. In the first round of search, the cost of a pairwise comparison is assumed to be 0.001. In round two, the cost of a pairwise comparison is assumed to be 0.05.

The relation between the preferences of DMs may lead to a complex correlation structure existing between the signals observed by the DMs. To keep this structure relatively simple, the following assumptions are made:

1. The coefficient of correlation between the two signals describing an offer observed by a single DM is ρ_1 (independently of the DM).
2. The coefficient of correlation between the value of a signal according to the two DMs is ρ_2 (independently of the signal). This is a measure of the coherence of the DMs' preferences.
3. For a given value of the initial signal as observed by a DM, the value of the second signal as observed by this DM is conditionally independent of the value of the initial signal as observed by the other DM.

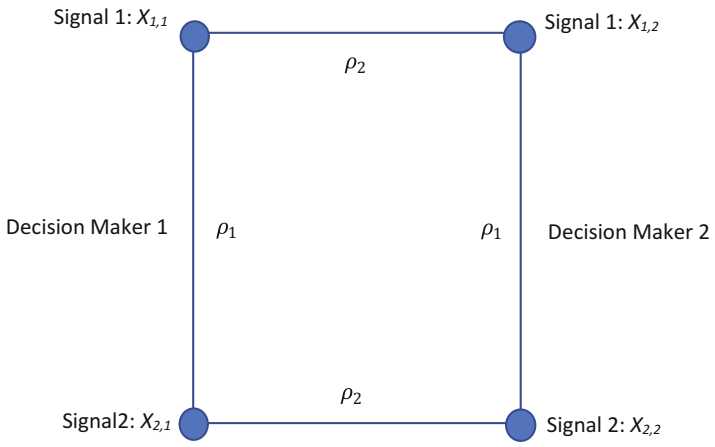


Fig. 1 Structure of the correlations between the signals observed by the decision makers

4. Analogously, for a given value of the second signal as observed by a DM, the value of the initial signal as observed by this DM is conditionally independent of the value of the second signal as observed by the other DM.

Let $X_{i,j}$ denote the value of the i-th signal as observed by the j-th DM, and let $\mathbf{X} = (X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2})$ be the set of signals of the value of an offer as observed by the two DMs. The correlation structure for these signals is illustrated in Fig. 1.

The correlation matrix describing the associations between these signals is

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_1\rho_2 \\ \rho_1 & 1 & \rho_1\rho_2 & \rho_2 \\ \rho_2 & \rho_1\rho_2 & 1 & \rho_1 \\ \rho_1\rho_2 & \rho_2 & \rho_1 & 1. \end{pmatrix} \tag{8}$$

As in the original model, the variance of the initial signal is assumed to be equal to one and the residual variance of the second signal is equal to σ. The overall value of an offer to a DM is assumed to be the sum of the two signals observed (independently of the DM).

The level of altruism shown by DM1 toward to DM2 is denoted by α. When DM1 obtains a payoff of y and DM2 obtains a payoff of z, the utility of DM1 is given by $u_1 = (1 - \alpha)y + \alpha z$. By assumption, $0 \leq \alpha \leq 0.5$. Here, $\alpha = 0$ corresponds to DM1 being economically rational (i.e., he is only interested in his own payoff). On the other hand, $\alpha = 0.5$ corresponds to DM1 assigning the same weight on the payoff of DM2 as on his own. By assumption, DM1 chooses the length of shortlist that maximizes his utility given that in the second round DM2 chooses the offer that is most attractive to her (the best response to DM2's decision).

Table 4 Effect of the level of good will shown by DM1 and the coherence of player's preferences on the efficiency and equilibrium length of the shortlist in games where $\sigma = 1$ and $\rho_1 = 0$ (the first value in each cell gives the efficiency and the second the empirically derived equilibrium length of the shortlist)

	$\rho_2 = 0$	$\rho_2 = 1/3$	$\rho_2 = 2/3$
$\alpha = 0$	(0.4002, 1)	(0.5433, 1)	(0.7996, 2)
$\alpha = 0.25$	(0.4895, 2)	(0.6457, 2)	(0.8360, 3)
$\alpha = 0.5$	(0.5257, 4)	(0.6904, 5)	(0.8501, 6)

The results from simulations indicate that the altruism shown by DM1 is implicit in his choice of the length of the shortlist. Intuitively, the equilibrium length of the shortlist is non-decreasing in α, i.e., DM1 shows good will by giving DM2 a range of choices. The equilibrium length of the shortlist also tends to increase as ρ_2 (a measure of the coherence of the player's choices). For the model of symmetric preferences presented here, when $\rho_2 = 1$, the problem reduces to one in which DM1 is a single decision maker, since the players give the same assessments of the attractiveness of the offers. On the other hand, when $\rho_2 = \alpha = 0$, there is no common interest between the DMs. In this case, DM1 acts as a form of dictator, since the length of the shortlist at equilibrium is one (i.e., DM1 chooses an offer on the basis of the initial signal).

In order to state when such a search procedure might be effective, one should look at the relative efficiency of the search procedure at equilibrium compared to the optimization problem with a single DM. One may define this efficiency as the mean of the payoffs obtained by the DMs divided by the optimal expected payoff of a single DM in the corresponding optimization problem (Table 4).

The results from the simulations indicate that such a procedure is effective when the preferences of the DMs are coherent, particularly when DM1 shows altruism toward DM2. Allowing the relative importance of signals to vary according to the DM, such procedures are particularly effective when the DM1 places more weight on the initial signal and DM2 places more weight on the second signal. However, in order to assess when such a search procedure is likely to be used, one should consider other procedures for first choosing a shortlist and then making the final selection. For example, in the first round, both DMs could place a number of offers on the shortlist and then some procedure is used to make the final selection given the rankings of the offers on the shortlist constructed by the DMs.

The assumption that DM1 shows altruism toward DM2, but DM2 does not show altruism toward DM1 might seem problematic. However, within the framework of the model, the level of altruism shown by DM1 is naturally reflected in the equilibrium length of the shortlist. On the other hand, in order to show DM1's good will, DM2 must have information about how DM1 assesses the offers. This would require an extension of the model.

4.2 Group Decision Procedures

Ramsey [16] considers a problem in which a group of m DMs must together choose one offer from N. On the basis of an initial signal, the DMs construct a shortlist of k offers to investigate more closely. After closely observing the offers on the shortlist, the DMs make their final selection. It is assumed that the overall attractiveness of an offer to the group is measured by a function of the ranks ascribed to that offer by the DMs individually. To adapt the concept of a group decision procedure to the shortlist heuristic, the following components of an appropriate decision rule are required:

1. the assessment function $g_1(r_1, r_2, \ldots, r_m)$ measuring the overall attractiveness of an offer based on initial information, where r_i is the rank ascribed to the offer by the i-th decision maker in the first round of inspection;
2. the length of the shortlist to be used, k;
3. the assessment function $g_2(s_1, s_2, \ldots, s_m)$ measuring the overall attractiveness of an offer on the shortlist based on all the information gained in both rounds, where s_i is the rank ascribed to the offer by the i-th decision maker in the second round of inspection.

In addition, the overall goal of the group should be defined. Here, we consider two possibilities. By definition, the functions g_1 and g_2 are non-increasing in each of their arguments. The k offers with the largest values of g_1 are selected to be on the shortlist. After close inspection of the offers on the shortlist, the group selects the offer with the largest value of g_2. The simplest to use assessment functions are symmetric and additive, as defined below.

Definition An assessment function $g(r_1, r_2, \ldots, r_m)$ is symmetric and additive when there exists a function g_c such that

$$g(r_1, r_2, \ldots, r_m) = \sum_{i=1}^{m} g_c(r_i).$$

Thus, the overall measure of attractiveness may be interpreted as a sum of the attractiveness measures ascribed by the individual DMs (based on the function g_c). This overall measure is independent of the labelling of the DMs (i.e., switching the assessments of any two players never has any effect on the decisions made at any stage). The function g_c is called the inducing function.

The three types of inducing functions described below are natural within this framework:

1. Linear: $g_c(r) = N_0 - r$, where N_0 is the number of offers currently under consideration.
2. Exponential: $g_c(r) = \alpha^{r-1}$, where $0 < \alpha < 1$.
3. Hyperbolic: $g_c(r) = \frac{1}{1+\beta(r-1)}$, where $\beta > 0$.

Note that a wide range of linear inducing functions are admissible for the problems considered here (the only requirement is that an inducing function is decreasing in r). However, it can be easily shown that all of these functions are equivalent, since maximization of the assessment function always reduces to minimizing the sum of the ranks ascribed by the DMs.

The choice of the inducing function should take into account whether it is assumed to be better that at least some of the DMs are very happy or that all the DMs are relatively happy.

In the first case, it would be more natural to use a convex inducing function (the overall attractiveness is larger when the ranks ascribed to an offer by two DMs are 1 and $r-1$ than when both DMs ascribe a rank of $\frac{r}{2}$, where r is an even number such that $r \geq 4$). In the second case, it would be more natural to use a concave inducing function (the overall attractiveness is lower when the ranks ascribed to an offer by two DMs are 1 and $r-1$ than when both DMs ascribe a rank of $\frac{r}{2}$). The exponential and hyperbolic functions given above are both convex.

The linear inducing function can be generalized to the following family of inducing functions: $g_c(r) = (N_0 - r)^\gamma$, where $\gamma > 0$. When $\gamma > 1$, this function is convex and when $\gamma < 1$, this function is concave.

Simulations of decision making procedures with two DMs were used to see what types of decision rule based on this family of inducing functions are best adapted to the following goals (as defined by the DMs). For a given inducing function, the optimal length of the shortlist was found empirically by simulation when 100 offers were available (the length of the shortlist was allowed to vary from one to twenty).

1. To maximize the sum of the payoffs of the DMs.
2. To maximize the minimum of the payoffs of the DMs.

The structure of the correlations between the signals as observed by the DMs is the same as that used for the game theoretic model. These simulations confirm the intuition given above that concave inducing functions are best adopted to maximizing the minimum payoff of the two DMs. However, the form of the inducing rule had only a very small effect on either the expected value of the sum of the payoffs obtained or the minimum of the payoffs obtained. In addition, given the inducing function used, the optimal lengths of the shortlist according to these two criteria are always very similar (most often, they are equal or the length of the shortlist that maximizes the minimum is one greater than the length of the shortlist maximizing the sum of the payoffs). Based on these results, the use of linear inducing functions can be recommended, since such a rule is both simple to implement and intuitive.

Table 5 gives results regarding the optimal length of the shortlist in a problem with two DMs using a linear inducing function ($\gamma = 1$) according to the correlation between the two signals associated with an offer (ρ_1) and the coherence of the DMs' preferences (ρ_2). As for the game theoretic model, the efficiency measures give the ratio of the mean payoff of the DMs to the payoff of an individual DM in the corresponding optimization problem. Again, the efficiency of such procedures

Table 5 Empirically derived optimal expected sum of rewards to two DMs (first entry), optimal length of shortlist (second entry), and efficiency of the group decision procedure (third entry) for search procedures with two DMs when 100 offers are available

	$\rho_2 = 0$	$\rho_2 = 1/3$	$\rho_2 = 2/3$
$\rho_1 = 0$	(4.7435, 6, 0.6040)	(5.8393, 6, 0.7436)	(6.8445, 7, 0.8716)
$\rho_2 = 1/3$	(6.7938, 5, 0.6296)	(8.2035, 5, 0.7603)	(9.4978, 6, 0.8802)
$\rho_3 = 2/3$	(10.0625, 4, 0.6460)	(12.0068, 4, 0.7709)	(13.7753, 4, 0.8844)

is clearly increasing in the coherence of the DMs preferences and also positively associated with the correlation between the two signals of an offer's value.

5 Conclusion

This paper has given an overview on recent work regarding models of decision making using the shortlist heuristic. The shortlist heuristic can be a valuable tool when some information about the large number of offers available is available at very little cost, while extra information is required to make an informed decision.

In the basic model, there are a fixed number of offers and the search costs are convex in the length of the shortlist. Under these assumptions, the optimal length of the shortlist is the smallest value for which the marginal gain from increasing the length of the shortlist (in terms of the increase in the expected value of the offer finally accepted by increasing the length of the shortlist) is greater than the marginal increase in the search costs from increasing the length of the shortlist. Empirical results show that the optimal length of the shortlist is very robust to changes in the total number of offers available. Also, when the amount of information gained from both signals is very similar, shortlists of moderate length (four to six) are close to optimal over a wide range of parameters determining the search costs.

Several extensions of the original model were considered in which there is a single DM. The first extension is based on a model for controlling the shortlist via pairwise comparisons. Under this procedure, the expected number of comparisons needed is of order $n \ln(n)$, i.e., this procedure does not carry out all the $\frac{n(n-1)}{2}$ pairwise comparisons. Under this model, the search costs are not convex in the length of the shortlist. However, numerical calculations show that the marginal search costs $[c(k, n) - c(k-1, n)]$ have, at least, a tendency to increase in k. Thus, under this model, any length of shortlist that satisfies the optimality condition in the basic model will be at least close to optimal.

The second extension considers the possibility that the pairwise comparisons used to control the shortlist may be imperfect. In the second round of inspection, the DM may choose the order in which the offers on the shortlist are inspected. Also, each successive offer is compared with the one assessed to be best of the previous offer. Under such a regime, it is natural to view the offers on the shortlist from the least attractive (according to the initial signal). Simulations indicate that

when the error rate is low, such a procedure is effective and the optimal length of the shortlist tends to increase slightly (compared to when there are no errors in comparisons). When pairwise comparisons are cheap, but somewhat inaccurate, it may well be more effective to carry out a full set of pairwise comparisons. This will be considered in future research.

The third extension considers the problem of selecting a shortlist when the initial signal gives multivariant quantitative information. This may prove very useful in practical problems, e.g., suggesting a shortlist of flats for sale that should be viewed using information from the Internet. Given such information, a shortlist should contain potentially attractive offers that show variety. An algorithm constructing such a shortlist by maximizing a weighted average of the mean attractiveness of offers and a measure of their variety (mean distance between them) is described. In practice, constructing such a shortlist may well be used as a first step in a three stage procedure. Firstly, quantitative information can be used to form an initial shortlist. Secondly, descriptions of offers often include qualitative information, e.g., photographs. Such information could then be used to form a second shortlist of offers to view in real life. The final choice is made after a close inspection of the offers on this shortlist. Future research will aim to study what weight should be placed on variety (depending on a DM's knowledge of the market) and implement a practical version of the algorithm.

In addition, models of decision process with multiple DMs were also presented. These models can be generally classified into game theoretic approaches and models of group decision making. A game theoretic model was presented in which one DM selects the shortlist and the second DM makes the final decision. Such an approach is practical when the DMs have common interests (i.e., show altruism toward each other and/or have coherent preferences) and one DM places a lot of weight on the initial information, while the other places more weight on the second signal. Future research will consider game theoretic models in which both players are active in both stages of the search process. In many ways, models of group decision making processes may be more realistic, since at each stage of the decision process the expressed preferences of the DMs are combined to define the appropriate decision. The model presented here assumes symmetry between the DMs (i.e., they all weight the signals in the same way and incur the same search costs). Future research will consider processes in which the DMs have different priorities and the possibility of collusion between DMs.

Acknowledgments The author is grateful to funds from the Polish National Science Centre for project number 2018/29/B/HS4/02857, "Logistics, Trade and Consumer Decisions in the Age of the Internet" that facilitated this research.

References

1. P.P. Analytis, A. Kothiyal, K. Katsikopoulos, Multi-attribute utility models as cognitive search engines. Judgm. Decis. Mak. **95**, 403–419 (2014)
2. J. Armouti-Hansen, C. Kops, This or that? Sequential rationalization of indecisive choice behavior. Theory Decis. **84**(4), 507–524 (2018)
3. S. Bobadilla-Suarez, B.C. Love, Fast or frugal, but not both: decision heuristics under time pressure. J. Exp. Psychol. Learn. Mem. Cogn. **44**(1), 24 (2018)
4. A. Borah, C. Kops, Rational choices: an ecological approach. Theory Decis. **86**(3–4), 401–420 (2019)
5. M. Kimya, Choice, consideration sets, and attribute filters. Am. Econ. J. Microecon. **10**(4), 223–247 (2018)
6. D.E. Knuth, *The Art of Computer Programming: Volume 3, Sorting and Searching* (Adison-Wesley, Reading, 1973)
7. G. Leitmann, On generalized Stackelberg strategies. J. Optim. Theory Appl. **26**(4), 637–643 (1978)
8. J.S. Lleras, Y. Masatlioglu, D. Nakajima, E.Y. Ozbay, When more is less: limited consideration. J. Econ. Theory **170**, 70–85 (2017)
9. M. Mandler, P. Manzini, M. Mariotti, A million answers to twenty questions: choosing by checklist. J. Econ. Theory. **147**(1), 71–92 (2012)
10. Mariański, A., Kędziora M., Ramsey D. M., Szczurowski L, On forming shortlists of attractive offers from large databases: the example of purchasing a flat. In: Proceedings of the 35th International Business Information Management Association Conference (IBIMA). Seville, Spain: International Business Information Management Association, pp. 13037–13047 (2020)
11. Y. Masatlioglu, D. Nakajima, E.Y. Ozbay, Revealed attention. Am. Econ. Rev. **102**(5), 2183–2205 (2012)
12. D.M. Ramsey, On a sequential decision process where offers are described by two traits. Mult. Criteria Decis. Mak. **10**, 141–154 (2015)
13. D.M. Ramsey, Optimal selection from a set of offers using a short list. Mult. Criteria Decis. Mak. **14**, 75–92 (2019)
14. D.M. Ramsey, On the effect of errors in pairwise comparisons during search based on a short list, in *Proceedings of the 38th International Conference on Mathematical Methods in Economics, Brno*, ed. by S. Kapounek, H. Vránová, 9–11 Sept 2020
15. D.M. Ramsey, A game theoretic model of choosing a valuable good via a short list heuristic. Mathematics **8**(2), 199 (2020)
16. D.M. Ramsey, Group decision making based on constructing a short list, in *Transactions on Computational Collective Intelligence XXXV*, ed. by N.T. Nguyen, R. Kowalczyk, J. Mercik, A. Motylska-Kuźma (Springer, Berlin, 2020)
17. A. Rubinstein, Y. Salant, A model of choice from lists. Theor. Econ. **1**(1), 3–17 (2006)
18. H.A. Simon, A behavioral model of rational choice. Q. J. Econ. **69**(1), 99–118 (1955)
19. H.A. Simon, Rational choice and the structure of the environment. Psychol. Rev. **63**(2), 129 (1956)
20. P.M. Todd, G. Gigerenzer, Précis of simple heuristics that make us smart. Behav. Brain Sci. **23**(5), 727–741 (2000)
21. K.P. Yoon, C.L. Hwang, *Multiple Attribute Decision Making: An Introduction* (Sage, Thousand Oaks, 1995)

Optimality Conditions for Vector Equilibrium Problems

Ali Farajzadeh and Sahar Ranjbar

Abstract Several optimality conditions for solutions of vector equilibrium problems are presented. Some examples in order to clear the main achievements are provided.

Keywords Separation theorem · Optimality conditions · Equilibrium problem

1 Introduction and Preliminaries

Inspired by the pioneer work of Giannessi [13], the theory of vector equilibrium problems was started during the last decade of the last century. The vector equilibrium problems (for short, VEPs) are among the most interesting and intensively studied classes of nonlinear problems. They include fundamental mathematical problems, namely, vector optimization problems, vector variational inequality problems, the Nash equilibrium problem for vector-valued mappings, and fixed point problems as special cases. A large number of research papers have been published on different aspects of vector equilibrium problems; see, for example, [1–10] and the references therein. There are several possible ways to generalize vector equilibrium problems for set-valued mappings; see, for example, [11, 15, 17] and the references therein. Such generalizations are based on the concepts, namely, weak efficient solutions, efficient solutions, strong efficient solutions, etc., of vector optimization problems. Some of the generalizations of vector equilibrium problems are listed below.

$$\text{Find } x \in K \text{ such that } F(x, y) \subseteq Y \backslash (-\operatorname{int} C), \quad \forall y \in K, \tag{1}$$

$$\text{Find } x \in K \text{ such that } F(x, y) \not\subseteq -\operatorname{int} C, \quad \forall y \in K, \tag{2}$$

A. Farajzadeh (✉) · S. Ranjbar
Department of Mathematics, Razi University, Kermanshah, Iran
e-mail: A.Farajzadeh@razi.ac.ir

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_14

$$\text{Find } x \in K \text{ such that } F(x, y) \cap Y \backslash (-\operatorname{int} C) \neq \emptyset \quad \forall y \in K, \tag{3}$$

$$\text{Find } x \in K \text{ such that } F(x, y) \subseteq C, \quad \forall y \in K, \tag{4}$$

where K is a nonempty set, $F : K \times K \to Y$ is a set-valued mapping with nonempty values, and C is a convex cone in a topological vector space Y with nonempty interior, denoted by $\operatorname{int} C$.

These problems are called generalized vector equilibrium problems (in short, GVEPs).

Let X, Y, and Z be ordered vector spaces and $P \subseteq Y$ and $Q \subseteq Z$ be pointed convex cones. We denote by Y^* and Z^* the algebraic dual spaces of Y and Z, respectively. If A is a nonempty subset of Y, then the generated cone of A is defined as $\operatorname{cone} A = \bigcup_{\lambda \geq 0} \lambda A = \{\lambda a : \lambda \geq 0, a \in A\}$. The algebraic dual cone P^* and strictly dual cone $P^{\#}$ of P are defined as

$$P^* = \{y^* \in Y^* : \langle y^*, p\rangle \geq 0 \text{ for all } p \in P\},$$

and

$$P^{\#} = \{y^* \in Y^* : \langle y^*, p\rangle > 0, \forall p \in P \setminus \{0\}\},$$

where $\langle y^*, p\rangle$ denotes the value of the linear functional y^* at the point p, and 0 denotes the zero vector of the corresponding vector space.

The algebraic interior of A, denoted by $\operatorname{cor} A$, is defined as

$$\operatorname{cor} A = \{a \in A : \forall y \in Y, \exists \delta_0 > 0, \forall \delta \in [0, \delta_0], a + \delta y \in A\}.$$

Let Y be a topological vector space and A be a nonempty subset of Y. Then, the topological interior of A, denoted by $\operatorname{int} A$, is a subset of $\operatorname{cor} A$.

The following lemma provides the equivalence between the topological interior and the algebraic interior of a set under certain conditions.

Lemma 1 ([16]) *Let A be a nonempty convex subset of a topological vector space X such that* $\operatorname{int} A \neq \emptyset$. *Then, the following assertions hold:*

(a) $\operatorname{int} A = \operatorname{cor} A$.
(b) $\operatorname{cl} A = \operatorname{cl}(\operatorname{int} A)$ *and* $\operatorname{int} A = \operatorname{int}(\operatorname{cl} A)$, *where* $\operatorname{cl} A$ *denotes the closure of the set* A.

The following lemma plays a key role in the sequel.

The following result is the main motivation of considering algebraic interior instead of topological interior.

Proposition 1 ([12]) *Let X be a topological vector space. For every discontinuous linear functional f on X, there exists a convex pointed cone C_f in X whose*

topological interior is empty, but its algebraic interior is nonempty, that is, $int C_f = \emptyset$ *and* $cor C_f \neq \emptyset$.

Remark 1 If X is a topological vector space whose topology is not Hausdorff, then there is a discontinuous linear functional on X, see [12].

Also note that if f is a discontinuous linear functional on X, then αf is a discontinuous linear function for each nonzero real number α. Then, the set of discontinuous linear functionals on X is an uncountable set, and so the set of convex cone with empty interior and nonempty algebraic interior is uncountable.

Theorem 1 ([16]) *Let A and B be nonempty convex subsets of a vector space X such that $cor A \neq \emptyset$. Then, $(cor A) \cap B = \emptyset$ if and only if there exist a nonzero linear functional $l \in X^*$ and a real number α such that*

$$l(s) \leq \alpha \leq l(t), \quad \text{for all } s \in A,\ t \in B,$$

and

$$l(s) < \alpha, \quad \text{for all } s \in cor A.$$

The next fact is a direct consequence of the previous theorem.

Corollary 1 *Let A be a nonempty convex subset of a vector space X. Then, $x \notin cor A$ if and only if there exist a nonzero linear functional l and a real number α such that*

$$l(s) \leq \alpha \leq l(x), \quad \text{for all } s \in A$$

and

$$l(s) < \alpha, \quad \text{for all } s \in cor A.$$

Definition 1 Let B be a nonempty convex subset of a vector space Y and P be a cone in Y. The set B is called a base of P if $P = cone B$ and there exists a balanced, absorbent, and convex set V in X such that $0 \notin B + V$.

Let X and Y be vector spaces, K be a nonempty subset of X, P be a pointed convex cone in Y, and $F : K \times K \to Y$ be a set-valued mapping with nonempty values. We consider the following generalized vector equilibrium problems.

Definition 2 (GWVEP) A vector $x \in K$ satisfying

$$F(x, y) \not\subseteq -\text{cor } P$$

for all $y \in K$ is called a weakly efficient solution to the VEP.

Definition 3 (GVEP) A vector $x \in K$ is called a globally efficient solution to the VEP if there exists a pointed convex cone $H \subset Y$ with $P \setminus \{0\} \subset cor\, H$ such that

$$F(x, K) \cap ((-H) \setminus \{0\}) = \emptyset,$$

where $F(x, K) = \bigcup_{y \in K} F(x, y)$.

Definition 4 (HGVEP) A vector $x \in K$ is called a Henig efficient solution to the VEP if there exists an algebraically open set U containing 0 with $U \subset V_B$ satisfying

$$cone F(x, K) \cap (-cor\, P_U(B)) = \emptyset.$$

Definition 5 (SGVEP) A vector $x \in K$ is called a superefficient solution to the VEP if for each algebraically open set V of 0, there exists an algebraic open set U of 0 satisfying

$$cone F(x, K) \cap (U - P) \subset V.$$

Clearly,

$$cone F(x, K) \cap (-cor\, P_U(B)) \subset cone F(x, K) \cap (U - P).$$

2 Optimality Conditions

Let X be a vector space, Y and Z be ordered vector spaces, $P \subseteq Y$ and $Q \subseteq Z$ be pointed convex cones, $K \subseteq X$ be a nonempty set, and $F : K \to Y$ and $G : K \to Z$ be set-valued mappings with nonempty values. Define

$$\langle F(x), y^* \rangle = \{\langle y, y^* \rangle : y \in F(x)\} \quad \text{and} \quad \langle F(K), y^* \rangle = \bigcup_{x \in K} \langle F(x), y^* \rangle.$$

We write

$$F(x) <_P y_0 \quad \text{if and only if} \quad y <_P y_0, \quad \forall y \in F(x),$$
$$F(x) \leq_P y_0 \quad \text{if and only if} \quad y \leq_P y_0, \quad \forall y \in F(x).$$

Let K be a nonempty convex subset of X. A set-valued mapping $F : K \to Y$ is said to be

(a) P-convex if for all $x_1, x_2 \in K$ and all $t \in [0, 1]$,

$$tF(x_1) + (1 - t)F(x_2) \subset F(tx_1 + (1 - t)x_2) + P;$$

(b) P-concave if for all $x_1, x_2 \in K$ and all $t \in [0, 1]$,

$$tF(x_1) + (1-t)F(x_2) \subset F(tx_1 + (1-t)x_2) - P.$$

Remark 2 A set-valued mapping $F : K \to Y$ is P-convex if and only if $F(K) + P$ is convex.

Theorem 2 *Let K be a nonempty convex subset of X. If $F : K \to Y$ is P-convex, $G : K \to Z$ is Q-convex, and the system*

$$\begin{cases} F(x) <_P 0, \\ G(x) <_Q 0 \end{cases}$$

has no solution in K, then there exists a nonzero element $(y^*, z^*) \in P^* \times Q^*$ such that for all $x \in K$,

$$\begin{aligned} &\langle y^*, F(x)\rangle + \langle z^*, G(x)\rangle \geq 0, \\ &\text{that is, } \langle y^*, y\rangle + \langle z^*, z\rangle \geq 0, \quad \text{for all } y \in F(x),\ z \in G(x), \end{aligned}$$

where $F(x) <_P 0$ and $G(x) <_Q 0$ mean that $F(x) \subset -cor\, P$ and $G(x) \subset -cor\, Q$, respectively.

Proof By Remark 2, the sets $F(K) + P$ and $G(K) + Q$ are convex in Y and Z, respectively. Define a set-valued mapping $H : K \times K \to Y \times Z$ by

$$H(e, w) = (F(e) + P) \times (G(w) + Q), \quad \text{for all } (e, w) \in K \times K.$$

Then, $H(K \times K) = (F(K)+P) \times (G(K)+Q)$ is convex, and also by the hypothesis, we have $cor(H(K \times K)) \cap ((-P) \times (-Q)) = \emptyset$. By the separation Theorem 1, there exist a nonzero element $(y^*, z^*) \in (Y^* \times Z^*)$ and a real number α such that

$$(y^*, z^*)(p, q) \leq \alpha \leq (y^*, z^*)(e, w),$$

for all $(p, q) \in (-P) \times (-Q)$, $(e, w) \in H(K \times K)$.

Since P and Q are convex pointed cones, $0_Y \in P$ and $0_Z \in Q$. Hence, by applying the last inequalities, we get $(y^*, z^*)(0, 0) = 0 \leq \alpha \leq (y^*, z^*)(e + 0, w + 0)$, where $(e, w) \in F(K) \times G(K)$. Consequently,

$$\langle y^*, F(x)\rangle + \langle z^*, G(x)\rangle \geq 0.$$

This completes the proof. □

Remark 3 When X, Y, and Z are topological vector spaces, $int\, P \neq \emptyset$, and $int\, Q \neq \emptyset$, then it follows from Lemma 1 that $int\, P = cor\, P$, $int\, Q = cor\, Q$, and so Theorem 2 collapses to Theorem 3.3 in [18] with a new proof. Consequently, Theorem 2 generalizes Theorem 3.3 in [18].

(Assumption C) For all $x \in K$, $F(x, x) = \{0\}$ and $F(x, y)$ is P-convex in y; G is Q-concave, and there exists $x_0 \in K$ such that $G(x_0) \subset cor\, Q$.

By using Theorem 2, we present the following result which is a set-valued version of Theorem 3.1 in [14].

Theorem 3 *Suppose that the Condition C is satisfied and* $cor\, P \neq \emptyset$. *If* $x \in K$ *is a solution of (GWVEP), then there exists* $(p^*, q^*) \in P^* \setminus \{0\} \times (-Q)^*)$ *such that* $\langle q^*, G(x)\rangle = \{0\}$ *and*

$$\langle p^*, F(x, x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y \in K} \left[\langle p^*, F(x, y)\rangle + \langle q^*, G(y)\rangle\right].$$

The converse is true when the range of G *is a subset of* Q, *that is,* $G(K) \subseteq Q$.

Proof Assume that $x \in K$ is a solution of GWVEP. Then, for any $y \in K$, we have $F(x, y) \not\subseteq -cor\, P$. Therefore, the system

$$\begin{cases} F(x, y) <_P 0, \\ -G(y) <_Q 0 \end{cases}$$

has no solution in K. Then, by Theorem 2, there exists a nonzero element $(p^*, q^*) \in (P \times (-Q))^* = P^* \times (-Q)^*$ such that

$$\left\langle p^*, F(x, y)\right\rangle + \left\langle q^*, G(y)\right\rangle \geq 0. \tag{5}$$

We claim that $p^* \neq 0$. Otherwise, if $p^* = 0$ in the last inequality, then $\langle q^*, G(y)\rangle \geq 0$ for all $y \in N$. It follows from the hypothesis that there exists $k \in N$ such that $G(k) \subset cor\, Q$. Since $-q^* \in Q^* \setminus \{0\}$, we have $\langle -q^*, G(k)\rangle > 0$, which implies that $\langle q^*, G(k)\rangle < 0$. This is contradicted by $\langle q^*, G(y)\rangle \geq 0$ for all $y \in N$. Thus, $p^* \neq 0$. Consequently, by setting $y = x$, for all $x \in K$, in (5), we obtain $\langle q^*, G(x)\rangle \geq 0$. Since $x \in K$ and $q^* \in (-Q)^*$, $\langle -q^*, G(x)\rangle \geq 0$. Thus,

$$\langle q^*, G(x)\rangle = 0. \tag{6}$$

It follows from (5) and (6) and $F(x, x) = \{0\}$ that

$$\langle p^*, F(x, x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y \in K} \left[\langle p^*, F(x, y)\rangle + \langle q^*, G(y)\rangle\right].$$

Conversely, assume that $x \in K$, $p^* \in P^* \setminus \{0\}$ and $q^* \in -Q^*$ with $\langle q^*, G(x)\rangle = 0$ and

$$\langle p^*, F(x, x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y \in K} \left[\langle p^*, F(x, y)\rangle + \langle q^*, G(y)\rangle\right]. \tag{7}$$

We now show that $x \in K$ is a solution of GWVEP. Suppose, on the contrary, that $x \in K$ is not a solution of GWVEP. Then, there exists $y_0 \in K$ such that $F(x, y_0) \subseteq$

$-cor\, P$. Hence, it follows from (7) that $\langle q^*, G(y_0)\rangle \leq 0$, which is contradicted to $G(K) \subseteq Q$. This completes the proof. □

Theorem 4 *Assume that the Condition C is satisfied and that P has a base B. Then, $x \in N$ is a solution of HGVEP if and only if there exists $(p^*, q^*) \in P^{\Delta}(B) \times -Q^*$ such that $\langle q^*, G(x)\rangle = 0$ and*

$$\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle = \min_{y\in K}\left[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle\right].$$

Proof Assume that $x \in N$ be a solution of HGVEP. Then, there exists an algebraically open set U containing 0 with $U \subset V_B$ such that $F(x,y) \cap -cor\, P_U(B) = \emptyset$. We replace the cone $P_U(B) = cone(U + B)$ by the cone P in Theorem 2. Then, $F(x,y)$ is $P_U(B)$-convex, and also the following system:

$$\begin{cases} F(x,y) \subset -cor\, P_U(B), \\ -G(y) \subset -cor\, Q \end{cases}$$

has no solution in K.

By the similar argument as in the proof of Theorem 3, we obtain that there exists $(0,0) \neq (p^*, q^*) \in P_U^*(B) \times (-Q)^* \subset P^{\Delta}(B) \times (-Q)^*$ such that

$$\langle p^*\ ,\ F(x,y)\rangle + \langle q^*\ ,\ G(y)\rangle \geq 0, \quad \text{for all } y \in K. \tag{8}$$

Taking $y = x$ in (8), we obtain $\langle q^*\ ,\ G(x)\rangle \geq 0$. Since $x \in N$, $q^* \in (-Q)^*$, we have $\langle q^*\ ,\ G(x)\rangle = 0$. From this and (8) and $F(x,x) = \{0\}$, we have

$$\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y\in K}\left[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle\right].$$

Conversely, let $x \in N$, and suppose that there exists $(p^*, q^*) \in P^{\Delta}(B) \times -Q^*$ such that $\langle q^*, G(x)\rangle = 0$ and

$$\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y\in K}\left[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle\right].$$

We show that x is a solution of HGVEP, that is, there exists some algebraically open set U containing $\{0\}$ with $U \subset V_B$ such that $F(x,N) \cap (-cor\, P_U(B)) = \emptyset$. Suppose on the contrary that this does not hold; that is, for all algebraically open set U containing 0 with $U \subset V_B$, we have

$$F(x,N) \cap (-cor\, P_U(B)) \neq \emptyset. \tag{9}$$

Since $p^* \in P^{\Delta}(B)$, by Lemma 1, there exists an algebraically open set $W \subset V_B$ such that $p^* \in (P_W(B))^* \setminus \{0\}$, and by (9), there exist $y_U \in N$, $y_W \in F(x, y_U)$, and $y_W \in -cor\, P_W(B)$. Therefore, $\langle p^*, y_W\rangle < 0$. Since $y_U \in N$, $q^* \in -Q^*$, we have $\langle q^*, G(y_U)\rangle \leq 0$. Hence, we have

$$\begin{aligned}\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle &= \{0\} \equiv 0\\ &= \min_{y\in K}[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle]\\ &\leq \langle p^*, y_W\rangle + \langle q^*, G(y_U)\rangle < 0,\end{aligned}$$

a contradiction. Hence, x is a solution of HGVEP. □

Theorem 5 *Assume that the Condition C is satisfied and that P has a base B. Then, $x \in N$ is a globally efficient solution of GVEP if and only if there exists $(p^*, q^*) \in P^{\#} \times -Q^*$ such that $\langle q^*, G(x)\rangle = 0$ and*

$$\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y\in K}\left[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle\right].$$

Proof Suppose that $x \in N$ is a globally efficient solution of VEP. Then, there exists a pointed convex cone $H \subset Y$ such that $P \setminus \{0\} \subset cor\, H$ and $F(x,N) \cap ((-H) \setminus \{0\}) = \emptyset$. We replace the cone H by the cone P in Theorem 2. We have that $F(x,y)$ is H-convex in Y and that the system

$$\begin{cases} F(x,y) \subset -cor\, H,\\ -G(y) \subset -cor\, Q\end{cases}$$

has no solution in K. By similar argument as in the proof of Theorem 3, there exists $(p^*, q^*) \in P^{\#} \times -Q^*$ such that

$$\left\langle p^*\ ,\ F(x,y)\right\rangle + \left\langle q^*\ ,\ G(y)\right\rangle \geq 0, \quad \text{for all } y \in K. \tag{10}$$

Taking $y = x$ in (10), then $\langle q^*\ ,\ G(x)\rangle \geq 0$. Since $x \in N$ and $q^* \in -Q^*$, we obtain $\langle q^*, G(x)\rangle \leq 0$ together with $\langle q^*, G(x)\rangle = 0$. From $F(x,x) = \{0\}$, $\langle q^*, G(x)\rangle = \{0\}$, and (10), we get

$$\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle = \{0\} \equiv 0 = \min_{y\in K}\left[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle\right].$$

Conversely, let $x \in N$, and suppose that there exists $p^* \in P^{\#}, q^* \in -Q^*$ such that $\langle q^*, G(x)\rangle = \{0\}$ and

$$\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle = \min_{y\in K}\left[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle\right]. \tag{11}$$

We show that x is a globally efficient solution of VEP, that is, there exists a pointed convex cone H such that $P \setminus \{0\} \subset cor\, H$ and

$$F(x,N) \cap ((-H) \setminus \{0\}) = \emptyset. \tag{12}$$

We give

$$H_0 = \{y \in Y : \langle y^*, y\rangle > 0\} \cup \{0\}.$$

Then, we have $P \setminus \{0\} \subset cor\, H_0$, and H_0 is a pointed convex cone. By (12), there exists $y_N \in N$, $y_{H_0} \in F(x, y_N)$, $y_{H_0} \in -H_0$.

By the definition of H_0, we have

$$\langle p^*, y_{H_0}\rangle < 0. \tag{13}$$

Notice that $y_N \in N$, $G(y_N) \subset Q$, we have

$$\langle q^*, G(y_N)\rangle \leq 0. \tag{14}$$

From (13) and (14), we obtain

$$\langle p^*, y_{H_0}\rangle + \langle q^*, G(y_N)\rangle < 0.$$

By (11), we have

$$\begin{aligned}\langle p^*, F(x,x)\rangle + \langle q^*, G(x)\rangle &= \{0\} \equiv 0\\ &= \min_{y\in K}[\langle p^*, F(x,y)\rangle + \langle q^*, G(y)\rangle]\\ &\leq \langle p^*, y_{H_0}\rangle + \langle q^*, G(y_N)\rangle < 0,\end{aligned}$$

a contradiction. Hence, x is a globally efficient solution of VEP.

□

We remark that all above results remain valid when the mappings P and Q are vector-valued.

References

1. Q.H. Ansari, I.V. Konnov, J.C. Yao, Characterizations of solutions for vector equilibrium problems. J. Optim. Theory Appl. **113**, 435–447 (2002)
2. Q.H. Ansari, A.P. Farajzadeh, S. Schaible, Existence of solutions of strong vector equilibrium problems. Taiwan. J. Math. **16**, 165–178 (2012)
3. M. Bianchi, N. Hadjisavvas, S. Schaible, Vector equilibrium problems with generalized monotone bifunctions. J. Optim. Theory Appl. **92**(3), 527–542 (1997)
4. A.P. Farajzadeh, A. Amini-Harandi, On the generalized vector equilibrium problems. J. Math. Anal. **344**, 999–1004 (2008)
5. A.P. Farajzadeh, J. Zafarani, Equilibrium problems and variational inequalities in topological vector spaces. Optimization **59**, 485–499 (2010)
6. A.P. Farajzadeh, J. Zafarani, Vector F-implicit complementarity problems in topological vector spaces. Appl. Math. Lett. **20**(10), 1075–1081 (2010)

7. A.P. Farajzadeh, A. Amini-Harandi, M.A. Noor, On the generalized vector F-implicit complementarity problems and vector F-implicit variational inequality problems. Math. Commun. **12**(2), 203–211 (2007)
8. A.P. Farajzadeh, M.A. Noor, K.I. Noor, Vector nonsmooth variational-like inequalities and optimization problems. Nonlinear Anal. Theory Methods Appl. **71**, 3471–3476 (2009)
9. A.P. Farajzadeh, A. Amini-Harandi, K.R. Kazmi, Existence of solutions to generalized vector variational-like inequalities. J. Optim. Theory Appl. **146**(1), 95–104 (2010)
10. A.P. Farajzadeh, B.S. Lee, S. Plubteing, On generalized quasi-vector equilibrium problems via scalarization method. J. Optim. Theory Appl. **168**, 584–599 (2016)
11. A.P. Farajzadeh, R. Wangkeeree, J. Kerdkaew, On the existence of solutions of symmetric vector equilibrium problems via nonlinear scalarization. Bull. Iran. Math. Soc. **45**, 35–58 (2019)
12. F.J. Garcia-Pacheco, Non-continuous linear functionals on topological vector spaces. Banach. J. Math. Anal. **2**, 11–15 (2008)
13. F. Giannessi, Theorems of alternative, quadratic programs and complementarity problems, in *Variational Inequalities and Complementarity Problems*, ed. by R.W. Cottle, F. Giannessi, J.L. Lions (Wiley, New York, 1980), pp. 151–186
14. X.H. Gong, Optimality conditions for vector equilibrium problems. J. Math. Anal. Appl. **342**, 1455–1466 (2008)
15. S.M. Halimi, A. Farajzadeh, On vector equilibrium problem with generalized pseudo monotonicity. Adv. Math. Finance Appl. **4**(2), 65–74 (2019)
16. J. Jahn, *Vector Optimization Theory, Applications, and Extensions*, 2nd edn. (Springer, Berlin, 2011)
17. M. Khonchaliew, A. Farajzadeh, N. Petrot, Shrinking extra gradient method for pseudo monotone equilibrium problems and quasi-non expansive mappings. Symmetry **11**, 1–18 (2019)
18. L.J. Lin, Optimization of set-valued functions. J. Math. Anal. Appl. **186**, 30–51 (1994)

Strong Pseudoconvexity and Strong Quasiconvexity of Non-differentiable Functions

Sanjeev Kumar Singh, Avanish Shahi, and Shashi Kant Mishra

Abstract In this chapter, we introduce the concept of strong pseudomonotonicity and strong quasimonotonicity of set-valued maps of higher order. Non-differentiable strong pseudoconvex/quasiconvex functions of higher order are characterized by the strong pseudomonotonicity/quasimonotonicity of their corresponding set-valued maps. As a by-product, we solve the open problem (converse part of Proposition 6.2) of Karamardian and Schaible (J. Optim. Theory Appl. 66:37–46, 1990) for the more general case as strong pseudoconvexity for non-smooth, locally Lipschitz continuous functions.

Keywords Generalized convexity · Generalized monotonicity · Clarke generalized subdifferential mappings

2010 Mathematics Subject Classification 90C25, 90C30, 90C99

1 Introduction

The concept of monotone maps was introduced by Minty [10] in 1964. Karamardian [5] extended the concept of monotonicity to strict and strongly monotone maps and also established the relationship between the strongly convex functions and strongly monotone maps. Furthermore, Karamardian and Schaible [6] discussed about seven kinds of monotone maps and established their relationships with corresponding convex functions.

Besides some penalty results for nonlinear programs, Lin and Fukushima [8] introduced the concept of strongly convex functions of order $\sigma > 0$ and established their relationship with strongly monotone maps of order $\sigma > 0$.

S. K. Singh (✉) · A. Shahi · S. K. Mishra
Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_15

It is very natural to see that a function is convex, and then its generalized subgradients are monotone (see [12]). The class of non-differentiable functions plays a very crucial role in the study of generalized convexity and generalized monotonicity. The theory of generalized gradients of non-smooth functions was given by Clarke [1], Rockaffelar [13], and Hiriart-Urruty [4].

Komlósi [7] proposed the relationship of quasi (pseudo, strict pseudo) convexity of lower semicontinuous bifunctions and multifunctions with quasi (pseudo, strict pseudo) monotonicity of its generalized derivatives. In 2003, Fan et al. [3] established the relationships between (strict, strong) convexity and quasiconvexity of non-differentiable functions and (strict, strong) monotonicity and quasimonotonicity of set-valued mappings. In addition to that, Fan et al. [3] also investigated the relationships between (strict, strong, and sharp) pseudoconvexity of non-smooth functions and (strict, strong, and sharp) pseudomonotonicity of set-valued mappings. Recently, Singh et al. [14] presented the first-order characterizations of strong pseudoconvex/quasiconvex functions of higher order. In addition to that, Mishra et al. [11] established the relationships between generalized convex functions and generalized monotone maps in case of semidifferentiability.

Motivated by the work of Karamardian and Schaible [6], Lin and Fukushima [8], and Fan et al. [3], we generalize the concepts of strong convexity/pseudoconvexity/quasiconvexity to strong convexity/pseudoconvexity/quasiconvexity of order $\sigma > 0$ for non-differentiable, locally Lipschitz continuous functions and establish their relationships with strong monotonicity/pseudomonotonicity/quasimonotonicity of order $\sigma > 0$ of set-valued mappings.

2 Preliminaries

Let X be a real Banach space with a norm $\|.\|$ and X^* be its dual space with a norm $\|.\|^*$. Let U be a non-empty open convex subset of X, $F : X \to 2^{X^*}$ be a set-valued mapping from a real Banach space to the family of non-empty subsets of X^*, and $f : X \to \mathbb{R}$ be a non-differentiable real-valued function.

Definition 2.1 ([1, 9]) Let f be locally Lipschitz continuous at a given point $x \in X$ and v be any other vector in X. The Clarke generalized directional derivative of f at x in the direction of v, denoted by $f^0(x; v)$, is defined by

$$f^0(x; v) = \limsup_{y \to x,\, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}.$$

Definition 2.2 ([1, 9]) Let f be locally Lipschitz continuous at a given point $x \in X$ and v be any other vector in X. The Clarke generalized subdifferential of f at x, denoted by $\partial^c f(x)$, is defined by

$$\partial^c f(x) = \{\xi \in X^* : f^0(x; v) \geq \langle \xi, v \rangle, \forall v \in X\}.$$

Lemma 2.1 ([1, 9]) *Let f be locally Lipschitz continuous with a constant L at $x \in X$. Then,*

(a) $\partial^c f(x)$ is a non-empty convex weak-compact subset of X^* and $\|\xi\|_* \leq L$ for every $\xi \in \partial^c f(x)$.*
(b) For every $v \in X$, $f^0(x; v) = max\{\langle \xi, v\rangle : \xi \in \partial^c f(x)\}$.

Lemma 2.2 ([1, 9]) *If f is convex on X and locally Lipschitz continuous at $x \in X$, then $\partial^c f(x)$ coincides with the subdifferential $\partial f(x)$ of f at x in the sense of convex analysis and $f^0(x; v)$ coincides with the directional derivative $f'(x; v)$ for each $v \in X$, where*

$$\partial f(x) = \{\xi \in X^* : f(y) - f(x) \geq \langle \xi, y - x\rangle, \forall y \in X\},$$

$$f'(x; v) = \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

Lemma 2.3 ([1] (Mean Value Theorem)) *Let x and y be points in X, and suppose that f is Lipschitz on an open set X containing the line segment $[x, y]$. Then, $\exists$ a point $u \in (x, y)$ such that*

$$f(x) - f(y) \in \langle \partial^c f(u), x - y\rangle.$$

Definition 2.3 ([2]) A function f is quasiconvex on a convex set X of $\mathbb{R}^n$ if $\forall x, y \in X, \lambda \in [0, 1]$, we have

$$f(x) \leq f(y) \Rightarrow f(\lambda x + (1 - \lambda)y) \leq f(y).$$

Proposition 2.1 ([2]) *Let f be a locally Lipschitz continuous function on X. Then, f is said to be quasiconvex if and only if for any $x, y \in X$ and any $\eta \in \partial^c f(y)$, we have*

$$f(x) \leq f(y) \Rightarrow \langle \eta, x - y\rangle \leq 0.$$

3 Strong Convexity and Monotonicity of Order σ

We collect some definitions related to strong convexity and strong monotonicity of order σ, where $\sigma > 0$ be any positive integer, that is, strong convexity and strong monotonicity of integer order $\sigma \geq 1$ [8].

Definition 3.1 ([8]) A function $f : X \to \mathbb{R}$ is said to be strongly convex of order $\sigma > 0$ on a non-empty open convex subset $X \subseteq \mathbb{R}^n$ if $\exists\ c > 0$ such that for any $x, y \in X$ and any $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - c\lambda(1-\lambda)\|x-y\|^{\sigma}.$$

Definition 3.2 ([8]) F is said to be strongly monotone of order $\sigma > 0$ on X if $\exists$ a constant $\alpha > 0$ such that for any $x, y \in X$ and any $u \in F(x)$, $v \in F(y)$, we have

$$\langle u - v, x - y\rangle \geq \alpha\|x-y\|^{\sigma}.$$

Proposition 3.1 *Let f be a locally Lipschitz continuous function on an open convex subset X. Then, f is strongly convex of order $\sigma > 0$ on X if and only if $\exists\, c > 0$ and $\eta \in \partial^c f(y)$ such that*

$$f(x) - f(y) \geq \langle \eta, x - y\rangle + c\|x-y\|^{\sigma}.$$

Proof Let f be strongly convex function of order $\sigma > 0$ on X. Then, for any $x, y \in X$ and any $\lambda \in [0,1]$, we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - c\lambda(1-\lambda)\|x-y\|^{\sigma}, \tag{1}$$

$$\frac{f(y + \lambda(x-y)) - f(y)}{\lambda} \leq f(x) - f(y) - c(1-\lambda)\|x-y\|^{\sigma}.$$

Taking $\limsup_{\lambda \downarrow 0}$, we have

$$f^0(y, x-y) \leq f(x) - f(y) - c(1-\lambda)\|x-y\|^{\sigma}. \tag{2}$$

Again, $\exists\, \eta \in \partial^c f(y)$ such that $\langle \eta, x-y\rangle \leq f^0(y, x-y)$, and then

$$\langle \eta, x-y\rangle \leq f(x) - f(y) - c(1-\lambda)\|x-y\|^{\sigma},$$

$$f(x) - f(y) \geq \langle \eta, x-y\rangle + c'\|x-y\|^{\sigma}, \qquad c' = c(1-\lambda).$$

Conversely, suppose that $f(x) - f(y) \geq \langle \eta, x-y\rangle + c\|x-y\|^{\sigma}$.

Let $x \neq y \in X$, $\lambda \in [0,1]$, $x_\lambda = y + \lambda(x-y) \in X$ as X is convex.

In particular, $\exists\, \eta_0 \in \partial^c f(x_\lambda)$ such that

$$f(x) - f(x_\lambda) \geq \langle \eta_0, x - x_\lambda\rangle + c\|x - x_\lambda\|^{\sigma}, \tag{3}$$

and

$$f(y) - f(x_\lambda) \geq \langle \eta_0, y - x_\lambda\rangle + c\|y - x_\lambda\|^{\sigma}. \tag{4}$$

Multiplying inequality (3) by λ and (4) by $(1-\lambda)$ and adding them, we obtain

$$\lambda f(x) + (1-\lambda) f(y) - f(x_\lambda) \geq c\lambda(1-\lambda)^{\sigma}\|x-y\|^{\sigma} + c\lambda^{\sigma}(1-\lambda)\|x-y\|^{\sigma}.$$

Consider $[(1-\lambda)^{\sigma-1}+\lambda^{\sigma-1}]$ for $0 < \lambda \leq 2$, $[(1-\lambda)^{\sigma-1}+\lambda^{\sigma-1}] \geq (1-\lambda)+\lambda = 1$, and for $\lambda > 2$, since the real function $\phi(\lambda) = \lambda^{\sigma-1}$ is convex on (0,1), then $[(1-\lambda)^{\sigma-1}+\lambda^{\sigma-1}] \geq (\frac{1}{2})^{\sigma-2}$.

It follows from the above argument that $\exists$ some constant $c' > 0$ independent of x, y, and λ such that

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - c'\lambda(1-\lambda)\|x-y\|^{\sigma}.$$

Therefore, f is strongly convex of order $\sigma > 0$ on X. □

Theorem 3.1 *Let f be a locally Lipschitz continuous function on X. Then, f is strongly convex of order $\sigma > 0$ on X if and only if $\partial^c f$ is strongly monotone of order $\sigma > 0$ on X.*

Proof Let f be strongly convex of order $\sigma > 0$, then for any $x, y \in X$ and $\eta \in \partial^c f(y)$, we have

$$f(x) - f(y) \geq \langle \eta, x-y \rangle + c\|x-y\|^{\sigma}. \tag{5}$$

Interchanging the role of x and y and for any $\xi \in \partial^c f(x)$, we have

$$f(y) - f(x) \geq \langle \xi, y-x \rangle + c\|y-x\|^{\sigma}. \tag{6}$$

Adding inequalities (5) and (6), we get

$$0 \geq \langle \eta - \xi, x-y \rangle + 2c\|x-y\|^{\sigma},$$

$$\langle \xi - \eta, x-y \rangle \geq \beta\|x-y\|^{\sigma}.$$

Therefore, $\partial^c f$ is strongly monotone of order σ on X.

Conversely, suppose that $\partial^c f$ is strongly monotone of order $\sigma > 0$ on X; that is, for any $x, y \in X$, $\exists\, \xi \in \partial^c f(x)$ and $\eta \in \partial^c f(y)$ such that

$$\langle \xi - \eta, x-y \rangle \geq \alpha\|x-y\|^{\sigma}.$$

By the mean value theorem, for any $x \neq y \in X$, $\exists\, z = \lambda x + (1-\lambda)y$ for some $\lambda \in (0, 1)$ and $\exists\, \eta_0 \in \partial^c f(z)$ such that

$$f(x) - f(y) = \langle \eta_0, x-y \rangle = \frac{1}{\lambda}\langle \eta_0, z-y \rangle. \tag{7}$$

Since $\partial^c f$ is strongly monotone of order $\sigma > 0$ on X,

$$\langle \eta_0 - \eta, z-y \rangle \geq \alpha\|z-y\|^{\sigma},$$

for any $z \neq y \in X$.

$$\langle \eta_0, z - y \rangle \geq \langle \eta, z - y \rangle + \alpha \|z - y\|^{\sigma}. \tag{8}$$

Using inequality (8) in inequality (7), we have

$$f(x) - f(y) \geq \frac{1}{\lambda}[\langle \eta, z - y \rangle + \alpha \|z - y\|^{\sigma}],$$

$$f(x) - f(y) \geq \langle \eta, x - y \rangle + \alpha \lambda^{\sigma - 1} \|x - y\|^{\sigma}.$$

Therefore,

$$f(x) - f(y) \geq \langle \eta, x - y \rangle + c\|x - y\|^{\sigma}.$$

Hence, f is strongly convex of order $\sigma > 0$. □

Remark 3.1 Proposition 3.1 and Theorem 3.1 generalize Proposition 3.1 and Theorem 3.4 of Fan et al. [3], respectively, which was given for $\sigma = 2$.

4 Strong Pseudoconvexity and Pseudomonotonicity of Order σ

We introduce the concept of strongly pseudoconvex functions of order $\sigma > 0$ for non-smooth locally Lipschitz continuous functions.

Definition 4.1 Let f be a locally Lipschitz continuous function on X. Then, f is said to be strongly pseudoconvex of order $\sigma > 0$ on X if for any $x, y \in X$ and for any $\eta \in \partial^c f(y)$ $\exists\, \alpha > 0$, we have

$$\langle \eta, x - y \rangle + \alpha \|x - y\|^{\sigma} \geq 0 \Rightarrow f(x) - f(y) \geq 0.$$

Remark 4.1 For $\sigma = 2$, the definition was given by Fan et al. [3].

We introduce the concept of strongly pseudomonotone of set-valued mappings of order $\sigma > 0$ for non-smooth locally Lipschitz continuous functions.

Definition 4.2 F is said to be strongly pseudomonotone of order $\sigma > 0$ on X if for any $x, y \in X$ and any $u \in F(x)$, $v \in F(y)$, $\exists$ a constant $\alpha > 0$, and we have

$$\langle v, x - y \rangle + \alpha \|x - y\|^{\sigma} \geq 0 \Rightarrow \langle u, x - y \rangle \geq 0.$$

Remark 4.2 For $\sigma = 2$, the definition was given by Karamardian and Schaible [6] for real-valued mappings.

We establish the relationship between strong pseudoconvexity of locally Lipschitz continuous functions and strong pseudomonotonicity of set-valued mappings of order $\sigma > 0$, which is the natural generalization of the locally Lipschitz strong pseudoconvex functions given by Fan et al. [3].

Remark 4.3 Fan et al. [3] have left an open problem as the converse of Theorem 4.3, and we prove necessary and sufficient both part for more general class as locally Lipschitz strong pseudoconvex functions of order $\sigma > 0$.

Theorem 4.1 *Let f be a locally Lipschitz continuous function on X. Then, f is strongly pseudoconvex of order $\sigma > 0$ on X if and only if $\partial^c f$ is strongly pseudomonotone of order $\sigma > 0$ on X.*

Proof Let f be strongly pseudoconvex of order $\sigma > 0$ on X, then for any $x, y \in X$ and $\eta \in \partial^c f(y)$ $\exists$ a constant $\alpha > 0$, such that

$$\langle \eta, x - y \rangle + \alpha \|x - y\|^\sigma \geq 0 \Rightarrow f(x) \geq f(y).$$

Since we know that every strongly pseudoconvex function of order $\sigma > 0$ is quasiconvex,

$$f(\lambda x + (1 - \lambda)y) \leq f(x). \tag{9}$$

Also, by the definition of non-smooth quasiconvex function if for any $x, y \in X$ and any $\xi \in \partial^c f(x)$, we have

$$f(\lambda x + (1 - \lambda)y) \leq f(x) \Rightarrow \langle \xi, (\lambda x + (1 - \lambda)y) - x \rangle \leq 0,$$

$$\Rightarrow \langle \xi, x - y \rangle \geq 0.$$

Therefore, we have

$$\langle \eta, x - y \rangle + \alpha \|x - y\|^\sigma \geq 0 \Rightarrow \langle \xi, x - y \rangle \geq 0.$$

Thus, $\partial^c f$ is strongly pseudomonotone of order σ on X.

Conversely, suppose that $\partial^c f$ is strongly pseudomonotone of order $\sigma > 0$, then for any $x, y \in X$ and $\xi \in \partial^c f(x)$, $\eta \in \partial^c f(y)$, $\exists$ a constant $\beta > 0$, such that

$$\langle \eta, x - y \rangle + \beta \|x - y\|^\sigma \geq 0 \Rightarrow \langle \xi, x - y \rangle \geq 0.$$

Equivalently,

$$\langle \xi, x - y \rangle < 0 \Rightarrow \langle \eta, x - y \rangle + \beta \|x - y\|^\sigma < 0. \tag{10}$$

We want to show that f is strongly pseudoconvex of order $\sigma > 0$; that is, for any $x, y \in X$ and $\eta \in \partial^c f(y)$, $\exists$ a constant $\alpha > 0$, and we have

$$\langle \eta, x - y \rangle + \alpha \|x - y\|^\sigma \geq 0 \Rightarrow f(x) \geq f(y). \tag{11}$$

Suppose, on contrary, $f(x) < f(y)$.

By the mean value theorem, $\exists\ z = \lambda x + (1 - \lambda)y$ for some $\lambda \in (0, 1)$ and $\eta_0 \in \partial^c f(z)$, such that

$$f(x) - f(y) = \langle \eta_0, x - y \rangle = \frac{1}{\lambda} \langle \eta_0, z - y \rangle < 0.$$

Since $\partial^c f$ is strongly pseudomonotone of order σ,

$$\langle \eta_0, z - y \rangle < 0 \Rightarrow \langle \eta, z - y \rangle + \beta \|z - y\|^\sigma < 0,$$

$$\langle \eta_0, z - y \rangle < 0 \Rightarrow \langle \eta, x - y \rangle + \beta \lambda^{\sigma - 1} \|x - y\|^\sigma < 0,$$

which contradicts to the left-side inequality of (11).

Hence, $f(x) \geq f(y)$, and f is strongly pseudoconvex of order $\sigma > 0$. □

Remark 4.4 Every strongly monotone map of order $\sigma > 0$ is strongly pseudomonotone of order $\sigma > 0$, but the converse is not necessarily true.

Example 4.1 Let $F : X \to \mathbb{R}$, where $X = [0, 4]$ defined by

$$F(x) = \begin{cases} 2 - x & \text{for } 0 \leq x < 1, \\ 1 & \text{for } 1 \leq x \leq 4. \end{cases}$$

This is an example of strongly pseudomonotone map of order $\sigma > 0$, but not strongly monotone map of order $\sigma > 0$.

5 Strong Quasiconvexity and Quasimonotonicity of Order σ

Definition 5.1 Let f be a locally Lipschitz continuous function on an open convex subset X. Then, f is said to be strongly quasiconvex of order $\sigma > 0$ on X if for any $x, y \in X$ and any $\eta \in \partial^c f(y)\ \exists\ \alpha > 0$, we have

$$f(x) \leq f(y) \Rightarrow \langle \eta, x - y \rangle + \alpha \|x - y\|^\sigma \leq 0.$$

Definition 5.2 F is said to be strongly quasimonotone of order $\sigma > 0$ on X if for any $x, y \in X$ and any $u \in F(x), v \in F(y)\ \exists\ \beta > 0$, we have

$$\langle v, x - y \rangle > 0 \Rightarrow \langle u, x - y \rangle \geq \beta \|x - y\|^\sigma.$$

Theorem 5.1 *Let f be a locally Lipschitz continuous function on X. Then, f is strongly quasiconvex of order $\sigma > 0$ on X if and only if $\partial^c f$ is strongly quasimonotone of order $\sigma > 0$ on X.*

Proof Let f be strongly quasiconvex of order $\sigma > 0$ on X, then for any $x \neq y \in X$ and $\eta \in \partial^c f(y)$, $\exists$ a constant $\alpha > 0$, such that

$$f(x) \leq f(y) \Rightarrow \langle \eta, x - y \rangle + \alpha \|x - y\|^\sigma \leq 0. \tag{12}$$

We have to show that $\partial^c f$ is strongly quasimonotone on X; that is, for any $\xi \in \partial^c f(x)$ and $\eta \in \partial^c f(y)$, $\exists$ a constant $\beta > 0$, such that

$$\langle \eta, x - y \rangle > 0 \Rightarrow \langle \xi, x - y \rangle \geq \beta \|x - y\|^\sigma.$$

As f is strongly quasiconvex, then it is also quasiconvex; that is, for any $\eta \in \partial^c f(y)$, we have

$$\langle \eta, x - y \rangle > 0 \Rightarrow f(x) > f(y).$$

By the definition of strongly quasiconvex function of order $\sigma > 0$, we have

$$f(y) < f(x) \Rightarrow \langle \xi, y - x \rangle + \alpha \|y - x\|^\sigma \leq 0,$$

$$f(y) < f(x) \Rightarrow \langle \xi, x - y \rangle \geq \alpha \|x - y\|^\sigma.$$

Therefore, we have $\langle \eta, x - y \rangle > 0 \Rightarrow \langle \xi, x - y \rangle \geq \alpha \|x - y\|^\sigma$.

Thus, $\partial^c f$ is strongly quasimonotone of order σ.

Conversely, suppose that $\partial^c f$ is strongly quasimonotone of order $\sigma > 0$, then for any $\xi \in \partial^c f(x)$ and $\eta \in \partial^c f(y)$, $\exists$ a constant $\beta > 0$, such that

$$\langle \eta, x - y \rangle > 0 \Rightarrow \langle \xi, x - y \rangle \geq \beta \|x - y\|^\sigma.$$

We want to show that f is strongly quasiconvex of order $\sigma > 0$; that is, $f(x) \leq f(y) \Rightarrow \langle \eta, x - y \rangle + \alpha \|x - y\|^\sigma \leq 0$.

Suppose that $f(x) \leq f(y)$.

By the mean value theorem, $\exists\ z = \lambda x + (1 - \lambda) y$ for some $\lambda \in (0, 1)$ and $\eta_0 \in \partial^c f(z)$, such that

$$f(x) - f(y) = \langle \eta_0, x - y \rangle = \frac{1}{\lambda} \langle \eta_0, z - y \rangle \leq 0.$$

By the use of strongly quasimonotone map, we have

$$\langle \eta_0, y - z \rangle > 0 \Rightarrow \langle \eta, y - z \rangle \geq \beta \|y - z\|^\sigma,$$

$$\langle \eta_0, y - z \rangle > 0 \Rightarrow \langle \eta, y - x \rangle \geq \beta \lambda^{\sigma - 1} \|y - x\|^\sigma,$$

$$\langle \eta_0, y - z \rangle > 0 \Rightarrow \langle \eta, x - y \rangle + \alpha \|x - y\|^{\sigma} \leq 0,$$

Hence, f is strongly quasiconvex of order $\sigma > 0$. □

Remark 5.1 Every strongly quasiconvex function of order $\sigma > 0$ is quasiconvex, but the converse is not always true.

Remark 5.2 The class of quasi-functions is the largest class, so every strongly pseudomonotone map of order $\sigma > 0$ is strongly quasimonotone of order $\sigma > 0$, but it is not always true in the converse case.

Example 5.1 Let $F : X \to \mathbb{R}$, where $X = [-2, 2]$ defined by

$$F(x) = \begin{cases} 0 & \text{for } -2 \leq x < 0, \\ x & \text{for } 0 \leq x < 1, \\ 2x - 1 & \text{for } 1 \leq x \leq 2. \end{cases}$$

This is an example of strongly quasimonotone map of order $\sigma > 0$, but not strongly pseudomonotone map of order $\sigma > 0$.

Acknowledgments The authors are indebted to anonymous referees for valuable comments and suggestions which led to the present improved version as it stands.

The first author is financially supported by CSIR-UGC JRF, New Delhi, India, through reference no. 1272/(CSIR-UGC NET DEC.2016). The second author is financially supported by UGC-BHU Research Fellowship, through sanction letter no. Ref. No./Math/Res/Sept.2015/2015-16/918. The third author is financially supported by the Department of Science and Technology, SERB, New Delhi, India, through grant no. MTR/2018/000121.

References

1. F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley-Interscience, New York, 1983)
2. R. Ellaia, A. Hassouni, Characterization of nonsmooth functions through their generalized gradients. Optimization **22**(3), 401–416 (1991)
3. L. Fan, S. Liu, S. Gao, Generalized monotonicity and convexity of non-differentiable functions. J. Math. Anal. Appl. **279**(1), 276–289 (2003)
4. J.B. Hiriart-Urruty, New concepts in non-differentiable programming. Soc. Math. France **60**, 57–85 (1979)
5. S. Karamardian, Complementarity problems over cones with monotone and pseudomonotone maps. J. Optim. Theory Appl. **18**(4), 445–454 (1976)
6. S. Karamardian, S. Schaible, Seven kinds of monotone maps. J. Optim. Theory Appl. **66**(1), 37–46 (1990)
7. S. Komlósi, Generalized monotonicity and generalized convexity. J. Optim. Theory Appl. **84**(2), 361–376 (1995)
8. G.H. Lin, M. Fukushima, Some exact penalty results for nonlinear programs and mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **118**(1), 67–80 (2003)
9. M.M. Mäkelä, P. Neittaanmäki, *Nonsmooth Optimization, Analysis and Algorithms with Applications to Optimal Control* (World Scientific, River Edge, 1992)

10. G.J. Minty, On the monotonicity of the gradient of a convex function. Pac. J. Math. **14**(1), 243–247 (1964)
11. S.K. Mishra, S.K. Singh, A. Shahi, On monotone maps: semidifferentiable case, in *World Congress on Global Optimization* (Springer, Berlin, 2019), pp. 182–190
12. R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)
13. R.T. Rockafellar, *The Theory of Subgradients and Its Applications to Problems of Optimization* (Heldermann, Berlin, 1981)
14. S.K. Singh, A. Shahi, S.K. Mishra, On strong pseudomonotone and strong quasimonotone maps, in *International Conference on Mathematics and Computing* (Springer, Singapore, 2018), pp. 13–22

Optimality and Duality of Pseudolinear Multiobjective Mathematical Programs with Vanishing Constraints

Jitendra Kumar Maurya, Avanish Shahi, and Shashi Kant Mishra

Abstract In this chapter, we establish necessary and sufficient optimality conditions for a special class of optimization problems called multiobjective mathematical programs with vanishing constraints under pseudolinear assumption. We propose Mond–Weir type dual model for the considered problem and establish usual duality results. Furthermore, we present some examples to validate our results.

Keywords Pseudolinear multiobjective programming · Optimality conditions · Duality · Vanishing constraints

Mathematics Subject Classification (2010) 90C29, 90C33, 90C46

1 Introduction

Mathematical programs with vanishing constraints (MPVCs) are an interesting subclass of nonlinear programming problems. It has many applications in truss topology optimization [1], robot pathfinding problem with logic communication constraints in robot motion planning [11], mixed-integer nonlinear optimal control problems [9], scheduling problems with disjoint feasible regions in power generation dispatch [8], etc.

In most of the cases, feasible region of the MPVC is nonconvex due to natural formation of the constraints. In general, the majority of basic constraint qualifications do not hold for the MPVC, this is why MPVCs are considered as a difficult class of optimization problems. Therefore, the traditional and most basic optimality conditions, that is, Karush–Kuhn–Tucker conditions, are not satisfied. Achtziger and Kanzow [1] proposed some constraint qualifications for the MPVC and obtained first-order stationary conditions. Hoheisel and Kanzow [5] established

J. K. Maurya (✉) · A. Shahi · S. K. Mishra
Department of Mathematics, Institute of Science, Banaras Hindu University, Varanasi, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_16

first-order sufficient optimality conditions as well as second-order necessary and sufficient optimality conditions. Furthermore, Hoheisel and Kanzow [6] derived various stationary conditions under weak constraint qualifications. Hoheisel and Kanzow [7] investigated the Abadie and Guignard type constraint qualifications and obtained necessary and sufficient optimality conditions.

We have to optimize several objectives simultaneously in real-life situations, the so-called multiobjective optimization problems. This chapter focuses on the study of the multiobjective mathematical programs with vanishing constraints for a particular class of functions known as pseudolinear. Pseudolinear functions contain several useful classes of functions. We establish necessary optimality conditions without any constraints qualifications, which is an additional merit of this class. Initially, Kortanek and Evans [12] noticed the existence of some functions that are both pseudoconvex and pseudoconcave. Furthermore, these functions called pseudolinear by Chew and Choo [3] characterize its behavior in optimality sense. For more details on pseudolinearity and its applications, see the monograph by Mishra and Upadhyay [15].

Recently, Mishra et al. [16] formulated Wolfe and Mond–Weir type dual models for MPVC and established many duality results, Benko and Gfrerer [2] proposed an algorithm for solving MPVC, Dussault et al. [4] introduced a new scheme for solving the MPVC, Khare and Nath [10] established an enhanced Fritz John type stationary condition for MPVC, which leads to enhanced M-stationarity under a new and weaker constraint qualification and also local error bound result established under MPVC-generalized quasinormality, and strong efficient S-stationary conditions for multiobjective mathematical programs with equilibrium constraints (MMPECs) have been studied by Zhang et al. [17]. Since MPVCs are closely related to mathematical programs with equilibrium constraints (MPECs), all of the above researches motivate us to think about strong efficient S-stationary conditions for pseudolinear multiobjective MPVC.

We consider the pseudolinear multiobjective mathematical programs with vanishing constraints (MMPVCs) as follows:

$$\begin{aligned} &\min\ (f_1(z), \ldots, f_p(z)) \\ &\text{subject to}\ \ g_i(z) \leqq 0, \quad \forall i = 1, \ldots, q, \\ &\qquad\quad h_i(z) = 0, \quad \forall i = 1, \ldots, r, \\ &\qquad\quad H_i(z) \geqq 0, \quad \forall i = 1, \ldots, m, \\ &\qquad\quad G_i(z)H_i(z) \leqq 0, \quad \forall i = 1, \ldots, m, \end{aligned} \tag{1}$$

where the functions $f_i,\ g_i,\ h_i,\ H_i,\ G_i : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable on $\mathbb{R}^n$.

The organization of this chapter is as follows: in Sect. 2, we recall the needful definitions and results. In Sect. 3, we establish necessary and sufficient optimality conditions, and in Sect. 4, the Mond–Weir type dual model and basic duality results are given.

2 Preliminaries

Throughout this chapter, we use the following notations, definitions, and some well-known results. Let the vectors $y,\ z \in \mathbb{R}^n$, then we shall use the following conventions of inequalities:

$$\begin{aligned} y \leqq z &\Longleftrightarrow y_i \leqq z_i,\ i = 1, \ldots, n, \\ y \leq z &\Longleftrightarrow y \leqq z \text{ and } y \neq z, \\ y < z &\Longleftrightarrow y_i < z_i,\ i = 1, \ldots, n. \end{aligned}$$

Let

$$\begin{aligned} S := \{z \in \mathbb{R}^n \mid g_i(z) \leqq 0, \quad & \forall i = 1, \ldots, q, \\ h_i(z) = 0, \quad & \forall i = 1, \ldots, r, \\ H_i(z) \geqq 0, \quad & \forall i = 1, \ldots, m, \\ G_i(z)H_i(z) \leqq 0, \quad & \forall i = 1, \ldots, m\} \end{aligned}$$

be the feasible region of the MMPVC (1), and let $z^* \in S$ be a feasible solution of the MMPVC (1). Then, the following index sets will be used in the sequel:

$$\begin{aligned} I_f &:= \{1, 2, \ldots, p\}, \\ \text{Set of active constraints } I_g(z^*) &:= \{i \in \{1, 2, \ldots, q\} | g_i(z^*) = 0\}, \\ I_h &:= \{1, 2, \ldots, r\}, \\ I_+(z^*) &:= \{i \in \{1, 2, \ldots, m\} | H_i(z^*) > 0\}, \\ I_0(z^*) &:= \{i \in \{1, 2, \ldots, m\} | H_i(z^*) = 0\}. \end{aligned} \tag{2}$$

We classify the index set $I_+(z^*)$ into the following subsets:

$$\begin{aligned} I_{+0}(z^*) &:= \{i | H_i(z^*) > 0,\ G_i(z^*) = 0\}, \\ I_{+-}(z^*) &:= \{i | H_i(z^*) > 0,\ G_i(z^*) < 0\}. \end{aligned} \tag{3}$$

Similarly, we classify the set I_0 in the following subsets:

$$\begin{aligned} I_{0+}(z^*) &:= \{i | H_i(z^*) = 0,\ G_i(z^*) > 0\}, \\ I_{00}(z^*) &:= \{i | H_i(z^*) = 0,\ G_i(z^*) = 0\}, \\ I_{0-}(z^*) &:= \{i | H_i(z^*) = 0,\ G_i(z^*) < 0\}. \end{aligned} \tag{4}$$

Definition 1 ([14])

1. A point $z^* \in S$ is said to be a weak efficient solution of the multiobjective optimization problem, if there is no $z \in S$, such that

$$f(z) < f(z^*).$$

2. A point $z^* \in S$ is said to be an efficient solution of the multiobjective optimization problem, if there is no $z \in S$, such that

$$f(z) \leq f(z^*) .$$

3. A point $z^* \in S$ is said to be a locally efficient solution of the multiobjective optimization problem, if there exists a neighborhood U of z^* and there is no $z \in S \cap U$, such that

$$f(z) \leq f(z^*) .$$

Definition 2 ([15]) Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ be a differentiable function on an open convex set S. The function f is said to be

1. pseudoconvex at $z^* \in S$, if $\forall z \in S$,

$$\langle \nabla f(z^*), z - z^* \rangle \geqq 0 \Rightarrow f(z) \geqq f(z^*),$$

2. pseudoconcave at $z^* \in S$, if $\forall z \in S$,

$$\langle \nabla f(z^*), z - z^* \rangle \leqq 0 \Rightarrow f(z) \leqq f(z^*).$$

The function is said to be pseudoconvex (pseudoconcave) on S if it is pseudoconvex (pseudoconcave) at every $z \in S$. Moreover, the function is said to be pseudolinear on X if it is both pseudoconvex and pseudoconcave on S. More precisely, a differentiable function $f : S \to \mathbb{R}$ on an open convex subset $S \subseteq \mathbb{R}^n$ is said to be pseudolinear if $\forall z_1, z_2 \in S$, one has

$$\langle \nabla f(z_1), z_2 - z_1 \rangle \geqq 0 \Rightarrow f(z_2) \geqq f(z_1),$$

and

$$\langle \nabla f(z_1), z_2 - z_1 \rangle \leqq 0 \Rightarrow f(z_2) \leqq f(z_1).$$

Theorem 1 ([3]) *Let $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ be an open convex set. Then, f is a differentiable pseudolinear function on X if and only if $\forall z_1, z_2 \in S$, there exists a function $p : S \times S \to \mathbb{R}_+$, where $\mathbb{R}_+$ denotes positive real number, such that*

$$f(z_2) = f(z_1) + p(z_1, z_2)\langle \nabla f(z_1), z_2 - z_1 \rangle.$$

The function p is called proportional function.

3 Optimality Conditions for the MMPVC

We define strong efficient S-stationary conditions for a feasible point of pseudolinear MMPVC (1) motivated by strong Pareto S-stationary point introduced by Zhang et al. [17].

Definition 3 A feasible point z^* is called strong efficient S-stationary point of the pseudolinear MMPVC (1) if there exist multipliers $(\eta^f, \eta^g, \eta^h, \eta^H, \eta^G) \in \mathbb{R}^p_+ \times \mathbb{R}^q \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^m$ satisfying the following conditions:

$$\begin{aligned} \nabla f(z^*)\eta^f + \nabla g(z^*)\eta^g + \nabla h(z^*)\eta^h - \nabla H(z^*)\eta^H + \nabla G(z^*)\eta^G = 0, \\ \eta^f > 0, \;\; g(z^*) \leqq 0, \;\; \eta^g \geqq 0, \;\; g(z^*)^T\eta^g = 0, \\ \eta_i^H = 0 \; (i \in I_+(z^*)), \;\; \eta_i^H \geqq 0 \; (i \in I_{00}(z^*) \cup I_{0-}(z^*)), \;\; \eta_i^H \geqq 0 \; (i \in I_{0+}(z^*)), \\ \eta_i^G = 0 \; (i \in I_{0+}(z^*) \cup I_{0-}(z^*) \cup I_{+-}(z^*) \cup I_{00}(z^*)), \;\; \eta_i^G \geqq 0 \; (i \in I_{+0}(z^*)). \end{aligned} \tag{5}$$

Now, we establish the main result as follows.

Theorem 2 *A feasible point z^* is an efficient solution of the pseudolinear MMPVC (1) if and only if z^* is a strong efficient S-stationary point of the MMPVC (1).*

Proof Let z^* be a strong efficient S-stationary point, then there exist multipliers $(\eta^f, \eta^g, \eta^h, \eta^H, \eta^G) \in \mathbb{R}^p_+ \times \mathbb{R}^q \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{aligned} \nabla f(z^*)\eta^f + \nabla g(z^*)\eta^g + \nabla h(z^*)\eta^h - \nabla H(z^*)\eta^H + \nabla G(z^*)\eta^G = 0, \\ \eta^f > 0, \;\; g(z^*) \leqq 0, \;\; \eta^g \geqq 0, \;\; g(z^*)^T\eta^g = 0, \\ \eta_i^H = 0 \; (i \in I_+(z^*)), \;\; \eta_i^H \geqq 0 \; (i \in I_{00}(z^*) \cup I_{0-}(z^*)), \;\; \eta_i^H \geqq 0 \; (i \in I_{0+}(z^*)), \\ \eta_i^G = 0 \; (i \in I_{0+}(z^*) \cup I_{0-}(z^*) \cup I_{+-}(z^*) \cup I_{00}(z^*)), \;\; \eta_i^G \geqq 0 \; (i \in I_{+0}(z^*)). \end{aligned} \tag{6}$$

Suppose that z^* is not an efficient solution. Then, there exists a feasible point $z \neq z^*$ such that $f_i(z) \leqq f_i(z^*)$ for all i except at least one k such that $f_k(z) < f_k(z^*)$. Now, from pseudolinearity, we have

$$f_i(z) - f_i(z^*) = p_i^f(z, z^*)\langle \nabla f_i(z^*), z - z^* \rangle \leqq 0 \; \forall \, i \in \{1, \ldots, p\} \setminus \{k\} \tag{7}$$

$$f_k(z) - f_k(z^*) = p_k^f(z, z^*)\langle \nabla f_k(z^*), z - z^* \rangle < 0, \tag{8}$$

$$g_i(z) - g_i(z^*) = p_i^g(z, z^*)\langle \nabla g_i(z^*), z - z^* \rangle \leqq 0, \;\; i \in I_g(z^*), \tag{9}$$

$$h_i(z) - h_i(z^*) = p_i^h(z, z^*)\langle \nabla h_i(z^*), z - z^* \rangle = 0,\ i \in I_h, \tag{10}$$

$$-H_i(z) + H_i(z^*) = p_i^H(z, z^*)\langle -\nabla H_i(z^*), z - z^* \rangle \leqq 0,\ i \in I_{0+}(z^*), \tag{11}$$

$$-H_i(z) + H_i(z^*) = p_i^H(z, z^*)\langle -\nabla H_i(z^*), z - z^* \rangle \leqq 0,\ i \in I_{00}(z^*) \cup I_{0-}(z^*), \tag{12}$$

$$G_i(z) - G_i(z^*) = p_i^G(z, z^*)\langle \nabla G_i(z^*), z - z^* \rangle \leqq 0,\ i \in I_{+0}(z^*). \tag{13}$$

Multiplying (7)–(12) by $\eta_i^f > 0\ (i \in I_f)$, $\eta_i^g \geqq 0\ (i \in I_g(z^*))$, $\eta_i^h\ (i \in I_h)$, $\eta_i^H \geqq 0\ (i \in I_{00}(z^*) \cup I_{0-}(z^*))$, $\eta_i^H \geqq 0\ (i \in I_{0+}(z^*))$, and $\eta_i^G \geqq 0\ (i \in I_{+0}(z^*))$, respectively, and using the fact that each $p_i > 0$, we get

$$\left\langle \sum_{i=1}^{p} \eta_i^f \nabla f_i(z^*) + \sum_{i=1}^{q} \eta_i^g \nabla g_i(z^*) + \sum_{i=1}^{r} \eta_i^h \nabla h_i(z^*) - \sum_{i=1}^{m} \eta_i^H \nabla H_i(z^*) + \sum_{i=1}^{m} \eta_i^G \nabla G_i(z^*),\ z - z^* \right\rangle < 0,$$

which contradicts the stationarity of z^*. Hence, the result.

Conversely, suppose that z^* is an efficient solution of pseudolinear MMPVC (1), then from pseudolinearity of all the functions, there does not exist any feasible point z different from z^* such that the following system has solution:

$$\begin{aligned}
\langle \nabla f_i(z^*), z - z^* \rangle &< 0,\ i = k,\\
\langle \nabla f_i(z^*), z - z^* \rangle &\leqq 0,\ i = \{1, \ldots, p\} \setminus \{k\},\\
\langle \nabla g_i(z^*), z - z^* \rangle &\leqq 0,\ i \in I_g(z^*),\\
\langle \nabla h_i(z^*), z - z^* \rangle &= 0,\\
\langle -\nabla H_i(z^*), z - z^* \rangle &\leqq 0\ (i \in I_{00}(z^*) \cup I_{0-}(z^*)),\\
\langle -\nabla H_i(z^*), z - z^* \rangle &\leqq 0\ (i \in I_{0+}(z^*)),\\
\langle \nabla G_i(z^*), z - z^* \rangle &\leqq 0\ (i \in I_{+0}(z^*)).
\end{aligned} \tag{14}$$

That is, the system of inequalities (14) has no solution. Therefore, from Tucker theorem [13, pp. 29], there exist $\eta = (\eta^f, \eta^g, \eta^h, \eta^H, \eta^G) \in \mathbb{R}_+^p \times \mathbb{R}^q \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^m$, and setting multipliers zero for inactive constraints as follows:

$$\eta_i^f > 0\ (i \in I = \{1, \ldots, p\}),\ \eta_i^g \geqq 0,\ g(z^*)^T \eta^g = 0,$$
$$\eta_i^H = 0\ (i \in I_+(z^*)),\ \eta_i^H \geqq 0\ (i \in I_{00}(z^*) \cup I_{0-}(z^*)),\ \eta^H \geqq 0\ (i \in I_{0+}),$$
$$\eta_i^G = 0\ (i \in I_{0+}(z^*) \cup I_{0-}(z^*) \cup I_{+-}(z^*) \cup I_{00}(z^*)),\ \eta_i^G \geqq 0\ (i \in I_{+0}(z^*)),$$

we get

$$\sum_{i=1}^{p} \eta_i^f \nabla f_i(z^*) + \sum_{i=1}^{q} \eta_i^g \nabla g_i(z^*) + \sum_{i=1}^{r} \eta_i^h \nabla h_i(z^*) - \sum_{i=1}^{m} \eta_i^H \nabla H_i(z^*) + \sum_{i=1}^{m} \eta_i^G \nabla G_i(z^*) = 0.$$

Hence, we get the required results. □

Example 3.1 Consider the problem

$$\begin{aligned} \min \quad & f(z) = (f_1(z), f_2(z)), \\ \text{s. t.} \quad & g(z) \leqq 0,\ H(z) \geqq 0,\ G(z)H(z) \leqq 0, \\ \text{where} \quad & f_1(z) = \exp z_1,\ f_2(z) = z_2,\ g(z) = -z_1 - z_2 \leqq 0,\ H(z) = z_1, \\ & G(z) = -z_2,\ z = (z_1, z_2) \in \mathbb{R}^2, \\ & \text{feasible set } S = \{z \in \mathbb{R}^2 : z_1 \geqq 0,\ z_2 \geqq 0\}, \end{aligned}$$

at a feasible point $z^* = (0, 0) \in S$. Then, for $\eta_1^f > 0,\ \eta_2^f > 0,\ \eta^g \geqq 0,\ \eta^H \geqq 0$, and $\eta^G \geqq 0$, the expression:

$$\begin{aligned} &\eta_1^f \nabla f_1(z^*) + \eta_2^f \nabla f_2(z^*) + \eta^g \nabla g(z^*) - \eta^H \nabla H(z^*) + \eta^G \nabla G(z^*) \\ &\quad = \eta_1^f \begin{bmatrix} \exp z_1 \\ 0 \end{bmatrix} + \eta_2^f \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \eta^g \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \eta^H \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \eta^G \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \end{aligned}$$

at point $z^* = (0, 0)$ for choosing $\eta_1^f = \eta^g + \eta^H,\ \eta_2^f = \eta^g,\ \eta^G = 0$. Thus, from Theorem 2, the point $z^* = (0, 0)$ is a strong efficient S-stationary point. Since no other feasible point can dominate $z^* = (0, 0)$, the point $z^* = (0, 0)$ is an efficient solution of the given problem by simple observation.

Remark 3.2 Strong efficient S-stationary conditions and efficiency can be satisfied without satisfying pseudolinearity. See the following example.

Example 3.3 Consider the problem

$$\begin{aligned} \min \quad & f(z) = (f_1(z), f_2(z)) \\ \text{s.t.} \quad & g(z) \leqq 0,\ H(z) \geqq 0,\ G(z)H(z) \leqq 0, \\ \text{where} \quad & f_1(z) = z_2 + \tan^{-1}(z_2),\ f_2(z) = z_1,\ H(z) = z_2, \\ & g(z) = -z_1 - z_2 - 1,\ G(z) = -z_1 \text{ and } z = (z_1, z_2) \in \mathbb{R}^2, \end{aligned}$$

at point $z^* = (-1, 0)$. Then, $z^* = (-1, 0)$ is a strong efficient S-stationary point as follows:

$$\eta_1^f \nabla f_1(z^*) + \eta_2^f \nabla f_2(z^*) + \eta^g \nabla g(z^*) - \eta^H \nabla H(z^*) + \eta^G \nabla G(z^*)$$
$$= \eta_1^f \begin{bmatrix} 0 \\ 1 + \frac{1}{1+z_1^2} \end{bmatrix} + \eta_2^f \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \eta^g \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \eta^H \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \eta^G \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

if $\eta^g + \eta^G = \eta_2^f$, $\eta^g + \eta^H = \eta_1^f (1 + \frac{1}{1+z_1^2})$, and $\eta^G = 0$. Also, $z^* = (-1, 0)$ is an efficient point, but all functions at feasible point $z^* = (-1, 0)$ are not pseudolinear.

4 Duality

In this section, we propose Mond–Weir type dual model to the MMPVC (1) and establish weak and strong duality results under pseudolinear assumptions. The Mond–Weir type dual model to the pseudolinear MMPVC (1) is defined as follows:

max $f(u)$

subject to $(u, \eta^f, \eta^g, \eta^h, \eta^H, \eta^G) \in P = \{(u, \eta^f, \eta^g, \eta^h, \eta^H, \eta^G) :$

$$\sum_{i=1}^{p} \eta_i^f \nabla f_i(u) + \sum_{i=1}^{q} \eta_i^g \nabla g_i(u) + \sum_{i=1}^{r} \eta_i^h \nabla h_i(u) - \sum_{i=1}^{m} \eta_i^H \nabla H_i(u)$$
$$+ \sum_{i=1}^{m} \eta_i^G \nabla G_i(u) = 0$$

$\eta_i^f > 0\ (i \in I = \{1, \ldots, p\})$, $\eta_i^g \geqq 0$, $g(u)^T \eta^g \geqq 0$, $G(u)^T \eta^G \geqq 0$,

$H(u)^T \eta^H \leqq 0$,

$h(u) = 0$, $\eta_i^H = 0\ (i \in I_+(u))$, $\eta_i^H \geqq 0\ (i \in I_{00}(u) \cup I_{0-}(u))$, $\eta^H \geqq 0\ (i \in I_{0+})$,

$\eta_i^G = 0\ (i \in I_{0+}(u) \cup I_{0-}(u) \cup I_{+-}(u) \cup I_{00}(u))$, $\eta_i^G \geqq 0\ (i \in I_{+0}(u))\}$.

Consider the following set:

$$P_u = \{u : (u, \eta^f, \eta^g, \eta^h, \eta^H, \eta^G) \in P\}.$$

Theorem 3 (Weak Duality) *Let z be a feasible point of the pseudolinear MMPVC (1) and $(u, \eta^f, \eta^g, \eta^h, \eta^H, \eta^G)$ be a feasible point of the Mond–Weir dual problem. Suppose that the given functions $g^T \eta^g$, h_i $-$ H_i $(i \in I_{00}(u) \cup I_{0+}(u) \cup I_{0-}(u))$, and G_i $(i \in I_{+0}(u))$ are pseudolinear at u. If any of the following holds:*

(a) $\eta_i^f > 0$ and $f_i(\cdot)(\forall i \in I_f)$ are pseudolinear at u;

(b) $\eta_i^f > 0$ $(\forall i \in I_f)$ and $\sum_{i=1}^{p} \eta_i^f f_i(\cdot)$ is pseudolinear at u,

then

$$f(z) \nleqq f(u).$$

Proof Assume that

$$f(z) \leq f(u).$$

Then,

$$f_i(z) \leqq f_i(u), \ \forall\, i \in I_f, \text{ except at least one } k, \text{ such that}$$

$$f_k(z) < f_k(u).$$

Multiplying by $\eta_i^f > 0$ and adding, we get

$$(\eta^f)^T f(z) < (\eta^f)^T f(u).$$

Using the feasibility and pseudolinearity assumptions, we get

$$\left\langle \sum_{i=1}^{p} \eta_i^f \nabla f_i(u), z - u \right\rangle < 0, \tag{15}$$

$$\sum_{i=1}^{q} \eta_i^g g_i(z) \leqq \sum_{i=1}^{q} \eta_i^g g_i(u) \implies \left\langle \sum_{i=1}^{q} \eta_i^g \nabla g_i(u), z - u \right\rangle \leqq 0, \tag{16}$$

$$\sum \eta_i^h h_i(z) = \sum \eta_i^h h_i(u) \implies \left\langle \sum \eta_i^h \nabla h_i(u),\ z - u \right\rangle = 0,\ i \in I_h, \tag{17}$$

$$-\sum \eta_i^H H_i(z) \leqq -\sum \eta_i^H H_i(u) \implies \left\langle -\sum \eta_i^H \nabla H_i(u), z - u \right\rangle \leqq 0, \tag{18}$$
$$i \in I_{00}(u) \cup I_{0+}(u) \cup I_{0-}(u),$$

$$\sum \eta_i^G G_i(z) \leqq \sum \eta_i^G G_i(u) \implies \left\langle \sum \eta_i^G \nabla G_i(u), z-u \right\rangle \leqq 0, \quad i \in I_{+0}(u). \tag{19}$$

Adding (15)–(19), we get

$$\left\langle \sum_{i=1}^{p} \eta_i^f \nabla f_i(u) + \sum_{i=1}^{q} \eta_i^g \nabla g_j(u) + \sum_{i=1}^{r} \eta_i^h \nabla h_i(u) - \sum_{i=1}^{m} \eta_i^H \nabla H_i(u) + \sum_{i=1}^{m} \eta_i^G \nabla G_i(u),\ z-u \right\rangle < 0.$$

which contradicts the feasibility of u. Hence, we get the required result. □

Theorem 4 (Strong Duality) *Let z^* be an efficient solution of the pseudolinear MMPVC (1). If weak duality Theorem 3 holds, then there exist $(\bar{\eta}^f, \bar{\eta}^g, \bar{\eta}^h, \bar{\eta}^H, \bar{\eta}^G) \in \mathbb{R}_+^p \times \mathbb{R}^q \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^m$ such that $(z^*, \bar{\eta}^f, \bar{\eta}^g, \bar{\eta}^h, \bar{\eta}^H, \bar{\eta}^G)$ is an efficient solution of the Mond–Weir dual problem, and the corresponding values of objective functions are same.*

Proof As z^* is an efficient solution of pseudolinear MMPVC (1), then from Theorem 2, there exist $(\bar{\eta}^f, \bar{\eta}^g, \bar{\eta}^h, \bar{\eta}^H, \bar{\eta}^G) \in \mathbb{R}_+^p \times \mathbb{R}^q \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^m$ such that z^* is a strong efficient S-stationary point. That is,

$$\sum_{i=1}^{p} \bar{\eta}_i^f \nabla f_i(z^*) + \sum_{i=1}^{q} \bar{\eta}_i^g \nabla g_i(z^*) + \sum_{i=1}^{r} \bar{\eta}_i^h \nabla h_i(z^*) - \sum_{i=1}^{m} \bar{\eta}_i^H \nabla H_i(z^*) + \sum_{i=1}^{m} \bar{\eta}_i^G \nabla G_i(z^*) = 0,$$

$$\bar{\eta}_i^f > 0,\ \bar{\eta}_i^g \geqq 0,\ g(z^*)^T \bar{\eta}^g = 0,\ \bar{\eta}_i^H = 0\ (i \in I_+(z^*))\quad \bar{\eta}_i^H \geqq 0\ (i \in I_{00}(z^*) \cup I_{0-}(z^*)),$$

$$\bar{\eta}^H \geqq 0\ (i \in I_{0+}(z^*)),\ \bar{\eta}_i^G = 0\ (i \in I_{+-}(z^*) \cup I_{0+}(z^*) \cup I_{0-}(z^*) \cup I_{00}(z^*)),\ h(z^*) = 0,$$

$$\bar{\eta}_i^G \geqq 0\ (i \in I_{+0}(z^*)).$$

Therefore, $(z^*, \bar{\eta}^f, \bar{\eta}^g, \bar{\eta}^h, \bar{\eta}^H, \bar{\eta}^G)$ is a feasible solution of the Mond–Weir dual problem, and from feasibility of weak duality Theorem 3, we have

$$f(z^*) \geqq f(u),$$

for any feasible solution $(u, \eta^f, \eta^g, \eta^h, \eta^H, \eta^G) \in \mathbb{R}^n \times \mathbb{R}_+^p \times \mathbb{R}^q \times \mathbb{R}^r \times \mathbb{R}^m \times \mathbb{R}^m$ of the Mond–Weir dual problem. Hence, $(z^*, \bar{\eta}^f, \bar{\eta}^g, \bar{\eta}^h, \bar{\eta}^H, \bar{\eta}^G)$ is an efficient solution of the Mond–Weir dual problem, and their values are equal. □

The following example verifies the Mond–Weir dual model and duality results as follows.

Example 4.1 Consider the following pseudolinear MMPVC problem:

$$\min f(z) = (f_1(z), f_2(z)), \text{ where } f_1(z) = \tan^{-1}(z_1),\ f_2(z) = \tan^{-1}(z_2),$$
$$\text{subject to } g_1(z) = -z_1 \leqq 0,\ g_2(z) = -z_2 \leqq 0,\ H(z) = z_1 + z_2 \geqq 0,$$
$$G(z)^T H(z) = (z_1 + z_2)(z_1 - z_2) \leqq 0 \text{ at feasible point } z^* = (0, 0).$$

Feasible set $S = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 \geqq 0,\ z_2 \geqq 0,\ z_1+z_2 \geqq 0,\ (z_1+z_2)(z_1-z_2) \leqq 0\}$. The Mond–Weir dual model is

$$\max f(u) = (\tan^{-1}(u_1),\ \tan^{-1}(u_2)),$$
$$\text{s. t. } \eta_1^f \nabla f_1(u) + \eta_2^f \nabla f_2(u) + \eta_1^g \nabla g_1(u) + \eta_2^g \nabla g_2(u) - \eta^H \nabla H(u) + \eta^G \nabla G(u)$$
$$= \eta_1^f \begin{bmatrix} \frac{1}{1+u_1^2} \\ 0 \end{bmatrix} + \eta_2^f \begin{bmatrix} 0 \\ \frac{1}{1+u_2^2} \end{bmatrix} + \eta_1^g \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \eta_2^g \begin{bmatrix} 0 \\ -1 \end{bmatrix} - \eta^H \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \eta^G \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$
$$\text{for } \eta_1^f > 0,\ \eta_2^f > 0,\ \eta_1^g + \eta^H = \frac{\eta_1^f}{1+u_1^2} + \eta^G,\ \frac{\eta_2^f}{1+u_2^2} = \eta_2^g + \eta^H + \eta^G,$$
$$\text{and } \eta_1^g g_1(u) = -\eta_1^g u_1 \geqq 0,\ \eta_1^g \geqq 0,\ \eta_2^g g_2(u) = -\eta_2^g u_2 \geqq 0,\ \eta_2^g \geqq 0,\ \eta^H \geqq 0,$$
$$\eta^G \geqq 0,\ \eta^H H(u) = \eta^H (u_1 + u_2) \leqq 0,\ \eta^G G(u) = \eta^G (u_1 - u_2) \geqq 0,\ u = (u_1, u_2) \in \mathbb{R}^2.$$

Solving above, we get

$$P_u = \{u \in \mathbb{R}^2 : u_2 \leqq 0,\ u_1 \leqq 0,\ u_1 - u_2 \geqq 0,\ u_1 + u_2 \leqq 0\}.$$

It is clear from the feasibility that

$$f(z) \geqq f(u).$$

Hence, the weak duality Theorem 3 is verified, and the strong duality theorem is obviously satisfied at point z^*.

5 Conclusions

In this chapter, we have established Karush–Kuhn–Tucker type optimality conditions for pseudolinear multiobjective mathematical programs with vanishing constraints under smooth assumptions. Moreover, we formulated Mond–Weir type

dual models and established duality results for pseudolinear multiobjective mathematical programs with vanishing constraints. We verify our results through some examples. In future, this chapter can be extended to nonsmooth case.

Acknowledgments We are really thankful to the anonymous referees for their insightful comments and suggestions which led to the present improved version as it stands. The first author is supported by the CSIR-New Delhi, Ministry of Human Resources Development, Government of India, with grant no. 09/013(0583)/2015-EMR-I. The second author is supported by UGC Research Fellowship, through sanction letter no. Ref. No./Math/ Res/Sept.2015/2015-16/918. The research of the third author is supported by the Department of Science and Technology, SERB, New Delhi, India, with grant no. MTR/2018/000121.

References

1. W. Achtziger, C. Kanzow, Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. Math. Program. **114**(1, Ser. A), 69–99 (2008)
2. M. Benko, H. Gfrerer, An SQP method for mathematical programs with vanishing constraints with strong convergence properties. Comput. Optim. Appl. **67**(2), 361–399 (2017)
3. K.L. Chew, E.U. Choo, Pseudolinearity and efficiency. Math. Program. **28**(2), 226–239 (1984)
4. J.P. Dussault, M. Haddou, T. Migot, Mathematical programs with vanishing constraints: constraint qualifications, their applications, and a new regularization method. Optimization **68**(2–3), 509–538 (2019)
5. T. Hoheisel, C. Kanzow, First- and second-order optimality conditions for mathematical programs with vanishing constraints. Appl. Math. **52**(6), 495–514 (2007)
6. T. Hoheisel, C. Kanzow, Stationary conditions for mathematical programs with vanishing constraints using weak constraint qualifications. J. Math. Anal. Appl. **337**(1), 292–310 (2008)
7. T. Hoheisel, C. Kanzow, On the Abadie and Guignard constraint qualifications for mathematical programmes with vanishing constraints. Optimization **58**(4), 431–448 (2009)
8. R. Jabr, Solution to economic dispatching with disjoint feasible regions via semidefinite programming. IEEE Trans. Power Syst. **27**(1), 572–573 (2012)
9. M.N. Jung, C. Kirches, S. Sager, *On Perspective Functions and Vanishing Constraints in Mixed-Integer Nonlinear Optimal Control* (Springer, Heidelberg, 2013), pp. 387–417
10. A. Khare, T. Nath, Enhanced Fritz John stationarity, new constraint qualifications and local error bound for mathematical programs with vanishing constraints. J. Math. Anal. Appl. **472**(1), 1042–1077 (2019)
11. C. Kirches, A. Potschka, H.G. Bock, S. Sager, A parametric active-set method for QPS with vanishing constraints arising in a robot motion planning problem. Pac. J. Optim. **9**(2), 275–299 (2013)
12. K.O. Kortanek, J.P. Evans, Pseudo-concave programming and Lagrange regularity. Oper. Res. **15**(5), 882–891 (1967)
13. O.L. Mangasarian, *Nonlinear Programming* (SIAM, Philadelphia, 1994)
14. K. Miettinen, *Nonlinear Multiobjective Optimization* (Springer, Boston, 1998)
15. S.K. Mishra, B.B. Upadhyay, *Pseudolinear Functions and Optimization* (CRC Press, Boca Raton, 2015)
16. S.K. Mishra, V. Singh, V. Laha, On duality for mathematical programs with vanishing constraints. Ann. Oper. Res. **243**(1–2), 249–272 (2016)
17. P. Zhang, J. Zhang, G.H. Lin, X. Yang, Some kind of Pareto stationarity for multiobjective problems with equilibrium constraints. Optimization **68**(6), 1245–1260 (2019)

The Solvability and Optimality for Semilinear Stochastic Equations with Unbounded Delay

Yadav Shobha and Surendra Kumar

Abstract The objective of this chapter is to study the standard optimal control problem for a new class of semilinear stochastic system with unbounded delay. We use the fundamental solution technique to write an expression for the mild solution. We then derive sufficient conditions that ensure the existence and uniqueness of mild solution. Under natural assumptions, the existence of an optimal state–control pair for the standard Lagrangian problem is also examined. Finally, the developed theory is validated with an example.

Keywords Stochastic differential equations · Unbounded delay · Fundamental solution · Mild solution · Optimal control

2010 Mathematics Subject Classification 34A12, 34K30, 34K35, 34K50, 47H10

1 Introduction

It is well known that deterministic model often fluctuates due to noise, so there is a need to switch from deterministic systems to stochastic systems. Stochastic differential equations having bounded or unbounded delay have attracted considerable attention due to their importance in areas such as economics, finance, population dynamics, and many more [9, 10]. Particularly, qualitative and quantitative attributes for delayed systems (bounded or unbounded) have been examined rather broadly since delayed dynamical systems are aplenty in nature; for instance, dynamical systems in biology [11, 31] and optics [17, 18] among others can be displayed as the first-order delayed differential equations. Stochastic partial differential equations (SPDEs) with unbounded delay can be seen in modeling of several phenomena of

Y. Shobha (✉) · S. Kumar
Department of Mathematics, University of Delhi, New Delhi, Delhi, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_17

natural and social sciences (see [12, 38] and the references therein). That is why, recently, more investigation is going on existence and uniqueness, optimality, and invariant measures among others of SPDEs with delays [2, 6, 16, 24, 27, 29, 33, 40, 41, 43].

In general, the control issue involves the minimization or maximization of performance index of the state and control variables of the system over a set of acceptable control functions. Mostly, we deal with the problems of finding an optimal state–control pair, time optimal controls for a given target set, and time optimal control to a point target set. The optimality of both deterministic and stochastic systems has attracted appreciable attention of researchers. Chen et al. [7] and Tanabe [39] studied the optimality of linear systems in finite-dimensional spaces. Lasiecka and Triggiani [22], Li and Yong [23], Ahmed and Teo [1], and Curtain and Pritchard [8] among others focused on the optimality of linear systems in infinite-dimensional spaces. And, for the stochastic case, we refer [6, 46] and references cited therein. Liu [26] and Nakagiri [35] used the fundamental solution theory and discussed the optimal control problems of delayed linear systems. Li et al. [25] examined an indefinite stochastic linear quadratic optimal control problem with delay and related forward–backward stochastic differential equations. Recently, the optimal control problem of semilinear systems becomes a vital area of research; see [13, 20, 28] for systems without delay and [44] for systems with finite delay. Jeong et al. [19] considered semilinear evolution equation and examined the existence of optimal control as well as maximal principles under the hypothesis that the nonlinear term is Lipschitz continuous. Furthermore, Buckdahn and Răscanu [5] discussed the existence of optimal control for parabolic semilinear stochastic differential equation by utilizing Ekeland's principle. Papageorgiou [36] obtained necessary conditions for the optimality of a system governed by nonlinear evolution equations with the help of penalty function. The existence of optimal controls for framework representing a semilinear parabolic equation with boundary conditions contained control variable has been studied by Wang et al. [42].

The optimal control issues of unbounded delayed systems are also an interesting area of research. Xiaoling and Huawu [45] generalized the Gronwall lemma with time lags and used them to examine nonlinear control systems with delay. Mokkedem and Fu [34] proved the existence of state–control pair for the semilinear system with unbounded delay by using the theory of fundamental solutions associated with the corresponding linear part. The main objective of this chapter is to extend the results of Mokkedem and Fu for delayed semilinear stochastic systems in a Hilbert space. This fact is the motivation for this chapter. We utilize the technique of stochastic calculus, fundamental solution, and successive approximation for the existence of solution of the system (2.2). We also show the existence of optimal state–control pair for the Lagrange problem subject to the semilinear stochastic functional differential equation with unbounded delay. The obtained results can be considered as a contribution to the literature of stochastic optimal control.

The rest of the chapter is prepared as follows: we introduce notations, definitions, and the phase space $\mathscr{G}$ in Sect. 2, which will be used throughout the chapter. Next, we give the fundamental solution for the unbounded delayed linear system with

$B \equiv f \equiv g = 0$ in system (2.2). In Sect. 3, we show the existence and uniqueness results for the mild solution of differential equation (2.2). Section 4 is devoted for the existence of optimal state–control pair of the system (2.1). An example is constructed in Sect. 5 to validate the developed theory.

2 Preliminaries

Let $(Y, \|\cdot\|)$ be a real separable Hilbert space and $(H, \|\cdot\|_H)$ another separable Hilbert space. Suppose that the space of all bounded linear operators from H to Y is denoted by $(\mathscr{L}_b(H; Y), \|\cdot\|)$, and $\mathscr{L}_b(Y)$ denotes the space $\mathscr{L}_b(Y; Y)$. For more details concerning the theory of semigroups, one can refer to Pazy [37].

Consider the integral cost functional given by

$$\mathcal{I}(y, v) = \mathbb{E}\left\{\int_0^a \mathcal{J}(r, y^v(r), y_r^v, v(r))dr\right\}, \tag{2.1}$$

subject to the equations

$$\begin{cases} dy(r) = [Ay(r) + L(y_r) + B(r)v(r) + f(r, y_r)]\,dr + g(r, y_r)dW(r), \ 0 < r \le a \\ y_0 = \phi \in L_p^{\mathscr{F}}(\Omega; \mathscr{G}), \end{cases} \tag{2.2}$$

where functional $\mathcal{J}$ in (2.1) is specified later. Here, the state $y(\cdot)$ is Y-valued stochastic process, and its histories $y_r : (-\infty, 0] \to Y$ given by $y_r(\eta) = y(r + \eta)$, for $\eta \le 0$, belong to $\mathscr{G}$, which is an abstract phase space; the control $v(\cdot)$ takes values in another separable reflexive Hilbert space V, and the operator $A : D(A) \subset Y \to Y$ is the infinitesimal generator of a strongly continuous semigroup $\{S(r)\}_{r\ge 0}$ on Y. Let $L \in \mathscr{L}_b(\mathscr{G}; Y)$, and for $r \ge 0$, $B(r) \in \mathscr{L}_b(V; Y)$. Furthermore, the functions $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ are nonlinear, and $W(r)$ denotes the Q-Wiener process.

Motivated by Hale and Kato [14] and Hino et al. [15], we consider the phase space $\mathscr{G}$, which is the collection of functions from $(-\infty, 0]$ to Y. It is a linear space with seminorm $\|\cdot\|_{\mathscr{G}}$. Moreover, it satisfies the following axioms:

(H_1) Let $\rho \ge 0$ and $d > 0$. If function y from $(-\infty, \rho + d]$ to Y is continuous on $[\rho, \rho + d]$ and $y_\rho \in \mathscr{G}$, then for $\rho \le r \le \rho + d$, we have the following:

(i) $y_r \in \mathscr{G}$;
(ii) $\|y(r)\| \le K\|y_r\|_{\mathscr{G}}$, where $K > 0$ is a constant and is independent of $y(\cdot)$; and
(iii) $\|y_r\|_{\mathscr{G}} \le N(r - \rho) \sup\{\|y(l)\| : \rho \le l \le r\} + \Gamma(r - \rho)\|y_\rho\|_{\mathscr{G}}$, where function $N : [0, \infty) \to [0, \infty)$ is continuous and $\Gamma : [0, \infty) \to [0, \infty)$ is locally bounded. Also, $N(\cdot)$ and $\Gamma(\cdot)$ are independent of $y(\cdot)$.

(H_2) For $y(r)$ in (H_1), $y_r \in \mathscr{G}$ is a continuous function on the interval $[\rho, \rho + d]$.
(H_3) $\mathscr{G}$ is complete.

To write an expression for the fundamental solution and to list some of its properties, we assume that the following assumptions hold in $\mathscr{G}$:

(b_1) For any $z \in Y$, define

$$\phi_z^0(\eta) = \begin{cases} z, & \eta = 0, \\ 0, & \eta < 0, \end{cases} \tag{2.3}$$

which is in $\mathscr{G}$ and $\|\phi_z^0\|_{\mathscr{G}} \leq \|z\|$.

(b_2) The functions $N(\cdot)$ and $\Gamma(\cdot)$ in Axiom $(H_1(iii))$ are bounded. That is, there exist constants N_a and Γ_a such that

$$N_a = \max_{r\in[0,a]} N(r) \text{ and } \Gamma_a = \sup_{r\in[0,a]} \Gamma(r).$$

Let $\Omega = (\Omega, \mathscr{F}, \{\mathscr{F}_r\}_{r\geq 0}, \mathbf{P})$ be a filtered complete probability space. Also, assume that $\mathscr{F}_r$ is the σ-algebra generated by the Wiener process W with $\mathscr{F}_a = \mathscr{F}$ and that $W(r)$ in H is the Wiener process defined on Ω with nuclear covariance operator Q such that $tr(Q) < \infty$. Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a complete orthonormal basis for H and $\{\alpha_k(r)\}_{k\in\mathbb{N}}$ be a sequence of independent Brownian motions such that

$$W(r) = \sum_{k=1}^{\infty} \sqrt{\ell_k}\xi_k\alpha_k(r), \quad r \geq 0,$$

where $\ell_k \geq 0$ for all $k \in \mathbb{N}$, and let $Q \in \mathscr{L}_b(H)$ given by $Q\xi_k = \ell_k\xi_k$ for all $k \in \mathbb{N}$ with trace $tr(Q) = \sum_{k=1}^{\infty} \ell_k < \infty$.

The norm of the operator $\chi \in \mathscr{L}_b(H; Y)$ is defined by

$$\|\chi\|_Q^2 = tr(\chi Q\chi^*) = \sum_{k=1}^{\infty} \|\sqrt{\ell_k}\chi\xi_k\|^2.$$

If $\|\chi\|_Q < \infty$, then χ is called a Q-Hilbert–Schmidt operator, and the space of all Q-Hilbert–Schmidt operators $\chi : H \to Y$ is denoted by $L_2^0(H; Y)$. If for $r \geq 0$, $y(r) : \Omega \to Y$ is a continuous $\mathscr{F}_r$-adapted stochastic process, then the process $y_r : \Omega \to \mathscr{G}$ generated by $y(r)$ is defined by $y_r(l)(\omega) = y(r + l)(\omega)$, $l \in (-\infty, 0]$.

Let $L_p^{\mathscr{F}}(\Omega; Y)$ be the closed subspace of $L_p([0, a] \times \Omega; Y)$ which consists of $\mathscr{F}_r$-adapted process, and the Banach space of all continuous functions from interval $[0, a]$ to $L_p^{\mathscr{F}}(\Omega; Y)$ is denoted by $C([0, a]; L_p^{\mathscr{F}}(\Omega; Y))$ with $\sup_{r\in[0,a]} \mathbb{E}\|y(r)\|^p < \infty$.

Now, introduce the set of admissible controls

$$\mathcal{A}_{ad} = \left\{ v : [0,a] \times \Omega \to V \text{ is } \mathscr{F}_r - \text{adapted proces with } \mathbb{E}\int_0^a \|v(r)\|^p dr < \infty \right\}.$$

Thus, our main problem can be expressed as follows:

Find a state–control pair $(\tilde{y}, \tilde{v})$, where $\tilde{y}$ is the mild solution (Definition 2.2) of system (2.2) with a control $\tilde{v} \in \mathcal{A}_{ad}$, such that

$$\mathcal{I}(\tilde{y}, \tilde{v}) \leq \mathcal{I}(y^v, v), \text{ for all } (y^v, v) \in C([0,a]; L_p^{\mathscr{F}}(\Omega; Y)) \times \mathcal{A}_{ad}.$$

We impose the following restrictions on the system parameters:

(P_1) The operator A generates a strongly continuous semigroup $\{S(r)\}_{r\geq 0}$ on the Hilbert space Y. Suppose there are constants $\theta \in \mathbb{R}$ and $R_\theta \geq 1$ such that for all $r \geq 0$,

$$\|S(r)\| \leq R_\theta e^{\theta r}.$$

(P_2) The operator B is in $L_\infty([0,a]; \mathscr{L}_b(V;Y))$, and let $M_B = \sup_{0\leq r\leq a} \|B(r)\|$.

(P_3) The operator $L \in \mathscr{L}_b(\mathscr{G}; Y)$ and $\|L\| = l_0$ for some $l_0 > 0$.

(P_4) Suppose that for measurable functions $f : [0,a] \times \mathscr{G} \to Y$ and $g : [0,a] \times \mathscr{G} \to L_2^0(H;Y)$, there exists a constant $N_1 > 0$ such that

$$\|f(r,\xi_1) - f(r,\xi_2)\|^p + \|g(r,\xi_1) - g(r,\xi_2)\|_Q^p \leq N_1 \|\xi_1 - \xi_2\|_{\mathscr{G}}^p,$$
$$\|f(r,\xi)\|^p + \|g(r,\xi)\|_Q^p \leq N_1(1 + \|\xi\|_{\mathscr{G}}^p),$$

for all $0 \leq r \leq a$ and $\xi, \xi_1, \xi_2 \in \mathscr{G}$.

For $B \equiv f \equiv g = 0$, the system (2.2) becomes

$$\begin{cases} \frac{d}{dr} y(r) = Ay(r) + L(y_r), \ r > 0, \\ y_0 = \phi \in \mathscr{G}. \end{cases} \tag{2.4}$$

Let $y(r,\phi)$ be the mild solution of the system (2.4). Then, Mokkedem and Fu [32] proved that under the hypotheses (P_1) and (P_3), the fundamental solution $\mathcal{D}(\cdot) \in \mathscr{L}_b(Y)$ is given by

$$\mathcal{D}(r)z = \begin{cases} y(r, \phi_z^0), & r \geq 0, \\ 0, & r < 0, \end{cases}$$

for any $z \in Y$. Moreover, $\mathcal{D}(r)$ satisfies the following:

$$\mathcal{D}(r) = \begin{cases} S(r) + \int_0^r S(r-l)L(\mathcal{D}_l)dl, & r \geq 0, \\ 0, & r < 0, \end{cases} \tag{2.5}$$

where $\mathcal{D}_r(\eta) = \mathcal{D}(r+\eta)$, $\eta \leq 0$, and the solution of (2.5) is unique (see [32]).

Remark 2.1 ([32, Theorem 3.2]) For $r \geq 0$, $\mathcal{D}(r)$ is a strongly continuous bounded linear operator on Y and

$$\|\mathcal{D}(r)\| \leq Ce^{\mu r}, \quad \text{where } C > 0, \ \mu \in \mathbb{R} \text{ are constants.}$$

Also, for all $0 \leq r \leq a$,

$$\|\mathcal{D}(r)\| \leq M, \ \text{for some } M \geq 1.$$

Now, define the mild solution of the system (2.2) as follows:

Definition 2.2 A stochastic process $y(\cdot) : (-\infty, a] \times \Omega \to Y$ is said to be a mild solution of the system (2.2) if

(i) $y(r, \omega)$ is measurable and $y(r)$ is $\mathscr{F}_r$-adapted;
(ii) for each $0 \leq r \leq a$, $\mathbb{E}\|y(r)\|^p < \infty$ and y_r is $\mathscr{G}$-valued stochastic process;
(iii) for each $v(\cdot) \in L_p([0, a]; V)$, the following is satisfied:

$$y(r) = \begin{cases} \mathcal{D}(r)\phi(0) + \int_0^r \mathcal{D}(r-l)\Big[L(\tilde{\phi}_l) + f(l, y_l) + B(l)v(l)\Big]dl \\ + \int_0^r \mathcal{D}(r-l)g(l, y_l)dW(l), & r \in [0, a], \\ \phi(r) \in L_p^{\mathscr{F}}(\Omega; \mathscr{G}), & r \in (-\infty, 0], \end{cases} \tag{2.6}$$

where $\tilde{\phi}(\cdot)$ is defined by

$$\tilde{\phi}(r) = \begin{cases} \phi(r), & r \leq 0, \\ 0, & r > 0. \end{cases}$$

We end the section by stating the following well-known lemma.

Lemma 2.3 ([9, Lemma 7.2]) *For any $p \geq 2$, $r \in [0, a]$, and $\psi \in L_p^{\mathscr{F}}(\Omega; L_2([0, a]; L_2^0(H; Y)))$, we have*

$$\mathbb{E}\Big(\sup_{s\in[0,r]} \Big\| \int_0^s \psi(l)dW(l) \Big\|^p\Big) \leq C_p \mathbb{E}\Big(\int_0^r \|\psi(l)\|_Q^2 dl\Big)^{\frac{p}{2}},$$

where $C_p = \left(\frac{p}{2}(p-1)\right)^{\frac{p}{2}} \left(\frac{p}{p-1}\right)^{\frac{p^2}{2}}$.

3 Existence and Uniqueness of Solution

This section is devoted to the study of existence results of mild solution of the system (2.2). We use the method discussed by Luo [30] with appropriate modifications.

Theorem 3.1 *Suppose that* $v \in L_p([0,a];V)$ *and* $\phi \in \mathscr{G}$. *If* (P_1)-(P_4) *are satisfied, then there is a unique mild solution for system* (2.2).

Proof Consider the iteration technique to construct the sequence $\{y^{(m)}(\cdot)\}_{m\in\mathbb{N}}$. Now, define for $m = 1, 2, \ldots$

$$y^{(m)}(r) = \begin{cases} \mathcal{D}(r)\phi(0) + \int_0^r \mathcal{D}(r-l)\Big[L(\tilde{\phi}_l) + B(l)v(l) + f(l, y_l^{(m-1)})\Big]dl \\ + \int_0^r \mathcal{D}(r-l)g(l, y_l^{(m-1)})dW(l), & r \in (0,a], \\ \phi(r) \in L_p^{\mathscr{F}}(\Omega;\mathscr{G}), & r \in (-\infty, 0], \end{cases}$$

and

$$y^{(0)}(r) = \begin{cases} \mathcal{D}(r)\phi(0) + \int_0^r \mathcal{D}(r-l)\Big[L(\tilde{\phi}_l) + B(l)v(l)\Big]dl, & r \in (0,a], \\ \phi(r) \in L_p^{\mathscr{F}}(\Omega;\mathscr{G}), & r \in (-\infty, 0]. \end{cases}$$

For $0 \le l \le r \le a$,

$$y^{(m)}(l) = \mathcal{D}(l)\phi(0) + \int_0^l \mathcal{D}(l-\zeta)\Big[L(\tilde{\phi}_\zeta) + B(\zeta)v(\zeta) + f(\zeta, y_\zeta^{(m-1)})\Big]d\zeta + \int_0^l \mathcal{D}(l-\zeta)g(\zeta, y_\zeta^{(m-1)})dW(\zeta).$$

Remark 2.1 yields that

$$\begin{aligned} &\mathbb{E}\|y^{(m)}(l)\|^p \\ &\le 5^{p-1}\Bigg[K^p\|\mathcal{D}(l)\|^p\|\phi\|_{\mathscr{G}}^p + \mathbb{E}\left\|\int_0^l \mathcal{D}(l-\zeta)L(\tilde{\phi}_\zeta)d\zeta\right\|^p \\ &\quad + \mathbb{E}\left\|\int_0^l \mathcal{D}(l-\zeta)B(\zeta)v(\zeta)d\zeta\right\|^p \\ &\quad + \mathbb{E}\left\|\int_0^l \mathcal{D}(l-\zeta)f(\zeta, y_\zeta^{(m-1)})d\zeta\right\|^p \\ &\quad + \mathbb{E}\left\|\int_0^l \mathcal{D}(l-\zeta)g(\zeta, y_\zeta^{(m-1)})dW(\zeta)\right\|^p\Bigg] \end{aligned}$$

$$
\begin{aligned}
&\leq 5^{p-1}\Bigg[K^p\|\mathcal{D}(l)\|^p\|\phi\|_{\mathscr{G}}^p+\mathbb{E}\bigg(\int_0^l\|\mathcal{D}(l-\zeta)L(\tilde{\phi}_\zeta)\|d\zeta\bigg)^p\\
&\quad+\mathbb{E}\bigg(\int_0^l\|\mathcal{D}(l-\zeta)B(\zeta)v(\zeta)\|d\zeta\bigg)^p\\
&\quad+\mathbb{E}\bigg(\int_0^l\|\mathcal{D}(l-\zeta)f(\zeta,y_\zeta^{(m-1)})\|d\zeta\bigg)^p\\
&\quad+\mathbb{E}\bigg\|\int_0^l\mathcal{D}(l-\zeta)g(\zeta,y_\zeta^{(m-1)})dW(\zeta)\bigg\|^p\Bigg]\\
&\leq 5^{p-1}\Bigg[K^pM^p\|\phi\|_{\mathscr{G}}^p+M^p\sup_{0\leq l\leq r}\mathbb{E}\bigg(\int_0^l\|L(\tilde{\phi}_\zeta)\|d\zeta\bigg)^p\\
&\quad+M^p\sup_{0\leq l\leq r}\mathbb{E}\bigg(\int_0^l\|B(\zeta)v(\zeta)\|d\zeta\bigg)^p\\
&\quad+M^p\sup_{0\leq l\leq r}\mathbb{E}\bigg(\int_0^l\|f(\zeta,y_\zeta^{(m-1)})\|d\zeta\bigg)^p\\
&\quad+\sup_{0\leq l\leq r}\mathbb{E}\bigg\|\int_0^l\mathcal{D}(l-\zeta)g(\zeta,y_\zeta^{(m-1)})dW(\zeta)\bigg\|^p\Bigg].
\end{aligned}
$$

Now, Lemma 2.3, Hölder's inequality, and assumptions (P_2)–(P_3) imply that

$$
\begin{aligned}
&\mathbb{E}\|y^{(m)}(l)\|^p\\
&\leq 5^{p-1}\Bigg[K^pM^p\|\phi\|_{\mathscr{G}}^p+M^p\sup_{0\leq l\leq r}\bigg(\int_0^l 1^q\bigg)^{\frac{p}{q}}\mathbb{E}\bigg(\int_0^l\|L(\tilde{\phi}_\zeta)\|^pd\zeta\bigg)\\
&\quad+M^p\sup_{0\leq l\leq r}\bigg(\int_0^l 1^q\bigg)^{\frac{p}{q}}\mathbb{E}\bigg(\int_0^l\|B(\zeta)v(\zeta)\|^pd\zeta\bigg)+M^p\sup_{0\leq l\leq r}\bigg(\int_0^l 1^q\bigg)^{\frac{p}{q}}\\
&\quad\times\mathbb{E}\bigg(\int_0^l\|f(\zeta,y_\zeta^{(m-1)})\|^pd\zeta\bigg)+C_p\mathbb{E}\bigg(\int_0^r\|\mathcal{D}(l-\zeta)g(\zeta,y_\zeta^{(m-1)})\|_Q^2d\zeta\bigg)^{\frac{p}{2}}\Bigg]\\
&\leq 5^{p-1}\Bigg[K^pM^p\|\phi\|_{\mathscr{G}}^p+M^pa^{\frac{p}{q}}l_0^p\sup_{0\leq l\leq r}\mathbb{E}\bigg(\int_0^l\|\tilde{\phi}_\zeta\|_{\mathscr{G}}^pd\zeta\bigg)\\
&\quad+M^pa^{\frac{p}{q}}N_1\sup_{0\leq l\leq r}\bigg(\int_0^l\big(1+\|y_\zeta^{(m-1)}\|_{\mathscr{G}}^p\big)d\zeta\bigg)\\
&\quad+M^pa^{\frac{p}{q}}M_B^p\sup_{0\leq l\leq r}\mathbb{E}\bigg(\int_0^l\|v(\zeta)\|^pd\zeta\bigg)
\end{aligned}
$$

$$
\begin{aligned}
&+ M^p C_p N_1 \Big(\int_0^r 1^{\frac{p}{p-2}} \Big)^{\frac{p-2}{2}} \mathbb{E} \Big(\int_0^r \big(1 + \|y_\zeta^{(m-1)}\|_{\mathscr{G}}^p\big) d\zeta \Big) \Big] \\
&\le 5^{p-1} M^p \Big[K^p \|\phi\|_{\mathscr{G}}^p + a^{\frac{p}{q}+1} l_0^p \Gamma_a^p \|\phi\|_{\mathscr{G}}^p + M_B^p a^{\frac{p}{q}} \|v\|_{L_p([0,a];V)}^p + a^{\frac{p}{q}+1} N_1 \\
&\quad + a^{\frac{p}{q}} N_1 \mathbb{E} \Big\{ \int_0^r \Big(N_a \sup_{0\le\tau\le\zeta} \|y^{(m-1)}(\tau)\| + \Gamma_a \|\phi\|_{\mathscr{G}} \Big)^p d\zeta \Big\} + C_p N_1 a^{\frac{p}{2}} \\
&\quad + C_p N_1 a^{\frac{p-2}{2}} \mathbb{E} \Big\{ \int_0^r \Big(N_a \sup_{0\le\tau\le\zeta} \|y^{(m-1)}(\tau)\| + \Gamma_a \|\phi\|_{\mathscr{G}} \Big)^p d\zeta \Big\} \Big] \\
&\le R_1 + R_2 \int_0^r \sup_{0\le\tau\le\zeta} \mathbb{E} \|y^{(m-1)}(\tau)\|^p d\zeta. \qquad (3.1)
\end{aligned}
$$

where $R_1 = 5^{p-1} M^p \big[K^p + a^{\frac{p}{q}+1} l_0^p \Gamma_a^p + 2^{p-1} a^{\frac{p}{q}+1} N_1 \Gamma_a^p + 2^{p-1} a^{\frac{p}{2}} C_p N_1 \Gamma_a^p \big] \|\phi\|_{\mathscr{G}}^p$ $+ 5^{p-1} M^p \big[a^{\frac{p}{q}+1} N_1 + C_p a^{\frac{p}{2}} N_1 + M_B^p a^{\frac{p}{q}} \|v\|_{L_p([0,a];V)}^p \big]$ and $R_2 = 5^{p-1} M^p \big(a^{\frac{p}{q}} + C_p a^{\frac{p-2}{2}} \big) N_1$.

For any $k \ge 1$, we have the following inequality:

$$
\max_{1\le m\le k} \mathbb{E} \sup_{0\le l\le r} \|y^{(m-1)}(l)\|^p \le \mathbb{E} \sup_{0\le l\le r} \|y^{(0)}(l)\|^p + \max_{1\le m\le k} \mathbb{E} \sup_{0\le l\le r} \|y^{(m)}(l)\|^p.
$$

Substitution of above inequality in (3.1) yields that

$$
\begin{aligned}
\max_{1\le m\le k} \mathbb{E} \sup_{0\le l\le r} \|y^{(m)}(l)\|^p \le\ & R_1 + R_2 3^{p-1} M^p a \big[K^p \|\phi\|_{\mathscr{G}}^p + a^{\frac{p}{q}+1} l_0^p \Gamma_a^p \|\phi\|_{\mathscr{G}}^p \\
& + M_B^p a^{\frac{p}{q}} \|v\|_{L_p([0,a];V)}^p \big] \\
& + R_2 \int_0^r \max_{1\le m\le k} \mathbb{E} \sup_{0\le\tau\le\zeta} \|y^{(m)}(\tau)\|^p d\zeta.
\end{aligned}
$$

Gronwall inequality and arbitrariness of k imply that

$$
\mathbb{E} \sup_{0\le l\le r} \|y^{(m)}(l)\|^p \le R_3 e^{R_2 a},
$$

where $R_3 = R_1 + R_2 3^{p-1} M^p a \big[K^p \|\phi\|_{\mathscr{G}}^p + a^{\frac{p}{q}+1} l_0^p \Gamma_a^p \|\phi\|_{\mathscr{G}}^p + M_B^p a^{\frac{p}{q}} \|v\|_{L_p([0,a];V)}^p \big]$.

Since $\|v\|_{L_p([0,a];V)}^p < \infty$, we deduce that

$$
\sup_{0\le l\le r} \mathbb{E} \|y^{(m)}(l)\|^p < \infty, \quad m \in \mathbb{N}.
$$

Thus, we assert that the sequence $\{y^{(m)}(r)\}_{m\in\mathbb{N}}$ is bounded. Next, we claim that the sequence $\{y^{(m)}(r)\}_{m\in\mathbb{N}}$ is Cauchy. For any $0 \le l \le a$,

$$y^{(1)}(l) - y^{(0)}(l) = \int_0^l \mathcal{D}(l-\zeta) f(\zeta, y_\zeta^{(0)}) d\zeta + \int_0^l \mathcal{D}(l-\zeta) g(\zeta, y_\zeta^{(0)}) dW(\zeta).$$

Applying the same procedure as we did to get (3.1), we infer that

$$\begin{aligned}
&\sup_{0\le l\le r} \mathbb{E}\|y^{(1)}(l) - y^{(0)}(l)\|^p \\
&\quad \le 2^{p-1} M^p \big(a^{\frac{p}{q}+1} + C_p a^{\frac{p}{2}}\big) + 4^{p-1} M^p a \big(a^{\frac{p}{q}} + C_p a^{\frac{p-2}{2}}\big) \Gamma_a^p \|\phi\|_{\mathscr{G}}^p \\
&\qquad + 12^{p-1} N_a^p M^{2p} a^2 \big(a^{\frac{p}{q}} + C_p a^{\frac{p-2}{2}}\big) \big[K^p \|\phi\|_{\mathscr{G}}^p + a^{\frac{p}{q}+1} l_0^p \Gamma_a^p \|\phi\|_{\mathscr{G}}^p \\
&\qquad + M_B^p a^{\frac{p}{q}} \|v\|^p_{L_p([0,a];V)}\big].
\end{aligned}$$

Also, for any $0 \le l \le a$,

$$\begin{aligned}
y^{(m)}(l) - y^{(m-1)}(l) = &\int_0^l \mathcal{D}(l-\zeta)\Big[f(\zeta, y_\zeta^{(m-1)}) - f(\zeta, y_\zeta^{(m-2)})\Big] d\zeta \\
&+ \int_0^l \mathcal{D}(l-\zeta)\Big[g(\zeta, y_\zeta^{(m-1)}) - g(\zeta, y_\zeta^{(m-2)})\Big] dW(\zeta).
\end{aligned}$$

Then, it follows that

$$\begin{aligned}
&\sup_{0\le l\le r} \mathbb{E}\|y^{(m)}(l) - y^{(m-1)}(l)\|^p \\
&\quad \le 2^{p-1} \sup_{0\le l\le r} \mathbb{E}\Big(\Big\| \int_0^l \mathcal{D}(l-\zeta)\Big[f(\zeta, y_\zeta^{(m-1)}) - f(\zeta,\ y_\zeta^{(m-2)})\Big] d\zeta \Big\|^p \Big) \\
&\qquad + 2^{p-1} \sup_{0\le l\le r} \mathbb{E}\Big(\Big\| \int_0^l \mathcal{D}(l-\zeta)\Big[g(\zeta, y_\zeta^{(m-1)}) - g(\zeta, y_\zeta^{(m-2)})\Big] dW(\zeta) \Big\|^p \Big) \\
&\quad \le 2^{p-1} \sup_{0\le l\le r} \mathbb{E}\Big(\int_0^l \Big\| \mathcal{D}(l-\zeta)\Big[f(\zeta, y_\zeta^{(m-1)}) - f(\zeta, y_\zeta^{(m-2)})\Big] \Big\| d\zeta \Big)^p \\
&\qquad + 2^{p-1} C_p \mathbb{E}\Big(\int_0^r \Big\| \mathcal{D}(l-\zeta)\Big[g(\zeta, y_\zeta^{(m-1)}) - g(\zeta, y_\zeta^{(m-2)})\Big] \Big\|_Q^2 d\zeta \Big)^{p/2} \\
&\quad \le 2^{p-1} M^p \sup_{0\le l\le r} \Big(\int_0^l 1^q \Big)^{p/q} \mathbb{E}\Big(\int_0^l \Big\| f(\zeta, y_\zeta^{(m-1)}) - f(\zeta, y_\zeta^{(m-2)}) \Big\|^p d\zeta \Big) \\
&\qquad + 2^{p-1} M^p C_p \Big(\int_0^r 1^{p/p-2} \Big)^{p-2/2} \mathbb{E}\Big(\int_0^r \Big\| g(\zeta, y_\zeta^{(m-1)}) - g(\zeta, y_\zeta^{(m-2)}) \Big\|_Q^p d\zeta \Big)
\end{aligned}$$

$$\leq 2^{p-1} M^p a^{\frac{p}{q}} N_1 \int_0^r \mathbb{E}\|y_\zeta^{(m-1)} - y_\zeta^{(m-2)}\|_{\mathscr{G}}^p d\zeta$$
$$+ 2^{p-1} M^p C_p a^{\frac{p-2}{2}} N_1 \int_0^r \mathbb{E}\|y_\zeta^{(m-1)} - y_\zeta^{(m-2)}\|_{\mathscr{G}}^p d\zeta$$
$$= \bar{b}(p,a) \int_0^r \mathbb{E}\|y_\zeta^{(m-1)} - y_\zeta^{(m-2)}\|_{\mathscr{G}}^p d\zeta.$$

On repeating the above process successively, we get

$$\sup_{0\leq l\leq r} \mathbb{E}\|y^{(m)}(l) - y^{(m-1)}(l)\|^p$$
$$\leq \frac{\left(a\,\bar{b}(p,a)\right)^{m-1} (N_a^p)^{m-1}}{(m-1)!} \sup_{0\leq l\leq r} \mathbb{E}\|y^{(1)}(l) - y^{(0)}(l)\|^p.$$

Thus, $\{y^{(m)}(\cdot)\}_{m\in\mathbb{N}}$ is a Cauchy sequence in Y. Hence, by the standard Borel–Cantelli lemma, $y^{(m)}(\cdot) \to y(\cdot)$ uniformly on $[0,a]$ as $m \to \infty$, and $y(\cdot)$ is the unique mild solution of (2.2). □

4 Existence of Optimal Control

This section deals with the existence of an optimal pair of the state and control functions. We also need the following assumption:

(P_5) For the functional $\mathcal{J} : [0,a] \times Y \times \mathscr{G} \times V \to \mathbb{R} \cup \{\infty\}$, the following hold:

(i) $\mathcal{J}$ is $\mathscr{F}_r$-measurable.
(ii) For almost all $0 \leq r \leq a$, $\mathcal{J}(r,\cdot,\cdot,\cdot)$ is sequentially lower semicontinuous on $Y \times \mathscr{G} \times V$.
(iii) For almost all $0 \leq r \leq a$ and for each $z \in Y$, $\varphi \in \mathscr{G}$, $\mathcal{J}(r,z,\varphi,\cdot)$ is convex on V.
(iv) Let $\sigma \in L_1([0,a];\mathbb{R})$ be a nonnegative function such that

$$\sigma(r) + b_1\|z\| + b_2\|\varphi\|_{\mathscr{G}} + d_1\|v\|_V^p \leq \mathcal{J}(r,z,\varphi,v),$$

where $b_1, b_2 \geq 0$ and $d_1 > 0$ are constants.

Theorem 4.1 *Suppose that all assumptions in Theorem 3.1 and (P_5) hold. If B is strongly continuous, then the problem (2.1) admits at least one optimal pair.*

Proof The proof is motivated by the research of Balasubramaniam and Tamilalagan [3] and Kumar [21]. Consider the set

$$\mathcal{P}_{ad} = \{(y, v) : y \text{ satisfies (2.6) with control function } v \in \mathcal{A}_{ad}\}.$$

Now, without loss of all inclusive statements, we can accept that

$$\inf\{\mathcal{I}(y, v) : (y, v) \in C([0, a]; L_p^{\mathcal{F}}(\Omega; Y)) \times \mathcal{A}_{ad}\} = \rho < \infty.$$

Then, assumption (P_5) implies that $\rho > -\infty$. Clearly, there is a minimizing sequence of feasible pairs $\{(y^k, v^k)\} \subset \mathcal{P}_{ad}$, which converges to ρ as $k \to \infty$. Since the set $\{v^k\}_{k\in\mathbb{N}} \subseteq \mathcal{A}_{ad}$ is bounded in $L_p([0, a]; V)$, there exists a subsequence, still represented by $\{v^k\}$ and some $\tilde{v} \in L_p([0, a]; V)$ such that v^k weakly converges to $\tilde{v}$ ($v^k \xrightarrow{w} \tilde{v}$) in $L_p([0, a]; V)$. It is readily verified that the set $\mathcal{A}_{ad}$ is closed and convex. Therefore, by the Marzur lemma, we assert that $\tilde{v} \in \mathcal{A}_{ad}$. Suppose that y^k and $\tilde{y}$ satisfy (2.6) with controls v^k and $\tilde{v}$, respectively. That is,

$$\begin{aligned} y^k(r) &= \mathcal{D}(r)\phi(0) + \int_0^r \mathcal{D}(r-l)\Big[L(\tilde{\phi}_l) + f(l, y_l^k) + B(l)v^k(l)\Big]dl \\ &\quad + \int_0^r \mathcal{D}(r-l)g(l, y_l^k)dW(l), \end{aligned}$$

$$\begin{aligned} \tilde{y}(r) &= \mathcal{D}(r)\phi(0) + \int_0^r \mathcal{D}(r-l)\Big[L(\tilde{\phi}_l) + f(l, \tilde{y}_l) + B(l)\tilde{v}(l)\Big]dl \\ &\quad + \int_0^r \mathcal{D}(r-l)g(l, \tilde{y}_l)dW(l). \end{aligned}$$

If $r \le 0$, then $\mathbb{E}\|y^k(r) - \tilde{y}(r)\|^p = 0$. For $r \in [0, a]$, by (P_2) and (P_4), Hölder's inequality, Remark 2.1, and Lemma 2.3, we obtain

$$\begin{aligned} &\mathbb{E}\|y^k(r) - \tilde{y}(r)\|^p \\ &\le 3^{p-1} \sup_{0\le r\le a} \mathbb{E}\left\|\int_0^r \mathcal{D}(r-l)\Big[f(l, y_l^k) - f(l, \tilde{y}_l)\Big]dl\right\|^p \\ &\quad + 3^{p-1} \sup_{0\le r\le a} \mathbb{E}\left\|\int_0^r \mathcal{D}(r-l)\Big[B(l)v^k(l) - B(l)\tilde{v}(l)\Big]dl\right\|^p \\ &\quad + 3^{p-1} \sup_{0\le r\le a} \mathbb{E}\left\|\int_0^r \mathcal{D}(r-l)\Big[g(l, y_l^k) - g(l, \tilde{y}_l)\Big]dW(l)\right\|^p \end{aligned}$$

$$
\begin{aligned}
&\le 3^{p-1}\Bigg[\sup_{0\le r\le a}\mathbb{E}\bigg(\int_0^r \|\mathcal{D}(r-l)\Big[f(l,y_l^k)-f(l,\tilde{y}_l)\Big]\|dl\bigg)^p\\
&\quad+\sup_{0\le r\le a}\mathbb{E}\bigg(\int_0^r \|\mathcal{D}(r-l)\Big[B(l)v^k(l)-B(l)\tilde{v}(l)\Big]\|dl\bigg)^p\\
&\quad+\sup_{0\le r\le a}\mathbb{E}\left\|\int_0^r \mathcal{D}(r-l)\Big[g(l,y_l^k)-g(l,\tilde{y}_l)\Big]dW(l)\right\|^p\Bigg]\\
&\le 3^{p-1}M^p\Bigg[a^{p/q}N_1\sup_{0\le r\le a}\int_0^r \left\|y_l^k-\tilde{y}_l\right\|_{\mathscr{G}}^p dl+a^{p/q}\left\|Bv^k-B\tilde{v}\right\|_{L_p([0,a];V)}^p\\
&\quad+C_p\bigg(\int_0^r 1^{\frac{p}{p-2}}dl\bigg)^{\frac{p-2}{2}}\mathbb{E}\bigg(\int_0^a \left\|g(l,y_l^k)-g(l,\tilde{y}_l)\right\|_{\mathcal{Q}}^p dl\bigg)\Bigg]\\
&\le 3^{p-1}M^p\Bigg[a^{p/q}N_1\int_0^a \left\|y_l^k-\tilde{y}_l\right\|_{\mathscr{G}}^p dl+a^{p/q}\left\|Bv^k-B\tilde{v}\right\|_{L_p([0,a];V)}^p\\
&\quad+C_pa^{p-2/2}N_1\int_0^a \left\|y_l^k-\tilde{y}_l\right\|_{\mathscr{G}}^p dl\Bigg]\\
&\le 3^{p-1}M^pa^{p/q}\left\|Bv^k-B\tilde{v}\right\|_{L_p([0,a];V)}^p\\
&\quad+3^{p-1}M^pN_1\Big(a^{p/q}+C_pa^{p-2/2}\Big)N_a^p\int_0^a\sup_{0\le z\le l}\left\|y^k(z)-\tilde{y}(z)\right\|^p dl.
\end{aligned}
$$

Now, Gronwall's lemma yields that

$$
\sup_{0\le r\le a}\mathbb{E}\|y^k(r)-\tilde{y}(r)\|^p \le M^*\left\|Bv^k-B\tilde{v}\right\|_{L_p([0,a];V)}^p, \tag{4.1}
$$

where M^* is independent of v, k, and r.

By the strong continuity of B, we infer that $\left\|Bv^k-B\tilde{v}\right\|_{L_p([0,a];V)}^p \xrightarrow{w} 0$ as $k\to\infty$.

From (4.1), $\mathbb{E}\|y^k(r)-\tilde{y}(r)\|^p \xrightarrow{w} 0$ as $k\to\infty$, and hence

$$
y^k \xrightarrow{w} \tilde{y} \text{ in } C([0,a];L_p^{\mathscr{F}}(\Omega;Y)) \text{ as } k\to\infty.
$$

Assumptions of Balder [4, Theorem 2.1] hold due to hypothesis (P_5), and hence we conclude that $(y^v,v)\longrightarrow \mathbb{E}\Big\{\int_0^a \mathcal{J}(r,y^v(r),y_r^v,v(r))dr\Big\}$ is sequentially lower semicontinuous in the weak topology of $L_p([0,a];V)\subset L_1([0,a];V)$ and

strong topology of $L_1([0,a];Y)$. Therefore, $\mathcal{I}$ is weakly lower semicontinuous on $L_p([0,a];V)$, and by $P_5(iv)$, $\mathcal{I}$ attains its infimum at $\tilde{v} \in \mathcal{A}_{ad}$. Thus,

$$\begin{aligned}\rho &= \lim_{k\to\infty} \mathbb{E}\left\{ \int_0^a \mathcal{J}(r,\ y^k(r),\ y_r^k,\ v^k(r))dr \right\} \\ &\geq \mathbb{E}\left\{ \int_0^a \mathcal{J}(r,\ \tilde{y}(r),\ \tilde{y}_r,\ \tilde{v}(r))dr \right\} = \mathcal{I}(\tilde{y},\tilde{v}) \geq \rho.\end{aligned}$$

The proof is complete. □

5 Example

Consider the following infinite-dimensional semilinear stochastic system with unbounded delay:

$$\begin{cases} du(r,y) = \Big(\frac{\partial^2}{\partial y^2}u(r,y) + \int_{-\infty}^{r-1}\int_0^\pi w(l-r,y,z)u(l,z)dz\,dl \\ \quad + \int_{-\infty}^{r} P(r,l)F_1(l,u(l,y))dl + \int_0^\pi E(y,l)v(l,r)dl\Big)dr \\ \quad + \int_{-\infty}^{r} R(r,l)G_1(l,u(l,y))dl\,d\beta(r), \quad 0<r\leq 2,\ \ 0\leq y\leq\pi, \\ u(r,0) = u(r,\pi) = 0, \quad 0\leq r\leq 2, \\ u(\eta,y) = \phi_0(\eta,y), \quad \eta\leq 0, \quad 0\leq y\leq\pi, \end{cases} \tag{5.1}$$

where $\phi_0(\cdot,\cdot)$ is $\mathscr{F}_0$-measurable, $w(\cdot,\cdot,\cdot)$, $P(\cdot,\cdot)$, $E(\cdot,\cdot)$, $R(\cdot,\cdot)$, $F_1(\cdot,\cdot)$, and $G_1(\cdot,\cdot)$ are functions to be defined later, and $\beta(r)$ is one-dimensional standard Brownian motion in Y on $(\Omega,\mathscr{F},\{\mathscr{F}_t\},\mathbf{P})$. The system (5.1) represents a Volterra stochastic integro-differential equation. Let $Y = L_2([0,\pi])$, $H = \mathbb{R}$, and $V = L_2([0,2])$. Then, define operator $A : D(A) \subset Y \to Y$ by $Av = -v''$, where $D(A) = \left\{v \in Y :\ \frac{\partial v}{\partial y},\ \frac{\partial^2 v}{\partial y^2} \in Y,\ \text{and}\ \ v(0) = v(\pi) = 0\right\}$.

Clearly, A generates an analytic C_0-semigroup $\{S(r)\}_{r\geq 0}$, which is compact and self-adjoint. Furthermore, A has the eigenvalues m^2, $m \in \mathbb{N}$, and $e_m(y) = \sqrt{\frac{2}{\pi}}\sin(my)$, $m \in \mathbb{N}$, are the corresponding normalized eigenvectors. Now, the following properties hold:

(i) For $\zeta \in D(A)$,

$$A\zeta = \sum_{m=1}^{\infty} m^2\langle\zeta, e_m\rangle e_m.$$

(ii) For every $\zeta \in Y$,

$$S(r)\zeta = \sum_{m=1}^{\infty} e^{-m^2 r}\langle \zeta, e_m\rangle e_m.$$

Thus, (P_1) is satisfied for the operator A.

Consider the phase space $\mathscr{G} = C_0 \times L_p(g_1 : Y), \quad 1 < p < \infty$, with the norm

$$\|\varphi\|_{\mathscr{G}} = \|\varphi(0)\| + \Big(\int_{-\infty}^{0} g_1(\eta)\|\varphi(\eta)\|^p d\eta\Big)^{\frac{1}{p}},$$

where g_1 and $g_1\|\varphi(\eta)\|^p$ are real-valued Lebesgue integrable functions on $(-\infty, 0)$ and φ is continuous at 0.

It is notable that $\mathscr{G}$ satisfies the axioms (H_1), (H_2), and (H_3) for a properly chosen function g_1. The assumptions (b_1) and (b_2) in Sect. 2 also hold (see [15]).

Now, suppose that for the system (5.1), the following hold:

(i) For $\eta \le 0$, $(\eta, \cdot, \cdot) \in C([0, \pi] \times [0, \pi])$ is measurable and $w(\eta, 0, \cdot) = w(\eta, \pi, \cdot) = 0$, $\eta \le 0$. Moreover, $l_0 = \int_0^{\pi}\Big(\int_{-\infty}^{-1} \frac{1}{(g_1(\eta))^{\frac{q}{p}}}\int_0^{\pi} |w(\eta, y, z)|^q dz\, d\eta\Big)^{\frac{2}{q}} dy < \infty$, with $\frac{1}{p} + \frac{1}{q} = 1$.

(ii) The functions $F_1, G_1 : \mathbb{R} \times Y \to \mathbb{R}$ are continuous, uniformly bounded, and Lipschitz continuous in the second variable with Lipschitz constant say C_1 and D_1, respectively.

(iii) The functions $P, R : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are continuous and satisfy the inequality $|P(r, r+\eta)| \vee |R(r, r+\eta)| < h_1(\eta)$ with $\Big(\int_{-\infty}^{0} \frac{1}{(g_1(\eta))^{\frac{q'}{p'}}} |h_1(\eta)|^{q'} d\eta\Big)^{\frac{2}{q'}} < \infty$, $\frac{1}{p'} + \frac{1}{q'} = 1$.

(iv) The function $\phi_0(\eta, y) \in L_p^{\mathscr{F}}(\Omega; \mathscr{G})$.

(v) The function $E : [0, \pi] \times [0, \pi] \to \mathbb{R}$ is continuous.

Let $v : \mathcal{T}y([0, \pi]) \to \mathbb{R}$ be such that the map $r \to v(\cdot, r)$ is measurable and $v \in L_2(\mathcal{T}y([0, \pi]))$ as the controls. Let $\mathcal{A} = \{v \in V : \|v\|_V \le \mu_1\}$, where $\mu_1 \in L_2([0, 2]; \mathbb{R}^+)$, and $A_{ad} = \{v \in L_2(\mathcal{T}y([0, \pi])) : \|v(\cdot, r)\|_{L_2[(0,2)]} \le \mu_1(r)$ a.e. $r \in [0, 2]\}$. To represent the system (5.1) as the abstract form given by (2.2), set $Y(r)(\cdot) = u(r, \cdot)$ and $\phi(r)(\cdot) = \phi_0(r, \cdot)$. Define $L : \mathscr{G} \to Y$, $B(r) : V \to Y$, $f(\cdot, \cdot) : [0, 2] \times \mathscr{G} \to Y$ and $g(\cdot, \cdot) : [0, 2] \times \mathscr{G} \to \mathbb{R}$ by

$$L(\varphi)(y) = L(\varphi(\cdot, y)) = \int_{-\infty}^{-1}\int_0^{\pi} w(\eta, y, z)\, \varphi(\eta, z) dz\, d\eta,$$

$$B(r)v(r)(y) = \int_0^\pi E(y,l)v(l,r)dl,$$

$$f(r,\varphi)(y) = f(r,\varphi(\cdot,y)) = \int_{-\infty}^0 P(r,r+\eta)F_1(r+\eta,\varphi(\eta,y))\,d\eta,$$

$$g(r,\varphi)(y) = g(r,\varphi(\cdot,y)) = \int_{-\infty}^0 R(r,r+\eta)G_1(r+\eta,\varphi(\eta,y))\,d\eta.$$

For any $r \in [0,2]$ and $\varphi_1,\ \varphi_2 \in \mathscr{G}$, assumptions (ii) and (iii) yield that

$$\begin{aligned}
&|f(r,\varphi_1)(y) - f(r,\varphi_2)(y)|^2 \\
&\le \left(\int_{-\infty}^0 |P(r,r+\eta)|\ |F_1(r+\eta,\varphi_1(\eta,y)) - F_1(r+\eta,\varphi_2(\eta,y))|\,d\eta\right)^2 \\
&\le \left(\int_{-\infty}^0 |h_1(\eta)|\ C_1\|\varphi_1(\eta)-\varphi_2(\eta)\|\,d\eta\right)^2 \\
&\le C_1^2\left(\int_{-\infty}^0 \frac{1}{g_1(\eta)^{\frac{1}{p}}}\ g_1(\eta)^{\frac{1}{p}}\ |h_1(\eta)|\ \|\varphi_1(\eta)-\varphi_2(\eta)\|\,d\eta\right)^2 \\
&\le C_1^2\left(\int_{-\infty}^0 \frac{1}{g_1(\eta)^{\frac{q}{p}}}\ |h_1(\eta)|^q\,d\eta\right)^{\frac{2}{q}} \left(\int_{-\infty}^0 g_1(\eta)\ \|\varphi_1(\eta)-\varphi_2(\eta)\|^p\,d\eta\right)^{\frac{2}{p}} \\
&\le C_2\|\varphi_1-\varphi_2\|_{\mathscr{G}}^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|f(r,\varphi_1) - f(r,\varphi_2)\|^p &= \left(\int_0^\pi |f(r,\varphi_1)(y) - f(r,\varphi_2)(y)|^2 dy\right)^{\frac{p}{2}} \\
&\le \left(\int_0^\pi C_2\|\phi_1-\phi_2\|_{\mathscr{G}}^2 dy\right)^{\frac{p}{2}} \\
&\le C_3\|\varphi_1-\varphi_2\|_{\mathscr{G}}^p,
\end{aligned}$$

for some constant $C_3 > 0$. Similarly, there exists a constant $C_4 > 0$ such that

$$\|g(r,\varphi_1) - g(r,\varphi_2)\|^p \le C_4\|\varphi_1-\varphi_2\|_{\mathscr{G}}^p.$$

The uniform boundedness of $P(\cdot,\cdot)$, $R(\cdot,\cdot)$, $F_1(\cdot,\cdot)$, and $G_1(\cdot,\cdot)$ implies that f and g are also uniformly bounded, and hence the hypothesis (P_4) is satisfied. Next, for any $\varphi \in \mathscr{G}$, by assumption (i),

$$\begin{aligned}
&\|L(\varphi)\|^2 \\
&= \int_0^\pi \left| \int_{-\infty}^{-1} \int_0^\pi w(\eta, y, z)\, \varphi(\eta, z) dz\, d\eta \right|^2 dy \\
&\le \int_0^\pi \left(\int_{-\infty}^{-1} \int_0^\pi |w(\eta, y, z)\, \varphi(\eta, z)|\, dz\, d\eta \right)^2 dy \\
&\le \int_0^\pi \left[\int_{-\infty}^{-1} \left(\int_0^\pi |w(\eta, y, z)|^q dz \right)^{\frac{1}{q}} \left(\int_0^\pi |\varphi(\eta, z)|^p dz \right)^{\frac{1}{p}} d\eta \right]^2 dy \\
&\le \int_0^\pi \left[\left(\int_{-\infty}^{-1} \frac{1}{(g_1(\eta))^{\frac{q}{p}}} \int_0^\pi |w(\eta, y, z)|^q dz\, d\eta \right)^{\frac{2}{q}} \right. \\
&\quad \left. \times \left(\int_{-\infty}^{-1} g_1(\eta) \int_0^\pi |\varphi(\eta, z)|^p dz\, d\eta \right)^{\frac{2}{p}} \right] dy \\
&\le \int_0^\pi \left(\int_{-\infty}^{-1} \frac{1}{(g_1(\eta))^{\frac{q}{p}}} \int_0^\pi |w(\eta, y, z)|^q dz\, d\eta \right)^{\frac{2}{q}} dy\, \|\varphi\|_{\mathscr{G}}^2,
\end{aligned}$$

where $\frac{1}{p} + \frac{1}{q} = 1$. This shows that (P_3) is satisfied.

Now, consider the cost function $\mathcal{I}(v) = \mathbb{E}\left\{ \int_0^2 \mathcal{J}(r,\, u^v(r),\, u_r^v,\, v(r))dr \right\}$, where

$$\begin{aligned}
\mathcal{J}(r,\, u^v(r),\, u_r^v,\, v(r)) \;=\; & \int_0^\pi \Big(\|u(r, y)\|^2 + \|v(r, y)\|^2 \Big) dy \\
& + \int_0^\pi \int_{-\infty}^0 \|u(r + l, y)\|^2 dl\, dy,
\end{aligned}$$

with respect to the system (5.1). Since all the hypotheses of Theorem 4.1 are satisfied, the system (5.1) has at least one optimal state–control pair.

6 Conclusion

The existence of solutions to a given system is a fundamental need to study the optimal control. Some of the suitable and effectively confirmed conditions to ensure the solvability of a nonlinear differential system are linear growth and Lipschitz condition. So, it is interesting to study the optimality results under these conditions.

The existence and uniqueness of mild solution of the system (2.2) are studied by using the theory of fundamental solution and the successive approximation method. It is additionally demonstrated that the Lagrangian problem has at least an optimal state–control pair under some natural hypotheses. Studies exhibit that in the modeling of several dynamical systems, it is essential to include both standard and fractional Brownian motions. Therefore, in future, we might want to tackle the above issue for mixed fractional Brownian motion.

Conflict of Interest Shobha Yadav and Surendra Kumar declare that there is no conflict of interest.

References

1. N.U. Ahmed, K.L. Teo, *Optimal Control of Distributed Parameter Systems* (Elsevier Science, New York, 1981)
2. P. Balasubramaniam, S.K. Ntouyas, Controllability for neutral stochastic functional differential inclusions with infinite delay in abstract space. Math. Anal. Appl. **324**(1), 161–176 (2006). https://doi.org/10.1016/j.jmaa.2005.12.005
3. P. Balasubramaniam, P. Tamilalagan, The solvability and optimal controls for impulsive fractional stochastic integro-differential equations via resolvent operators. J. Optim. Theory Appl. **174**(1), 139–155 (2017). https://doi.org/10.1007/s10957-016-0865-6
4. E.J. Balder, Necessary and sufficient conditions for L_1-strong weak lower semicontinuity of integral functionals. Nonlinear Anal. **11**(12), 1399–1404 (1987). https://doi.org/10.1016/0362-546X(87)90092-7
5. R. Buckdahn, A. Rascanu, On the existence of stochastic optimal control of distributed state system. Nonlinear Anal. **52**(4), 1153–1184 (2003). https://doi.org/10.1016/S0362-546X(02)00158-X
6. S. Chen, J. Yong, Stochastic linear quadratic optimal control problems. Appl. Math. Comput. **43**(1), 21–45 (2001). https://doi.org/10.1007/s002450010016
7. S. Chen, X. Li, S. Peng, J. Yong, A linear quadratic optimal control problem with disturbances-an algebraic Riccati equation and differential games approach. Appl. Math. Comput. **30**, 267–305 (1994). https://doi.org/10.1007/BF01183014
8. R.F. Curtain, A.J. Pritchard, *Infinite Dimensional Linear Systems Theory*. Lecture Notes in Control and Information Science (Springer, Berlin, 1978)
9. G. Da Prato, J. Zabczyk, *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and Its Applications (Cambridge University Press, Cambridge, 1992)
10. G. Da Prato, J. Zabczyk, *Second Order Partial Differential Equations in Hilbert Spaces*. London Mathematical Society Lecture Note Series (Cambridge University Press, Cambridge, 2002)
11. L. Glass, M.C. Mackey, *From Clocks to Chaos, The Rhythms of Life* (Princeton University Press, Princeton, 1988)
12. W. Grecksch, C. Tudor, *Stochastic Evolution Equations: A Hilbert Space Approach* (AKademic-Verlag, Berlin, 1995)
13. H.F. Guliyev, H.T. Tagiyev, An optimal control problem with nonlocal conditions for the weakly nonlinear hyperbolic equation. Optimal Control Appl. Methods **34**(2), 216–235 (2013). https://doi.org/10.1002/oca.2018
14. J.K. Hale, J. Kato, Phase space for retarded equations with infinite delay. Funk. Ekvac. **21**, 11–41 (1978)

15. Y. Hino, S. Murakami, T. Naito, *Functional Differential Equations with Infinite Delay, Lecture Notes in Mathematics*, vol. 1473 (Springer, Berlin, 1991)
16. Y. Hu, F. Wu, C. Huang, Stochastic stability of a class of unbounded delay neutral stochastic differential equations with general decay rate. Int. J. Syst. Sci. **43**(2), 308–318 (2012). https://doi.org/10.1080/00207721.2010.495188
17. K. Ikeda, K. Matsumoto, High-dimensional chaotic behaviour in systems with time-delayed feedback. Phys. D. **29**, 1–2 (1987). https://doi.org/10.1016/0167-2789(87)90058-3
18. K. Ikeda, H. Daido, O. Akimoto, Optical turbulence: chaotic behaviour of transmitted light from a ring cavity. Phys. Rev. Lett. **45**(9), 709–712 (1980). https://doi.org/10.1103/PhysRevLett.45.709
19. J.M. Jeong, J.R. Kim, H.H. Roh, Optimal control problems for semilinear evolution equations. J. Korean Math. Soc. **45**(3),757–769 (2008)
20. J.M. Jeong, E.Y. Ju, S.J. Cheon, Optimal control problems for evolution equations of parabolic type with nonlinear perturbations. J. Optim. Theory Appl. **151**(3), 573–588 (2011). https://doi.org/10.1007/s10957-011-9866-7
21. S. Kumar, Mild solution and fractional optimal control of semilinear system with fixed delay. J. Optim. Theory Appl. **174**(1), 108–121 (2017). https://doi.org/10.1007/s10957-015-0828-3
22. I. Lasiecka, R. Triggiani, *Differential and Algebraic Riccati Equations with Applications to Boundary/Point Control Problems: Continuous Theory and Approximation Theory*. Lecture Notes in Control and Information Sciences, vol. 164 (Springer, Berlin, 1991)
23. X.J. Li, J.M. Yong, *Optimal Control Theory for Infinite Dimensional Systems, Systems & Control: Foundations & Applications* (Birkhäuser, Boston, 1995)
24. Z. Li, K. Liu, J. Luo, On almost periodic mild solutions for neutral stochastic evolution equations with infinite delay. Nonlinear Anal. **110**, 182–190 (2014). https://doi.org/10.1016/j.na.2014.08.005
25. N. Li, Y. Wang, Z. Wu, An indefinite stochastic linear quadratic optimal control problem with delay and related forward–backward stochastic differential equations. J. Optim. Theory Appl. **179**(2), 722–744 (2018). https://doi.org/10.1007/s10957-018-1237-1
26. K. Liu, The fundamental solution and its role in the optimal control of infinite dimensional neutral systems. Appl. Math. Optim. **60**(1), 1–38 (2009). https://doi.org/10.1007/s00245-009-9065-1
27. K. Liu, Existence of invariant measures of stochastic systems with delay in the highest order partial derivatives. Stat. Prob. Lett. **94**, 267–272 (2014). https://doi.org/10.1016/j.spl.2014.07.028
28. J. Liu, M. Xiao, A leapfrog semi-smooth Newton-multigrid method for semilinear parabolic optimal control problems. Comput. Optim. Appl. **63**(1), 69–95 (2016). https://doi.org/10.1007/s10589-015-9759-z
29. J. Luo, Stability of stochastic partial differential equations with infinite delays. J. Comput. Appl. Math. **222**(2), 364–371 (2008). https://doi.org/10.1016/j.cam.2007.11.002
30. J. Luo, Exponential stability for stochastic neutral partial functional differential equations. J. Math. Anal. Appl. **355**(1), 414–425 (2009). https://doi.org/10.1016/j.jmaa.2009.02.001
31. M.C. Mackey, L. Glass, Oscillation and chaos in physiological control systems. Science **197**, 287–289 (1977)
32. F.Z. Mokkedem, X. Fu, Approximate controllability for a semilinear evolution system with infinite delay. J. Dyn. Control Syst. **22**(1), 71–89 (2016). https://doi.org/10.1007/s10883-014-9252-5
33. F.Z. Mokkedem, X. Fu, Approximate controllability for a semilinear stochastic evolution system with infinite delay in L_p space. Appl. Math. Optim. **75**(2), 253–283 (2017). https://doi.org/10.1007/s00245-016-9332-x
34. F.Z. Mokkedem, X. Fu, Optimal control problems for a semilinear evolution system with infinite delay. Appl. Math. Optim. **79**(1)(2017). https://doi.org/10.1007/s00245-017-9420-6
35. S. Nakagiri, Optimal control of linear retarded systems in Banach spaces. J. Math. Anal. Appl. **120**(1), 169–210 (1986). https://doi.org/10.1016/0022-247X(86)90210-6

36. N.S. Papageorgiou, On the optimal control of strongly nonlinear evolution equations. J. Math. Anal. Appl. **164**, 83–103 (1992). https://doi.org/10.1016/0022-247X(92)90146-5
37. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences, vol. 44 (Springer, New York, 1983)
38. K. Sobczyk, *Stochastic Differential Equations with Applications to Physics and Engineering* (Kluwer Academic, London, 1991)
39. H. Tanabe, *Equations of Evolution, Monographs and Studies in Mathematics* (Pitman (Advanced Publishing Program), London, 1979)
40. T. Taniguchi, Almost sure exponential stability for stochastic partial functional differential equations. Stoch. Anal. Appl. **16**(5), 965–975 (1998). https://doi.org/10.1080/07362999808809573
41. T. Taniguchi, K. Liu, A. Truman, Existence, uniqueness, and asymptotic behaviour of mild solutions to stochastic functional equations in Hilbert spaces. J. Differ. Equ. **181**(1), 72–91 (2002). https://doi.org/10.1006/jdeq.2001.4073
42. W. Wang, B. Wang, *Existence of the Optimal Control for Stochastic Boundary Control Problems Governed by Semilinear Parabolic Equations*. Math. Probl. Eng. **2014**, 534604 (2014). Hindawi Publishing Corporation
43. J.R. Wang, Y. Zhou, M. Medved', On the solvability and optimal controls of fractional integrodifferential evolution systems with infinite delay. J. Optim. Theory Appl. **152**(1), 31–50 (2012). https://doi.org/10.1007/s10957-011-9892-5
44. W. Witayakiattilerd, Nonlinear fuzzy differential equation with time delay and optimal control problem. Abstr. Appl. Anal. **14**, 659072 (2015). https://doi.org/10.1155/2015/659072
45. X. Xiaoling, K. Huawu, Delay systems and optimal control. Acta Math. Appl. Sin. **16**(1), 27–35 (2000). https://doi.org/10.1007/BF02670961
46. X.Y. Zhou, On the necessary conditions of optimal controls for stochastic partial differential equations. SIAM J. Control Optim. **31**(6), 1462–1478 (1993). https://doi.org/10.1137/0331068

Part III
High Performance and Scientific Computing

The Role of Machine Learning and Artificial Intelligence for High-Performance Computing

Michael M. Resch

Abstract High-performance computing has recently been challenged by the advent of data analytics, machine learning and artificial intelligence. In this chapter, we explore the role that these technologies can play when coming together. We will look into the situation of HPC and into how DA, ML and AI can change the scientific and industrial usage of simulation on high-performance computers.

Keywords High-performance computing · Data analytics · Machine learning · Artificial intelligence · Simulation

1 Introduction

Machine learning (ML) and artificial intelligence (AI) have become more visible over the last years and have developed into fields that show a huge potential for using computers in a variety of applications. Areas of usage range from improving and speeding up medical image processing to optimizing urban planning processes and to a standardized and high-speed handling of banking processes even in the usually heavily personalized consumer market. Some economists assume that ML and AI will change the world so dramatically that millions of jobs will be lost and we need to speak of a "second machine age" [1]. But this is not the scope of a scientific investigation.

In this chapter, we have a look at the merger of high-performance computing (HPC) with ML and AI. The situation of HPC has been described before [2, 3] and is considered to be interesting but also limited by the technical problems that we face with the end of Moore's law. We will argue that ML and AI have to be seen as two different technologies that are part of a chain of technologies that naturally lead

M. M. Resch (✉)
High Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Stuttgart, Germany
e-mail: resch@hlrs.de

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_18

us from HPC to AI. We will also argue that HPC is both a facilitating technology for ML and AI and a heavy user of ML and AI technologies. In this chapter, we will not address ethical issues that come with ML and especially AI. This is an important topic but should not be part of a technical overview chapter.

Finally, we give an outlook of how HPC, ML and AI will be merged into a new world of technologies. This merger will allow to tackle new research questions and will help to substantially improve our way of doing simulation on supercomputers.

2 The Status of High-Performance Computing

The future of HPC is widely discussed in the scientific community and has also attracted substantial political interest in the last years. The topic around which these discussions evolve is the race for the first Exaflop system. The US has announced to build such systems in the coming years [4]. Japan has announced to build a system as a follow-on national project for the current RIKEN K-Computer system [5]. China has started a program to build such a system and is planning to have up to eight systems, each capable of Exaflop performance [6]. Europe has started an initiative called EuroHPC [7] which has decided to fund three European pre-Exaflop systems in 2021 and plans to fund two Exaflop systems later on. All in all, the world of HPC seems to be set for moving from the era of Petaflops computing to the new era of Exaflops computing.

Having a look at the list of fastest supercomputer in the world [8], we see that in November 2019, three systems can be considered to be in the pre-Exaflop range. The fastest system in the world (Summit) shows a peak performance of 200 PF and a Linpack performance of 143 PF. Assuming a growth rate as expected by Moore's law [9], we should see a performance increase of two every 18 months. That would be a 400 PF system in June 2020 and an 800 PF system in November 2021. Assuming that Moore's law still holds and that the budget of 500 million US$ is an increase in system cost of about 50% compared to the Summit system, an Exaflop system seems to be feasible in 2021.

It becomes obvious in this calculation that the main commonality of all international projects aiming at Exaflop systems is a substantial increase in budget. One reason is that the energy costs are still increasing. For an Exaflop system, we expect to see a power consumption in the range of 30 to 50 MW. Costs for electricity vary from country to country but such a high-power consumption substantially increases total cost of ownership. On the other hand, the increase in investment is necessary to make up for the slowdown in Moore's law as computers gain speed mainly by increasing the number of processors used.

3 Slowing Down the Speedup

Even though the question of architecture is perhaps no longer the most pressing one, the first problem that HPC will have to handle in the future is a hardware problem: the end of Moore's law [9]. The prediction of Moore in 1965 that we would be able to cram ever more transistors on the same surface, with the numbers doubling every 12—later Moore shifted this to 18—months, did hold for about 50 years. As of today, it is unclear how far we still can get with miniaturization. Seven nanometres are already claimed to be used. Five nanometres and finally three nanometres might be achievable in the foreseeable future. However, it is unclear whether this is economically feasible. As a consequence, we need to assume that by the mid-2020s increase in performance for supercomputers will be difficult to achieve through a further increase in transistors on a chip [10].

This triggers the need for new solutions. Quantum computing has started to carry the hopes of funding agencies to the point that departments and divisions are renamed to be responsible for "Quantum Computing and HPC". Given that there are no real quantum computers available yet, this chapter will not cover the topic of quantum computing in detail. What we see when we look at first systems that are similar to quantum computers is that these systems will most likely not provide higher performance in the sense of more floating point operations but will rather open up a new field of simulation.

Two key questions seem to be important from the point of view of HPC for quantum computing as of today.

- Is there a simple way to transform classical HPC simulation problems into equivalent problems to be solved by quantum computers?
- Will users accept the fact that with quantum computing the notion of a deterministic solution is lost or at least weakened?

Quantum computing will remain a field of research for the coming decades. It will take some time before they can be made available for scientific usage and still more time to turn them into a widely spread device to be used in science and industry alike.

With traditional hardware making it much more difficult to squeeze more performance from a given HPC architecture, the focus of attention will have to be shifted towards software and towards mathematical methods for HPC simulation. Over the last decades, various investigations show that mathematical methods—at least for the solution of systems of equations—have contributed substantially to the overall increase in simulation performance [11]. The focus of attention for future algorithms will have to be on speed and on power consumption. What we find is that for both targets, optimum usage of memory is the key as access to memory both slows down computation and increases power consumption.

Another trend that has a huge impact on HPC is what can best be described as the transition from "big data" through machine learning towards artificial intelligence. Even though usage of data goes way beyond the original idea of handling large

amount of data, the term "big data" still in a sense is useful as it describes well how HPC may be overwhelmed by the data problem in the coming years. HPC may well become a smaller part of a larger digital infrastructure that is focusing around data rather than around compute power. We will address, how this will impact HPC.

4 From Big Data to Artificial Intelligence

Over the last years, a number of new paradigms and one old paradigm have grown in importance. All of these are based on data. Big data was already introduced more than a decade ago and for a while was considered to bring a new paradigm to science [12]. Some even went further to claim that with big data science would reach the "end of theory" [13]. However, correlation and causality are two different things, and hence the simple analysis of data will always show correlation but never causality. Big data was soon further developed into a concept that brings together data and insight and which is usually called machine learning. But at the same time, a new wave of artificial intelligence projects has hit the high-performance computing community. From an HPC expert point of view, there is a logical path from big data to artificial intelligence that can be seen as a new chance for simulation on HPC. In the following, we will describe how we can find a continuous spectrum of applications ranging from classical simulation to artificial intelligence.

Classical Simulation In the classical simulation approach, the simulation expert goes through a series of steps which are handled sequentially. The results of a simulation are analyzed post-mortem in a visualization environment. So, data are created and each data set is considered individually. Visualization provides the necessary techniques for analysis. Usually, all simulation runs are independent. The simulation expert has a clear understanding of the job she is running and also knows which features or values to look for in the computed results. For a Computational Fluid Dynamics (CFD) simulation, this usually means to look for velocities and pressure and to visually identify spots where special flow phenomena—like turbulence, recirculation, stagnation—appear [14]. A global view of all simulation runs or a deep dive into the data is usually not undertaken.

Big Data The concept of big data evolves from traditional data analytics and looks at data from the point of view of harvesting information that may be buried and hidden in too many data for human beings to analyze. HPC simulation is currently moving from traditional simulation to big data in the sense that simulations create huge amounts of data. These data can still be visualized but the human eye is unable to grasp all details and to identify interesting spots. The promise of the "end of theory" [13] will most probably not materialize in HPC simulation as analyzing simulated data requires a deep understanding of the overall simulation process. However, concepts of big data may help to create awareness in the HPC community that classical visualization methods may not be enough to fully exploit the knowledge created by an HPC simulation. For our CFD example, big data may

lead the simulation expert to explore several simulation results at a time. It may also make the user want to start to search for features (stagnation, turbulence, vortices, etc) automatically based on improved evaluation methods.

Machine Learning Machine learning is a technology that not necessarily evolves from big data. However, it can be seen as a logical continuation of the idea to extract information from large amounts of data. If we assume that we still need some theory to extract knowledge from data, we need to be able to use the data we have to improve the theory. The learning process, however, now goes beyond the pure analysis of data. It makes use of the data to improve our understanding and leads us to improved or new theories. When we now look at our example from flow simulation, machine learning can help to use existing simulation data to learn how to design future simulation runs or to learn how to interpret a large number of simulation results in a coherent view.

Artificial Intelligence The notion of artificial intelligence (AI) is said to have been first introduced by Alan Turing back in 1950 [15]. Intelligence is a concept that is basically not a technical one. Over the last decades, it has seen a change in meaning and understanding. It is hence a bit difficult to clearly judge the technical merits of AI. While Alan Turing was referring to AI as a computer system that is able to fully imitate the logical behaviour of a human being modern interpretation of AI is looking at two main features. On the one hand, AI is considered to be a way to create humanoid robots. The focus of this approach is to create an artificial human being including the physical body. On the other hand, AI is considered to be able to replace human beings in the decision-making process. During the 1970s and the 1980s, there was substantial investment in AI research, and expectations to achieve both goals were high. The most recent wave of enthusiasm about AI has a more realistic focus. It usually aims at integrating software and hardware solutions with enough data to create a system that is able to unburden the human being from complex but standardized decisions. Typical examples are decision-making in medical treatment and in the analysis of human faces. This is certainly far away from the original human-like machine. However, the potential for this technology is high. When we come back to our CFD example, AI can help to learn from previous simulations to make decisions about the future simulations that have to be done to solve a given problem. The decision-making in the simulation would practically entirely be offloaded to an AI system.

What we see when looking at these technologies are two things:

- There is no clear distinction to be made between HPC, big data, ML and AI. These technologies are a gradual advance from a process purely controlled by the human being towards a process nearly entirely controlled by what we might call machine intelligence.
- HPC is not a technology separate from big data, ML and AI but all these technologies rely heavily on the availability of both compute power and theoretical knowledge.

4.1 What Does This Mean for HPC?

Even though a traditional look at HPC already shows some dramatic change, there is something that might be even more important for HPC. Considering current trends, we find that HPC is going to be part of something bigger—which is driven by data but not only data. It is meanwhile well accepted that there is value in the data. However, there is much more value in the right learning processes and algorithms.

HPC simulations might be one source for such data. Sources of data can however be manifold:

- The traces each person is generating each day using systems in the internet, when shopping, when communicating, when watching movies, when visiting other webpages.
- The data of business operations which are digitally available and stored for years.
- The increasing amount of sensors everywhere especially powered by the Internet of Things, going from production lines to personal homes—smart meters are a good example for that.

Two main scenarios for HPC in such a data-driven world evolve.

HPC Needs Data HPC simulation will increasingly make use of modern methods to handle, explore, interpret and turn data into decision. The simulation community will move from classical batch processing or co-simulation with visualization and simulation running in parallel towards a setup that is driven by data. Simulations will bring in more data from fields other than simulation. Meteorology is an example where measured data combine with simulation data in order to improve the quality of the picture. Simulations will bring in data analytics methods in order to better understand computed results. This will take away control of visualization from the human being and put it more in the "hands" of the computer system. But the change will go even further. AI systems will help to analyze simulation runs and learn from the results in order to make suggestions for future simulations. In a mid-term perspective, simulations could even be entirely taken off the hands of human beings and be done by AI systems that access simulation data and theory repositories automatically responding to user questions through simulation and their interpretation. As strange as this may sound to traditional simulation experts, it would only be a continuation of a process in which the behaviour of computers is hidden from the user. And it would be the logical evolvement of all technology that is supposed to replace human beings in order to improve and/or speedup a process that can be standardized.

Data Needs HPC When looking carefully at the requirements and the potential of data analytics, ML and AI, it is obvious that these technologies will not replace HPC but will rather give a new boost to HPC. One of the key aspects in ML and AI is the learning phase. While it is obvious that data are required to learn, it is less obvious that compute power is a must for this learning phase. It is hence not surprising that the fastest Japanese supercomputing system in November 2018 [16] was exclusively

devoted to artificial intelligence. The AI Bridging Cloud Infrastructure (ABCI) has a focus on applications from AI and will serve the Japanese research and industrial community for the coming years. In that sense, HPC will have a new user community that will increase the need for large-scale systems and Exaflop performance.

4.2 Who Might Benefit?

We can find a number of interesting cases that will benefit from a merger of HPC with data technologies. Some of them are rather obvious. Others do not seem to be good candidates in the first place.

Banks are one potential group of customers that may move even further into the field of HPC. They already have a history of analyzing data when it comes to stock exchange analysis. There are further topics that might be interesting. Fraud detection is one field that might benefit both by increasing the speed of a detection and by increasing the level of accuracy. Permanent and individualized portfolio analysis both for institutional and private customers is a field that will need HPC performance.

In many of these business cases where the analytics is done in large in-memory data bases, not many are thinking about HPC. However, after the analysis of business data, a next step would be to change and improve the situation. In several cases, this could lead to the requirement of large simulations and parameter studies which will naturally require HPC systems. A good example for this is railroad companies. In case of delays, simulations are used to decide between different options to improve the difficult traffic situation.

The increasing use and number of linked sensors is another area where data volumes are exploding. This leads to the idea of in-time analytics to detect events before they actually occur, for example, with machine learning technologies. This may lead to new insights and better understanding of existing dependencies. In order to extract such information, inverse problems will have to be solved. This will require HPC to a much bigger extent than today.

Another example with even higher impact on HPC is the usage of sensors to detect major natural disasters which might lead to damage and loss of lives. In case of a marine earthquake, Japan has set up a system to automatically analyze data, simulate the impact of a tsunami and take measures to protect its people. This is an example where data simulation and AI have to work together to come up with a solution that could never be achieved with classical simulation approaches. Given the time-critical situation and at the same time the financial impact and the risk for human lives, only such an integrated approach can help to come to acceptable solutions.

4.3 What Does This Mean for HPC Environments and Architectures?

The development described above already has an impact on architectures and overall HPC environments. I/O and the handling of large data sets are considered to be a critical topic in HPC procurements and in systems offered. Specialized I/O nodes are part of any HPC system already today. They will become more important in the future. Large memory nodes to be able to handle larger data sets have become a standard component. The size of the memory is continuously increasing.

In several cases, a direct connection to include up-to-date input data into the ongoing simulations will require a change in the HPC environment setup and will require to solve new security issues. Additionally, there is the upcoming requirement for "urgent" computing which needs to be solved administratively as well as technically as many HPC systems are not prepared for such a requirement. The main problem for HPC operation in urgent computing is the fact that jobs will have to be interrupted such that users may lose their jobs or results.

5 Conclusion

Summarizing our findings, we see a number of trends which will have an impact on HPC and AI in the coming years. It is getting ever more clear that the main driving force of HPC in the last decades will go away. Moore's law is coming to an end and will not help us increase HPC performance in the future. Improved algorithms and mathematical methods will still have the potential to increase sustained performance but will only extend the race in HPC without being able to overcome the stagnation in peak performance to be expected.

At the same time, we see a shift away from pure HPC to an integration of technologies. Big data, machine learning and artificial intelligence are added to the set of tools that help to solve many of the traditional problems much better and to tackle new problems. For the coming 10 years, this convergence of technology will be the most important aspect in high-performance computing.

References

1. E. Brynjolfsson, A. McAfee, *The Second Machine Age: Work, Progress, Prosperity in a Time of Brilliant Technologies* (W. W. Norton, New York, 2016)
2. M.M. Resch, T. Boenisch, M. Gienger, B. Koller, High performance computing—Challenges and risks for the future, in *Advances in Mathematical Methods and High Performance Computing*, ed. by V. K. Singh, D. Gao, A. Fischer, (Springer, Berlin, 2019)
3. M.M. Resch, T. Boenisch, High performance computing—Trends, opportunities and challenges, in *Advances in Parallel, Distributed, Grid and Cloud Computing for Engineering*, ed. by P. Ivanyi, B. H. V. Topping, G. Varady, (Saxe-Coburg, Kippen, Scotland, 2017), pp. 1–8

4. https://www.energy.gov/articles/us-department-energy-and-intel-build-first-exascale-supercomputer. Accessed 20 Nov 2019
5. https://www.r-ccs.riken.jp/en/postk/project. Accessed 20 Nov 2019
6. Private communication with Chinese colleagues, January 2019
7. https://eurohpc-ju.europa.eu/. Accessed 20 Nov 2019
8. www.top500.org. Accessed 20 Nov 2019
9. G.E. Moore, Cramming more components onto integrated circuits. Electronics **38**(8), 114–117 (1965)
10. R. Courtland, Transistors could stop shrinking in 2021, IEEE Spectrum, http://spectrum.ieee.org/semiconductors/devices/transistors-could-stop-shrinking-in-2021. Accessed 20 Nov 2019
11. V. Marra, On Solvers: Multigrid methods, https://www.comsol.com/blogs/on-solvers-multigrid-methods/. Accessed 20 Nov 2019
12. T. Hey, K.M. Tolle, S. Tansley, *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research, Redmond, VA, 2009)
13. C. Anderson, The end of theory: The data deluge makes the scientific method obsolete*, Wired Magazine*, June 23 (2008)
14. K. Perktold, M. Resch, R. Peter, Three-dimensional numerical analysis of pulsatile flow and wall shear stress in the carotid artery bifurcation. J. Biomech. **24**(6), 409–420 (1991)
15. A.M. Turing, Computing machinery and intelligence. Mind **59**, 433–460 (1950)
16. https://www.top500.org/system/179393. Accessed 20 Nov 2019

Slip Effect on an Unsteady Ferromagnetic Fluid Flow Toward Stagnation Point Over a Stretching Sheet

Kaushik Preeti, Mishra Upendra, and Vinai Kumar Singh

Abstract In this paper the heat transfer characteristics of ferromagnetic fluid flow towards stagnation point has been investigated numerically. In this study we deal with the slip boundary condition in the presence of electromagnetic field over a stretching sheet considering the Brownian motion impacts on ferrofluid viscosity. The mathematical model is presented in the form of partial differential equations. The governing equations determine the flow conditions, and these equations are reduced by similarity transformations. Finite difference method is implemented to acquire the solution of the problem. The effect of various physical parameters on the flow is also investigated. Graphs are plotted to examine the influence of pertinent flow parameters involved, such as velocity profile, temperature profile. The important physical quantities of skin friction coefficient and the local Nusselt number are also studied. It is observed that increasing value of ferromagnetic interaction parameter enhances the velocity field and reverse observation holds for temperature field.

Keywords Magnetic dipole · Ferromagnetic fluid · Stagnation point flow · Viscoelastic parameter · Stretching sheet

1 Introduction

A ferrofluid is a liquid that becomes strongly magnetized in the presence of a magnetic field. These fluids are liquids such as kerosene, heptane, or water. Mechanics of ferrofluid motions is influenced by strong forces of magnetization.

K. Preeti (✉) · V. K. Singh
Inderprastha Engineering College, Ghaziabad, UP, India
e-mail: pre.kaushik@ipec.org.in; deanacademics@ipec.org.in

M. Upendra
Amity University Rajasthan, Kant Kalwar, Jaipur, India
e-mail: umishra@jpr.amity.edu

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_19

Ferrohydrodynamics usually deals the nonconducting liquids with magnetic properties. Ferrofluids are used to image magnetic domain structures on the surface of ferromagnetic materials. Many researchers analyze heat transfer through boundary layers over a stretching surface. This field has received a significant attention due to its useful engineering applications such as solar collectors, designing building, and thermal insulation and cooling of electronic components. Pioneer works have been done by Crane et al. [14] on the boundary layer flow of an electrically conducting viscous incompressible fluid over a stretching sheet. Elbashbeshy et al. [16] analyzed the laminar flow and heat transfer over an unsteady stretching surface when the surface temperature is constant. In the presence of variable surface temperature, Chakrabarti et al. [11] studied the magnetohydrodynamic MHD flow with uniform suction over a stretching sheet at different temperatures. Grubka et al. [21] studied the heat transfer analysis over a stretching surface in the presence of heat flux. Ellahi et al. [18] studied the influence of temperature-dependent viscosity on MHD flow of non-Newtonian fluid. The fact that velocity of pseudoplastic fluids decreases with decrease in Hartmann number was found by Khan et al. [32]. Hayat et al. [23] investigated the heat transfer effect of Eyring–Powell fluid considering exponentially stretching sheet. Narayana et al. [42] studied the influence of unsteadiness parameter on the flow of thin film over an unsteady stretching sheet. Abdel-Wahed and Emam [3] studied the MHD flow of nanofluid over a moving surface in a nanofluid under thermal radiation and convective boundary layer conditions. Abdelwahed et al. [2] and [4] inspected the variation of the thermal conductivity and viscosity on the MHD flow. Heat transfer in a Newtonian fluid in the presence of thermal conductivity was analyzed by Chiam et al. [13]. Khan and Pop et al. [30] studied the behavior of laminar flow of nanofluid over a stretching surface and investigated the influence of Brownian motion and thermophoresis parameters have inclination to the fluid temperature. Raju et al. [44] analyzed the heat and mass transfer on MHD flow over a permeable stretching sheet. Similar type of study of unsteady flow through a stretching sheet was performed by Mustafa et al. [41]. Aziz et al. [7] investigated the problem of mixed convective fluid flow along a stretching sheet with variable viscosity. Dutta et al. [15] determined the temperature distribution of heat flux over a stretching surface. Khan M. et al. [34] studied the Brownian motion and thermophoresis effect on heat and mass transfer. Hayat et al. [22] studied the boundary layer flow at stagnation point through a porous medium in the presence of thermal radiation over a stretching vertical plate. Ishak et al. [27] presented the concept of unsteadiness in mixed convection boundary layer flow and heat transfer through vertical stretching surface. Ibrahim and Bhandari et al. [25] analyzed the heat transfer on a permeable stretching surface due to a nanofluid with the influence of magnetic field and slip boundary conditions. Andersson et al. [5] examined impact of magnetic field on the flow of viscoelastic fluid over the stretching surface. Elbashbeshy et al. [17] obtained an analytical solution for the boundary layer flow over a moving plate. Khan M. et al. [33] studied the two-dimensional incompressible Casson nanofluid in the presence of magnetic field. Bachok et al. [9] inspected the flow of a nanofluid at stagnation point over a stretching or shrinking sheet. Effects of heat transfer in the presence of magnetic

field on ferrofluid flow was reported by Sheikholeslami et al. [46]. Chen et al. [12] studied the effect of continuous surface on heat transform in laminar flow. Viscoelastic fluid characteristics were investigated in the presence of temperature viscosity by Faraz et al. [20]. Abbas et al. [1] investigated the flow of a viscous fluid at stagnation point over an unsteady surface. Stagnation point flow of Maxwell nanofluid was investigated by Khan et al. [35]. Mukhopadhyay and Battacharyya et al. [40] studied the influence of Maxwell fluid in the heat transfer across a stretching sheet. Bachok et al. [8] studied the two-dimensional stagnation point flow of a nanofluid over a stretching or shrinking sheet. Partha et al. [43] tackled the heat transfer over an exponential stretching vertical sheet with dissipation effect. Khan et al. [31] discussed the heat transform reactions of nanofluid with the effect of viscous dissipation and thermal radiation along a stretching sheet under the action of thermophoresis with the help of finite difference scheme. Maxwell fluid is one of the examples of non-Newtonian fluid. Mukhopadhyay and Bhattacharyya et al. [39] determined the influence of Maxwell parameters on the unsteady flow of Maxwell fluid with chemical reaction. Heat transfer analysis on boundary layer flow with specific entropy generation was studied by Ellahi et al. [19]. Khan M. et al. [36] developed a Cattaneo–Christov model by using Fourier's and Fick's laws and solved by numerical method. Bovand et al. [10] investigated the two-dimensional MHD flow in the porous medium in different laminar flows. The study showed that the steady flow depends on magnetic fields. The heat transfer and fluid flow investigation of different kinds of base fluids on a stretching sheet was performed by Makarem et al. [38]. Numerical investigation of heat transfer enhancement by utilizing the properties of nanofluids was conducted by Sheri and Thuma et al. [47]. Numerical solution of Maxwell fluid with the condition of viscous dissipation was obtained by Khan M. et al. [37]. Unsteady magnetohydrodynamics mixed convection flow in a rotating medium with double diffusion was studied by Jian and Ismail et al. [29]. Computation and physical aspects of MHD Prandtl–Eyring fluid flow analysis over a stretching sheet were investigated by Hussain and Malik et al. [24]. Jafer et al. [28] studied the effects of external magnetic field, viscous dissipation, and Joule heating on MHD flow and heat transfer over a stretching or shrinking sheet. Analysis of modified Fourier law in the flow of ferromagnetic, Powell-Eyring fluid considering two equal magnetic dipoles was performed by Ijaz and Zubair et al. [26]. The characteristic of dust particles in a ferromagnetic fluid with thermal convection in a porous medium was analyzed by Sharma et al. [45].

Most of the studies depict that properties of ferromagnetic fluid such as higher thermal conductivity and strong influence of magnetic field are useful in the analysis of numerous fluid problems. The ferromagnetic fluids have the properties of both liquid and magnetized solid particles. Ferromagnetic fluids flow toward magnetic field if the external magnetic field is applied, and the flow resistance increases using ferrofluids under an applied magnetic field for enhancement of heat transfer, and hence ferrofluids are more useful compared with conventional nanofluids.

2 Mathematical Formulation

In this study we consider a two-dimensional unsteady ferromagnetic fluid flow over a stretching sheet in the presence of magnetic field with the dipole effects. Existence of magnetic field develops the higher intensity of the ferrofluid particles. Heat produced by the internal friction of the fluid, which is caused by the increase in temperature, affects the viscosity of the fluid, and so the viscosity of the fluid cannot be taken as constant. The rise of temperature leads to a local increase in the transport phenomenon by reducing the viscosity across the momentum boundary layer and so the heat transfer rate at the wall is also affected significantly. The study also characterized the phenomena of stagnation point. Coordinate X is taken along the stretching surface where the sheet is placed at $y = 0$. The velocity of the sheet $u_w = c_p x$, where $c_p > 0$ is stretching rate presented in Fig. 1. Y is normal to the stretching surface where the fluid flow is restricted by $y > 0$. Flow takes place by the two equal and opposite forces in the direction of X-axis. A magnetic dipole is placed in the center with distance "a" from the surface. The temperature of the surface is T_w and Curie temperature is taken as T_c, and the temperature of ferrofluid from the surface of the sheet is $T_\infty = T_c$; when the ferrofluid reaches the curie temperature,
magnetization ends at this point. The obtained boundary layer equations that govern the flow and heat transfer of ferrofluid are written as

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{1}$$

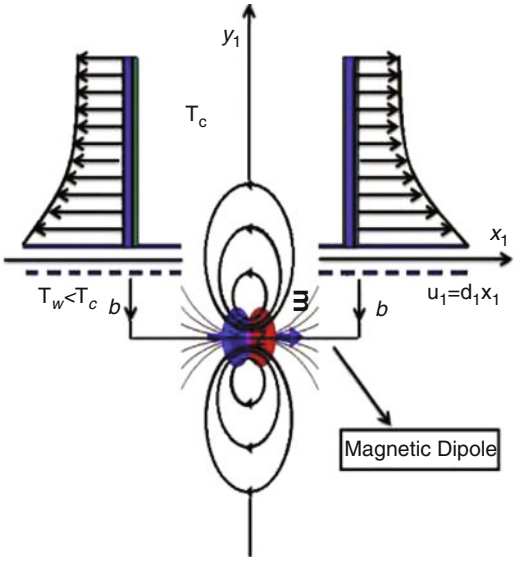


Fig. 1 Geometry of the flow

$$\frac{1}{\mu_e}\left(u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}\right)=\frac{M}{\widehat{p}}\frac{\partial M}{\partial x}-\frac{1}{\widehat{p}}\frac{\partial H}{\partial x}+\frac{\mu}{\widehat{p}\mu_e}+\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)+\frac{g}{\widehat{p}\mu_e}g(T_c-T_f) \tag{2}$$

$$\frac{1}{\mu}\left(u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}\right)=\frac{M}{\widehat{p}}\frac{\partial M}{\partial y}-\frac{1}{\widehat{p}}\frac{\partial H}{\partial y}+\frac{\mu}{\widehat{p}\mu_e}\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right)+\frac{g\beta}{\widehat{p}\mu_e}g(T_c-T_f) \tag{3}$$

$$\frac{c_p}{\mu_e}\left(u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right)\frac{\partial M}{\partial x}+\frac{T}{\widehat{p}}\left(u\frac{\partial H}{\partial x}+v\frac{\partial H}{\partial y}\right)\frac{\partial M}{\partial T}=\frac{1}{\widehat{p}}\frac{\partial^2 T}{\partial y^2}+\frac{2\mu}{\widehat{p}\mu_e}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2+\frac{\partial v}{\partial x}\frac{\partial u}{\partial y}\right] \tag{4}$$

In the above set of equations (u, v) are velocity components along x- and y-axis, respectively. T_f is the fluid temperature, μ represents the dynamic viscosity, $\widehat{p}$ signifies the fluid density, μ_e denotes the magnetic permeability, c_p signifies the specific heat, H represents the magnetic field, and M denotes the magnetization. The terms $\mu_e M\frac{\partial H}{\partial x}$ and $\mu_e M\frac{\partial H}{\partial y}$ given in Eqs. (2) and (3) represent the magnetic force and magnetic gradient corresponding to x- and y-coordinates, respectively. The corresponding boundary conditions expressed as

$$u=u_w+S\frac{\partial u}{\partial y}, v=-v_w, T_f=T_w+d_1 at y=0 \tag{5}$$

$$u=0, T_f=T_c+d_2 x, P+\frac{1}{2}\widehat{p}(u^2+v^2)=Cat y\Rightarrow\infty \tag{6}$$

where u_w and v_w are surface velocity along x- and y-direction, S represents the velocity slip factor, C is positive constant, and d_1 and d_2 are dimensionless constants.

3 Mathematical Analysis

$$\zeta=\frac{\lambda}{2\pi}\left(\frac{x}{x^2+(y+d)^2}\right) \tag{7}$$

where λ represents the magnetic field strength. We know that body force is directly proportionate to gradient of the magnitude, and the magnitude H of the magnetic strength is represented as

$$H = \sqrt{\left(\frac{\partial \zeta}{\partial x}\right)^2 + \left(\frac{\partial \zeta}{\partial y}\right)^2} \tag{8}$$

Components of magnetic field H are

$$\frac{\partial H}{\partial x} = -\frac{\lambda}{2\pi}\left(\frac{2x}{(y+d)^4}\right) \tag{9}$$

$$\frac{\partial H}{\partial x} = \frac{\lambda}{2\pi}\left(\frac{-2}{(y+d)^3} + \frac{4x^2}{(y+d)^5}\right) \tag{10}$$

Magnetization M leads to the expression of temperature given by Anderson et al. [6]

$$M = K^c(T_c - T_f). \tag{11}$$

4 Solution Procedure

It is pertinent to introduce the dimensionless variables and transformation considered by [6]

$$\psi(\tau, \eta) = \frac{\mu}{\widehat{p}}\tau f(\eta) \tag{12}$$

$$\alpha(\tau, \eta) = \frac{T_c - T_f}{T_c - T_w} = \theta(\eta) + \tau^2\Phi(\eta) \tag{13}$$

where $\psi(\tau, \eta)$ and $\alpha(\tau, \eta)$ are dimensionless steam function and temperature, respectively, and dimensionless coordinates τ and η are as follows:

$$\tau = \sqrt{\frac{c\widehat{p}}{\mu}}x, \eta = \sqrt{\frac{c\widehat{p}}{\mu}}y \tag{14}$$

$$u = \frac{\partial \psi}{\partial y} = cx \cdot f'(\eta) \tag{15}$$

$$v = -\frac{\partial \psi}{\partial x} = -(cv)^{\frac{1}{2}} \cdot f(\eta) \tag{16}$$

Substituting Eqs. (12)–(16) into Eqs. (2)–(4) and comparing coefficients up to n^2, we get the reduced nonlinear ordinary differential equations:

$$f''' + ff'' - f'^2 + \frac{2\beta\theta}{(\eta+\delta_1)^4} N[2ff'''' - (f'')^2] = 0 \tag{17}$$

$$\theta'' + Pr(f\theta' - 2f'\theta) + \frac{2N\beta(\theta - w)f}{(\eta+\delta_1)^3} - 4N(f')^2 + 2(w^2 - 1) = 0 \tag{18}$$

$$\phi'' + \frac{2N\beta f\theta_2}{(\eta+\delta_1)^3} - Pr(4f'\phi' - f\phi) - N\beta(\theta - w)\left[\frac{4f}{(\eta+\delta_1)^5} + \frac{2f'}{(\eta+\delta_1)^4}\right] - N(f'')^2 = 0 \tag{19}$$

Also boundary conditions (5) and (6) are converted as

$$f = S,\ f' = 1, \theta = 1 + \alpha f''(0), \phi = 0, \text{at}\ \eta = 0 \tag{20}$$

$$f' \to 0, \theta \to 0, \phi \to 0\ \text{at}\ \eta \to \infty \tag{21}$$

In the above system of nonlinear equations, β (ferromagnetic interaction parameter), N (viscoelastic parameter), Pr (Prandtl number), w (dimensionless curie temperature ratio), λ (dimensionless distance), and γ (viscous dissipation) are defined as

$$\beta = \frac{\lambda\widehat{p}K^c}{2\pi\mu_e^2}\mu_e(T_c - T_w),\ N = \frac{c\mu^2}{\widehat{p}(T_c - T_w)},\ \text{Pr} = \frac{\mu c_p}{k},$$

$$w = \frac{T_c}{T_c - T_w}, \lambda = \sqrt{\frac{k\widehat{p}d^2}{\mu}}, \gamma = \frac{c\mu^2}{\widehat{p}k(T_c - T_w)}$$

The skin friction coefficient and Nusselt number are defined as

$$C_{f_x} = -\frac{2\tau_w}{\rho(cx)^2} \tag{22}$$

$$Nu_x = \frac{x}{-k(T_c - T_w)}\frac{\partial T}{\partial y}\bigg|_{y=o} \tag{23}$$

where

$$\tau_w = \mu\left(\frac{\partial u}{\partial y}\bigg|_{y=0}\right) + \delta_1\left(u\frac{\partial^2 u}{\partial x\partial y} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial u}{\partial x}\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}\frac{\partial v}{\partial y}\right)$$

$$C_f Re_x^{\frac{1}{2}} = -2(1 - N*)f''(0)$$

$$Nu_x = -[(\theta'(0) + \xi^2\phi'(0)]Re_x^{\frac{1}{2}}$$

We first transformed differential equations (12)–(16) together with boundary conditions into a system of set of first-order ODE, which must be solved numerically by finite difference method. The step size is taken as $\nabla\eta = 0.01$. We choose $\eta(max) = 15$ with simulation error chosen as 10^5 in order to assure asymptotic convergence criteria. Trial values of $f'''(0)$, $f''(0)$, $\theta'(0)$, and $\Phi'(0)$ were adjusted iteratively in order to satisfy the far-field boundary condition.

5 Results and Discussion

The Influence of Ferromagnetic Interaction Parameter (β)

The fixed values of these physical parameters are taken as $\Pr = 7, N = 0.01, \epsilon = 2.0, and \delta_1 = 0$. Figures 2 and 3 are plotted to examine the influence of ferromagnetic parameter β on velocity profile $f(\eta)$ and temperature profile $\theta(\eta)$, respectively. As we increase the value of ferromagnetic parameter β, viscosity of the ferrofluid rises up and as a result velocity profile shows the decreasing behavior. This behavior occurs due to microsized particles in ferrofluid. It is observed that temperature profile increases significantly as β (ferromagnetic interaction parameter) increases. This phenomenon happened due to the interaction between

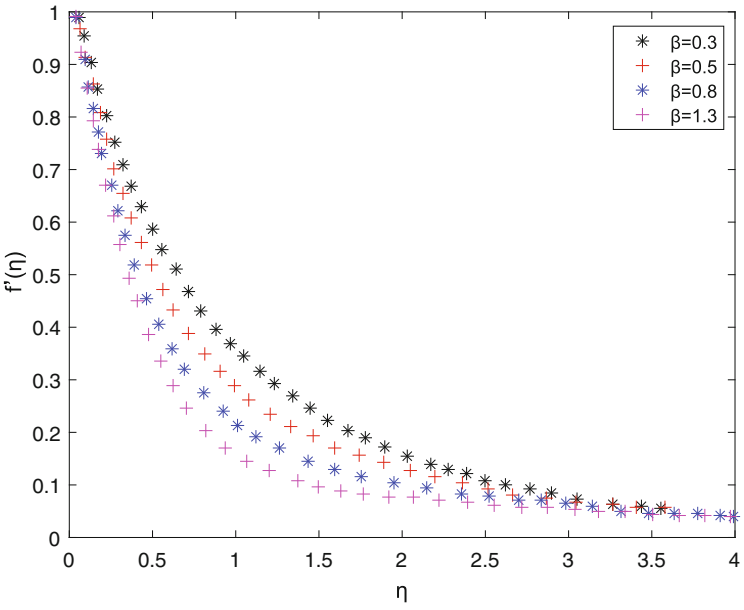


Fig. 2 Impact of β on $f'(\eta)$

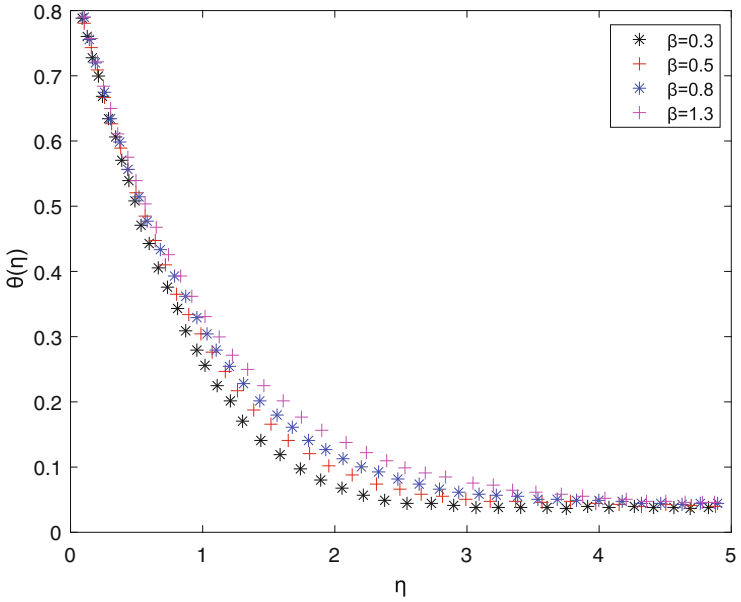


Fig. 3 Impact of β on $\theta(\eta)$

ferrofluid particles. Thus velocity profile $f'(\eta)$ reduces due to contact of ferrofluid particles and magnetic field, but the reverse condition is observed in temperature profile $\theta(\eta)$.

The Influence of Viscoelastic Parameter (N)

Figure 4 shows the impact of viscoelastic parameter (N) on velocity profile, as the increasing value of N enhances the velocity profile gradually. From this graph, it is confirmed that rising the values of viscoelastic parameter N restricts the fluid motion near the stretching sheet, while it assists the fluid motion far away from the stretching sheet. Increasing values of N permit the fluid to flow at a faster rate, because there is a decrease in the heat transfer. So by enlarging the values of N, the dimensionless stream function and velocity increase. Figure 5 illustrates the effect of viscoelastic parameter N on temperature profile $\theta(\eta)$ against η and thickness of temperature profile decreases with increase in viscoelastic parameter.

Analysis of Skin Friction Coefficient and Local Nusselt Number (Effect of Ratio ϵ)

Figure 6 designated the impact of ϵ on skin friction coefficient. The enhancement in the value ϵ reduces the skin friction coefficient. For upper values of ϵ, the velocity of ferrofluid controls the velocity of plate, and due to this, skin friction coefficient decreases. We can observe in Fig. 7 that the local Nusselt number reduces with the increasing value of β.

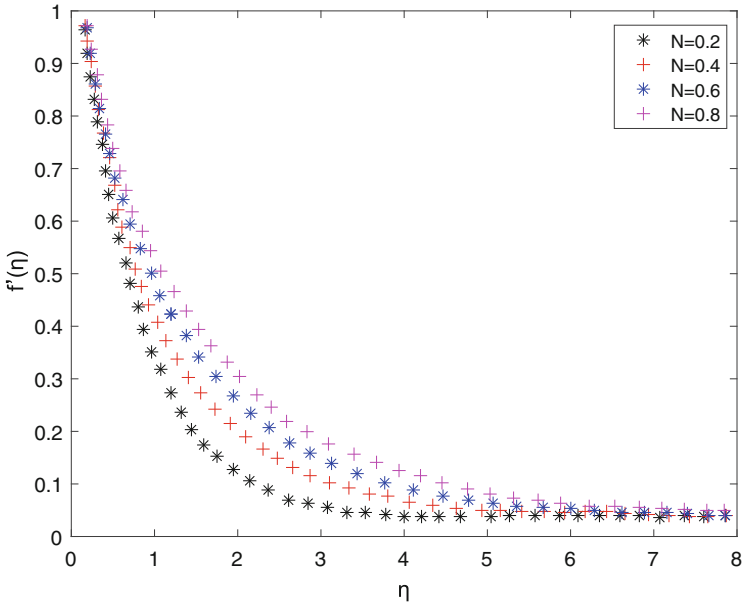


Fig. 4 Impact of N on $f'(\eta)$

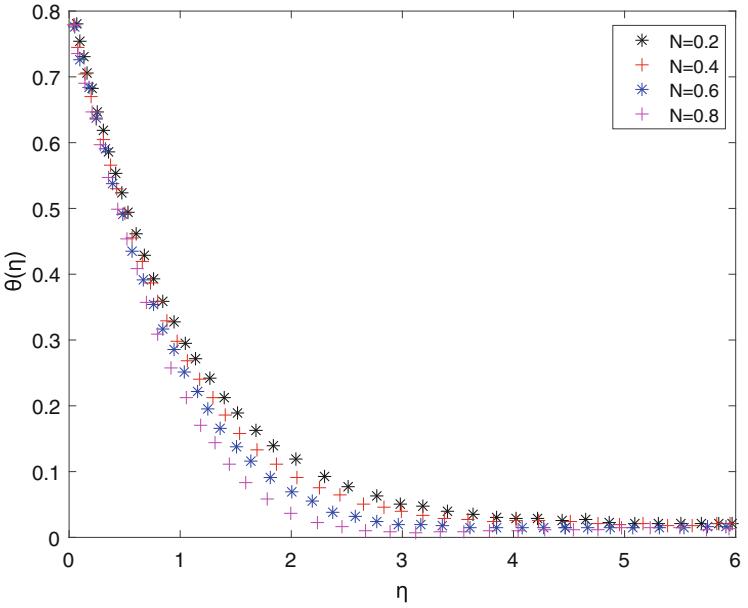


Fig. 5 Impact of N on $\theta(\eta)$

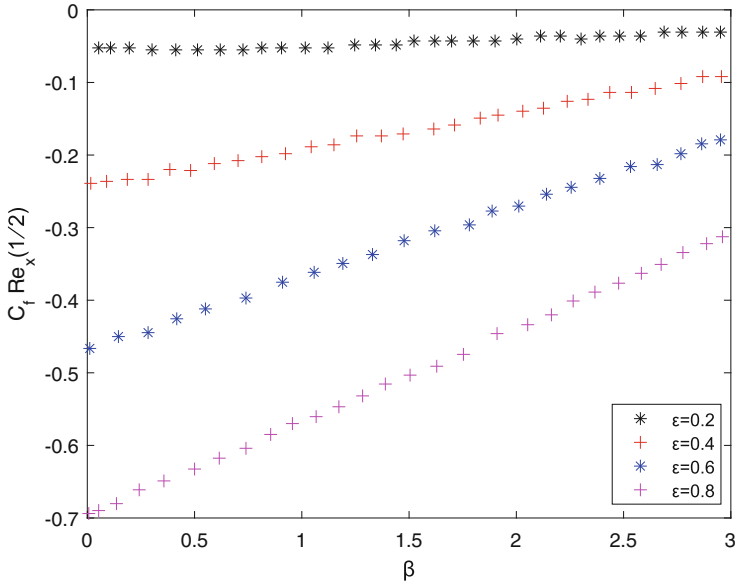


Fig. 6 Impact of ε on Skin friction coefficient

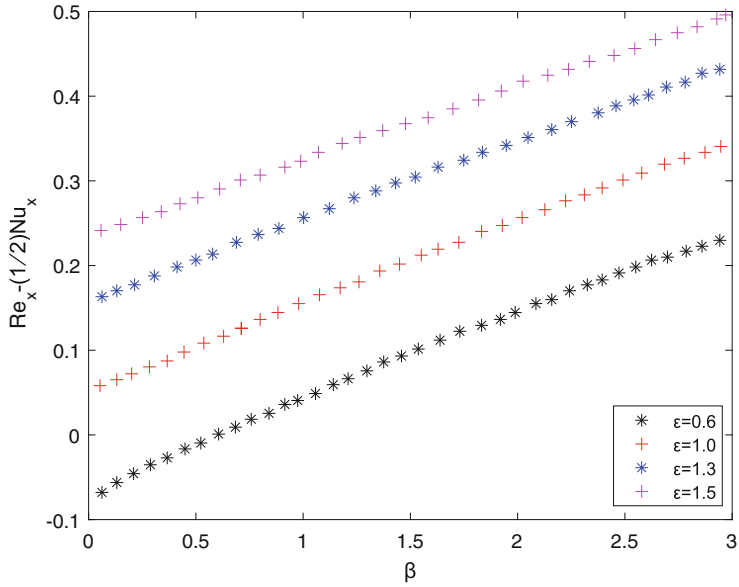


Fig. 7 Impact of ε on Local Nusselt number

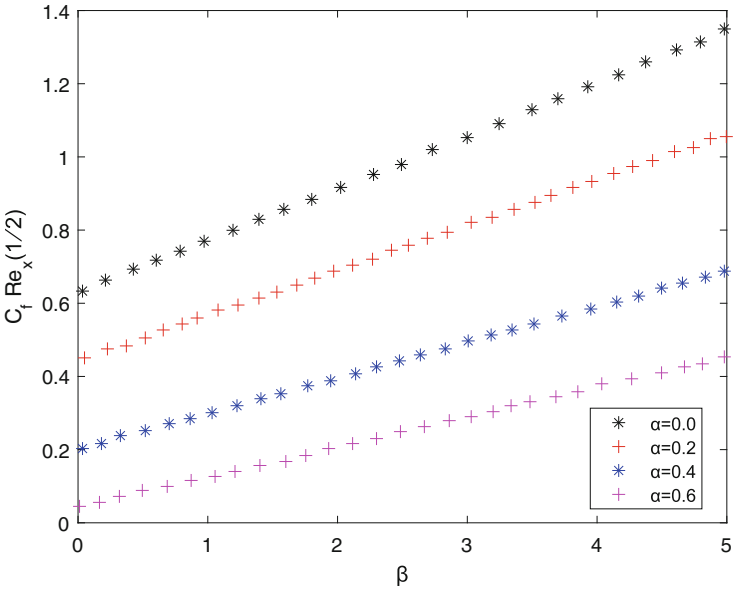


Fig. 8 Impact of α on Skin friction coefficient

Effect of Velocity Slip Parameter

Figure 8 shows that skin friction coefficient increases with variation of β; however, the reverse is true for slip parameter α. Skin friction coefficient decreases with the increase in slip parameter. The maximum surface sheer stress occurs in no slip condition ($\alpha = 0$). The local Nusselt number presented in Fig. 9 decreases for both slip parameter α and ferromagnetic field β with the increasing value of slip parameter.

6 Concluding Remarks

The two-dimensional ferrofluid problem towards stagnation point with slip boundary condition has been studied in this paper. Using similarity transformations, the governing equations were converted into nonlinear ordinary differential equations and the equations were solved numerically. The major findings of this study are as follows:

1. Velocity profile $f'(\eta)$ decreases with the effect of ferromagnetic interaction parameter β. Thus temperature profile increases with the increase in β.
2. Velocity profile increases with the increase in viscoelastic parameter N. Temperature profile decreases with the increase in N.

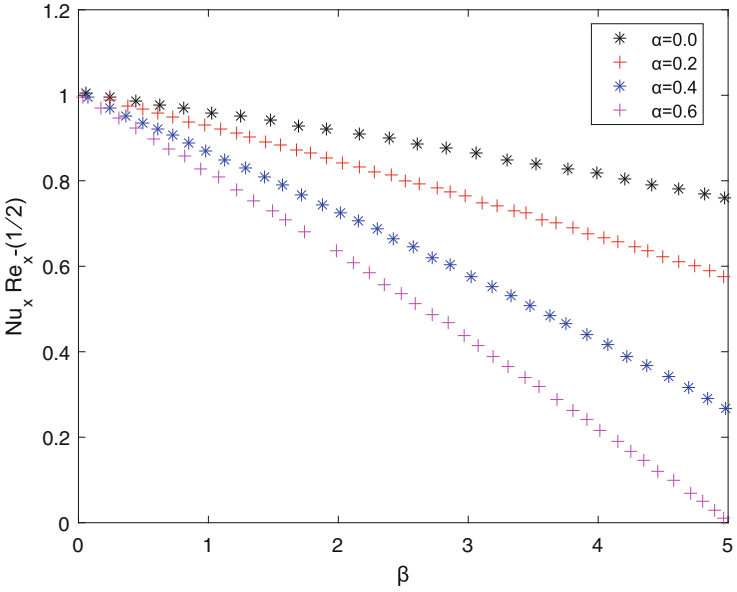


Fig. 9 Impact of α on Local Nusselt number

3. The skin friction coefficient reduces with the increase in ratio (ϵ). Similar impact of ferromagnetic parameter β is observed on local Nusselt number.
4. The skin friction coefficient and local Nusselt number both decrease with the increase in slip parameter α.

References

1. Z. Abbas, N. Muhammad, G. Mustafa, MHD stagnation slip flow over an unsteady stretching surface in porous medium. Sci. Iran. **21**, 1355–1366 (2014)
2. M.S. Abdel-Wahed, Nonlinear Rosseland thermal radiation and magnetic field effect on flow and heat transfer over a moving surface with variable thickness in a nanofluid. Can. J. Phys. **95**(3), 267–273 (2017)
3. M.S. Abdel-Wahed, T. Emam, MHD boundary layer behavior over a moving surface in a nanofluid under the influence of convective boundary conditions. J. Mech. Eng. **63**(2), 119–128 (2017)
4. M.S. Abdel-Wahed, A.K.L. Mohamed, Effect of hall current on MHD flow of a nanofluid with variable properties due to a rotating disk with viscous dissipation and nonlinear thermal radiation. AIP Adv. **6**, 095308 (2016)
5. H.I. Andersson, MHD flow of viscoelastic fluid past a stretching surface. Acta Mech. **95**, 227–230 (1992)
6. H.I. Anderson, O.A. Valnes, Flow of a heated ferrofluid over a stretching sheet in the presence of a magnetic dipole. Acta Mech. **128**, 39–44 (1998)
7. M.A.E. Aziz, Unsteady mixed convection heat transfer along a vertical stretching surface with variable viscosity and viscous dissipation. J. Egypt. Math. Soc. **22**, 529–537 (2014)

8. N. Bachok, A. Ishak, I. Pop, Unsteady boundary layer flow and heat transfer of a nanofluid over a permeable stretching/shrinking sheet. Int. J. Heat Mass Transf. **55**, 2102–2109 (2012)
9. N. Bachok, A. Ishak, I. Pop, Stagnation point flow over a permeable stretching/shrinking sheet in a copper-water nanofluid. Boundary Value Probl. **39** (2013). https://doi.org/10.1186/1687-2770-2013-39
10. M. Bovand, S. Rashidi, J.A. Esfahani, S.C. Saha, Y.T. Gu, M. Dehesht, Control of flow around a circular cylinder wrapped with a porous layer by magneto hydrodynamic. J. Magn. Magn. Mater. **401**, 1078–1087 (2016)
11. A. Chakrabarti, A.S. Gupta, Hydro magnetic flow and heat transfer over a stretching sheet. Q. Appl. Math. **37**, 73–78 (1979)
12. C.K. Chen, M. Char, Heat transfer of a continuous stretching surface with suction or blowing. J. Math. Anal. Appl. **135**, 568–580 (1988)
13. T.C. Chiam, Heat transfer in a fluid with variable thermal conductivity over a linear stretching sheet. Acta Mech. **129**, 63–72 (1998)
14. L.J. Crane, Flow past a stretching plate. J. Appl. Math. Phys. (ZAMP) **21**, 645–647 (1970)
15. B.K. Dutta, P. Roy, A.S. Gupta, Temperature field in flow over a stretching sheet with uniform heat flux. Int. Commun. Heat Mass Transf. **12**, 89–94 (1985)
16. E.M.A. Elbashbeshy, M.A.A. Bazid, Heat transfer over an unsteady stretching surface. Heat Mass Transf. **41**, 1–4 (2004)
17. E.M.A. Elbashbeshy, T.G. Emam, M.S. Abdel-Wahed, An exact solution of boundary layer flow over a moving surface embedded into an of fluid in the presence of magnetic field and suction/injection. Heat Mass Transf. **50**, 57–64 (2014)
18. R. Ellahi, M.M. Bhati, I. Pop, Effect of MHD and temperature dependent viscosity on the flow of non newtonian fluid. Appl. Math. Model. **37**, 1451–1467 (2013)
19. R. Ellahi, S.Z. Almri, A. Basit, A. Majeed, Effects of MHD and slip on heat transfer boundary layer flow over a moving plate based on specific entropy generation. J. Talibah Univ. Sci. **12**(10), 1–7 (2018)
20. N. Faraz, Y. Khan, Study of the rate type fluid with temperature dependent viscosity. Z. Naturforsch. **67a**, 460–468 (2012)
21. L.J. Grubka, K.M. Bobba, Heat transfer characteristic of a continuous stretching surface with variable temperature. J. Heat Transf. **107**, 248–250 (1985)
22. T. Hayat, Z. Abbas, I. Pop, S. Asghar, Effect of radiation and magnetic field on the mixed convection stagnation point flow over a vertical stretching sheet in porous medium. Int J. Heat Mass Transf. **53**, 466–474 (2010)
23. T. Hayat, Y. Saeed, A. Alsaedi, S. Asad, Effect of convective heat and mass transfer in flow of Powell-Eyring fluid past an exponentially stretching sheet. PLoS ONE **10**, e0133831 (2015)
24. A. Hussain, M.Y. Malik, M. Awais, T. Salahuddine, S. Bilal, Computational and physical aspects of MHD Prandtl-Eyring fluid flow analysis over a stretching sheet. Neural Comput. Appl. **31**, 425–433 (2017)
25. W. Ibrahim, B. Shankar, MHD Boundary layer flow and heat transfer of a nanofluid past a permeable stretching sheet with velocity, thermal and solutal slip boundary conditions. Comput. Fluids **75**, 1–10 (2013)
26. M. Ijaz, M. Zubair, T. Abbas, A. Riaz, Analysis of modified Fourier law in flow of ferromagnetic Powell-Eyring fluid considering two equal magnetic dipoles. Can. J. Phys. **9**(7), 772–776 (2019)
27. A. Ishak, R. Nazar, I. Pop, Boundary layer flow and heat transfer over an unsteady stretching vertical surface. Meccanica **44**, 369–375 (2009)
28. K. Jafar, R. Nazar, A. Iskak, I. Pop, MHD flow and heat transfer over stretching/shrinking sheet with external magnetic field, viscous dissipation and joule effects. Can J. Chem. Eng. **99**, 1–11 (2011)
29. L.Y. Jian, Z. Ismail, I. Khan, S. Shafie, Unsteady magneto hydrodynamics mixed convection flow in a rotating medium with double diffusion. AIP Conf. Proc. **1660**(1), 050082 (2015)
30. W.A. Khan, I. Pop, Boundary layer flow of a nanofluid past a stretching sheet. Int. J. Heat Mass Transf. **53**, 2477–2483 (2010)

31. M.S. Khan, I. Karim, L.E. Ali, A. Islam, Unsteady MHD free convection boundary layer flow of a nanofluid along a stretching sheet with thermal radiation and viscous dissipation effects. Int. Nano Lett. **2**, 1–9 (2012)
32. A.A. Khan, S. Muhammad, R. Ellahi, Q.M.Z. Zia, Bionic study of variable viscosity on MHD peristaltic fluid in an asymmetric channel. J. Magn. **21**, 273–280 (2016)
33. M. Khan, A. Shahid, T. Salahuddin, M.Y. Malik, M. Mushtaq, Heat and mass diffusions for Casson nanofluid flow over a stretching surface with variable viscosity and convective boundary conditions. J. Braz. Soc. Mech. Sci. Eng. **40**, 533 (2018)
34. M. Khan, T. Salahuddin, M.Y. Malik, A. Hussain, Change in internal energy of thermal diffusion stagnation point Maxwell nanofluid flow along with solar radiation and thermal conductivity. Chin. J. Chem. Eng. **27**(10), 2352–2358 (2019)
35. M. Khan, M.Y. Malik, T. Salahuddin, F. Khan, Generalized diffusion effects on Maxwell nanofluid stagnation point flow over a stretchable sheet with slip conditions and chemical reaction. J. Braz. Soc. Mech. Sci. Eng. **41**, 138 (2019)
36. M. Khan, A. Hussain, M.Y. Malik, T. Salahuddin, S. Aly, Numerical analysis of Carreau fluid flow for generalized Fourier's and Fick's laws. Appl. Numer. Math. **144**, 100–117 (2019)
37. M. Khan, M.Y. Malik, T. Salahuddin, S. Saleem, A. Hussain, Change in viscosity of Maxwell fluid flow due to thermal and solutal stratifications. J. Mo. Liq. **288**, 110970 (2019)
38. M.A. Makarem, A. Bakhtyari, M.R. Rahimpour, A numerical investigation on the heat and fluid flow of various nanofluid on a stretching sheet. Heat Transfer Asian Res. **47**, 347–365 (2017)
39. S. Mukhopadhyay, K. Bhattacharya, Unsteady flow of a Maxwell fluid over a stretching surface in presence of chemical reaction. J. Egypt Math. Soc. **20**, 229–234 (2012)
40. S. Mukhopadhyay, K. Bhattacharya, Heat transfer analysis of unsteady flow of a Maxwell fluid over a stretching surface in the presence of heat source/sink. Chin. Phy. Lett. **29**, 054703 (2012)
41. M. Mustafa, T. Hayat, A. Alsaedi, Unsteady boundary layer flow of nanofluid past an impulsively stretching sheet. J. Mech. **29**, 423–432 (2013)
42. M. Naryana, P. Sibanda, Laminar flow of a nanofluid film over an unsteady stretching sheet. Int. J. Heat Mass Transf. **55**, 7552–7560 (2012)
43. M.K. Partha, P.V.S.N. Murthy, G.P. Rajashekhar, Effect of viscous dissipation on the mixed convection heat transfer from an exponentially stretching surface. Heat Mass Transf. **41**, 360–366 (2005)
44. C.S.K. Raju, N. Sandeep, V. Sugunamma, M.J. Babu, J.V. Ramanareddy, Heat and Mass transfer in magneto hydrodynamic Casson fluid over an exponentially permeable stretching surface. Eng. Sci. Tech. Int. J. **19**, 45–52 (2016)
45. D. Sharma, R.C. Sharma, Effect of dust particles on the thermal convection in ferromagnetic fluid saturating a porous medium. J. Magn. Mater. **288**,183–195 (2005)
46. M. Sheikholeslami, Effect of spatially variable magnetic field on ferrofluid flow and heat transfer considering constant heat flux boundary condition. Eurphys. J. Plus **129**(11), 1–12 (2014)
47. S.R. Sheri, T. Thuma, Numerical study of heat transfer enhancement in MHD free convection flow over vertical plate utilizing nanofluid. Ain Shams Eng. J. **9**, 1169–1180 (2016)

Parallelization of Local Diagonal Extrema Pattern Using a Graphical Processing Unit and Its Optimization

B. Ashwath Rao and N. Gopalakrishna Kini

Abstract The incorporation of medical imaging devices in diagnosis has resulted in huge collection of medical images in hospitals and health centres. A search for a similar image from this image collection corresponding to a new medical image is a much needed help for junior doctors or students. This task involves describing each image in the image collection and also new image. After this, a similarity measure is applied between the image in the image collection and new image. Texture features have been found to be efficient in describing medical images owing to their high discerning capability. Various texture features have been introduced by researchers. Local Diagonal Extrema Pattern (LDEP) is a texture feature that uses only local diagonal neighbours, and hence the dimensionality of resulting feature vector is reduced. In this chapter we discuss parallel LDEP extractor on a GPU using CUDA. A constant kernel execution time for medical images of various sizes has been obtained on a GeForce GTX 1050 GPU.

1 Introduction

A new and efficient feature descriptor named Local Diagonal Extrema Pattern (LDEP) is introduced in [1]. In this descriptor, only diagonal neighbours are considered, as the diagonal neighbours contain most of the local information [2]. LDEP is a non-parametric visual descriptor and is useful in many applications. LDEP can be used in real-time applications due to its computational efficiency and simplicity. LDEP can be used in many domains, namely medical image analysis and understanding, object recognition, biometrics, content-based image retrieval, remote sensing, industrial inspection and document classification.

B. A. Rao · N. G. Kini (✉)
Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India
e-mail: ng.kini@manipal.edu

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_20

The LDEP has been defined in terms of first-order local diagonal derivative. The detailed steps of computing LDEP feature are described in the following subsection.

1.1 First-Order Local Diagonal Derivative

Let us consider an image M having $m1$ rows and $m2$ columns. Let $P^{i,j}$ be any centre pixel. The tth diagonal neighbour of $P^{i,j}$ at a distance R be $P_t^{i,j}$, where $t \in [1, 4]$. Let $I_t^{i,j}$ and $I_{i,j}$ be pixel intensities of pixels at $P_t^{i,j}$ and $P^{i,j}$, respectively. For each neighbour $t \in [1, 4]$, we define α and β as follows:

$$\alpha, \beta = \begin{cases} -R, +R & t = 1 \\ -R, +R & t = 2 \\ +R, -R & t = 3 \\ +R, +R & t = 4 \end{cases} \tag{1.1}$$

The first-order diagonal derivative for $\gamma = 0, 1, 2$ is

$$I_{t,\gamma}^{i,j} = I_{(1+mod(t,4))}^{i,j} - I_t^{i,j} \tag{1.2}$$

where $mod(x, y)$ is the remainder when x is divided by y.

A function sign is defined as follows:

$$sign(\lambda) = \begin{cases} 0 & \lambda < 0 \\ 1 & \lambda >= 0 \end{cases} \tag{1.3}$$

Two variables τ_{max} and τ_{min} are defined as follows:

$$\tau_{max} = argmin_t \left(sign(I_{t,\gamma}^{i,j} = 0, \forall \in [0, 2] \right) \tag{1.4}$$

$$\tau_{min} = argmax_t \left(sign(I_{t,\gamma}^{i,j} = 1, \forall \in [0, 2] \right) \tag{1.5}$$

1.1.1 Local Diagonal Extrema Pattern

The Local Diagonal Extrema Pattern for the pixel $P_{i,j}$ is defined as follows:

$$LDEP^{i,j} = (LDEP_1, LDEP_2 \ldots LDEP_{dim}) \tag{1.6}$$

where dim is the length of the pattern. The kth element of LDEP is defined as follows:

$$LDEP_k^{i,j} = \begin{cases} 1 & \text{if } k = \tau_{max} + 8\delta \text{ or k} = \tau_{min} + 4 + 8\delta \\ 0 & \text{otherwise} \end{cases} \tag{1.7}$$

We define local extrema difference factor $\Delta_{max}^{i,j}$ and $\Delta_{min}^{i,j}$ as follows:

$$\Delta_{max}^{i,j} = I_{\tau_{max}}^{i,j} - I_{i,j} \tag{1.8}$$

$$\Delta_{min}^{i,j} = I_{\tau_{min}}^{i,j} - I_{i,j} \tag{1.9}$$

We define δ as follows:

$$\delta = \begin{cases} 0 & \text{if } \left(sign(\Delta_{max}^{i,j}) = 0 \text{ and } sign(\Delta_{min}^{i,j}) = 0\right) \\ 1 & \text{if } \left(sign(\Delta_{max}^{i,j}) = 1 \text{ and } sign(\Delta_{min}^{i,j}) = 1\right) \\ 2 & \text{else} \end{cases} \tag{1.10}$$

The steps involved in computing LDEP pattern are shown in Fig. 1.

1.1.2 Few Sample Images and Corresponding LDEP Images

LDEP pattern can be obtained for any image. In this study, we have considered medical images. The amount of computational work involved in determining LDEP for an image is proportional to the size of image. We have considered medical images of size 256 x 256, 512 x 512 and 1024 x 1024. The set of medical images and their LDEP pattern is shown in Fig. 2.

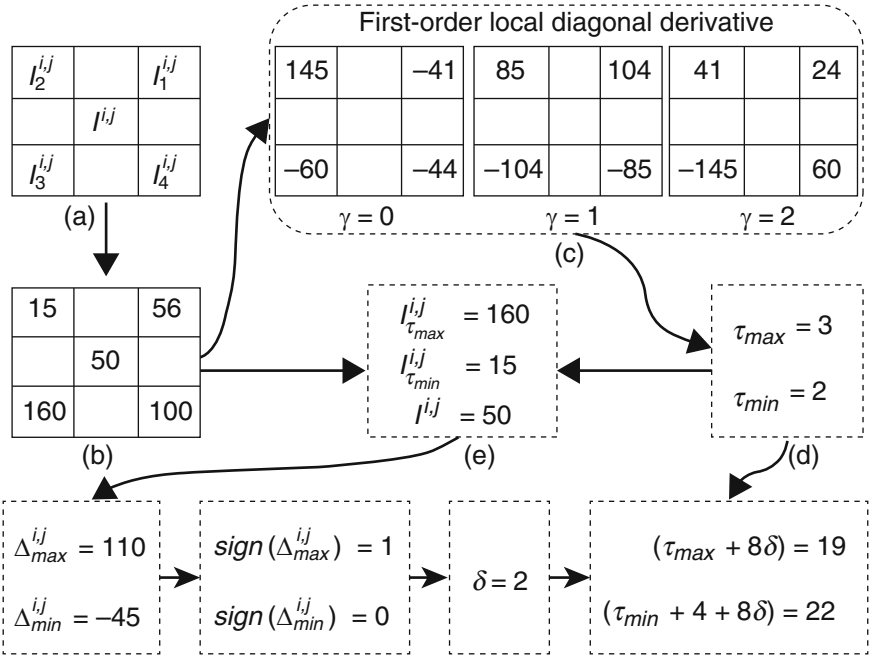


Fig. 1 An example showing LDEP feature computation

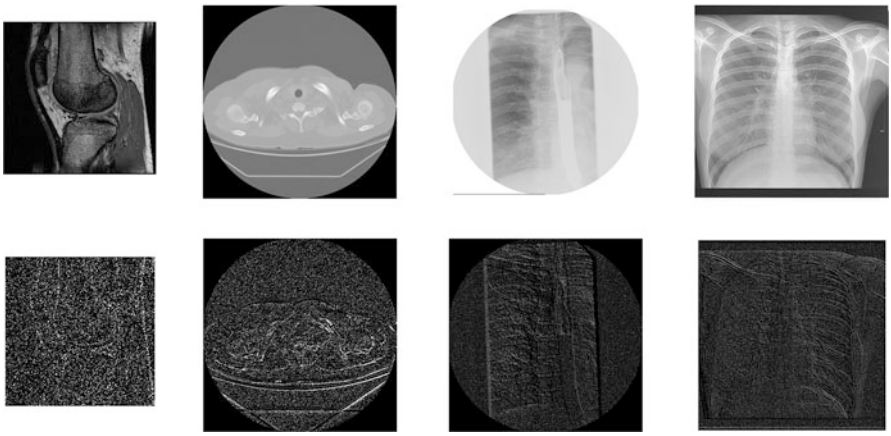


Fig. 2 Original medical images and their corresponding LDEP image

1.2 GPU Programming Using CUDA

A lot of general-purpose or compute-intensive scientific computations can be performed on a GPU after the release of general-purpose programmable GPUs. Prior to this, only graphical computations can be performed on a GPU. NVIDIA released CUDA programming for GPUs in 2007.

Ever since the release of CUDA, many general-purpose and scientific or industrial applications have been solved using it. CUDA applications when run, generate threads on the GPU. The developer can control the timing of threads, number of threads and what subtask will be performed by each thread.

GPUs contain multiple streaming multiprocessors (SMs). Each SM contains streaming processors (SPs). Threads are executed within SPs. Threads are organized in hierarchy. Each thread belongs to a thread block. A collection of thread blocks are organized into grids. A three-dimensional block structure and a two-dimensional grid structure can be specified during running. Each thread will execute a function called kernel.

The input data need to be loaded into GPU memory first. A kernel function will read input from this memory. Similarly an output from kernel will be written to global memory. The time for running kernel and loading or unloading data from or to global memory can be computed with the help of CUDA events.

The organization of this chapter is as follows. In Sect. 2, we present literature review. In Sect. 3 we discuss methodology. The results of experiments are discussed in Sect. 4. Conclusion and future scope is discussed in Sect. 5.

2 Literature Review

LDEP feature has been found useful in medical diagnosis. Emphysema medical image classification has been carried out by utilizing LDEP features [3]. LDEP features are found to be useful in distinguishing intra-class and inter-class regions

in a fingerprint. A fingerprint classification using LDEP features has been proposed in [4].

Subsequent to the introduction of LDEP, several variants of LDEP has been proposed. Few variants are Local Directional Extrema Number Pattern [5] and Local Ternary Quantized Extrema Pattern [6]. Another feature Local Diagonal Laplacian Pattern [7] has been devised in the similar manner as LDEP.

3 Methodology

In this section a description to parallel extraction of Local Diagonal Extrema feature is provided. The image is placed in the memory after decoding. We have considered medical images in our experiments. The DCM Toolkit [8] is used for reading DICOM images. The sequential algorithm for computing LDEP feature for an image is provided in Algorithm 1.

3.1 Algorithms

As shown in Algorithm 1, we take diagonal neighbours around a pixel. Then we compute first-order local diagonal derivative. From this, we derive the feature vector.

We have used CUDA for our implementation, as CUDA implementation has shown better results than OpenCL [9, 10]. The parallel algorithm using CUDA on a GPU is provided in Algorithm 2.

The parallel version of LDEP feature extraction in Algorithm 2 is similar to its sequential counterpart. However, since it is run parallel, the loop for navigating in the x and y directions is not present. Each CUDA thread will handle computation of feature corresponding to a centre pixel.

3.2 System Configuration

The host configuration and device configuration are provided in the following subsections.

3.2.1 Host Configuration

The host configuration is detailed as follows in Table 1.

Algorithm 1 Sequential Local Diagonal Extrema Pattern algorithm

```
1: procedure GETLDEP(pdata, w, h, out)
2:     for ty ←1, h − 1 do
3:         for tx ← 1, w − 1 do
4:             centrePx ← pdata[ty ∗ w + tx]
5:             diagNeigh[0] ← pdata[(ty − 1) ∗ w + tx + 1]
6:             diagNeigh[1] ← pdata[(ty − 1) ∗ w + tx − 1]
7:             diagNeigh[2] ← pdata[(ty + 1) ∗ w + tx − 1]
8:             diagNeigh[3] ← pdata[(ty + 1) ∗ w + tx + 1]
9:             maxNegDiff ← ∞
10:            minPosDiff ← ∞
11:            for g ← 0, 2 do
12:                for t ← 0, 3 do
13:                    diff ← diagNeigh[(t + g + 1)mod4] − diagNeigh[t]
14:                    if diff < 0 and diff < maxNegDiff then
15:                        maxNegDiff ← diff
16:                        tauMax ← t + 1
17:                    end if
18:                    if diff >= 0 and diff < minPosDiff then
19:                        minPosDiff ← diff
20:                        tauMin ← t + 1
21:                    end if
22:                end for
23:            end for
24:            itaumax ← diagNeigh[tauMax − 1]
25:            itaumin ← diagNeigh[tauMin − 1]
26:            if itaumax − centrePx >= 0 and itaumin − centrePx >= 0 then
27:                delta ← 1
28:            else if itaumax − centrePx < 0 and itaumin − centrePx < 0 then
29:                delta ← 0
30:            else
31:                delta ← 2
32:            end if
33:            out[(ty − 1) ∗ w + tx − 1] ← 1 << (tauMax + 8 ∗ delta) + 1 << (tauMin +
   4 + 8 ∗ delta)
34:        end for
35:    end for
36: end procedure
```

3.2.2 Device Configuration

The device configuration is detailed as follows in Table 2.

3.3 Optimizations on the Algorithm

As stated in [11], memory access involving a data in global memory will take a considerable amount of time. Hence one must reduce accessing global memory.

Algorithm 2 Parallel Local Diagonal Extrema Pattern algorithm on a GPU using CUDA

procedure GETLDEPPARALLEL($pdata, w, h, out$)
  $tx \leftarrow$ blockIdx.x*blockDim.x+threadIdx.x
  $ty \leftarrow$ blockIdx.y*blockDim.y+threadIdx.y
  if ty=0 or ty=h-1 or tx=0 or tx=w-1 **then**
    return
  end if
  $centrePx \leftarrow pdata[ty * w + tx]$
  $diagNeigh[0] \leftarrow pdata[(ty - 1) * w + tx + 1]$
  $diagNeigh[1] \leftarrow pdata[(ty - 1) * w + tx - 1]$
  $diagNeigh[2] \leftarrow pdata[(ty + 1) * w + tx - 1]$
  $diagNeigh[3] \leftarrow pdata[(ty + 1) * w + tx + 1]$
  $maxNegDiff \leftarrow \infty$
  $minPosDiff \leftarrow \infty$
  for $g \leftarrow 0, 2$ **do**
    for $t \leftarrow 0, 3$ **do**
      $diff \leftarrow diagNeigh[(t + g + 1) mod 4] - diagNeigh[t]$
      if $diff < 0$ and $diff < maxNegDiff$ **then**
        $maxNegDiff \leftarrow diff$
        $tauMax \leftarrow t + 1$
      end if
      if $diff >= 0$ and $diff < minPosDiff$ **then**
        $minPosDiff \leftarrow diff$
        $tauMin \leftarrow t + 1$
      end if
    end for
  end for
  $itaumax \leftarrow diagNeigh[tauMax - 1]$
  $itaumin \leftarrow diagNeigh[tauMin - 1]$
  if $itaumax - centrePx >= 0$ and $itaumin - centrePx >= 0$ **then**
    $delta \leftarrow 1$
  else if $itaumax - centrePx < 0$ and $itaumin - centrePx < 0$ **then**
    $delta \leftarrow 0$
  else
    $delta \leftarrow 2$
  end if
  $out[(ty-1)*w+tx-1] \leftarrow 1 << (tauMax+8*delta)+1 << (tauMin+4+8*delta)$
end procedure

A GPU provides shared memory that has considerably less access time. Shared memory is shared by all threads in a thread block. In each thread the pixel intensities around a centre pixel are accessed. All the pixel intensities are stored in global memory. A shared memory copy of pixel intensities is maintained to reduce the memory access time. In the experiments the kernel execution time is reduced when pixel intensities are accessed from shared memory rather than global memory.

However, for all block sizes the shared memory cannot be set up due to the limited shared memory. We have utilized shared memory when block size is 16×16.

Table 1 Host configuration

CPU	Intel Xeon X5550 2.67 GHz × 16
Main memory	47.2 GB

Table 2 Device configuration

GPU	GeForce GTX 1050 Ti
Global memory size	4 GB
GPU clock rate	1.62 GHz
Constant memory size	64 KB
Maximum threads per block	1024

Table 3 Sequential LDEP feature extraction time

Size (in pixels)	Function time for non-optimized (ms)	Program time for non-optimized (ms)	Function time for optimized (ms)	Program time for optimized (ms)
256 × 256	14.656512	28.268448	3.934208	23.712831
512 × 512	47.779839	63.436897	12.325888	27.480064
1024 × 1024	209.745926	231.110748	53.884930	73.078781

4 Experimental Results

In this section, the results of experiments are presented.

4.1 Sequential LDEP Feature Extraction Results

The sequential feature extraction time is shown in Table 3. We present the time for images of various sizes and with, without compiler optimization.

4.2 Parallel LDEP Feature Extraction Time

The parallel feature extraction time is shown in Table 4. We present the time for images of various sizes and with varying thread block sizes.

4.3 Performance Analysis

In this section analysis of performance is discussed. In Tables 3 and 4 the sequential and parallel time for LDEP feature extraction is shown. The efficiency of parallelization is measured using speedup. Speedup is the ratio of sequential time

Table 4 Parallel LDEP feature extraction time

Size (in pixels)	Thread block size	Using shared memory	Kernel time (ms)	Program time (ms)
256 × 256	16 × 16	No	0.049472	14.982080
256 × 256	16 × 16	Yes	0.039072	15.138848
256 × 256	256 × 256	No	0.001984	16.839680
512 × 512	16 × 16	No	0.171072	17.307360
512 × 512	16 × 16	Yes	0.126880	20.827040
512 × 512	256 × 256	No	0.002080	15.524512
512 × 512	512 × 512	No	0.001952	17.130079
1024 × 1024	16 × 16	No	0.739840	23.474913
1024 × 1024	16 × 16	Yes	0.499616	23.178207
1024 × 1024	256 × 256	No	0.008448	22.589567
1024 × 1024	512 × 512	No	0.007680	25.194176
1024 × 1024	1024 × 1024	No	0.007744	26.014751

Table 5 Performance improvement measure (Speedup)

Size (in pixels)	Function speedup with non-optimized	Program speedup with non-optimized	Function speedup with optimized	Program speedup with optimized
256 × 256	7387.35	1.67	1982.96	1.40
512 × 512	24477.37	3.70	6314.49	1.60
1024 × 1024	27084.95	8.88	6958.28	2.80

and parallel time. Since parallel time is in the denominator, very low parallel time will increase the speedup.

It is a practice to present the speedup considering compiler-optimized sequential code running time. Apart from this, we have also shown speedup measured with regular compilation (non-optimized). The speedup obtained in our study is presented in Table 5. We have also shown overall program speedup and kernel speedup for images of different sizes. The size of an image has a direct impact on the amount of computation and hence directly on the time.

Graphically, the kernel speedup is shown in Fig. 3.

The overall program speedup is shown in Fig. 4.

We have obtained near constant kernel time for extracting LDEP feature. Fig. 5 shows the relative amount of work involved in extracting LDEP feature for 256×256, 512×512 and 1024×1024 size images relative to 256×256 size image. The relative kernel execution time for the same image sizes is shown on the right. Hence using a Graphical Processing Unit, the kernel execution time is reduced significantly, and we are able to achieve constant kernel execution time.

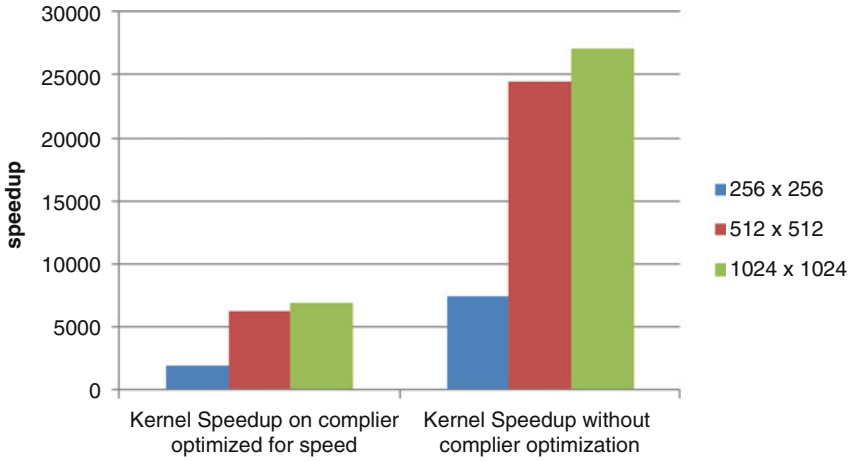


Fig. 3 Kernel speedup with compiler optimized for speed and non-optimized

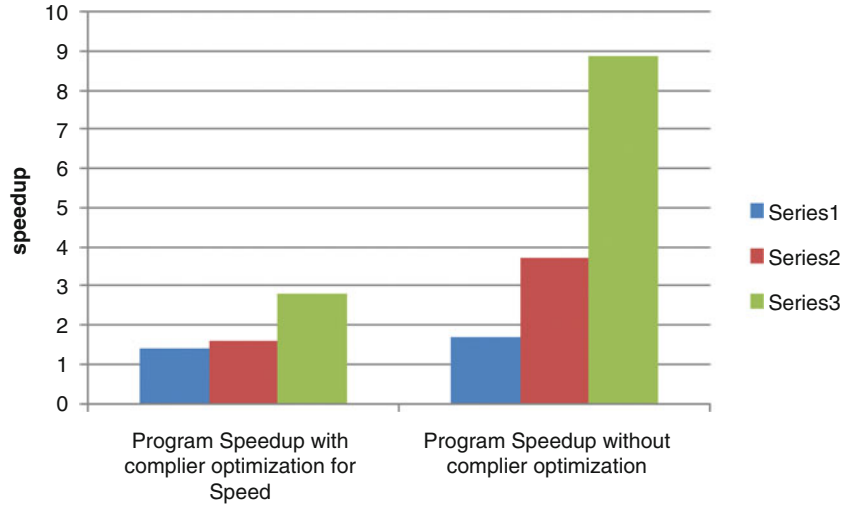


Fig. 4 Program speedup with compiler optimized for speed and non-optimized

5 Conclusion

Local Diagonal Extrema Pattern feature extraction has been parallelized and run using a GPU. The feature extraction time is directly proportional to the size of the image. The highest speedup of 27084.95 is obtained when measured with non-optimized sequential function code and a speedup of 6958.28 when measured with compiler-optimized sequential code for an image size of 1024×1024. We have obtained near constant kernel execution time. Usage of shared memory improved

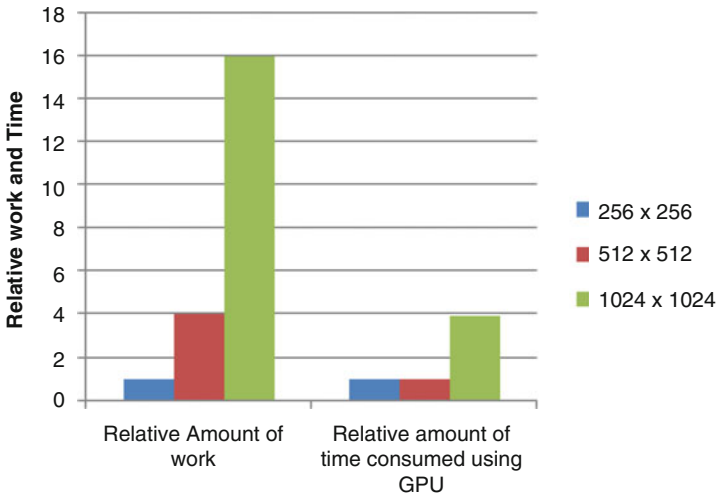


Fig. 5 Relative amount of work involved in feature extraction and relative time obtained

the speed of execution. However, since shared memory within a thread block is limited, its usage is not possible for every thread block size. By setting appropriate grid size and thread block size, it is possible to obtain constant feature extraction time for Local Diagonal Extrema Pattern.

References

1. S.R. Dubey, S.K. Singh, R.K. Singh, Local diagonal extrema pattern: a new and efficient feature descriptor for CT image retrieval. IEEE Signal Process. Lett. **22**(9), 1215–1219 (2015)
2. R. Gupta, H. Patil, A. Mittal, Robust order-based methods for feature description, in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, Piscataway, 2010), pp. 334–341
3. S.J. Narayanan, R. Soundrapandiyan, B. Perumal, C.J. Baby, Emphysema medical image classification using fuzzy decision tree with fuzzy particle swarm optimization clustering, in *Smart Intelligent Computing and Applications* (Springer, Berlin, 2019), pp. 305–313
4. A. Manickam, R. Haldar, S.M. Saqlain, V. Sellam, R. Soundrapandiyan, Fingerprint image classification using local diagonal and directional extrema patterns. J. Electron. Imag. **28**(3), 033027 (2019)
5. A.R. Rivera, Jo.R. Castillo, O.O. Chae, Local directional number pattern for face analysis: face and expression recognition. IEEE Trans. Image Process. **22**(5), 1740–1752 (2012)
6. G. Deep, L. Kaur, S. Gupta, Directional local ternary quantized extrema pattern: a new descriptor for biomedical image indexing and retrieval. Eng. Sci. Technol. Int. J. **19**(4), 1895–1909 (2016)
7. P.K.R. Yelampalli, J. Nayak, Local diagonal laplacian pattern a new MR and CT image feature descriptor, in *Progress in Advanced Computing and Intelligent Engineering* (Springer, Berlin, 2018), pp. 69–78

8. OFFIS Computer Science Institute, DCMTK - DICOM toolkit [Online]. Available: http://dicom.offis.de/dcmtk.php.en
9. J. Fang, A.L. Varbanescu, H. Sips, A comprehensive performance comparison of CUDA and OpenCL, in *2011 International Conference on Parallel Processing* (IEEE, Piscataway, 2011), pp. 216–225
10. K. Karimi, N.G. Dickson, F. Hamze, A performance comparison of CUDA and OpenCL (preprint, 2010). arXiv:1005.2581
11. D.B. Kirk, W.H. Wen-Mei, *Programming Massively Parallel Processors: A Hands-on Approach* (Morgan Kaufmann, Burlington, 2016)

On the Recommendations for Reducing CPU Time of Multigrid Preconditioned Gauss–Seidel Method

Abdul Hannan Faruqi, M. Hamid Siddique, Abdus Samad, and Syed Fahad Anwer

Abstract Gauss–Seidel method is one of the simplest available iterative methods for solving systems of linearized equations. It can effectively reduce high-frequency errors but performs poorly with errors of low frequency. Multigrid (MG) utilizes this quality of the point-wise methods by successively coarsening the grid, so that the lowest frequency errors appear as high frequency and can be easily reduced. In this work, optimization study was performed to lower the CPU time of the Multigrid method. We have considered several parameters, such as the number of grid levels used, the number of inner iterations (iterations at each intermediate grid), the overall coarsening and interpolation cycle (V and W), and the number of these cycles in each iteration. A surrogate model is used to predict optimum value for these parameters. In this chapter, MG is used with a Gauss–Seidel solver for a 2D conduction problem with Dirichlet boundary condition on a 256×256 structured grid. The results suggest that a W cycle is more efficient than a V cycle and should be executed to the penultimate grid level during both restriction (coarsening) and prolongation.

1 Introduction

Mathematical modelling of dynamic systems forms an essential part of modern engineering design. However, the governing differential equations are often complex and difficult to solve analytically. Therefore, they are discretized on a finite number of grid points, and the resultant algebraic equations are solved using

A. H. Faruqi (✉) · S. F. Anwer
Department of Mechanical Engineering, Aligarh Muslim University, Aligarh, India

M. H. Siddique
Department of Mechanical Engineering, ADCET, Ashta, Maharashtra, India

A. Samad
Department of Ocean Engineering, IIT Madras, Chennai, Tamil Nadu, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_21

iterative solvers. Starting from an initial or guessed solution, these solvers are able to quickly drop the residuals (a measure of error) to a few orders of magnitude below zero. However, once the high-frequency errors have been smoothed out, they become ineffective. This is because the low-frequency errors, whose wavelength spans across almost the entire solution domain, are not felt by the point-wise solvers. This inadequacy may be overcome by using Multigrid methods.

The Multigrid method stemmed from the pioneering work of Brandt [1] for elliptic partial differential equations and was later applied to Euler equations by Jameson [2]. A comprehensive review of Multigrid Schemes (particularly, Algebraic Multigrid) can be found in the paper by Stüben [3]. It is a smart technique that utilizes the effectiveness of the point-wise solvers for high-frequency errors, by successively coarsening the grid and smoothing the residuals at each intermediate grid level. In this way, even the errors of the lowest frequency can be made to appear as high frequency and are easily reduced by the simple iterative solvers.

However, to utilize the full potential of this technique, it needs to be optimized with respect to several inherent parameters. Vakili and Darbandi [4] have carried out optimization study on Algebraic Multigrid used as a preconditioner to GMRES. Suero et al. [5] have carried out extensive study of various AMG parameters for 2D steady-state heat diffusion equations. The same test problem has been taken up in this chapter, and the results are compared with the analytical solution for validation. We have used the CPU time as the objective of optimization as against iteration count used in other studies and have included crucial factors like V and W cycles and inner iterations. Unlike the manual optimization carried out in previous studies, we have developed surrogate models to approximate the data collected from test runs of the original code for carrying out multi-parametric optimization. The surrogate model is then used to identify the reason of feasible solution known as Pareto-optimal front (POF). These feasible solutions are again validated at clustered points until the approximate POF points converge with the program results.

2 The Test Problem and Its Modelling

The two-dimensional heat conduction problem with Dirichlet boundary condition has been taken as the test problem for this work. The governing differential equation in Cartesian coordinates is given as

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = S, \tag{1}$$

which is Poisson's equation with source term

$$S = -2\left(1 - 6x^2\right) y^2 \left(1 - y^2\right) + \left(1 - 6y^2\right) x^2 \left(1 - x^2\right). \tag{2}$$

Table 1 Problem specifications

Governing Equation	Boundary Conditions	Analytical Solution
Two-Dimensional Poisson's equation $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = S$ where $S = -2\left(1-6x^2\right) y^2 \left(1-y^2\right) + \left(1-6y^2\right) x^2 \left(1-x^2\right)$	T(0,y) = T(x,0) = 0 T(1,y) = T(x,1) = 0	$T(x,y) = (x^2 - x^4)(y^4 - x^2)$

The equation is solved over a unitary square solution domain using a 256 × 256 structured grid and discretized using the finite difference (FD) method. The derivates are approximated by second-order accurate central differencing scheme resulting in a system of linear equations of the form **AT=b**. The system is solved explicitly using Gauss–Seidel iterative solver with Multigrid as a preconditioner. Table 1 summarizes the problem along with the boundary conditions.

3 The Multigrid Methodology

The Multigrid cycle essentially serves the purpose of reducing the low-frequency errors that persist in the solutions obtained by using basic iterative solvers. The computational effort required to directly reduce these errors is extremely high. The use of Multigrid however drastically reduces the effort required to achieve the same level of residual tolerance. This reduction is achieved by utilizing the effectiveness of the solver in reducing high-frequency errors by successive coarsening of the original fine grid. The complete cycle is illustrated using an example. Figure 1 shows a hypothetical error distribution on the original fine grid.

Restriction The strategy is to carry out a few iterations on the fine grid, then calculate the residuals, and transfer it to successively coarser grids. The low-frequency residuals on the original grid appear as high frequency on the coarser grid (Fig. 2). A few sweeps of the smoother at each coarser grid level (inner iterations) effectively bring down these residuals.

Prolongation The error estimate from the coarse grid is interpolated back onto the finer grid and added to the previous iteration's solution. This cycle rapidly brings down the low-frequency component of the error, thereby, greatly reducing the computational effort.

Figure 3 shows an alternative way of looking at the effect of the Multigrid algorithm. Using normal Gauss–Seidel solver, the solution reaches its final value in a bounded manner. The use of Multigrid, however, relaxes the bound (by smoothening errors on coarser grids), allowing for oscillations about the final value which get damped with every successive iteration.

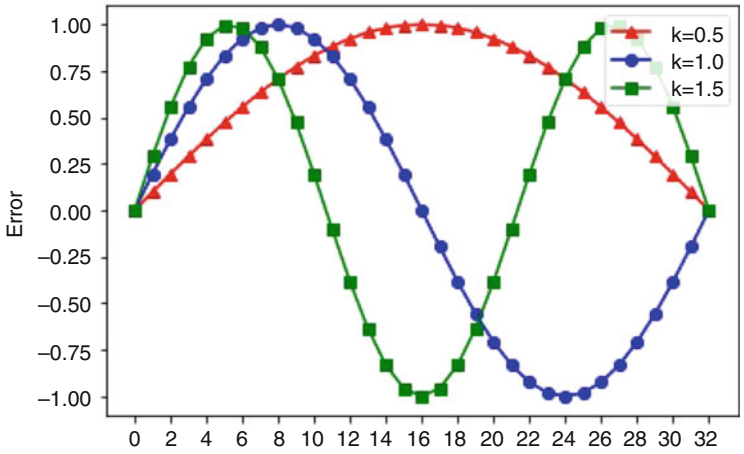


Fig. 1 Error distribution on the fine grid

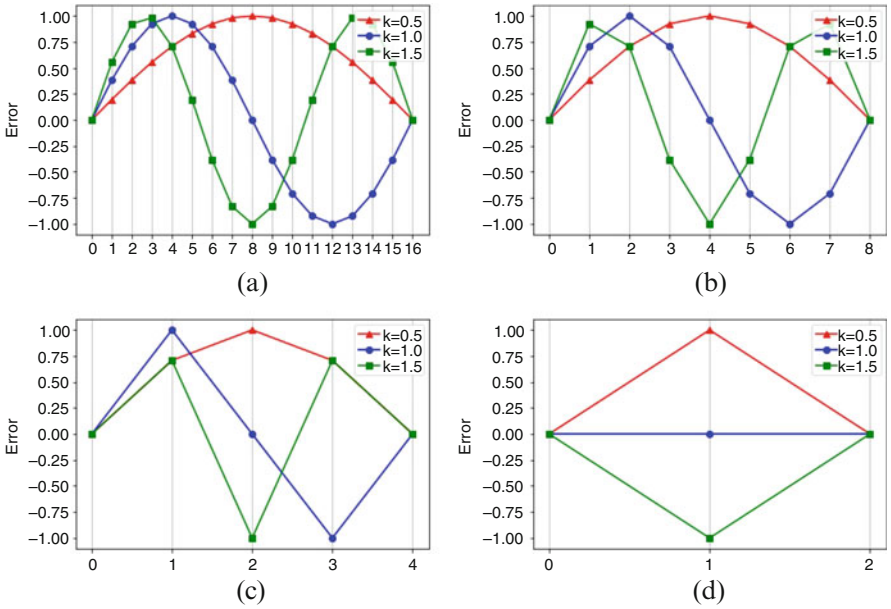


Fig. 2 Effect of coarsening on smoothness of error. (**a**) Level 1. (**b**) Level 2. (**c**) Level 3. (**d**) Level 4

The overall performance of this method depends on the value chosen for the various parameters, which include

- the number of grid levels,
- the number of inner iterations,

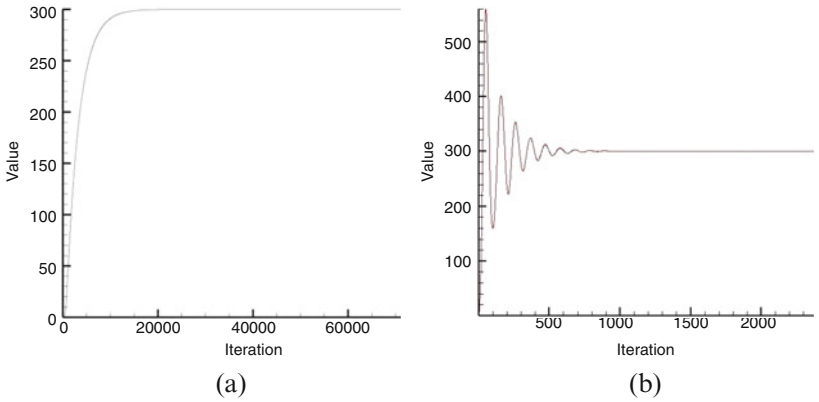


Fig. 3 Solution convergence at the central point of the domain. (**a**) Normal Gauss–Seidel. (**b**) Multigrid preconditioned Gauss–Seidel

- the cycle (V or W), and
- the number of cycles in each iteration.

To evaluate the effect of these parameters on the performance of the Multigrid method, we solved the test problem using an indigenously developed Multigrid code on an 8-core i7-6th generation machine with 8 GB of RAM. The results of the code were first validated using the analytical solution, then the computational time was recorded for different test runs, and the optimum value was sought using surrogate-based analysis.

4 Code Validation

We have used a 256 × 256 structured grid for solving the discretized system of equations. The results are validated by comparing with the analytical solution. Figure 4 shows the plot of the results for comparison.

The two plots are almost identical. For a more quantitative analysis, the values from the two solutions are compared at a set of discrete points along the central line, y = 0.5. The results are tabulated in Table 2.

As can be seen from the table, the maximum relative error between the actual and the calculated values is of the order of 10^{-4}, which is about 0.01%. Figure 5 shows the entire error distribution along the central line with a maximum near the left boundary ($x \cong 0$), where the actual solution is the smallest in magnitude. This leads to a large relative error ($\sim$0.4%), which drops steeply towards the right.

Hence, the numerical solution is considered sufficiently accurate to proceed for further analysis.

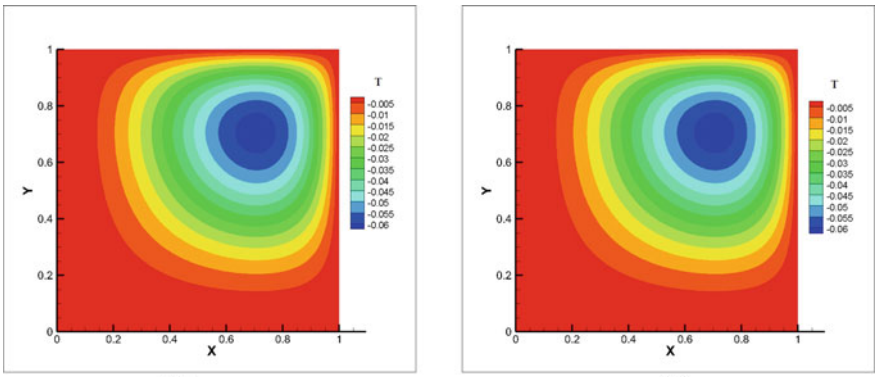


Fig. 4 Comparison of results. (**a**) Analytical solution. (**b**) Numerical solution

Table 2 Comparison of numerical and analytical results

x	T_{act}	T_{calc}	Error	Relative error
Error distribution at y= 0.5				
0	0	0	0	0
0.125	−0.00288391	−0.00288363	−2.80E−07	9.71E−05
0.25	−0.0109863	−-0.0109858	−5.00E−07	4.55E−05
0.375	−0.0226593	−0.0226587	−6.00E−07	2.65E−05
0.5	−0.0351563	−0.0351555	−8.00E−07	2.28E−05
0.625	−0.044632	−0.0446312	−8.00E−07	1.79E−05
0.75	−0.0461426	−0.0461419	−7.00E−07	1.52E−05
0.875	−0.0336456	−0.0336453	−3.00E−07	8.92E−06
1	0	0	0	0

4.1 Results

The data collected in the test runs shows some identifiable trends which shall be discussed next. We will consider each parameter one by one.

Figures 6, 7, and 8 give an idea of what each parameter means.

1. Effect of the number of grid levels: the number of grid levels refers to the number of restriction and prolongation operations done in a cycle. Given below is the plot of CPU time with the number of levels along a single line in the multidimensional design space.

   It can be seen that the minimum value is reached by going up to the 7th (penultimate) grid level.
2. Effect of the number of inner iterations: the smoothing iterations performed at each intermediate grid level are termed as inner iterations. Their effect can be studied by fixing the other parameters and varying the number of inner iterations.

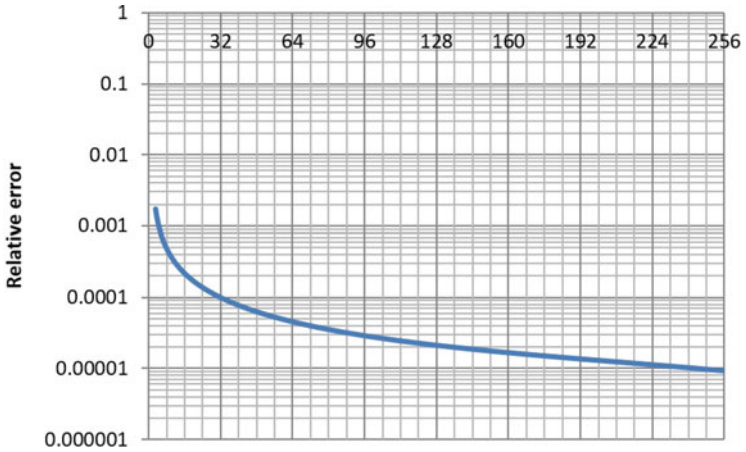


Fig. 5 Error distribution along the central line

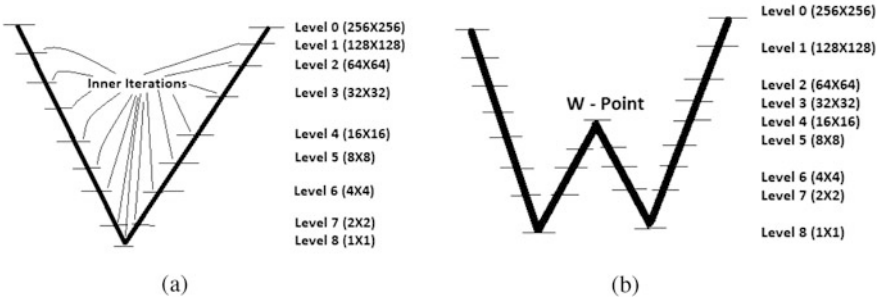
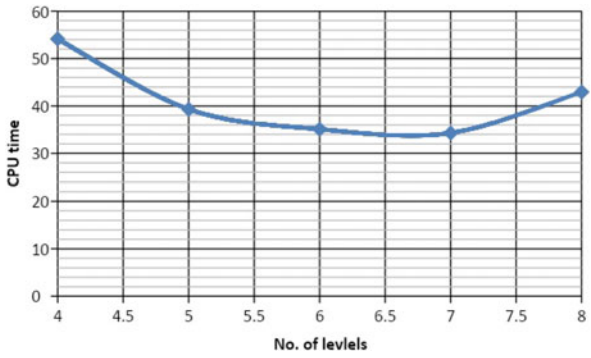


Fig. 6 Graphical description of the parameters. (**a**) V cycle. (**b**) W cycle

Fig. 7 Effect of number of grid levels



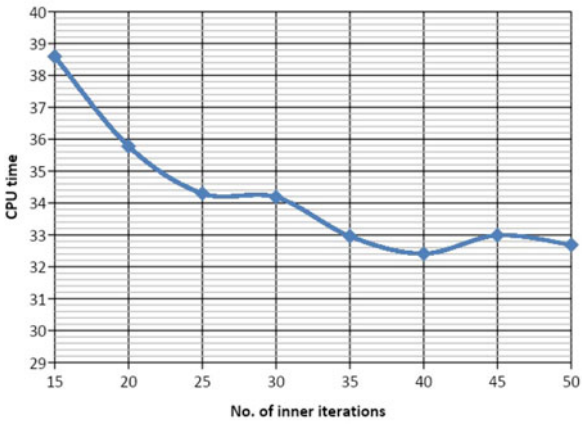


Fig. 8 Effect of number of inner iterations

The trend suggests that the CPU time decreases with an increase in the number of inner iterations up to a certain point after which increasing the iterations does not have any effect. If we go on increasing the iterations further, the CPU time even begins to increase.

3. Effect of the cycle: the V and W cycles were considered for comparison. Table 3 gives the mean CPU time and the standard deviation of all test runs of both the cycles (with varying values of other parameters).

   Clearly, the W cycle gives better performance than the V cycle, as shall be made clearer from the following plots.

4. Effect of the number of cycles: effect of the number of cycles is a parameter that cannot be directly quantified. It needs to be assessed with various combinations of the other parameters. The plots in Fig. 9 provide some useful information in this regard and sum up the effects of the different parameters.

5 Surrogate Modelling and Analysis

To optimize a design influenced by multiple parameters, each point in the multidimensional design space needs to be explored. This process is time-consuming and computationally expensive. An efficient way to solve this issue is to develop low-fidelity surrogate models that mimic the actual experiment. A thorough discussion on surrogate-based analysis and optimization (SBAO) is provided in the paper by Queipo et al. [6]. An excellent review of surrogate-based optimization of centrifugal pumps is given by Siddique et al. [7].

Surrogate models are low-fidelity regression models constructed using data drawn from high-fidelity models. These models can generate thousands of approximate results from a few samples, thereby simplifying the process of finding the optimal solution.

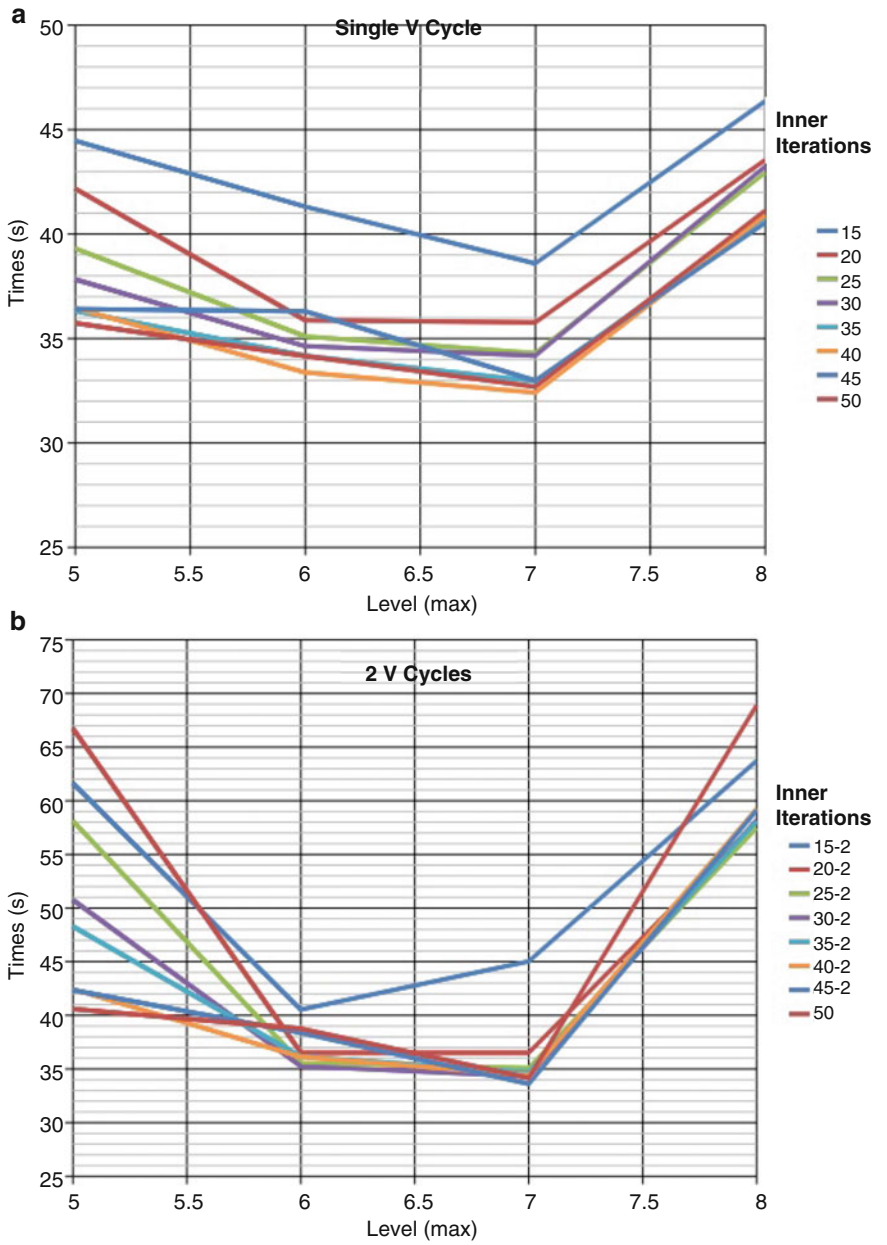


Fig. 9 **(a), (b)** V cycle **(c), (d)** W cycle

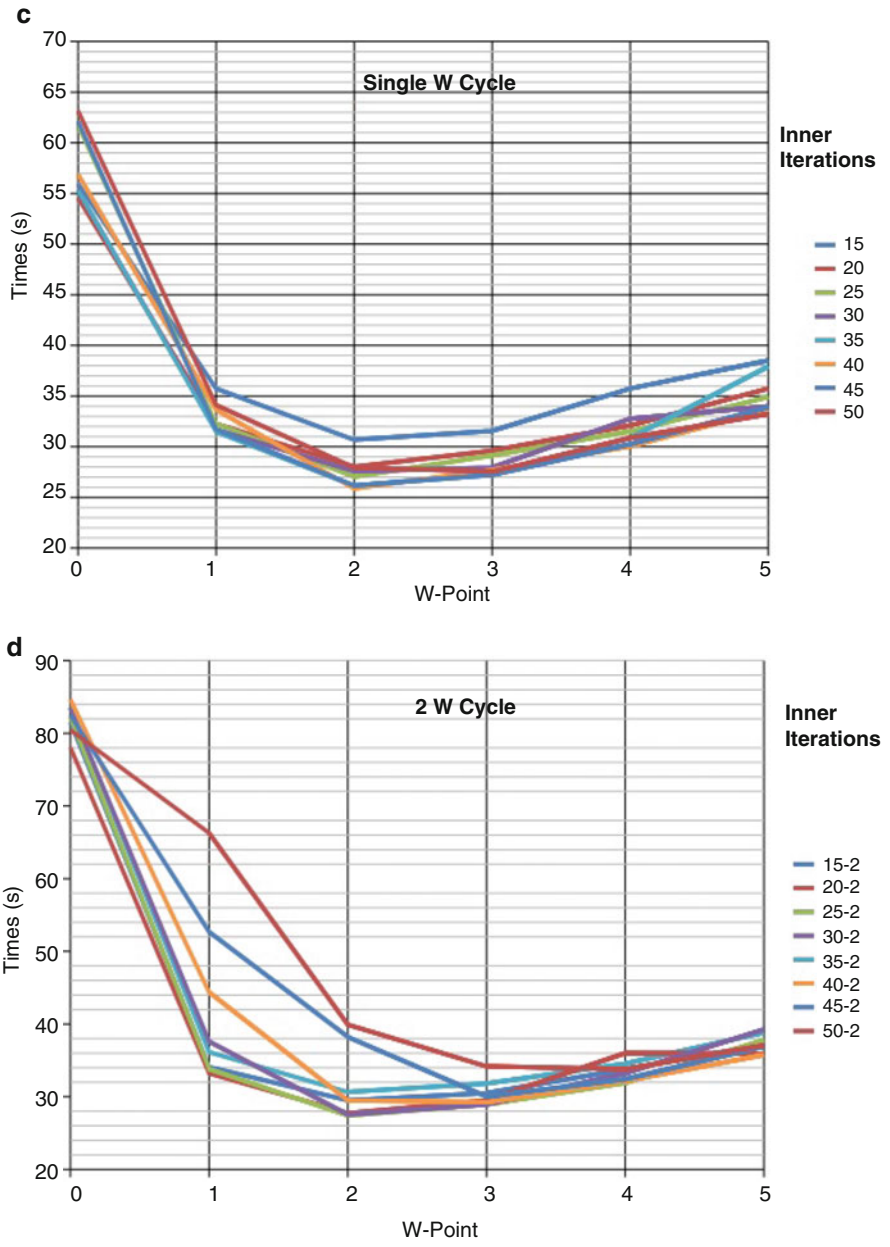


Fig. 9 (continued)

Table 3 Comparison of V and W cycles

CPU time (in s)	V cycle	W cycle
Mean	47.537804	39.17872
Standard deviation	19.362992	15.52617

5.1 Methodology

Design of Experiment The first step in surrogate modelling is the Design of Experiment, which consists of selecting sample points from the design space for the purpose of determining the relationship between the objectives and the design variables. This step is the most crucial as it effects the overall output of the model.

Surrogate-Based Modelling The objective function is then evaluated at these sample points, and this database is used to train the surrogates. Various surrogate models have been used by researchers, and each of them is problem-dependent. Some important examples are Response Surface Approximation (RSA), Kriging Model (KRG), and Radial Basis Neural Network (RBNN). A multiple surrogate technique introduced by Goel et al. [8] is used to improve the reliability and robustness of the surrogate approximation. This technique, termed as weighted-average surrogate, assigns a weight (ω) to each surrogate model and then determines the approximate model based on the weights assigned to the individual models. The most commonly used weighing method is based on the magnitude of the errors. This scheme can be expressed as

$$\omega_i = \frac{\sum_{j=1, j\neq i}^{m} e_j}{\sum_{j=1}^{m} e_j}, \tag{3}$$

where e_j is the global database error measured for the jth surrogate model and m is the number of models.

5.2 Optimization

To create the surrogate models, we evaluated our objective function (CPU time) at a set of carefully selected data points. Using this as the design space, full-factorial sampling was carried out, and the surrogate models were trained. A sample response surface (for single V cycle) is shown in Fig. 10. Then, from the feasible solution region of each model, the Pareto-optimal front (POF) was generated using a weighted-average surrogate method. This POF was validated by carrying out new tests runs at the suggested points, and the process was repeated by including the values of the new test points, until the solution converged.

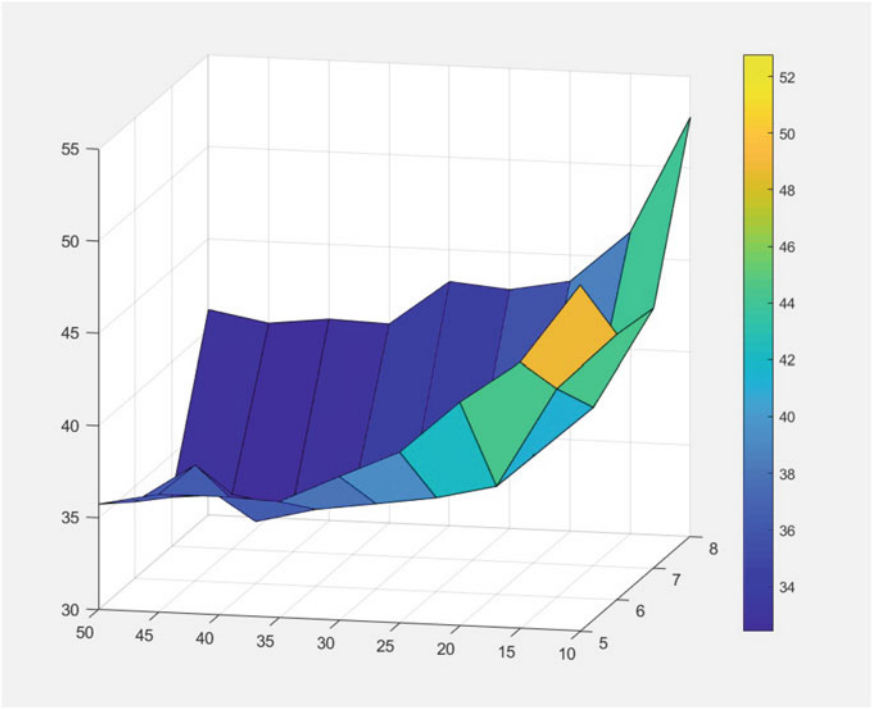


Fig. 10 Response surface for the two-dimensional design space (for single V cycle)

6 Results and Discussion

The results obtained from surrogate modelling are summarized in Table 4. For the same problem solved by Suero et al. [5], the CPU time of the optimized AMG (algebraic Multigrid) algorithm on a 4097×4097 grid was 69.7 s. Due to computing resource constraints, we were not able to use such fine grids, and hence, direct quantitative comparisons could not be made. Nevertheless, our objective was to perform a parametric optimization of the Multigrid algorithm that is independent of the implementation, and the same has been achieved in the form of the optimized values for each parameter.

The table clearly indicates that a W cycle performs better than a V cycle and should be carried out to the penultimate grid level for best performance. It is also observed that repeating the cycle over without transferring the error corrections does not cause any improvement, which signifies that all existing errors have been smoothed out in the first sweep itself. New errors will emerge only after recalculating the solution field, and the solution will continue to oscillate about its exact value in each sweep, but with decreasing amplitude. This can be thought to emerge directly from the principle of stationarity of total potential [9].

Table 4 Optimization results

Cycle	Parameters			CPU time (s) from surrogate model	Actual CPU time
	Levels	Inner iterations	W point		
Single V cycle	7	47	–	30.9	32.44
Double V cycle	5	38	–	32.7	36
Single W cycle	7	36	2	26.3	25.68
Double W cycle	7	28	2	27	27

The governing Poisson's equation defines the gradient of heat flux (a conservative field) and enforces a continuity over the potential function (temperature) under the action of the source. For equilibrium, the potential of the system should be minimum, and therefore, any deviation from the exact function would cause a rise in potential. This causes the system to fall back to a state of lower potential. We can think of it as the classical example of dropping a ball inside a bowl. The ball moves up and down the bowl until it dissipates all its kinetic energy and settles at the bottom. In a similar fashion, when using Multigrid, the solution shoots up and down as new error values are obtained from the coarser grids. But, upon enforcing the fine grid continuity (energy dissipation of the ball), a fall in the total potential of the system occurs, and the amplitude of the deviations comes down. This process is repeated in every iteration until the deviations become small enough. It can be understood that this process is much faster than the ball dissipating all its kinetic energy in the first half-cycle itself. It would lose a large amount of velocity on its way and would therefore take a very long time to reach the bottom of the bowl.

7 Future Work

Based on the analogy discussed in this chapter, we hope to model the Multigrid algorithm as a control system whose response is governed by the equivalent of inertia, stiffness, and damping forces. Then, using the principles of control theory, we would like to deduce the exact set of parameters for optimum performance of the algorithm and also derive a general method to optimize the algorithm for any given governing equation.

References

1. A. Brandt, Multi-level adaptive solutions to boundary-value problems. Math. Comput. **31**, 333–390 (1977)
2. A. Jameson, Multigrid algorithms for compressible flow calculations, in *Multigrid Methods II*, ed. by W. Hackbusch, U. Trottenberg. Lecture Notes in Mathematics, vol. 1228 (Springer, Berlin, 1986)

3. K. Stüben, A review of algebraic multigrid. J. Comput. Appl. Math. **128**, 281–309 (2001). https://doi.org/10.1016/S0377-0427(00)00516-1
4. S. Vakili, M. Darbandi, Recommendations on enhancing the efficiency of algebraic multigrid preconditioned GMRES in solving coupled fluid flow equation. Numer. Heat Transf. B Fundam. **553**(3), 232–256 (2009)
5. R. Suero, M.A.V. Pinto, C.H. Marchi, L.K. Araki, A.C. Alves, Analysis of algebraic multigrid parameters for two-dimensional steady state heat diffusion equations. Appl. Math. Model. **36**, 2996–3006 (2012)
6. N.V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P.K. Tucker, *Surrogate-Based Analysis and Optimization*. NASA (Createspace Independent Publishing Platform, Scotts Valley, 2005)
7. M.H. Siddique, A. Afzal, A. Samad, Design Optimization of the centrifugal pumps via low fidelity models. Math. Problems Eng. **2018**, 3987594 (2018)
8. T. Goel, R. Haftka, W. Shyy, N. Queipo, Ensemble of surrogates. Struct. Multidisciplinary Optim. **33**, 199–216 (2007). https://doi.org/10.1007/s00158-006-0051-9
9. A. Poceski, A variational approach in the finite element method. Comput. Struct. **33**, 395-402 (1989)

Fragment Production and Its Dynamics Using Spatial Correlations and Monte-Carlo Based Analysis Code

Rohit Kumar and Ishita Puri

Abstract A study is carried out to see the influence of using spatial-based clusterization algorithm and the one based on the Monte-Carlo technique coupled with simulated annealing procedure on the fragment–fragment correlations for the reactions of ^{40}Ca + ^{40}Ca at an incident energy of 35 MeV/nucleon. The phase space of the nucleons is generated using the quantum molecular dynamics (QMD) model. We checked and found a significant difference in the results of the multiplicity probability of the fragments, fragment's radii in coordinate and momentum space from the center-of-mass of the system, relative difference between radii of the fragments in coordinate and momentum space within the events, and correlations between the largest fragment charge and the charge bound in the fragments per event. A comparison of our calculations with the experimental data is also presented.

1 Introduction

The Monte-Carlo based techniques are extensively used to solve complex problems in various different fields including physics, mathematics, and engineering. One generally solves the problems via repeated random sampling and statistically analyzing the results. Among various fields, the field of physics has enjoyed a tremendous gain in understanding the vast range of phenomena ranging from cosmological events such as supernovae explosions and formation of neutron stars to sub-atomic phenomena such as fragment formation and quark–gluon plasma.

In theoretical nuclear physics, the Monte-Carlo simulations are used to model various phenomena in heavy-ion collisions. These phenomena include the fission and fusion of nuclei, multifragmentation (i.e., breaking of nuclei into many chunks),

R. Kumar (✉)
Department of Physics, Panjab University, Chandigarh, India

I. Puri
Department of Information Technology, UIET, Panjab University, Chandigarh, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_22

flow of particles/fragments, etc. [1–4]. Among these, the phenomenon of multi-fragmentation, considered to provide vital information of nuclear matter away from normal conditions, has been studied extensively using Monte-Carlo based computer models. The most widely used models are quantum molecular dynamics (QMD) model and Boltzmann–Uheling–Uhelenbeck (BUU) model [5, 6]. These models are generally termed as primary models as they provide us the information of each nucleon at various stages of the heavy-ion collision. It is well understood now that as soon as the compressed phase of the system is over, the system expands and cools down; thereafter, the phase space information of nucleons is used to construct fragments. The obtained fragment information is compared with experimental observations to testify the reliability of the theoretical approach.

Over the time, one realizes that to develop a computer program that constructs the realistic fragments is very tedious task. In the last few decades, the list of fragment recognition algorithms goes on increasing [5, 7–17]. This list includes simple computer programs based on the spatial correlations and/or momentum correlations among nucleons and binding energy cuts on fragments as well as the complex ones based on the binding energy minimization of the fragments using metropolis procedure. The use of one or the other algorithm helps to improve the consistency of theoretical calculations with the experimental observations. Among all the clusterization algorithms discussed above, the most widely accepted and successful ones are based on spatial correlations, i.e., minimum spanning tree (MST) method [5, 10] and energy minimization of the fragmenting System, i.e., simulated annealing clusterization algorithm (SACA) [16]. Though the results on fragmentation using these two algorithms are frequently compared with experimental observation in a wide range of entrance channels [11, 18], how the fragment–fragment correlations differ in the two remains untouched. The success of these clusterization algorithms to reproduce the experimental data motivates us to look for fragment–fragment correlations on an event-by-event basis. Therefore, this chapter will be dedicated to understand the change in fragment correlations if one uses the simple spatial correlations (i.e., MST method) or correlations among all nucleons at the same time (i.e., SACA method). We plan to shed light on the sensitivity of event-by-event constructed exclusive observables like the multiplicity probability of fragments, correlation between the largest fragment charge (Z_{max}) and the bound charge in fragments (Z_{bound}), $Z \geq 3$, fragment's correlations with respect to the center-of-mass and fragment–fragment correlation in coordinate and momentum space toward the MST and SACA methods.

The chapter is organized as follows: in the next section, we will give brief details of the primary model QMD along with the clusterization algorithms used in the study, i.e., MST and SACA methods. In Sect. 3, we discuss the results obtained using the MST and SACA methods. Lastly, in Sect. 4, we will give a summary of our work.

2 Methodology

The quantum molecular dynamics (QMD) model is a Monte-Carlo based many-body simulation program, in which each individual nucleon is represented by a Gaussian wave packet in coordinate and momentum space [5]. In this model, the centroid of each nucleon propagates in phase space using classical Hamilton's equations of motion:

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i}; \ \dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i}, \tag{1}$$

where H consists of kinetic energy and potential terms. During the propagation, the nucleons follow curved trajectories under the combined effect of mean field and collisions. This model is found to explain experimental results in a wide entrance channel domain. As far as the energy of projectile is concerned, this model is applicable starting from approximately 10 MeV/nucleon to 2 GeV/nucleon. For fine details of the model, the reader is referred to Ref. [5].

As mentioned earlier, the QMD model, being a many-body model, generates the collision information in the form of phase space of individual nucleons only. Depending on the problem in hand, this phase space information of nucleons is stored at various different time steps in the course of a reaction. This information acts as a raw information to form fragments utilizing the clusterization algorithms. In this chapter, we will be constructing fragments using the spatial correlations among the centroids of the nucleons and simultaneously using spatial+momentum correlations among the nucleons. The former one is known as the minimum spanning tree (MST) method [5], and the latter mentioned is dubbed as simulating annealing clusterization algorithm (SACA) [16]. In the MST method, only local correlations are considered and a nucleon is part of a fragment, if it is closer to any other nucleon by at least 4 fm in coordinate space. On the other hand, in the SACA method, the correlations among nucleons in coordinate and momentum space are considered on the global level, and the fragments are constructed using simulated annealing technique coupled with the metropolis procedure. Within the SACA method, the total binding energy of the fragments is calculated at each step by aiming to obtain the fragment distribution with minimum sum of the binding energies of the fragment or the most stable fragment configuration.

Although the MST method is suitable for certain entrance channels to explain the experimental results, its simple structure leads to a wide acceptability and utility [5, 10, 11]. On the other hand, the SACA method is more complex and used only in limited studies but helps to explain the physics of heavy-ion collisions in the projectile incident energy ranging from 10 MeV/nucleon to 25 GeV/nucleon [4, 17, 18]. Very recently, this model has also been extended in the direction of isotopes and hyperons in the clusters at the HypHI energies and is named as fragment recognition in general application (FRIGA) [17]. We will be utilizing the MST and SACA methods to understand correlations among the fragments and for the comparison with the experimental data.

3 Results and Discussion

For the present study, we have simulated the reactions of ^{40}Ca+^{40}Ca at an incident energy of 35 MeV/nucleon for all geometries (b) using soft equation of state. The energy-dependent nucleon–nucleon cross section is used in the present work. Throughout the chapter, the discussion of fragments is done at the freeze-out times only. Generally, the freeze-out time refers to the time after which the fragment structures do not change significantly. For the SACA and MST methods, the freeze-out times are 60 and 300 fm/c, respectively.

In many previous studies, the results of the MST and SACA methods are compared with the experimental observations and succeed partially or fully to explain the experimental results [4, 10, 11, 18]. This motivates us to look for the radii of the fragments in the coordinate (r) and momentum space (p) from the center-of-mass of the system as a function of their charge and the relative difference among their radii in coordinate (r_{ij}) and momentum space (p_{ij}) as a function of product of the corresponding charges of the fragments. In Fig. 1, we display the results of fragments for the ^{40}Ca+^{40}Ca reactions at an impact parameter b = 3 fm and at an incident energy of 35 MeV/nucleon. The left (right) panels correspond to the results of the MST (SACA) method. From the figure, we see that the spatial radius (r) of fragments from the center-of-mass of the system is larger for the MST method compared to the SACA method. The results can be understood as follows: the MST method is based on the spatial correlations among nucleons and therefore can be applicable at the moment when nucleons are well separated from each other. The definition of the MST method also makes it inappropriate for the earlier reaction times. On the other hand, the SACA method that uses the global spatial+momentum correlations among nucleons, therefore, identifies the fragments much earlier in reaction times. The smaller spatial radii for the SACA method compared to the MST method reflect this aspect. The difference in the freeze-out time remains the major reason for the observed behavior. At the same time, the values of momentum radii (p) of the fragments identified using the MST and SACA methods have opposite behavior. Here, again the difference in freeze-out times of the MST and SACA methods explains the results. One can see larger momentum values for the fragments of the SACA method pointing toward the fact that the fragments are very close to the compressed phase and expanding rapidly. Thus, the SACA method has more capability compared to the MST method to give answers regarding the origin of fragments and the involved dynamics. Interestingly, the trends of radii (both in coordinate space and in momentum space) for the SACA method and MST method are quite similar to each other when plotted against the charges of the fragments (although very different in magnitude). With the increase in the size of the fragment, the radii have comparatively lesser values (larger values) in coordinate (momentum) space. At the moment, it looks like that the fragments have same nature for both the algorithms!

In Fig. 1e–h, we display the results of the relative difference between the radii of the fragments within each event in coordinate (r_{ij}) (Fig. 1e, g) and momentum

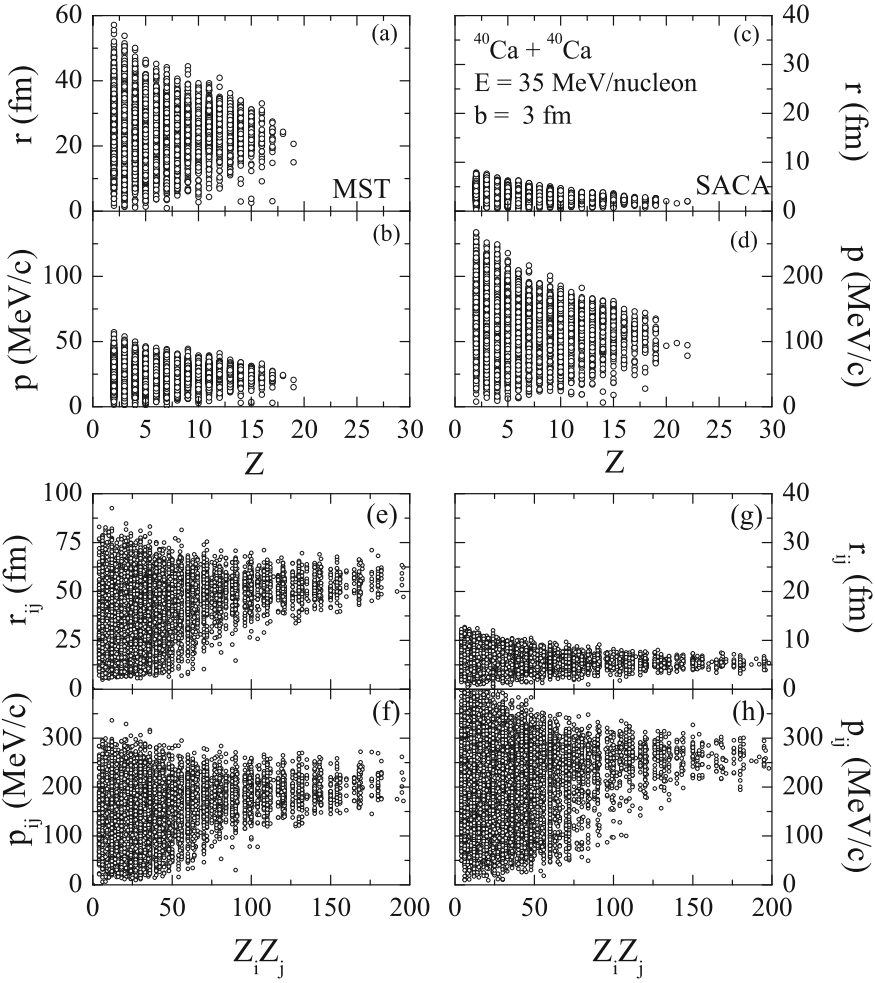


Fig. 1 (Top four panels) The radii of the fragments from the center-of-mass of the system in coordinate space (r-fm) and momentum space (p-MeV/c) as a function of the fragment charges and (bottom four panels) relative radii among fragments in coordinate (r_{ij}) and momentum space (p_{ij}) as a function of the product of their charges for the reactions of ^{40}Ca+^{40}Ca (b = 3 fm) at an incident energy of 35 MeV/nucleon. The results of fragments with the MST and SACA methods are displayed in left and right panels, respectively

space (p_{ij}) (Fig. 1f, h) as a function of the product of their corresponding charges. The small and large values of relative radii correspond to the closest neighbors and fragments originated from the target or projectile remnant, respectively. We see that the trends of the distributions are horizontal, showing that irrespective of the size of the fragments, the average relative distances (in coordinate space) between them are almost same. We also note that the minimum values of relative spatial radii are more

in case of the MST method compared to the SACA method. Again showing that the SACA fragments are identified much earlier in coordinate space. Interestingly, the relative momentum radii have same distribution for the MST and SACA methods. The results of SACA method are simple to understand and are just the outcome of the momentum radii (p) of the fragments (see Fig. 1d). Unexpectedly, the values of the relative momentum radii (p_{ij}) of fragments in the MST method are larger and comparable to the SACA method. These results may look surprising but reflect the formation of fragments from a non-equilibrated source. Earlier, this kind of non-equilibrium condition is also observed by Furuta and Ono for the same reactions by studying the kinetic energy and radial size of the reaction system [19]. Now, let us understand the correlation among the charge of the largest fragment (Z_{max}) and the bound charge in the fragments (Z_{bound}) on an event-by-event basis.

In Fig. 2, the higher order correlations among the largest fragment charge and the total charge bound in the fragments ($Z \geq 3$) are displayed for the semi-central (b = 3 fm) reactions of ^{40}Ca+^{40}Ca at an incident energy of 35 MeV/nucleon. We see that the range of Z_{max} varies between 4 and 17 in the case of MST method and 3 and 22 in the case of SACA method. At the same time, the values of Z_{bound} vary from 13 to 30 and from 5 to 35 for the MST and SACA methods, respectively. We see the largest cross section of (Z_{max}, Z_{bound}) for the MST and SACA methods at values of (10, 20) and (10, 21), respectively. Though the largest values are appeared at almost the same values of Z_{max} and Z_{bound}, their probability differs by more than a factor of 2. We also find that the probability is more uniform for all values of Z_{max} and Z_{bound} in the case of SACA method compared to the MST method. Thus, the SACA method isolates the overlapping fragments (or free nucleons and light charged particles). The larger values of the Z_{bound} in SACA method appear due to the reason that at earlier times, the nucleons are close to each other; thus, if any structured effect exists (as in the present case, ^{40}Ca is magic nuclei), its effect will be reduced. This is not the case in the MST method that identifies the fragments at later times. Next, we will look for the multiplicity probability of fragments and the total charge distribution of fragment charges.

In Fig. 3 (top panels), the results of the multiplicity probability of the light charged particles (LCPs) [2≤ A≤4] and intermediate mass fragments (IMFs) [5≤ A ≤13] obtained using the MST and SACA methods are displayed for the semi-central (b = 3 fm) reactions of ^{40}Ca+^{40}Ca at an incident energy of 35 MeV/nucleon. The solid lines and dash-dotted lines represent the results obtained using the SACA and MST methods, respectively. From the figure, we see that, one obtains peak in the multiplicity probability at lower values with the MST approach compared to the SACA method. Also, with the MST algorithm, one cannot obtain higher multiplicity events due to the reason that many different fragments are counted part of one fragment due to overlapping. For example, if we have two fragments (or free nucleons or light charged fragments near a fragment) that have different identities but due to low excitation energy lie closer to each other in space, the MST method by definition will identify these fragments as one fragment, therefore, giving more number of events with lower multiplicity values. On the other hand, the SACA method minimizes binding energy of the total system and therefore can

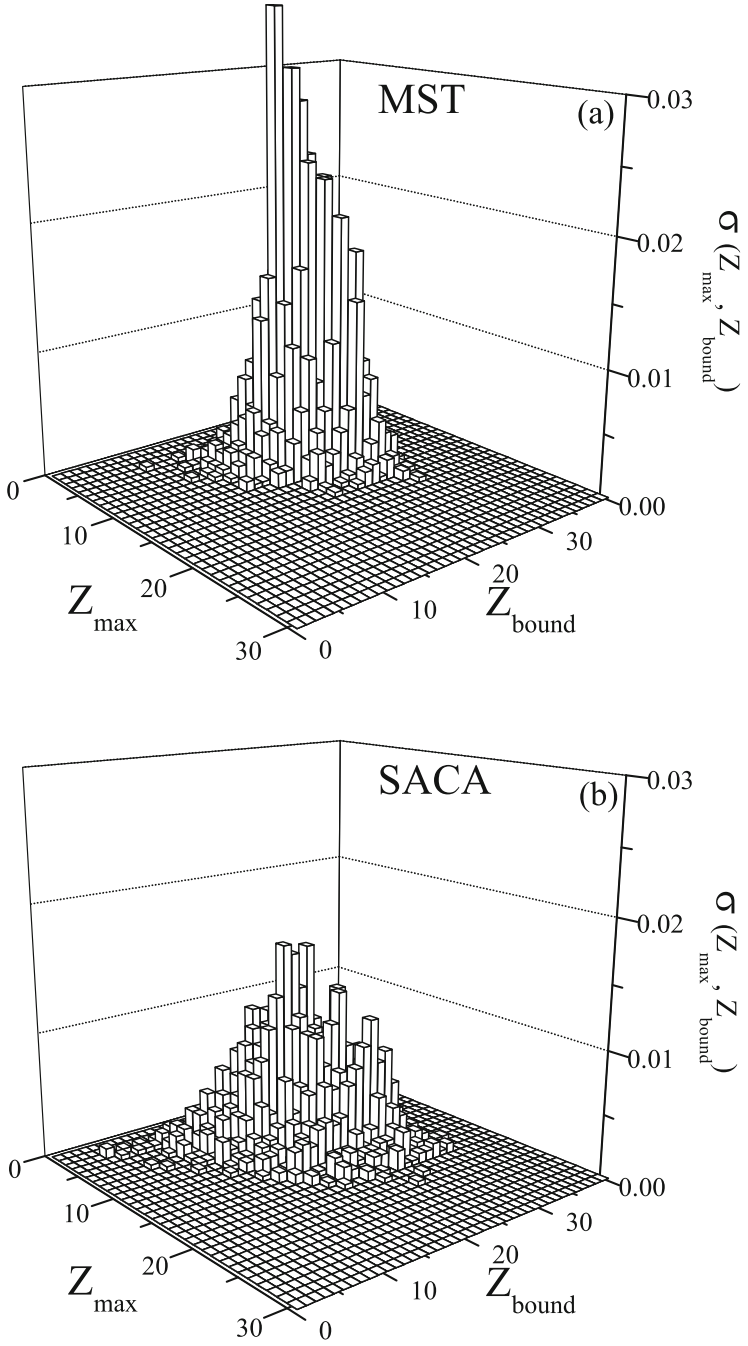


Fig. 2 Correlation between the largest fragment charge and the charge bound in the fragments ($Z \geq 3$) using the MST (top) and SACA (bottom) methods for the reactions of ^{40}Ca + ^{40}Ca at semi-central geometries (b = 3 fm) and at an incident energy of 35 MeV/nucleon

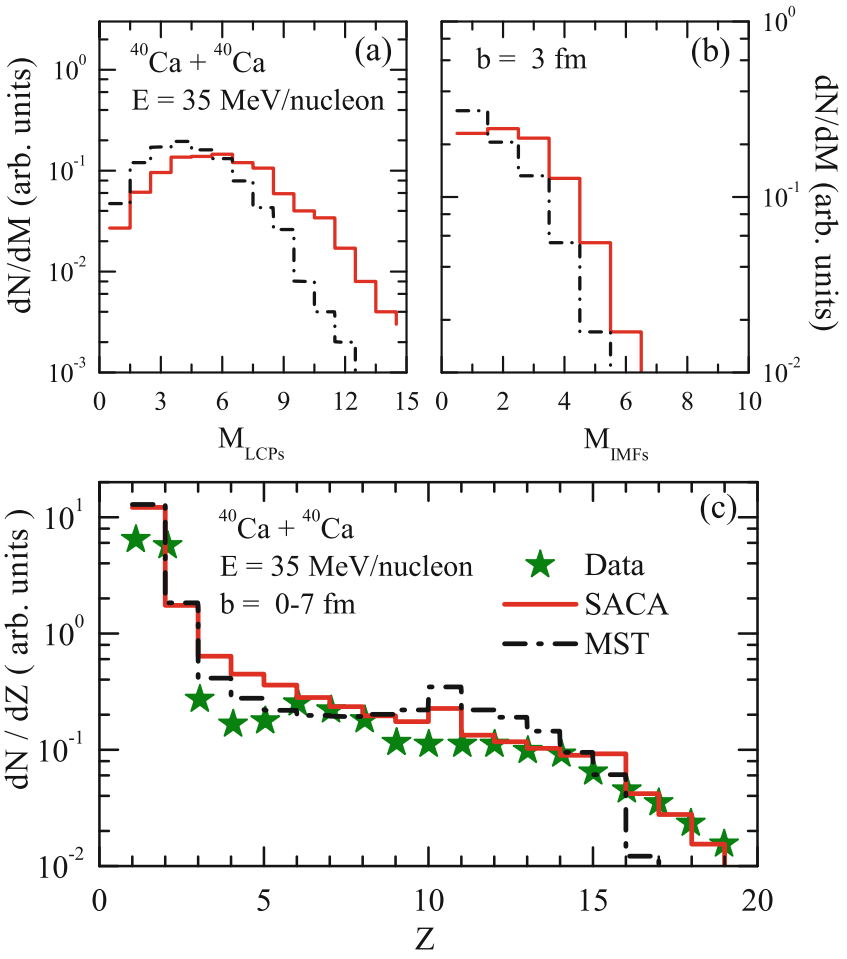


Fig. 3 (Upper panels) The multiplicity distribution of LCPs [$2 \leq A \leq 4$] and IMFs [$5 \leq A \leq 13$] for the reactions of ^{40}Ca + ^{40}Ca at semi-central geometries (b = 3 fm) and at incident energy of 35 MeV/nucleon, and (lower panel) comparison of our theoretical calculations with experimental data (lower panel) for b = 0–7 fm; other reaction conditions are same as in Fig. 3a, b. The meaning of different lines and symbols is described in the text

even separate out the fragments that lie closer to one another in space. This ability of the SACA method leads to shift the peaks of multiplicity probabilities to higher multiplicity values compared to the MST algorithm.

In Fig. 3 (lower panel), we compare our theoretical results obtained using the MST and SACA approaches with the experimental data in the impact parameter range of b=0–7 fm for the reactions of ^{40}Ca+^{40}Ca at an incident energy of 35 MeV/nucleon. The stars represent the experimental data from Ref. [20]. The comparison shows that the results obtained with SACA method are more consistent

with the experimental data compared to MST calculations. We observe a slight discrepancy with the measurements at higher fragment charge (Z) values using MST method. On the other hand, the results obtained using SACA method are in agreement with the experimental data. The discrepancy for fragments with charge values of 3-5 is due to the uncertainty in the measurement of these fragments in experiments. Lastly, the results from Figs. 1, 2, and 3 signify that the fragments obtained by the MST and SACA methods have originated from different phase space regions and effect the fragment correlations to a great extent.

4 Summary

We studied the influence of secondary algorithms, namely the MST and SACA methods on fragment–fragment correlations. The collisions of ^{40}Ca + ^{40}Ca at an incident energy of 35 MeV/nucleon at semi-central geometries are studied. We found a significant difference on the results of the multiplicity probability of the fragments, fragment radii in coordinate and momentum space from the center-of-mass of the system, relative difference between the radii of fragments in coordinate and momentum space within the events, and correlations among the largest fragment charge and the charge bound in the fragments. A comparison of our calculations with the experimental data for complete impact parameter range predicts that SACA method can give more realistic picture of reaction dynamics compared to MST method.

Acknowledgments This work is supported by the Council of Scientific and Industrial Research (CSIR), Government of India, via Grant No. 03 (1388)/16/EMR-II.

References

1. M.B. Tsang et al., Onset of nuclear vaporization in ^{197}Au+ ^{197}Au collision. Phys. Rev. Lett. **71**, 1502 (1993)
2. A. Schüttauf et al., Universality of spectator fragmentation at relativistic bombarding energies. Nucl. Phys. A **607**, 457 (1996)
3. B. Borderie et al., Nuclear multifragmentation and phase transition for hot nuclei. Prog. Part. Nucl. Phys. **51**, 551 (2008)
4. Y.K. Vermani et al., Microscopic approach to the spectator matter fragmentation from 400 to 1000 MeV/nucleon. Euro. Phys. Lett. **85**, 62001 (2009)
5. J. Aichelin et al., Quantum molecular dynamic: a dynamical microscopic n-body approach to investigate fragment formation and the nuclear equation of state in heavy-ion collisions. Phys. Rep. **202**, 233 (1991)
6. W. Bauer et al., Energetic photons from intermediate energy proton- and heavy-ion-induced reactions. Phys. Rev. C **34**, 2127 (1986)
7. S. Kumar, R.K. Puri, Role of momentum correlations in fragment formation. Phys. Rev. C **58**, 320 (1998)

8. S. Kumar, R.K. Puri, Stability of fragments formed in the simulations of central heavy ion collisions. Phys. Rev. C **58**, 2858 (1998); S. Goyal, R.K. Puri, Formation of fragments in heavy-ion collisions using a modified clusterization method. Phys. Rev. C **83**, 047601 (2011)
9. R. Kumar, Shivani, S. Gautam, Influence of different liquid-drop based bindings on lighter mass fragments and entropy production. Eur. Phys. J. A **52**, 112 (2016)
10. R. Kumar, S. Gautam, R.K. Puri, Multifragmentation within a clusterization algorithm based on thermal binding energies. Phys. Rev. C **89**, 064608 (2014)
11. R. Kumar, S. Gautam, R.K. Puri, Influence of different binding energies in clusterization approach: fragmentation as an example. J. Phys. G Nucl. Part. Phys. **43**, 025104 (2016)
12. S. Sood, R. Kumar, A. Sharma, R.K. Puri, Cluster formation and phase transition in nuclear disassembly using a variety of clusterization algorithms. Phys. Rev. C **99**, 054612 (2019)
13. Y. Zhang et al., Effect of isospin-dependent cluster recognition on the observables in heavy ion collisions. Phys. Rev. C **85**, 051602 (2000)
14. C. Ngo et al., Dynamical instability of hot and compressed nuclei. Nucl. Phys. A **499**, 148 (1989)
15. C. Dorso, et al., Early recognition of clusters in molecular dynamics model. Phys. Lett. B **301**, 328 (1993)
16. R.K. Puri et al., Simulating annealing clusterization algorithm for studying the multifragmentation. J. Comput. Phys. **162**, 245 (2000); J. Phys. G: Nucl. Part. Phys. **37**, 015105 (2010); R.K. Puri et al., Early fragment formation in heavy-ion collisions. Phys. Rev C **54**, R28 (1996)
17. A.L. Fevre, A. Aichelin, C. Hartnack, Y. Leifels, FRIGA: a new approach to identify isotopes and hypernuclei in n-body transport models. Phys. Rev. C **100**, 034624 (2019)
18. R. Kumar, R.K. Puri, Using experimental data to test an n-body dynamical model coupled with an energy-based clusterization algorithm at low incident energies. Phys. Rev. C **97**, 034624 (2018)
19. T. Furuta, A. Ono, Relevance of equilibrium in multifragmentation. Phys. Rev. C **79**, 014608 (2009)
20. K. Hagel et al., Violent collisions and multifragment final states of ^{40}Ca+^{40}Ca at 35 MeV/nucleon. Phys. Rev. C **50**, 2017 (1994)

Effect of Halo Structure in Nuclear Reactions Using Monte-Carlo Simulations

Sucheta, Rohit Kumar, and Rajeev K. Puri

Abstract In the present study, we have shown the effect of halo structure of nuclei on the fragment production at projectile energy of 100 MeV/nucleon. The present study is carried out using an n-body dynamical model that simulates the reactions on an event-by-event basis, and fragments are constructed using spatial correlations among nucleons. We show the quantities averaged over events and the correlation function of fragments constructed on an event-by-event basis for halo and stable nuclei induced reactions. Both average quantities and correlation function have slight variation toward halo structure of nuclei at this incident energy. Therefore, one should study the fragmentation at lower incident energies to understand the effect of halo structure of nuclei.

1 Introduction

With the tremendous progress of the radioactive ion beam facilities, it becomes possible to understand the exotic phenomena in the field of nuclear physics. Here, the topic of our interest is to study the nuclei toward the drip-line physics, i.e., halo nuclei, which has received a great attention in recent years. One defines halo nuclei as the weakly bound nuclei having outer one or two nucleons spatially decoupled from a tightly bound nuclear core. These fascinating nuclei were first found in 1985 in Berkely experiments by Tanihata et al. [1]. In their experiment, they observed enormous value of root mean square (rms) radii for ^{11}Li and ^{9}Be nuclei as estimated by standard $A^{1/3}$ dependence while measuring the interaction cross section. Thus far, the community of the nuclear physicists showed curiosity in the nuclear structure and reactions followed by halo nuclei. In later studies, the halo structure is also observed for the nuclei of ^{9}He, ^{14}Be, ^{17}B, ^{19}C, ^{22}C, ^{22}N, ^{23}O, ^{24}F, ^{29}Ne, ^{31}Ne, and ^{37}Mg [2–6].

Sucheta · R. Kumar (✉) · R. K. Puri
Department of Physics, Panjab University, Punjab, Chandigarh, India
e-mail: rkpuri@pu.ac.in

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_23

In many previous studies, the structure of halo nuclei is included to look for the behavior change in various phenomena such as fusion, fission at low incident energies, and multifragmentation at intermediate incident energies. To mention a few, in Ref. [7], the fusion cross section is studied for various stable and halo induced reactions using different proximity-based potentials. The study revealed that for the halo nuclei, the barrier heights are reduced effectively, and the fusion cross section is enhanced compared to stable nuclei. In another study, Sharma et al. [8] have examined the reactions of $^{24-40}$Mg + ^{12}C at a projectile energy of 240 MeV/nucleon to explore the role of halo structure on cross section and various other properties. They have used Glauber model with the conjunction of densities from the mean field formulation. The obtained results were also compared with the experimental observations. They were able to reproduce experimental results via using extended radius instead of actual halo structure of nucleus. In another study, Liu et al. [9] studied the breaking of colliding nuclei into many fragments, i.e., multifragmentation that occurs at intermediate energies. The study was done using the Isospin-dependent Quantum Molecular Dynamics (IQMD) model. They have used stable nucleus of ^{19}F and halo structured nucleus of ^{19}B in the incident energy range of 20 to 150 MeV/nucleon. They showed that the halo structured nuclei increases the fragment multiplicity at low incident energies and that the halo structure effect gradually disappears with the increase in incident energy. Opposite behavior was reported for the momentum dissipation. Interestingly, no other study has been reported on this topic which makes halo induced reactions a potential candidate to provide a new physics of multifragmentation phenomenon. In the present work, we will emphasize the role of halo structured nuclei on fragment production in nuclear reactions using the n-body dynamical model, i.e., Quantum Molecular Dynamics (QMD) model [10]. We will also compare the results with the outcomes of the reactions induced by stable mass nuclei.

Our paper is organized as follows: we briefly discuss the model in Sect. 2. We will discuss the results in Sect. 3. The summary of our work will be given in Sect. 4.

2 Quantum Molecular Dynamics (QMD) Model

The Quantum Molecular Dynamics (QMD) [10] model is an event generator that provides the information of the reaction in the form of phase space information of individual nucleon. Here, each nucleon is represented by a Gaussian wave function of the form:

$$\phi_i(\mathbf{r}, \mathbf{r}_i(t), \mathbf{p}_i(t)) = \frac{1}{(2\pi L)^{3/4}} e^{\left[\frac{i}{\hbar}\mathbf{p}_i(t)\cdot\mathbf{r} - \frac{(\mathbf{r}-\mathbf{r}_i(t))^2}{4L}\right]}. \quad (1)$$

The centroid of each nucleon is followed using Hamilton's equations of motion. The information of phase space of nucleons is stored on an event-by-event basis at various time steps during an event. This information is converted into fragments

using spatial correlations among nucleons at freeze-out times [10, 11]. Generally, the freeze-out time of a reaction is 300 fm/c.

3 Results and Discussions

In the present study, we have generated a sample of thousands of independent events for both halo nuclei induced reactions, i.e., ^{37}Mg + ^{27}Al and stable nuclei induced reactions, i.e., ^{36}Mg + ^{27}Al at an incident energy of 100 MeV/nucleon. The sample of events is generated for central geometries. The model parameters are fixed to soft equation of state (EoS) along with the energy-based nucleon–nucleon cross section. The sample of events has varying multiplicities, bound charges, and momentum distribution. As it is well known that the structure of halo nuclei itself is an open question till date, and therefore, in many previous studies, the extended radius is used instead of actual halo structure [7, 8]. In the present study, we also followed these studies, and therefore, the present study will only give us the upper limit of the behavior change if one introduces the actual halo structure.

In Fig. 1, the time evolution of the largest fragment ($< A_{max} >$) and multiplicities of free nucleons ($< N_{FNs} >$) [$1 \leq A_f \leq 1$], light charged particles ($< N_{LCPs} >$) [$2 \leq A_f \leq 4$], and intermediate mass fragments ($< N_{IMFs} >$) [$5 \leq A_f \leq A_{tot}/3$] are displayed. Here, we use two projectiles, i.e., halo (^{37}Mg) and stable (^{36}Mg) on a target of ^{27}Al. Note that the radius of (^{37}Mg) is considered much larger than (^{36}Mg) due to which the nucleus has lower value of Fermi momentum at initial stages. When the reaction happens, the different values of radii and Fermi momentum imply to different expansion rate. The same can be seen from the figure. The expansion of ^{36}Mg + ^{27}Al (solid lines) happens much faster compared to ^{37}Mg + ^{27}Al (dash-dotted lines), and this behavior is clearly seen for the size of the largest fragment and other fragment multiplicities. Though the expansion rate is different but up to freeze-out time, we do not see much difference in fragment structures. To depict this behavior much clearly, in Fig. 2, we have displayed the position and momentum space of nucleons in the reaction plane (i.e., x–z and p_x–p_z plane). For clarity of the figure, only ten isolated events are superimposed on one another at freeze-out time. From the figure, we see that the nucleons expand to almost same distance in coordinate and momentum space. There are slightly more number of nucleons at higher position and momentum values in case of halo induced reactions compared to stable nuclei induced reactions. By looking at this figure, one may say that the role of halo structure is minimal on reaction dynamics at freeze-out time for the present entrance channel. But it may happen that the fragments have originated from different phase space regions. Therefore, in Fig. 3, we display the centroids of fragments (LCPs and IMFs) for the same ten events. We see that the LCPs are distributed all over the space in coordinate and momentum space for ^{36}Mg + ^{27}Al compared to ^{37}Mg + ^{27}Al, where LCPs have slightly larger contribution from the projectile regions due to the loose structure of halo nuclei. We also see few LCPs for ^{37}Mg + ^{27}Al reactions at larger position and momentum values. These fragments

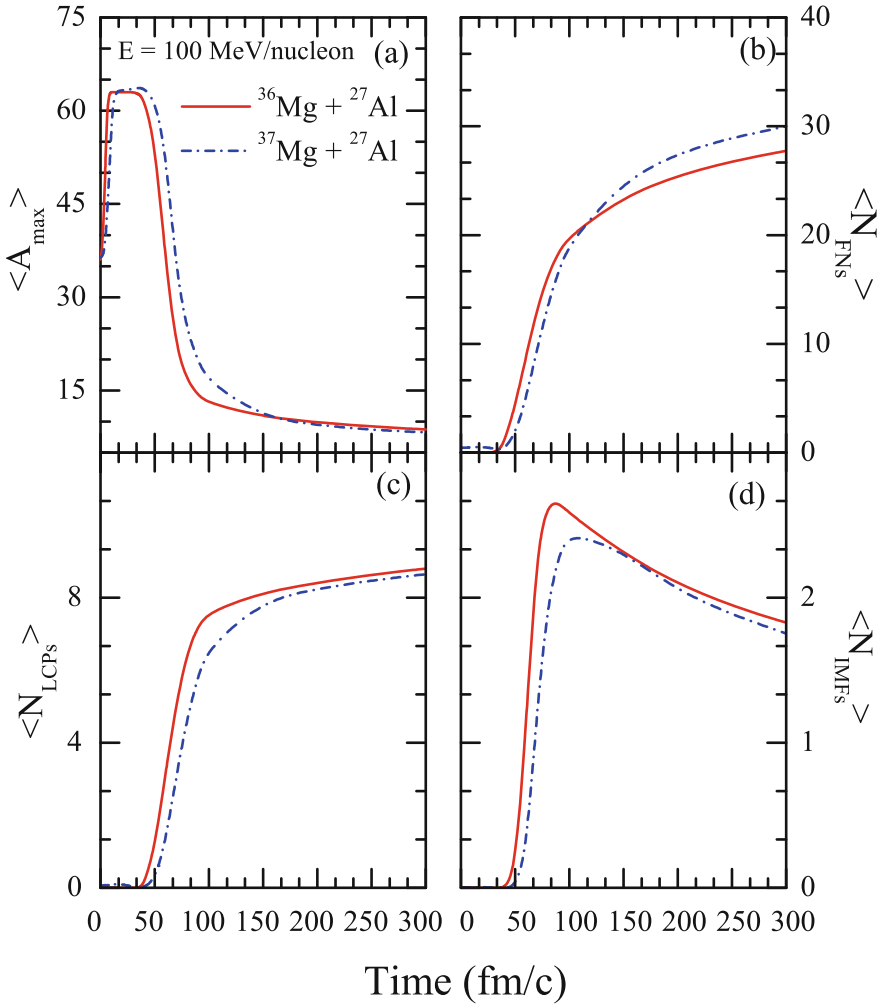


Fig. 1 The time evolution of the largest fragment ($< A_{max} >$) and multiplicities of free nucleons ($< N_{FNs} >$), light charged particles ($< N_{LCPs} >$), and intermediate mass fragments ($< N_{IMFs} >$) for the central collisions of ^{36}Mg + ^{27}Al (solid lines) and ^{37}Mg + ^{27}Al (dash-dotted lines) at an incident energy of 100 MeV/nucleon

are due to the structural change of nucleus. But, on an average, the picture is quite similar for IMFs. Now, if we combine the results of Figs. 1, 2, and 3, we can see that the halo structure affects the expansion rate to a great extent at initial stages of a reaction, but due to large energy pumped to the system, it has no significant role at final stages for the entrance channel considered in the present work. It is worth mentioning here that the above results combined with our earlier results [12] are in accordance with the results reported by Liu et al. [9].

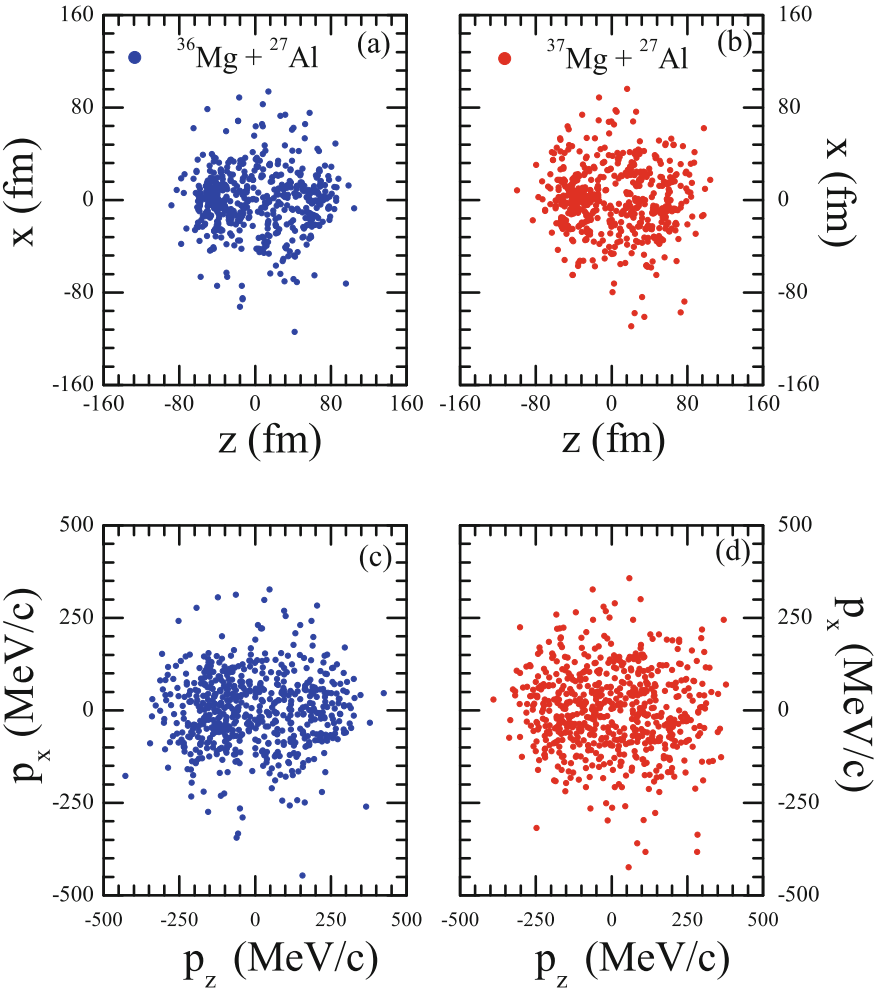


Fig. 2 The centroids of the nucleons are presented for ten events in coordinate (x–z) (top panels) and momentum (p_x-p_z) (bottom) reaction plane for the reactions of ^{36}Mg + ^{27}Al (left panels) and ^{37}Mg + ^{27}Al (right panels).

In the last two decades, both experimental and theoretical studies have also shown that the multiplicity of IMFs shows a rise and fall behavior with increase in incident energy of projectile and has a connection with the observation of liquid–gas-like behavior of nuclear matter [13–15]. Now, if we look at Figs. 1, 2, and 3, we may say that the multiplicity of IMFs at freeze-out stage is not much altered at this energy. In the next paragraph of this chapter, we plan to understand the event-by-event correlations among the intermediate mass fragments (i.e., IMFs). It may happen that the size of the fragments differs from one another in case of stable and halo induced reactions, but due to averaging over events, the final multiplicity

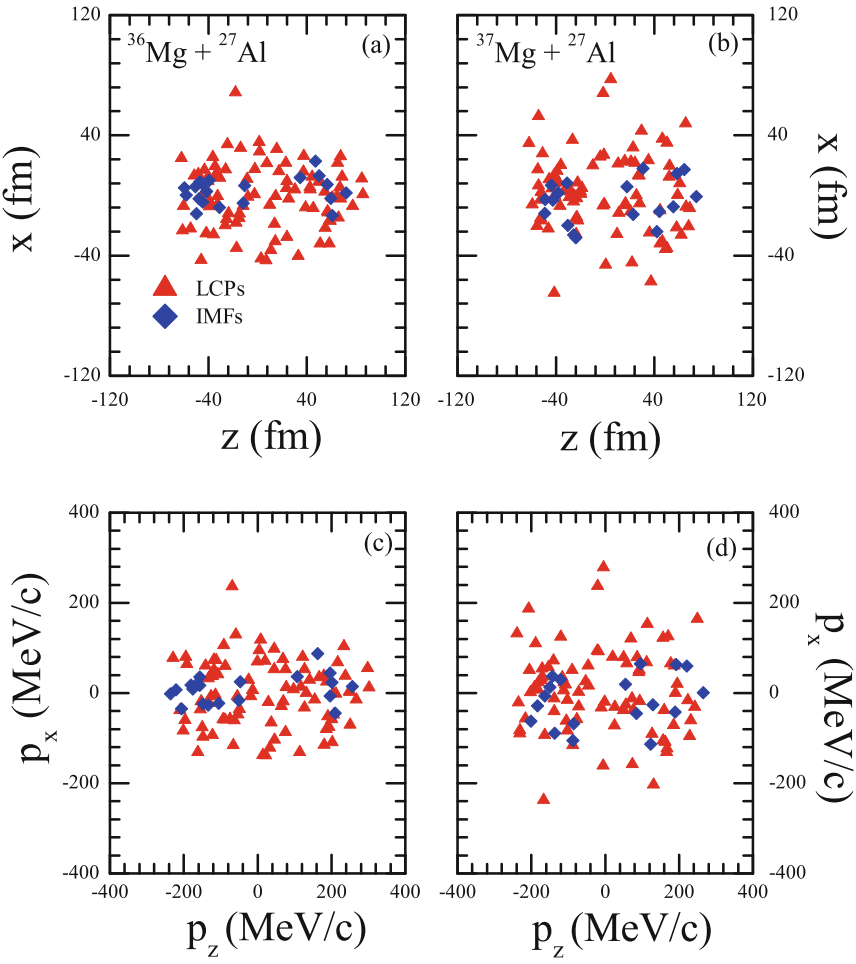


Fig. 3 Same as Fig. 2, but for the centroids of the fragments

values came out to be same. For this, we have constructed a sample of five thousand events with different multiplicities of IMFs and the nucleons bound in the IMFs and calculated the yield of the events where fragments are correlated with one another. If $Y(A_1, A_2)$ is the total yield of the correlated and uncorrelated events and $Y'(A_1, A_2)$ is the yield of the correlated events, then the normalized yield is

$$\sigma(A_1, A_2) = \frac{Y'(A_1, A_2)}{Y(A_1, A_2)}; \tag{2}$$

here, A_1 and A_2 are the masses of the fragments (existing in IMFs range). This helps to understand what is happening within the events instead of what is happening on

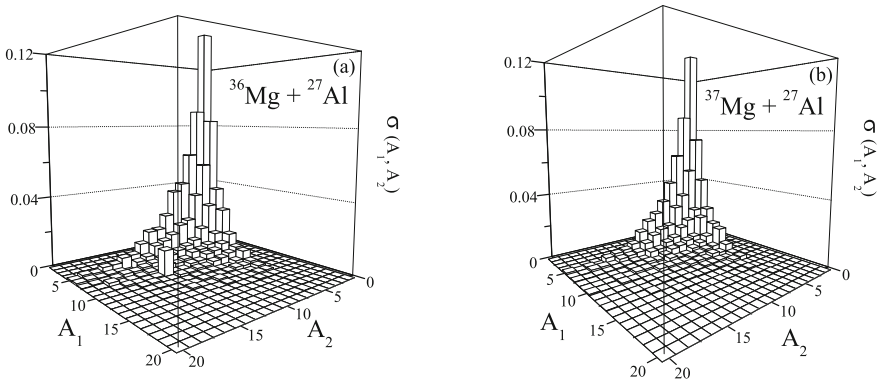


Fig. 4 The correlation among the fragments' mass on event-by-event is shown for the reactions of ^{36}Mg + ^{27}Al (top panel) and ^{37}Mg + ^{27}Al (bottom panel). Here, only intermediate mass fragments (IMFs) are used to construct the correlation function

an average. The results are plotted in Fig. 4. From the figure, we see that the event-by-event distribution of fragments has more cross sections of fragments in ^{36}Mg induced reactions compared to ^{37}Mg. These cross sections have link with the liquid–gas phase transition in nuclear matter. Here, the results reflect that the signals of liquid–gas phase transition or criticality are more pronounced for ^{36}Mg compared to ^{7}Mg induced reactions. As the energy considered in the present work is large, therefore, the difference is less significant here. We expect larger differences in cross sections at lower incident energies for stable and halo induced reactions. Lastly, if we combine the results of Figs. 1, 2, 3, and 4, the fragmentation results have only slight variation. If one is looking for the structure effect of halo nuclei on fragments, one should look for quantities at lower incident energies ($\leq$ 100 MeV/nucleon). Therefore, it will be very interesting to study peak fragment energies and signals of phase transition in nuclear matter which occur at lower incident energies than the energies studied in the present work. This will be presented in future studies.

4 Summary

In the present study, we have analyzed the average quantities related to fragment production and the ones that are based on an event-by-event basis for halo and stable nuclei induced reactions. The work is done using the Quantum Molecular Dynamics model, and fragments are obtained using the spatial correlations among nucleons. Our present study showed that the free nucleons and lighter charged particles show change in their multiplicities at freeze-out times, whereas intermediate mass fragments have almost same values. The correlation function constructed on an event-by-event basis showed slightly weaker signals for halo induced reactions than

the stable ones. We found that one should study reactions at lower incident energies to find the role of structural effects.

Acknowledgments This work is supported by the Council of Scientific and Industrial Research (CSIR), Government of India, via Grant No. 03 (1388)/16/EMR-II.

References

1. I. Tanihata et al., Measurements of interaction cross sections and nuclear radii in the light p-shell region. Phys. Rev. Lett. **55**, 2676 (1985)
2. I. Tanihata et al., Measurement of interaction cross sections using isotope beams of Be and B and isospin dependence of the nuclear radii. Phys. Lett. B **206**, 592 (1988)
3. A. Ozawa et al., Measurements of interaction cross sections for light neutron-rich nuclei at relativistic energies and determination of effective matter radii. Nucl. Phys. A **691**, 599 (2001)
4. M. Takechi et al., Interaction cross sections for Ne isotopes towards the island of inversion and halo structures of ^{29}Ne and ^{31}Ne. Phys. Lett. B **707**, 357 (2012)
5. M. Takechi et al., Evidence of halo structure in ^{37}Mg observed via reaction cross sections and intruder orbitals beyond the island of inversion. Phys. Rev. C **90**, 061305 (2014)
6. N. Kobayashi et al., Observation of a p-wave one-Neutron halo configuration in ^{37}Mg. Phys. Rev. Lett. **112**, 242501 (2014)
7. R. Kumari, Study of fusion probabilities with halo nuclei using different proximity based potentials. Nucl. Phys. A **917**, 85 (2013)
8. M.K. Sharma et al., Search for halo structure in ^{37}Mg using the Glauber model and microscopic relativistic mean-field densities. Phys. Rev. C **93**, 014322 (2016)
9. J.Y. Liu et al., Special roles of loose neutron-halo nucleus structure on the fragmentation and momentum dissipation in heavy ion collisions. Phys. Lett. B **617**, 24 (2005)
10. J. Aichelin Quantum molecular dynamics a dynamical microscopic n-body approach to investigate fragment formation and the nuclear equation of state in heavy ion collisions. Phys. Rep. **202**, 233 (1991)
11. R. Kumar, S. Gautam, R.K. Puri, Multifragmentation within a clusterization algorithm based on thermal binding energies. Phys. Rev. C **89**, 064608 (2014); R. Kumar, R.K. Puri, Using experimental data to test an n-body dynamical model coupled with an energy-based clusterization algorithm at low incident energies. Phys. Rev. C **97**, 034624 (2018)
12. Sucheta, R. Kumar, R.K. Puri, On the study of fragmentation of loosely bound nuclei using dynamical model. Proc. DAE Symp. Nucl. Phys. **63**, 556 (2018)
13. Y.K. Vermani, R.K. Puri, Mass dependence of the onset of multifragmentation in low energy heavy-ion collisions. J. Phys. G Nucl. Part. Phys. **36**, 105103 (2009); S. Kaur, R.K. Puri, Isospin effects on the energy of peak mass production. Phys. Rev. C **87**, 014620 (2013)
14. S. Sood, R. Kumar, A. Sharma, R.K. Puri, Cluster formation and phase transition in nuclear disassembly using a variety of clusterization algorithms. Phys. Rev. C **99**, 054612 (2019)
15. R. Kumar et al., On the multifragmentation and phase transition in the perspectives of different -body dynamical models. Act. Phys. Pol. B **49**, 301 (2018)

Part IV
Stochastic Models and Statistics

Performance Analysis of a Two-Dimensional State Multiserver Markovian Queueing Model with Reneging Customers

Neelam Singla and Sonia Kalra

Abstract In this chapter, a multiserver Markovian retrial queueing system with reneging customers is studied. If all or some of the servers are idle, then entering customer is admitted to join the system and receives his service immediately. Primary calls arrive according to a Poisson process. On the other hand, if all, some, or none of servers are busy, then all the admitted customers join the orbit. Upon retrial, the customer immediately receives his service if the servers are idle; otherwise, he may enter the orbit again or leave the system because of impatience. The repeating calls also follow the same fashion (Poisson process). Service times for all servers are same which follow exponential distribution. Recursive approach is followed to derive the system's time-dependent probabilities of exact number of arrivals and departures from the system at when all, some, or none servers are busy. Various measures of effectiveness are discussed, and some special cases are also deduced.

Keywords Multiserver · Probability · Queueing · Reneging · Retrial · System

1 Introduction

Retrial queues are pervasive among most of the real-life practical situations. In general, the retrial queueing systems are characterized by the fact that upon arrival, a customer on finding all servers busy must leave the system, but some time later the customer will come back to reinitiate his demand. In the process of making retrials, a customer is said to be in orbit and is called a retrial customer. Such queueing models with retrials can be considered as the most important tool for

Subject Classification Codes: 60K25; 90B22; 68M20.

N. Singla · S. Kalra (✉)
Department of Statistics, Punjabi University, Patiala, Punjab, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_24

the analysis of transportation networks, operating system, communication system, etc. One application of this model can be found in ticket booking service using telephone facility, where multiple employees are available to process the request of ticket booking. If a busy signal is received, the caller makes repeated attempts until the connection is made and when the caller makes a successful phone call, then he demands for booking tickets. A good account of retrial queues is given in Falin [1], Falin and Templeton [2], Kulkarni and Liang [3], Artalejo [4], and Choi and Chang [5].

The investigation of a retrial queueing systems with many servers is essentially more difficult than single-server queueing systems. Explicit results are available only in a few special cases. A customer will be said to have reneged if after joining the orbit, he gets impatient and leaves without receiving service. The concept of reneging was first studied by Haight [6]. Al-Seedy et al. [7] studied M/M/c queue with balking and reneging and derived its transient-state solution by using probability generating function techniques and properties of Bessel function. Shin [8] studied M/M/c/K retrial queues with impatient customers, retrial queue with two parallel service facilities, and retrial queue with two types of customers which can be modeled by a level-dependent quasi-birth-death process (LDQBD) with linear transition rates of the form $\lambda_k = \alpha + \beta_k$. Parthasarathy and Sudhesh [9] considered a time-dependent single-server retrial system with state-dependent arrival rate λ_n, service rate μ_n, and retrial rate θ_n. They found time-dependent system size probabilities using continued fraction and presented some numerical illustrations. Nathaniel Grier, William A. Massey, Tyrone McKoy, and Ward Whitt [10] studied the time-dependent erlang loss model with retrials.

Pegden and Rosenshine [11] gave an initial idea about two-state for a classical queueing model M/M/1/∞. They obtained the probability of exact number of arrivals in the system and exact number of departures from the system by a given time t. This measure supplies better insight into the behavior of a queueing system than the probability of the exact number of units in the system at a given time. Indra and Ruchi [12] obtained two-dimensional state time-dependent probabilities along with some interesting particular cases for a single-server Markovian queueing system where the service mechanism was non-exhaustive. Kumar et al. [13] described some new results for a two-state batch departure multiple vacation queueing model. Garg and Kumar [14] obtained explicit time-dependent probabilities of exact number of arrivals and departures from the orbit of a single-server retrial queue with impatient customers.

In this chapter, we obtain explicit time-dependent probabilities for the exact number of arrivals and departures from the system by a given time t when all, some, or none servers are busy for a multiserver retrial queueing system with reneging customers.

To examine this model, the remaining work is explained in the following manner. Section 2 presents the difference-differential equation of our queueing model with a description of the model. The time-dependent solution of our model is obtained in Sect. 3. Section 4 consists of the performance measures of the model along with some special cases. In Sect. 5, the analytical results are validated numerically on

the system performance and plotted graphically also. Finally, the chapter ends with Sect. 6, which presents a suitable conclusion.

2 System Model

An initial model description at Markovian level is as follows—Customers arrive in a multiserver system according to a Poisson process with rate λ. The service facility consists of "c" identical servers. An arriving customer, who finds all the servers busy, is blocked and temporarily leaves the service area. Such customers join a group of unattended customers called orbit. A customer in the orbit repeats his request for service with Poisson retrial rate θ. The service times follow exponential distribution with rate μ both for primary customers and successful repeated attempts. When the service is not available for a long time, the customers in a queue may become impatient and decide to abandon the orbit with probability $(1-\alpha)$ or to remain in the orbit with probability α. The input flows, intervals between repeated attempts, and service times are assumed to be mutually independent.

Laplace transformation $\overline{f}(s)$ of $f(t)$ is given by:

$$\overline{f}(s) = \int_0^\infty e^{-st} f(t)\mathrm{d}t, \quad \mathrm{Re}(s) > 0.$$

The Laplace inverse of $\frac{Q(p)}{P(p)}$ is:

$$\sum_{k=1}^{n} \sum_{l=1}^{m_k} \frac{t^{m_k-l} e^{a_k t}}{(m_k-l)!\,(l-1)!} \times \frac{\mathrm{d}^{l-1}}{\mathrm{d}p^{l-1}} \left(\frac{Q(p)}{P(p)}\right) (p-a_k)^{m_k}, \quad \forall p = a_k, a_i \neq a_k, \quad \text{for } i \neq k.$$

where, $P(p) = (p-a_1)^{m_1}(p-a_2)^{m_2} \ldots\ldots\ldots (p-a_n)^{m_n}$.

$Q(p)$ is a polynomial of degree $< m_1 + m_2 + m_3 + \ldots \ \ldots \ \ldots\ m_n - 1$.

The Laplace inverse of $\overline{N}^{a,b,c}_{n_1,n_2,n_3}(s) = \frac{1}{(s+a)^{n_1}\ (s+b)^{n_2}\ (s+c)^{n_3}}$ is:

$$N_{n_1,n_2,n_3}^{a,b,c}(t) = \sum_{l=1}^{n_3}\sum_{m=1}^{l} \frac{e^{-at}\ t^{n_3-l}\ (-1)^{m+1}\binom{l-1}{m-1}\left(\prod_{g_1=0}^{l-m-1}(n_1+g_1)\right)\left(\prod_{g_2=0}^{m-2}(n_2+g_2)\right)}{(n_3-l)!(m-1)!(b-a)^{n_2+m-1}\ (c-a)^{n_1+l-m}}$$
$$+\sum_{l=1}^{n_2}\sum_{m=1}^{l} \frac{e^{-bt}\ t^{n_2-l}\ (-1)^{m+1}\binom{l-1}{m-1}\left(\prod_{g_1=0}^{l-m-1}(n_1+g_1)\right)\left(\prod_{g_2=0}^{m-2}(n_3+g_2)\right)}{(n_2-l)!(m-1)!(a-b)^{n_3+m-1}\ (c-b)^{n_1+l-m}}.$$
$$+\sum_{l=1}^{n_1}\sum_{m=1}^{l} \frac{e^{-ct}\ t^{n_1-l}\ (-1)^{m+1}\binom{l-1}{m-1}\left(\prod_{g_1=0}^{l-m-1}(n_2+g_1)\right)\left(\prod_{g_2=0}^{m-2}(n_3+g_2)\right)}{(n_1-l)!(m-1)!(a-c)^{n_3+m-1}\ (b-c)^{n_2+l-m}}$$

If $L^{-1}\{f(s)\} = F(t)$ and $L^{-1}\{g(s)\} = G(t)$, then $L^{-1}\{f(s)\ g(s)\} = \int_0^t F(u)G(t-u)\ \mathrm{d}u = F * G$, $F * G$ is called the convolution of F and G.

2.1 The Two-Dimensional State Model

Definitions $P_{i,j,0}(t)$ = Probability that there are exactly i arrivals in the system and j departures from the system by time t when server is idle.

$P_{i,j,m}(t)$ = Probability that there are exactly i arrivals in the system and j departures from the system by time t when m servers are busy. $1 \le m \le c-1$.

$P_{i,j,c}(t)$ = Probability that there are exactly i arrivals in the system and j departures from the system by time t when all the c servers are busy.

$P_{i,j}(t)$ = Probability that there are exactly i arrivals in the system and j departures from the system by time t.

$$P_{i,j}(t) = P_{i,j,0}(t) + \sum_{m=1}^{c-1} P_{i,j,m}(t) + P_{i,j,c}(t) \quad \forall i, j \quad i \ge j$$

also

$$P_{i,j,c}(t)=0 \quad \text{and} \quad P_{i,j,m}(t)=0 \quad \text{for } i \le j, 1 \le m \le c-1; \quad P_{i,j,0}(t)=0, \quad i < j.$$

Initially

$$P_{0,0,0}(0) = 1; \quad P_{i,j,0}(0) = 0, \quad P_{i,j,c}(0) = 0 \quad \text{and} \quad P_{i,j,m}(t) = 0, \quad \forall i, j \ne 0$$
and $\quad 1 \le m \le c-1.$

2.2 The Difference-Differential Equations Governing the System Are

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{i,j,0}(t) = -(\lambda + (i-j)\theta) P_{i,j,0}(t) + \mu P_{i,j-1,1}(t) \quad i \geq j \geq 0 \tag{1}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} P_{i,j,m}(t) = &-(\lambda + m\mu + (i-j-m)\theta) P_{i,j,m}(t) + \lambda P_{i-1,j,m-1}(t) \\ &+ (i-j-(m-1))\theta P_{i,j,m-1}(t) + (m+1)\mu P_{i,j-1,m+1}(t) \\ &\qquad i > j \geq 0, \quad 1 \leq m < c\end{aligned} \tag{2}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} P_{i,j,c}(t) = &-(\lambda + c\mu + (i-j-c)\theta(1-\alpha)) P_{i,j,c}(t) + \lambda P_{i-1,j,c-1}(t) \\ &+\lambda\left(1-\delta_{i-c,j}\right) P_{i-1,j,c}(t) + (i-j-(c-1))\theta P_{i,j,c-1}(t) \\ &+ (i-j-(c-1))\theta(1-\alpha) P_{i,j-1,c}(t) \qquad i > 1, \quad i > j \geq 0\end{aligned} \tag{3}$$

where $\delta_{i-c,j} = \begin{cases} 1, \text{ when } i-c=j \\ 0, \text{ otherwise} \end{cases}$.

Using the Laplace transformation $\overline{f}(s)$ of $f(t)$ given by:

$$\overline{f}(s) = \int_0^\infty e^{-st} f(t)\mathrm{d}t, \quad \mathrm{Re}(s) > 0$$

in Eqs. (1)–(3) along with the initial conditions, we have:

$$(s + \lambda + (i-j)\theta)\overline{P}_{i,j,0}(s) = \mu \overline{P}_{i,j-1,1}(s) \quad i \geq j \geq 0 \tag{4}$$

$$\begin{aligned}&(s + \lambda + m\mu + (i-j-m)\theta)\overline{P}_{i,j,m}(s) \\ &\quad = \lambda \overline{P}_{i-1,j,m-1}(s) + (i-j\,(m-1))\theta \overline{P}_{i,j,m-1}(s) \\ &\qquad + (m+1)\mu \overline{P}_{i,j-1,m+1}(s) \qquad i > j \geq 0, \quad 1 \leq m < c\end{aligned} \tag{5}$$

$$\begin{aligned}(s+\lambda+c\mu+(i-j-c)\,\theta\,(1-\alpha))\,\overline{P}_{i,j,c}(s) \\ =\lambda\,\overline{P}_{i-1,j,c-1}(s)+\lambda\left(1-\delta_{i-c,j}\right)\,\overline{P}_{i-1,j,c}(s) \\ +(i-j-(c-1))\,\theta\,\overline{P}_{i,j,c-1}(s) \\ +(i-j-(c-1))\,\theta\,(1-\alpha)\,\overline{P}_{i,j-1,c}(s) \qquad i>j\geq 0\end{aligned} \tag{6}$$

where $\delta_{i-c,j}=\begin{cases}1, \text{ when } i-c=j\\ 0, \text{ otherwise}\end{cases}$.

3 Transient Solution of the Model

To obtain time-dependent probabilities of our model, we solved Eqs. (4)–(6) recursively.

$$\overline{P}_{0,0,0}(s)=\frac{1}{s+\lambda} \tag{7}$$

$$\overline{P}_{i,i,0}(s)=\frac{\mu}{(s+\lambda)}\overline{P}_{i,i-1,1} \quad i\geq 1 \tag{8}$$

$$\overline{P}_{m,0,m}(s)=\frac{\lambda}{s+\lambda+m\mu}\,\overline{P}_{m-1,0,m-1} \quad 1\leq m\leq c-1 \tag{9}$$

$$\begin{aligned}\overline{P}_{i,i-m,m}(s)=&\frac{\lambda}{s+\lambda+m\mu}\overline{P}_{i-1,i-m,m-1}(s) \\ &+\frac{(m+1)\,\mu}{s+\lambda+m\mu}\overline{P}_{i,i-m-1,m+1}(s) \quad m{=}1 \text{ to } c-2, \quad i{=}m{+}1 \text{ to } c-1\end{aligned} \tag{10}$$

$$\overline{P}_{c,1,c-1}(s)=\frac{\lambda}{(s+\lambda+(c-1)\,\mu\,)}\overline{P}_{c-1,1,c-2}(s)+\frac{c\mu}{(s+\lambda+(c-1)\,\mu\,)}\overline{P}_{c,0,c}(s) \tag{11}$$

$$\overline{P}_{i,1,c-1}(s) = \frac{c\mu}{\left(s+\lambda+(c-1)\,\mu+(i-j-(c-1))\,\theta\right.} \times \prod_{p=0}^{i-c} \frac{\lambda^{i-(c-1)}}{(s+\lambda+c\mu+p\theta\,(1-\alpha))} \overline{P}_{c-1,0,c-1}(s) \qquad i > c \tag{12}$$

$$\overline{P}_{i,0,c}(s) = \prod_{p=0}^{i-c} \frac{\lambda^{i-(c-1)}}{(s+\lambda+c\mu+p\theta\,(1-\alpha))}\, \overline{P}_{c-1,0,c-1}(s) \qquad i \geq c \tag{13}$$

$$\overline{P}_{i,j,c}(s) = \left[\sum_{k=1}^{i-j-(c-2)} \left\{\prod_{p=k-1}^{i-j-c} \left(\frac{\lambda^{i-j-(c-2)-k}}{\left(s+\lambda+c\mu+p\theta(1-\alpha)\right.}\right)\right\} \eta'_k(s)\overline{P}_{j+k+(c-2),j,c-1}(s)\right] + \left[\sum_{k=1}^{i-j-(c-1)} \left\{\prod_{p=k-1}^{i-j-c} \left(\frac{(\lambda)^{i-j-(c-1)-k}\ k\theta(1-\alpha)}{\left(s+\lambda+c\mu+p\theta(1-\alpha)\right.}\right)\right\}\ \overline{P}_{j+k+(c-1),j-1,c}(s)\right] \qquad i \geq j+c, \quad j \geq 1 \tag{14}$$

where $\eta'_k(s) = \begin{cases} 1 & \text{for } k = 1 \\ \left(1 + \frac{(k-1)\theta}{(s+\lambda+c\mu+(k-2)\theta(1-\alpha))}\right) & \text{for } k = 2 \text{ to } i-j-(c-1) \\ \frac{(k-1)\theta}{(s+\lambda+c\mu+(k-2)\theta(1-\alpha))} & \text{for } k = i-j-(c-2) \end{cases}$.

$$\begin{aligned}\overline{P}_{i,j,c-1}(s) = {} & \frac{\lambda}{s+\lambda+(c-1)\mu+(i-j-(c-1))\theta}\overline{P}_{i-1,j,c-2}(s) \\ & + \frac{(i-j-(c-2))\theta}{\left(s+\lambda+(c-1)\mu+(i-j-(c-1))\theta\right.}\overline{P}_{i,j,c-2}(s) \\ & + \frac{c\mu}{\left(s+\lambda+(c-1)\mu+(i-j-(c-1))\theta\right.} \\ & \left[\left\{\sum_{k=1}^{i-j-(c-3)} \left(\prod_{p=k-1}^{i-j-(c-1)} \frac{\lambda^{i-j-(c-3)-k}}{\left(s+\lambda+c\mu+p\theta(1-\alpha)\right.}\right)\right\} \eta'_k(s)\overline{P}_{j+(k+1),j-1,c-1}(s)\right. \\ & \left.+ \left\{\sum_{k=1}^{i-j-(c-2)} \left(\prod_{p=k-1}^{i-j-(c-1)} \frac{\lambda^{i-j-(c-2)-k}\ k\theta(1-\alpha)}{\left(s+\lambda+c\mu+p\theta(1-\alpha)\right.}\right)\right\} \overline{P}_{j+k+(c-2),j-2,c}(s)\right] \\ & \qquad i \geq c-1+j, \quad j > 1\end{aligned} \tag{15}$$

$$\text{where } \eta_k'(s) = \begin{cases} 1 & \text{for } k = 1 \\ \left(1 + \frac{(k-1)\theta}{(s+\lambda+c\mu+(k-2)\theta(1-\alpha))}\right) & \text{for } k = 2 \text{ to } i-j-(c-2) \\ \frac{(k-1)\theta}{(s+\lambda+c\mu+(k-2)\theta(1-\alpha))} & \text{for } k = i-j-(c-3) \end{cases}.$$

$$\begin{aligned} \overline{P}_{i,j,m}(s) = & \frac{\lambda}{\left(s+\lambda+m\ \mu+(i-j-m)\theta\right.} \overline{P}_{i-1,j,m-1}(s) \\ & + \frac{(i-j-(m-1))\theta}{\left(s+\lambda+m\mu+(i-j-m)\theta\right.} \overline{P}_{i,j,m-1}(s) \\ & + \frac{(m+1)\mu}{\left(s+\lambda+m\mu+(i-j-m)\theta\right.} \\ & \left\{ \begin{array}{l} \frac{\lambda}{\left(s+\lambda+(m+1)\mu+(i-j-m)\theta\right.} \overline{P}_{i-1,j-1,m} \\ + \frac{(i-j-(m-1))\theta}{\left(s+\lambda+(m+1)\mu+(i-j-m)\theta\right.} \overline{P}_{i,j-1,m}(s) \\ + \frac{(m+2)\mu}{\left(s+\lambda+(m+1)\mu+(i-j-m)\theta\right.} \overline{P}_{i,j-2,m+2}(s) \end{array} \right\} \\ & 1 \le m \le c-2, \quad i \ge j+m, \quad j > c-m \end{aligned} \tag{16}$$

$$\overline{P}_{i,j,0}(s) = \frac{(m+1)\,\mu}{s+\lambda+(i-j)\,\theta} \left[\begin{array}{l} \frac{\lambda}{s+\lambda+\mu+(i-j)\theta} \overline{P}_{i-1,j-1,0}(s) \\ + \frac{(i-j+1)\theta}{s+\lambda+\mu+(i-j)\theta} \overline{P}_{i,j-1,0}(s) \\ + \frac{(m+2)\mu}{s+\lambda+\mu+(i-j)\theta} \\ \left[\begin{array}{l} \frac{\lambda}{s+\lambda+2\mu+(i-j)\theta} \overline{P}_{i-1,j-2,m+1}(s) \\ + \frac{(i-j+1)}{s+\lambda+2\mu+(i-j)\theta} \overline{P}_{i,j-1,m+1}(s) \\ + \frac{(m+3)\mu}{\left(s+\lambda+2+(i-j)\right.} \overline{P}_{i,j-3,m+3}(s) \end{array} \right] \end{array} \right] \quad i > j \ge c \tag{17}$$

Taking the Inverse Laplace transform of Eqs. (7)–(17), we have:

$$P_{0,0,0}(t) = e^{-\lambda t} \tag{18}$$

$$P_{i,i,0}(t) = \mu e^{-\lambda t} * P_{i,i-1,1}(t) \quad i \ge 1 \tag{19}$$

$$P_{m,0,m}(t) = \lambda e^{-(\lambda+m\mu)t} * P_{m-1,0,m-1}(t) \quad 1 \le m \le c-1 \tag{20}$$

$$P_{i,i-m,m}(t) = \lambda e^{-(\lambda+m\mu)t} * P_{i-1,i-m,m-1}(t) + (m+1)\,\mu\, e^{-(\lambda+m\mu)t} * P_{i,i-1-m,m+1}(t) \quad m = 1 \text{ to } c-2, \quad i = m+1 \text{ to } c-1 \tag{21}$$

$$P_{c,1,c-1}(t) = \lambda e^{-(\lambda+(c-1)\mu)t} * P_{c-1,1,c-2}(t) + c\mu e^{-(\lambda+(c-1)\mu)t} * P_{c,0,c}(t) \tag{22}$$

$$P_{i,1,c-1}(t) = c\mu\lambda^{i-(c-1)}\, e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t} \times \left\{ \prod_{p=0}^{i-c} \frac{1}{(c\mu+p\theta\,(1-\alpha))} - \frac{e^{-(c\mu+p\theta(1-\alpha))t}}{(c\mu+p\theta\,(1-\alpha))} \right\} * P_{c-1,0,c-1}(t) \quad i > c \tag{23}$$

$$P_{i,0,c}(t) = \lambda^{i-(c-1)} \left\{ \prod_{p=0}^{i-c} e^{-(\lambda+c\mu+p\theta(1-\alpha))} \right\} * P_{c-1,0,c-1}(t) \quad i \geq c \tag{24}$$

$$\begin{aligned} P_{i,j,c}(t) = {} & \lambda^{i-j-(c-1)} \left\{ \prod_{p=0}^{i-j-c} e^{-(\lambda+c\mu+p\theta(1-\alpha))t} \right\} * P_{j+c-1,j,c-1}(t) \\ & + \sum_{k=2}^{i-j-(c-1)} \lambda^{i-j-(c-2)-k} \left\{ \prod_{p=k-1}^{i-j-c} e^{-(\lambda+c\mu+p\theta(1-\alpha))t} \right\} * P_{j+k+c-2,j,c-1}(t) \\ & + \sum_{k=2}^{i-j-(c-1)} \lambda^{i-j-(c-2)-k}\,(k-1)\,\theta\, e^{-(\lambda+c\mu+(k-2)\theta(1-\alpha))t} \\ & \times \left\{ \prod_{p=k-1}^{i-j-c} \frac{1}{(c\mu+p\theta(1-\alpha))} - \frac{e^{-(c\mu+p\theta(1-\alpha))t}}{(c\mu+p\theta(1-\alpha))} \right\} * P_{j+k+c-2,j,c-1}(t) \\ & + (i-j-c+1)\,\theta e^{-(\lambda+c\mu+(i-j-c)\theta(1-\alpha))t} * P_{i,j,c-1}(t) \\ & + \sum_{k=1}^{i-j-(c-1)} (\lambda)^{i-j-(c-1)-k} k\theta\ (1-\alpha) \\ & \times \left\{ \prod_{p=k-1}^{i-j-c} e^{-(\lambda+c\mu+p\theta(1-\alpha))t} \right\} * P_{j+k+c-1,j-1,c} \\ & \qquad i \geq j+c, \quad j \geq 1 \end{aligned} \tag{25}$$

$$
\begin{aligned}
P_{i,j,c-1}(t) = {} & \left(\lambda e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t}\right) * P_{i-1,j,c-2}(t) + (i-j-(c-2))\,\theta \\
& \left(e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t}\right) * P_{i,j,c-2}(t) + (c\mu)\,\lambda^{i-j-(c-2)} \\
& e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t} \left\{ \prod_{p=0}^{i-j-(c-1)} \frac{1}{(c\mu+p\theta(1-\alpha))} - \frac{e^{-(c\mu+p\theta(1-\alpha))t}}{(c\mu+p\theta(1-\alpha))} \right\} \\
& * P_{j+2,j-1,c-1}(t) + (c\mu)\,e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t} \sum_{k=2}^{i-j-c+2} \lambda^{i-j-(c-3)-k} \\
& \left\{ \prod_{p=k-1}^{i-j-(c-1)} \frac{1}{(c\mu+p\theta(1-\alpha))} - \frac{e^{-(c\mu+p\theta(1-\alpha))t}}{(c\mu+p\theta(1-\alpha))} \right\} * P_{j+k+1,j-1,c-1}(t) \\
& + \left[\begin{array}{l} (c\mu) \sum_{k=2}^{i-j-(c-2)} (k-1)\,\theta\lambda^{i-j-(c-3)-k} \\ \left\{ \prod_{p=k-1}^{i-j-(c-1)} \frac{e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t}}{\{\mu+\theta\{p(1-\alpha)-(i-j-(c-1))\}\}\{\mu+\theta\{(k-2)(1-\alpha)-(i-j-(c-1))\}\}} \right\} \\ + \left\{ \prod_{p=k-1}^{i-j-(c-1)} \frac{e^{-(\lambda+c\mu+(k-2)\theta(1-\alpha))t}}{\{-\mu+\theta\{(i-j-(c-1))-(k-2)(1-\alpha)\}\}\{\theta(1-\alpha)\{p-(k-2)\}\}} \right\} \\ + \left\{ \prod_{p=k-1}^{i-j-(c-1)} \frac{e^{-(\lambda+c\mu+p\theta(1-\alpha))t}}{\{\theta(1-\alpha)\{(k-2)-p\}\}\{\theta\{(i-j-(c-1))-p(1-\alpha)\}-\mu\}} \right\} \end{array} \right] \\
& * P_{j+k+1,j-1,c-1}(t) + c\mu\,(i-j-c+2)\,e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t} \\
& \left\{ \frac{1}{(c\mu+(i-j-c+1)\theta(1-\alpha))} - \frac{e^{-(c\mu+(i-j-c+1)(1-\alpha))t}}{(c\mu+(i-j-c+1)\theta(1-\alpha))} \right\} \\
& * P_{i-c+4,j-1,c-1}(t) + c\mu e^{-(\lambda+(c-1)\mu+(i-j-(c-1))\theta)t} \sum_{k=2}^{i-j-c+2} \\
& \quad (\lambda)^{i-j-(c-2)-k} k\theta\,(1-\alpha) \\
& \left\{ \prod_{p=k-1}^{i-j-(c-1)} \frac{1}{(c\mu+p\theta(1-\alpha))} - \frac{e^{-(c\mu+p\theta(1-\alpha))t}}{(c\mu+p\theta(1-\alpha))} \right\} * P_{j+k+c-2,j-2,c}(t) \\
& \quad i \geq j+(c-1), \quad j \geq 1
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
P_{i,j,m}(t) = {} & \lambda e^{-\{\lambda+m\mu+(i-j-m)\theta\}t} * P_{i-1,j,m-1}(t) + (i-j-(m-1))\,\theta \\
& e^{-\{\lambda+m\mu+(i-j-m)\theta\}t} * P_{i,j,m-1}(t) + \lambda\,((m+1)\,\mu)\Big)\; e^{-\{\lambda+m\mu+(i-j-m)\theta\}t} \\
& \left\{ \frac{1}{\{(m+1)\mu+(i-j-m)\theta\}} - \frac{e^{-\{(m+1)\mu+(i-j-m)\theta\}t}}{\{(m+1)\mu+(i-j-m)\theta\}} \right\} * P_{i-1,j-1,m}(t) \\
& + (m+1)\,\mu\;(i-j-(m-1))\,\theta e^{-\{\lambda+m\mu+(i-j-m)\theta\}t} \\
& \left\{ \frac{1}{\{(m+1)\mu+(i-j-m)\theta\}} - \frac{e^{-\{(m+1)\mu+(i-j-m)\theta\}t}}{\{(m+1)\mu+(i-j-m)\theta\}} \right\} * P_{i,j-1,m}(t) + \\
& (m+1)\,(m+2)\,\mu e^{-\{\lambda+m\mu+(i-j-m)\theta\}t} \\
& \left\{ \frac{1}{\{(m+1)\mu+(i-j-m)\theta\}} - \frac{e^{-\{(m+1)\mu+(i-j-m)\theta\}t}}{\{(m+1)\mu+(i-j-m)\theta\}} \right\} * P_{i,j-2,m+2}(t) \\
& 1 \leq m \leq c-2, \quad i \geq j+m, \quad j > c-m
\end{aligned}
\tag{27}
$$

$$
\begin{aligned}
P_{i,j,0}(t) = & \left[\lambda\,(m+1)\,\mu e^{-(\lambda+(i-j)\theta\}t}\left\{\frac{1}{\mu+(i-j)\theta\}}-\frac{e^{-\{\mu+(i-j)\theta\}t}}{\{\mu+(i-j)\theta\}}\right\}\right] \\
& *P_{i-1,j-1,0}(t) \\
& +\left[((i-j+1)\,\theta)\,(m+1)\,\mu e^{-(\lambda+(i-j)\theta\}t}\left\{\frac{1}{\mu+(i-j)\theta\}}-\frac{e^{-\{\mu+(i-j)\theta\}t}}{\{\mu+(i-j)\theta\}}\right\}\right] \\
& *P_{i,j-1,0}(t) \\
& +\left[(m+1)\,(m+2)\,\mu^2\lambda\left\{\frac{e^{-\{\lambda+(i-j)\theta\}t}}{2\mu^2\lambda}+\frac{e^{-\{\lambda+\mu+(i-j)\theta\}t}}{\mu^2}-\frac{e^{-\{\lambda+(i-j)\theta+2\mu\}t}}{2\mu^2}\right\}\right] \\
& *P_{i-1,j-2,m+1}(t) \\
& +\left[(m+1)\,(m+2)\,\mu^2\,((i-j+1)\,\theta)\left\{\frac{e^{-\{\lambda+(i-j)\theta\}t}}{2\mu^2}+\frac{e^{-\{\lambda+\mu+(i-j)\theta\}t}}{\mu^2}-\frac{e^{-\{\lambda+(i-j)\theta+2\mu\}t}}{2\mu^2}\right\}\right] \\
& *P_{i,j-1,m+1}(t) \\
& +\left[(m+1)\,(m+2)\,(m+3)\,\mu^3\left\{\frac{e^{-\{\lambda+(i-j)\theta\}t}}{2\mu^2}+\frac{e^{-\{\lambda+\mu+(i-j)\theta\}t}}{\mu^2}-\frac{e^{-\{\lambda+(i-j)\theta+2\mu\}t}}{2\mu^2}\right\}\right] \\
& *P_{i,j-3,m+3}(t) \\
& i>j\geq c
\end{aligned}
\tag{28}
$$

4 Performance Indices

1. The Laplace transform of the probability $P_{i.}(t)$ that exactly i units arrive by time t is:

$$\overline{P}_{i.}(s)=\sum_{j=0}^{i}\overline{P}_{i,j}(s)=\frac{\lambda^i}{(s+\lambda)^{i+1}}\quad i>0; \tag{29}$$

And its inverse Laplace transform is:

$$P_{i.}(t)=\frac{e^{-\lambda t}(\lambda t)^i}{i!}. \tag{30}$$

The basic assumption on primary arrivals is that it forms a Poisson process and above analysis of abstract solution also verifies the same.

2. The probability that exactly j customers have been served by time t, $P_{.j}(t)$ in terms of $P_{i,j}(t)$ is given by:

$$P_{.j}(t)=\sum_{i=j}^{\infty}P_{i,j}(t).$$

3. From the abstract solution of our model, we verified that the sum of all possible probabilities is one, i.e., taking summation over i and j on Eqs. (7)–(17) and adding, we get:

$$\sum_{i=0}^{\infty}\sum_{j=0}^{i}\left\{\overline{P}_{i,j,0}(s)+\overline{P}_{i,j,m}(s)+\overline{P}_{i,j,c}(s)\right\}=\frac{1}{s}.$$

After taking the inverse Laplace transformation, we get ($m = 1, 2, \ldots c-1$):

$$\sum_{i=0}^{\infty}\sum_{j=0}^{i}\left\{P_{i,j,0}(t)+P_{i,j,m}(t)+P_{i,j,c}(t)\right\}=1.$$

which is a verification of our results.

4. Define $Q_{n,m}(t)$ as the probability that there are n customers in the system at time t and m ($m = 1, 2, \ldots c$) servers are busy.

   When m servers are busy, it is defined by probability $Q_{n,m}(t)$:

$$Q_{n,m}(t)=\sum_{j=0}^{\infty}P_{j+n+m,j,m}(t)\quad(m=1,2,\ldots c).$$

   The number of customers, i.e., "n" in the orbit is obtained by using the relation:

$$n=(\text{number of arrivals}-\text{number of departures}-m).$$

   Using the above relation and letting $\mu = 1$ from the Eqs. (1)–(3), the sets of equations in statistical equilibrium are:

$$(\lambda+m+n\theta)\,Q_{n,m}=\lambda\,Q_{n,m-1}+(n+1)\,\theta\,Q_{n+1,m-1}+(m+1)\,Q_{n,m+1}$$
$$0\le m\le c-1,\quad n\ge 0 \tag{31}$$

$$(\lambda+n\theta\,(1-\alpha)+c)\,Q_{n,c}=\lambda\,Q_{n,c-1}+(n+1)\,\theta\,Q_{n+1,c-1}$$
$$+\lambda\,Q_{n-1,c}\left(1-\delta_{n,0}\right)+(n+1)\,\theta\,(1-\alpha)\,Q_{n+1,c}$$
$$(\text{case } m=c),\quad n\ge 0 \tag{32}$$

   where $\delta_{n,0}=\begin{cases}1, & \text{when } n=0\\ 0, & \text{when } n\ge 1\end{cases}$.

5. *Special Cases*:

   (a) Put $\alpha = 1$ in Eqs. (31) and (32) for getting following equations:

$$(\lambda+m+n\theta)\,Q_{n,m}=\lambda\,Q_{n,m-1}+(n+1)\,\theta\,Q_{n+1,m-1}+(m+1)\,Q_{n,m+1}$$
$$0\le m\le c-1,\quad n\ge 0 \tag{33}$$

$$(\lambda + c)\, Q_{n,c} = \lambda\, Q_{n,c-1} + (n+1)\,\theta\, Q_{n+1,c-1} + \lambda\, Q_{n-1,c}\left(1 - \delta_{n,0}\right) \quad (\text{case } m = c), \quad n \geq 0 \tag{34}$$

and these equations coincide with the Equations of (2.17) and (2.18) of Falin and Templeton [2].

(b) Considering the units are singly served, i.e., $c = 1$ and service times of all the units are exponentially distributed for Eqs. (18)–(28), then we get various probabilities and these results matches with Singla and Kalra [15].

$$P_{0,0,0}(t) = e^{-\lambda t} \tag{35}$$

$$P_{i,1,0}(t) = \mu e^{-(\lambda+(i-1)\theta)t} * P_{i,0,1}(t) \quad i \geq 1 \tag{36}$$

$$P_{i,i,0} = \left[\begin{array}{l} (\lambda\mu)\, e^{-\lambda t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu t}}{\mu} \right\} * P_{i-1,i-1,0}(t) + (\mu\theta)\, e^{-\lambda t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu t}}{\mu} \right\} * P_{i,i-1,0}(t) \\ + (\mu\theta)\,(1-\alpha)\, e^{-\lambda t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu t}}{\mu} \right\} * P_{i,i-2,1}(t) \end{array} \right] \; i > 1 \tag{37}$$

$$P_{1,0,1}(t) = \lambda e^{-\lambda t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu t}}{\mu} \right\} \tag{38}$$

$$P_{i,0,1}(t) = (\lambda)^{i-1} \left\{ \prod_{m=1}^{i-1} e^{-(\lambda+\mu+m\theta(1-\alpha))t} \right\} * P_{1,0,1}(t) \quad i > 1 \tag{39}$$

$$P_{i,i-1,1} = \Big(\lambda e^{-(\lambda+\mu)t} * P_{i-1,i-1,0} + \theta e^{-(\lambda+\mu)t} * P_{i,i-1,0} + \theta\,(1-\alpha)\, e^{-(\lambda+\mu)t} * P_{i,i-2,1} \Big) \quad i > 1 \tag{40}$$

$$
\begin{aligned}
P_{i,j,1}(t) = {} & \lambda^{i-j-1}\left\{\prod_{m=1}^{i-j-1} e^{-(\lambda+\mu+m\theta(1-\alpha))t}\right\} \frac{t^{m-1}}{(m-1)!} * P_{j+1,j,0}(t) \\
& \sum_{k=2}^{i-j-1}\left[\lambda^{i-j-k}\left\{\prod_{m=k}^{i-j-1} e^{-(\lambda+\mu+m\theta(1-\alpha))t}\right\} \frac{t^{m-k}}{(m-k)!} * P_{j+k,j,0}(t)\right] \\
& + \sum_{k=2}^{i-j-1}\left[\lambda^{i-j-k}\,(k\theta)\left\{\prod_{m=k-1}^{i-j-1} e^{-(\lambda+\mu+m\theta(1-\alpha))t}\right\} \frac{t^{m-k}}{(m-k)!}\right. \\
& \qquad \left. * P_{j+k,j,0}(t)\right] \\
& + (i-j)\,\theta e^{-(\lambda+\mu+(i-j-1)\theta(1-\alpha))t} * P_{i,j,0}(t) \\
& + \sum_{k=1}^{i-j-1}\left[(\lambda)^{i-j-k-1}\,(k+1)\,\theta\,(1-\alpha)\left\{\prod_{m=k}^{i-j-1} e^{-(\lambda+\mu+m\theta(1-\alpha))t}\right\}\right. \\
& \qquad \left. \frac{t^{m-k}}{(m-k)!} * P_{j+k+1,j-1,1}(t)\right] \\
& + \left[(\lambda)^{i-j-1}\left\{\prod_{p=1}^{i-j-1} e^{-(\lambda+\mu+p\theta(1-\alpha))t}\right\} \frac{t^{p-1}}{(p-1)!} * P_{j+1,j,1}(t)\right] \\
& \qquad\qquad i \geq j+2, \quad j \geq 1
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
P_{i,j,0}(t) = {} & \mu\lambda^{i-j} e^{-(\lambda+(i-j)\theta)t}\left\{\prod_{m=1}^{i-j} \frac{1}{(\mu+m\theta(1-\alpha))^m}\right. \\
& \left. -e^{-(\mu+m\theta(1-\alpha))t}\sum_{r=0}^{m-1} \frac{t^r}{r!}\frac{1}{(\mu+m\theta(1-\alpha))^{m-r}}\right\} * P_{j,j-1,0}(t) \\
& +\lambda\mu e^{-(\lambda+(i-j)\theta)t}\left[\sum_{k=2}^{i-j}(\lambda)^{i-j-k}\left\{\prod_{m=k}^{i-j}\frac{1}{(\mu+m\theta(1-\alpha))^{m-k+1}}\right.\right. \\
& \left.\left. -\ e^{-(\mu+m\theta(1-\alpha))t}\sum_{r=0}^{m-k}\frac{t^r}{r!}\frac{1}{(\mu+m\theta(1-\alpha))^{m-k+1-r}}\right\} * P_{j+k-1,j-1,0}(t)\right] \\
& +\lambda\mu e^{-(\lambda+(i-j)\theta)t}\left[\sum_{k=2}^{i-j}(\lambda)^{i-j-k}\,(k\theta)\left\{\prod_{m=k-1}^{i-j}\frac{1}{(\mu+m\theta(1-\alpha))^{m-k+2}}\right.\right. \\
& \left.\left. -e^{-(\mu+m\theta(1-\alpha))t}\sum_{r=0}^{m-k+1}\frac{t^r}{r!}\frac{1}{(\mu+m\theta(1-\alpha))^{m-k+2-r}}\right\} * P_{j+k-1,j-1,0}(t)\right] \\
& +\mu\,(i-j+1)\,\theta e^{-(\lambda+(i-j)\theta)t}\left\{\frac{1}{\mu+(i-j)\theta(1-\alpha)} - \frac{e^{-(\mu+(i-j)\theta(1-\alpha))t}}{\mu+(i-j)\theta(1-\alpha)}\right\} \\
& \qquad * P_{i,j-1,0}(t) \\
& +\mu e^{-(\lambda+(i-j)\theta)t}\left[\sum_{k=1}^{i-j}(\lambda)^{i-j-k}\,(k+1)\theta(1-\alpha)\left\{\prod_{m=k}^{i-j}\frac{1}{(\mu+m\theta(1-\alpha))^{m-k+1}}\right.\right. \\
& \left.\left. -e^{-(\mu+m\theta(1-\alpha))t}\sum_{r=0}^{m-k}\frac{t^r}{r!}\frac{1}{(\mu+m\theta(1-\alpha))^{m-k+1-r}}\right\} * P_{j+k,j-2,1}(t)\right] \\
& +\mu(\lambda)^{i-j} e^{-(\lambda+(i-j)\theta)t}\left\{\prod_{p=1}^{i-j}\frac{1}{(\mu+p\theta(1-\alpha))^p}\right. \\
& \left. -\ e^{-(\mu+p\theta(1-\alpha))t}\sum_{r=0}^{p-1}\frac{t^r}{r!}\frac{1}{(\mu+p\theta(1-\alpha))^{p-r}}\right\} * P_{j,j-1,1}(t) \qquad\qquad i > j > 1
\end{aligned}
\tag{42}
$$

(c) Letting $c = 1$ and $\alpha = 1$ in Eqs. (18)–(28), we get following results and these results coincide with that of Singla and Kalra [16].

$$P_{0,0,0}(t) = e^{-\lambda t} \tag{43}$$

$$P_{i,1,0}(t) = \mu e^{-(\lambda+(i-1)\theta)t} * P_{i,0,1}(t) \quad i \geq 1 \tag{44}$$

$$\begin{aligned} P_{i,i,0}(t) = &\left[(\lambda\mu)\, e^{-\lambda t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu t}}{\mu} \right\} * P_{i-1,i-1,0}(t) + \right. \\ &\left. (\mu\theta)\, e^{-\lambda t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu t}}{\mu} \right\} * P_{i,i-1,0}(t) \right] \quad i > 1 \end{aligned} \tag{45}$$

$$P_{i,0,1}(t) = \lambda^i e^{-\lambda t} \left\{ \frac{1}{(\mu)^i} - e^{-\mu\, t} \sum_{r=0}^{i-1} \frac{(t)^r}{r!} \frac{1}{(\mu)^{i-r}} \right\} \quad i \geq 1 \tag{46}$$

$$P_{i,i-1,1}(t) = \left(\lambda e^{-(\lambda+\mu)t} * P_{i-1,i-1,0}(t) + \theta e^{-(\lambda+\mu)t} * P_{i,i-1,0}(t) \right) \quad i > 1 \tag{47}$$

$$\begin{aligned} P_{i,j,0}(t) = {} & \mu\lambda^{i-j} e^{-(\lambda+(i-j)\theta)t} \left\{ \frac{1}{(\mu)^{i-j}} - e^{-\mu t} \sum_{r=0}^{i-j-1} \frac{(t)^r}{r!} \frac{1}{(\mu)^{i-j-r}} \right\} * P_{j,j-1,0}(t) \\ & + e^{-(\lambda+(i-j)\theta)t} \sum_{k=2}^{i-j} \mu\lambda^{i-j-k+1} \left\{ \frac{1}{(\mu)^{i-j-k+1}} \right. \\ & \left. - e^{-\mu t} \sum_{r=0}^{i-j-k} \frac{(t)^r}{r!} \frac{1}{(\mu)^{i-j-k-r+1}} \right\} * P_{j+k-1,j-1,0}(t) \\ & + e^{-(\lambda+(i-j)\theta)t} \sum_{k=2}^{i-j} (\mu k\theta)\, \lambda^{i-j-k+1} \left\{ \frac{1}{(\mu)^{i-j-k+2}} \right. \\ & \left. - e^{-\mu t} \sum_{r=0}^{i-j-k+1} \frac{(t)^r}{r!} \frac{1}{(\mu)^{i-j-k-r+2}} \right\} * P_{j+k-1,j-1,0}(t) \\ & + e^{-(\lambda+(i-j)\theta)t} \left\{ \frac{1}{\mu} - \frac{e^{-\mu\, t}}{\mu} \right\} ((i-j+1)\,\mu\theta) * P_{i,j-1,0}(t) + \mu\lambda^{i-j} \\ & e^{-(\lambda+(i-j)\theta)t} \left\{ \frac{1}{(\mu)^{i-j}} - e^{-\mu\, t} \sum_{r=0}^{i-j-1} \frac{(t)^r}{r!} \frac{1}{(\mu)^{i-j-r}} \right\} * P_{j,j-1,1}(t) \\ & i > j > 1 \end{aligned} \tag{48}$$

$$
\begin{aligned}
P_{i,j,1}(t) = {} & \lambda^{i-j-1} e^{-(\lambda+\mu)t} \frac{(t)^{i-j-2}}{(i-j-2)!} * P_{j+1,j,0}(t) \\
& e^{-(\lambda+\mu)t} \sum_{k=2}^{i-j-1} \lambda^{i-j-k} \frac{(t)^{i-j-k-1}}{(i-j-k-1)!} * P_{j+k,j,0}(t) \\
& + e^{-(\lambda+\mu)t} \sum_{k=2}^{i-j-1} k\theta\lambda^{i-j-k} \frac{(t)^{i-j-k}}{(i-j-k)!} * P_{j+k,j,0}(t) + (i-j)\,\theta \\
& e^{-(\lambda+\mu)t} * P_{i,j,0}(t) + \lambda^{i-j-1} \\
& e^{-(\lambda+\mu)t} \frac{(t)^{i-j-2}}{(i-j-2)!} * P_{j+1,j,1}(t) \\
& \qquad\qquad i \geq j+2, \quad j \geq 1
\end{aligned}
\tag{49}
$$

5 Numerical Solution and Graphical Representation

Using MATLAB programming, numerical results have been generated to demonstrate how various parameters of the model influence the behavior of the system. The numerical results are generated for the case $\rho = \left(\frac{\lambda}{\mu}\right) = 0.3$, $\eta = \left(\frac{\theta}{\mu}\right) = 0.6$. The probabilities against time are graphically represented through Figs. 1, 2, and 3.

To study the effect of an increasing number of servers on probability $P_{5,5,0}$ of the model, the data are generated. Figure 1 shows a plot of the probability $P_{5,5,0}$ against time t for $c = 2, 3, 4$. From the initial condition, it is seen that with time the probability $P_{5,5,0}$ start increasing from the initial value at $t = 0$ and finally attained maximum value, i.e., 1. The behavior of the probability $P_{5,5,0}$ is the same for all the

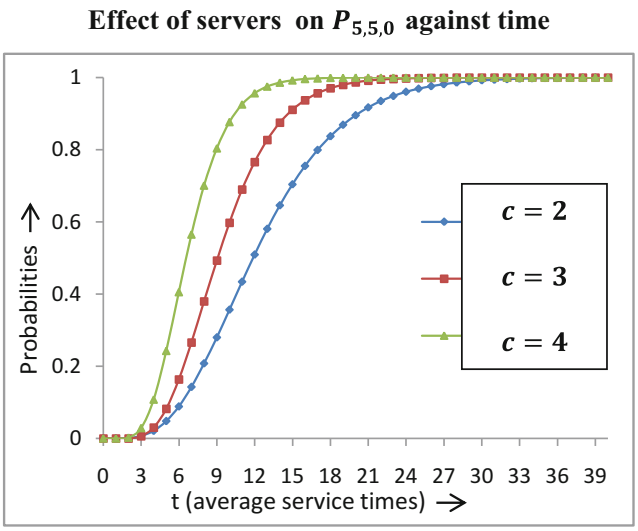


Fig. 1 Effect of servers on $P_{5,5,0}$ against time

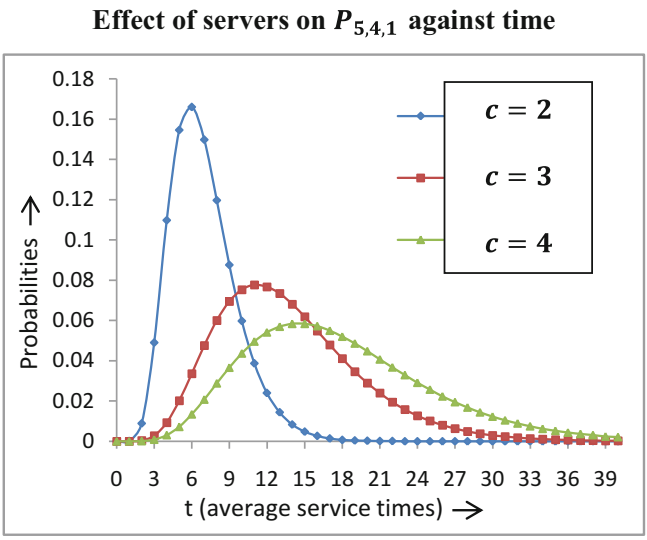


Fig. 2 Effect of servers on $P_{5,4,1}$ against time

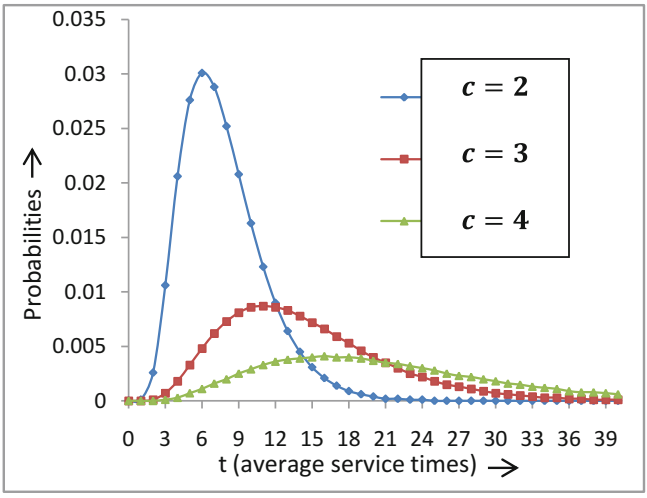


Fig. 3 Effect of servers on $P_{5,3,2}$ against time

values of c. From the figure, it is concluded that as the number of servers increases, the probability $P_{5,5,0}$ increases. So we can interpret that with increase in the number of servers, the probability of exactly the same number of arrivals and departures attain probability at smaller values of t.

In Figs. 2 and 3, the probabilities $P_{5,4,1}$ and $P_{5,3,2}$ are plotted against time t for the number of servers $c = 2, 3, 4$. From both the figures, it is concluded that due

to their particular behavior, the probabilities $P_{5,4,1}$ and $P_{5,3,2}$ increase in the starting times and then start decreasing for higher values of time. From these figures, it is concluded that as the number of servers increases the highest value attained by the probability decreases.

6 Conclusion

This system (multiserver retrial queueing system with reneging customers) can be considered as a generalized form of many existing queueing systems provided with many more features and related to many practical situations. The time-dependent probabilities of exact number of arrivals and departures at when all, some, or none servers are busy are obtained and expression for some important performance measures gives an exhaustive picture of various systems.

References

1. G.I. Falin, A survey of retrial queues. Queue. Syst. **7**, 127–167 (1990)
2. G.I. Falin, J.G.C. Templeton, *Retrial Queues* (Chapman & Hall, London, 1997)
3. V.G. Kulkarni, H.M. Liang, Retrial queues revisited, in *Frontiers in Queueing*, ed. by J. H. Dshalalow, (CRC Press, Boca Raton, FL, 1997), pp. 19–34
4. J.R. Artalejo, Accessible bibliography on retrial queues. Math. Comput. Model. **30**, 1–6 (1999)
5. B.D. Choi, Y. Chang, Single server retrial queues with priority calls. Math. Comput. Model. **30**, 7–32 (1999)
6. F.A. Haight, Queuing with reneging. Metrika **2**, 186–197 (1959)
7. R.O. Al-Seedy, A.A. El-Sherbiny, S.A. El-Shehawy, S.I. Ammar, Transient solution of the M/M/c queue with balking and reneging. Comput. Math. Appl. **57**, 1280–1285 (2009)
8. Y.W. Shin, Transient distributions of level dependent quasi-birth-death processes with linear transition rates. Korean J. Comput. Appl. Math. **7**, 83–100 (2000)
9. P.R. Parthasarathy, R. Sudhesh, Time-dependent analysis of a single-server retrial queue with state-dependent rates. Oper. Res. Lett. **35**, 601–611 (2007)
10. N. Grier, W.A. Massey, T. Mckoy, W. Whitt, The time-dependent Erlang loss model with retrials. Telecommun. Syst. **7**, 253–265 (1997)
11. C.D. Pegden, M. Rosenshine, Some new results for the M/M/1 queue. Manag. Sci. **28**, 821–828 (1982)
12. Indra, Ruchi, Transient analysis of two-dimensional M/M/1 queueing system with working vacations. JMASS. **5**, 110–128 (2010)
13. Kumar, Indra, Some new results for a two-state batch departure multiple vacation queueing models. Am. J. Operat. Res. **3**, 26–33 (2013)
14. P.C. Garg, Kumar, A single server retrial queues with impatient customers. Math. J. Interdiscipl. Sci. **1**, 67–82 (2012)
15. N. Singla, S. Kalra, Performance analysis of a two-state queueing model with retrials. J. Rajasthan Acad. Phys. Sci. **17**, 81–100 (2018a)
16. N. Singla, S. Kalra, A two-state retrial queueing model with reneging customers. Int. J. Manag. Technol. Eng. **8**, 2650–2663 (2018b)
17. B.D. Bunday, *Basic Queueing Theory* (Edward Arnold (Publishers) Ltd, London, 1986)
18. H. Bateman, *Tables of Integral Transform* (McGraw-Hill Book Company, New York, NY, 1954)

Performance Modelling of a Discrete-Time Retrial Queue with Preferred and Impatient Customers, Bernoulli Vacation and Second Optional Service

Geetika Malik and Shweta Upadhyaya

Abstract This study deals with analyzing a discrete-time retrial queue with Bernoulli vacation. We have concentrated on analyzing a Geo/G/1 retrial queueing model wherein server provides optional service in addition to compulsory service to fulfil customer's satisfaction and to improve the grade of service. On arrival of a customer, if the server is unavailable then either that customer enters the orbit to retry for the same service after a certain period of time or it leaves the system without being served. Once the first essential service is completed, it is up to the customer to opt for the second optional service or not. Also, after the completion of each service, the server may wait for another customer or it may leave for a vacation of random length. The steady state probabilities for different server states and queue size of the considered model are established. Further, some numerical experiments and results are presented.

Keywords Discrete-time queue · Preferred and impatient customers · Bernoulli vacation · Second optional service

1 Introduction

We are all aware that retrial queues have been one of the favourite topics for researchers from more than two decades. It has been widely explored and studied by them but in continuous time frame. In case of discrete-time retrial queue, still a lot of work needs to be done. In areas like call centres, mobile communication, packet switching networks and ATMs, the transfer of units is in the form of bits or packets of some specified length which are more suitable to model and analyze via discrete-time systems. Yang and Li [1] were the one to explore retrial queue

G. Malik (✉) · S. Upadhyaya
Amity Institute of Applied Sciences, Amity University, Noida, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_25

in discrete environment. They have calculated the distribution of queue size for Geo/G/1 discrete-time retrial queue. Atencia and Moreno [2], Wang and Zhao [3], Wang and Zhang [4] and Upadhyaya [5] are few other authors who are responsible to generalize the same model. Atencia and Moreno [2] calculated the steady state distribution of orbit and system sizes for a Geo/G/1 retrial model under discrete-time scenario. Thereafter, Wang and Zhao [3] examined this model with starting failures and general retrial time. Also, a Geo/G/1 retrial queue with negative customers as well as an unreliable server was investigated by Wang and Zhang [4]. Upadhyaya [5] was the one who gave various performance measures and a numerical analysis for discrete-time queuing system wherein the server follows J-vacation policy and may face breakdown.

It is not always necessary that an arriving customer leaves the system after completion of its service. In real life, there are a lot of cases that customer leaves without being served. For instance, if we make a call in a call centre and wait for the executive to come online then we often hang up before they serve us. This type of behaviour is very common in queueing analysis and such types of customers are *impatient customers*. Another case may happen that a customer has a sort of urgent demand on arrival and it interrupts the ongoing service to initiate its own service. Such situation arises because of *prioritized customer*. Hassan et al. [6] have investigated a single-server discrete-time queue wherein the inter-retrial times are generally distributed and customer may balk on arrival. Rabia [7] applied direct truncation technique and derived some very important performance measures for discrete-time system under the condition of phase-type service. Due to the fast pace of advancement in communication and network sector in this twenty-first century, it is very common that that server offers an *additional service* which is not mandatory and totally depends upon the customer to choose it or not. We must mention the work done by Jain and Agarwal [8], Zhang and Zhu [9] and Chen et al. [10] for incorporating this concept in discrete-time environment. They all have considered service in two parts viz first essential service and second optional service. While Zhang and Zhu [9] and Chen et al. [10] studied the Geom/G/1 model, Jain and Agarwal included geometric batch arrival process.

As per the literature survey conducted, we can conclude that there is no work done on retrial queue under Bernoulli vacation and second optional service with priority and impatient customers in discrete-time environment. This is what motivated us to analyze this model. The rest of the chapter is organized as follows: The next section describes model details and all the suitable notations. There after Sect. 3 gives the probability generating functions of the steady state probabilities. Further, the performance characteristics are derived in Sect. 4 by using the results obtained in Sect. 3. Section 5 explores some numerical examples and the effect of some parameters on the performance measures. Finally, future scope and conclusion is discussed in Sect. 6.

2 Model Description

In queueing model under discrete-time phenomenon, time is presumed to be a discrete random variable which is uniformly segmented in *slot*. Also, all the exercises like arrival, retrial, departure or vacation take place at these slot boundaries only. The assumptions made to examine this model are:

- We follow an *early arrival system (EAS)* according to which the arrival or retrials occur at an epoch just after a slot boundary, say (n, n^+) whereas the departures or ending of vacation take place at an epoch just prior to a slot boundary, say (n^-, n).
- The arrival process is geometrically distributed with arrival rate γ.
- If an arriving customer finds the server idle, it immediately receives the service with first come first serve discipline (FCFS), else if the server is busy or on vacation then either it waits with probability η or it exits the system completely with probability $\overline{\eta} = 1 - \eta$ (impatient behaviour).
- Server may willingly stop the ongoing service on demand of the customer just arrived and rather provide service to it with probability β (prioritized customer).
- Once essential service is finished, it completely relies on the customer whether to go for the second optional service (SOS) or not. The probability of choosing this option is s where as the customer leaves the system after receiving the first essential service (FES) with probability $\overline{s} = 1 - s$.
- Once all the customers present in the system have completely received their service, the server may hold back for next customer with probability v or heads towards vacation with probability $\overline{v} = 1 - v$.
- The two service times, i.e. first essential service and second optional service are independent and identically distributed and follows a general distribution $\{b_{1,i}\}$ and $\{b_{2,i}\}$ with corresponding generating functions $B_1(x) = \sum_{i=1}^{\infty} b_{1,i} x^i$ and $B_2(x) = \sum_{i=1}^{\infty} b_{2,i} x^i$.
- Vacation time is also assumed to be generally distributed with parameter $\{b_{3,i}\}$ and generating function $B_3(x) = \sum_{i=1}^{\infty} b_{3,i} x^i$.
- It may happen that on arriving when the customer finds the server busy or on vacation and observe that there is no waiting space available in the system. In such cases, the former enters the 'orbit' (pool of blocked customers) where they can wait and retry for their service for a random period of time. Retrial process can occur only after service completion or once the vacation is over. Retrial time too follows general distribution $\{r_i\}$ with generating function $R(x) = \sum_{i=0}^{\infty} r_i x^i$.
- The inter-arrival time, retrial time, service time and vacation time are all mutually independent.

3 Probability Generating Functions

In this section of our work, we have calculated the probability generating functions of different server states and for the system size as well. Firstly, the state governing equations are formed and thereafter the expressions are obtained by applying generating function technique.

Let the model under study be represented by $Z_n = (C_n, t_{i,n}, N_n)$, where

$$C_n \begin{cases} 0, & \text{if server is idle} \\ 1, & \text{if server is busy with FES.} \\ 2, & \text{if server is busy with SOS.} \\ 3, & \text{if server is on Vacation.} \end{cases}$$

and N_n denotes the number of repeated customers.

If $C_n = 0$ and $N_n > 0$, $t_{0,n}$ denotes the remaining retrial time. When $C_n = 1, 2$ or 3 then $t_{1,n}$, $t_{2,n}$ and $t_{3,n}$ denote remaining time during FES, SOS and vacation period, respectively.

Thus $\{Z_n, n \geq 1\}$ forms a Markov chain with state space $\{(0,0); (0,i,k): i \geq 1, k \geq 1; (1,i,k): i \geq 1, k \geq 0; (2,i,k): i \geq 1, k \geq 0; (3,i,k): i \geq 1, k \geq 0\}$.

Next, we define the stationary probabilities as:

$$\begin{aligned} \xi_{0,0} &= \lim_{n\to\infty} \Pr\{C_n = 0, N_n = 0\}, \\ \xi_{0,i,k} &= \lim_{n\to\infty} \Pr\left\{C_n = 0, t_{j,n} = i, N_n = k\right\};\ i, k \geq 1 \\ \xi_{j,i,k} &= \lim_{n\to\infty} \Pr\left\{C_n = j, t_{j,n} = i, N_n = k\right\};\ i \geq 1, \quad k \geq 0, \quad j = 1, 2, 3 \end{aligned}$$

The Kolmogorov equations thus formed from above are as follows:

$$\xi_{0,0} = \overline{\gamma}\xi_{0,0} + \overline{\gamma s}v\xi_{1,1,0} + v\overline{\gamma}\xi_{2,1,0} + \overline{\gamma}\xi_{3,1,0} \tag{1}$$

$$\xi_{0,i,k} = \overline{\gamma}\xi_{0,i+1,k} + \overline{\gamma}r_i\overline{s}v\xi_{1,1,k} + \overline{\gamma}r_i v\xi_{2,1,k} + \overline{\gamma}r_i\xi_{3,1,k}; \qquad i, k \geq 1 \tag{2}$$

$$\begin{aligned} \xi_{1,i,k} = {} & \delta_{0,k}\gamma b_{1,i}\xi_{0,0} + \overline{\gamma}b_{1,i}\xi_{0,1,k+1} + \left(1-\delta_{0,k}\right)\gamma b_{1,i}\sum_{j=1}^{\infty}\xi_{0,j,k} + \overline{s}\gamma v b_{1,i}\xi_{1,1,k} \\ & + \overline{s\gamma}v r_0 b_{1,i}\xi_{1,1,k+1} + \left(\overline{\gamma} + \gamma\overline{\eta}\right)\xi_{1,i+1,k} + \left(1-\delta_{0,k}\right)\gamma\eta\overline{\beta}\xi_{1,i+1,k-1} \\ & + \left(1-\delta_{0,k}\right)\sum_{j=2}^{\infty}\gamma\eta\beta b_{1,i}\xi_{1,j,k-1} + \gamma v b_{1,i}\xi_{2,1,k} \\ & + \overline{\gamma}v r_0 b_{1,i}\xi_{2,1,k+1} + \gamma b_{1,i}\xi_{3,1,k} + \overline{\gamma}r_0 b_{1,i}\xi_{3,1,k+1}; \qquad i \geq 1, k \geq 0 \end{aligned} \tag{3}$$

$$\xi_{2,1,k} = \left(1-\delta_{0,k}\right)\gamma s\eta\overline{\beta}b_{2,i}\xi_{1,1,k-1} + s\left(\overline{\gamma}+\gamma\overline{\eta}\right)b_{2,1}\xi_{1,1,k} + \left(\overline{\gamma}+\gamma\overline{\eta}\right)\xi_{2,i+1,k} + \left(1-\delta_{0,k}\right)\gamma\eta\overline{\beta}\xi_{2,i+1,k-1} + \left(1-\delta_{0,k}\right)\sum_{j=2}^{\infty}\gamma\eta\beta b_{2,i}\xi_{2,j,k-1}; \quad i \geq 1, k \geq 0 \tag{4}$$

$$\xi_{3,1,k} = \left(1-\delta_{0,k}\right)\gamma\overline{s}v\eta\overline{\beta}b_{3,i}\xi_{1,1,k-1} + \overline{s}\left(\overline{\gamma}+\gamma\overline{\eta}\right)\overline{v}b_{3,i}\xi_{1,1,k} + \left(1-\delta_{0,k}\right)\gamma\overline{s}\eta\overline{\beta}b_{3,i}\xi_{2,1,k-1} + \left(\overline{\gamma}+\gamma\overline{\eta}\right)\overline{v}b_{3,i}\xi_{2,1,k} + \left(1-\delta_{0,k}\right)\gamma\eta\overline{\beta}\xi_{3,i+1,k-1} + \left(\overline{\gamma}+\gamma\overline{\eta}\right)\xi_{3,i+1,k}; \quad i \geq 1, k \geq 0 \tag{5}$$

where $\delta_{ij} = \begin{cases} 1; & i = j \\ 0; & i \neq j \end{cases}$

The normalizing condition is given by:

$$\xi_{0,0} + \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\xi_{0,i,k} + \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\sum_{k=0}^{\infty}\xi_{j,i,k} = 1$$

For solving the above equations, we define the following auxiliary functions and generating functions:

$$\chi_{0,i}(z) = \sum_{k=1}^{\infty}\xi_{0,i,k}z^k; \qquad \chi_{1,i}(z) = \sum_{k=0}^{\infty}\xi_{1,i,k}z^k$$

$$\chi_{2,i}(z) = \sum_{k=0}^{\infty}\xi_{2,i,k}z^k; \qquad \chi_{3,i}(z) = \sum_{k=0}^{\infty}\xi_{3,i,k}z^k; \qquad i \geq 1$$

$$\chi_0(x,z) = \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}\xi_{0,i,k}z^k x^i; \qquad \chi_1(x,z) = \sum_{i=1}^{\infty}\sum_{k=0}^{\infty}\xi_{1,i,k}z^k x^i$$

$$\chi_2(x,z) = \sum_{i=1}^{\infty}\sum_{k=0}^{\infty}\xi_{2,i,k}z^k x^i; \qquad \chi_3(x,z) = \sum_{i=1}^{\infty}\sum_{k=0}^{\infty}\xi_{3,i,k}z^k x^i; \quad i \geq 1$$

We can rewrite Eq. (1) as:

$$\gamma\xi_{0,0} = \overline{\gamma s}v\xi_{1,1,0} + v\overline{\gamma}\xi_{2,1,0} + \overline{\gamma}\xi_{3,1,0} \tag{6}$$

Multiply Eqs. (2)–(5) by z^k and summing over k, we have:

$$\chi_{0,i}(z) = \overline{\gamma}\chi_{0,i+1}(z) + \overline{\gamma s}vr_i\chi_{1,1}(z) + \overline{\gamma}vr_i\chi_{2,1}(z) + \overline{\gamma}r_i\chi_{3,1}(z) - \gamma r_i\xi_{0,0} \tag{7}$$

$$\chi_{1,i}(z) = \left(\tfrac{z-r_0}{z}\right)\gamma b_{1,i}\xi_{0,0} + \tfrac{\overline{\gamma}}{z}b_{1,i}\chi_{0,1}(z) + \gamma b_{1,i}\chi_0(1,z) + z\gamma\eta\beta b_{1,i}\left(\chi_1(1,z) - \chi_{1,1}(z)\right) + \left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)\chi_{1,i+1}(z) + \left(\tfrac{\gamma z+\overline{\gamma}r_0}{z}\right)b_{1,i}\left(\overline{s}v\chi_{1,1}(z) + v\chi_{2,1}(z) + \chi_{3,1}(z)\right) \tag{8}$$

$$\chi_{2,i}(z) = s\left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)b_{2,i}\chi_{1,1}(z) + \left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)\chi_{2,i+1}(z) + z\gamma\eta\beta b_{2,i}\left(\chi_2(1,z) - \chi_{2,1}(z)\right) \tag{9}$$

$$\chi_{3,i}(z) = \overline{sv}\left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)b_{3,i}\chi_{1,1}(z) + \overline{v}\left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)b_{3,i}\chi_{2,1}(z) + \left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)\chi_{3,1}(z) \tag{10}$$

Now multiplying both sides of Eqs. (7)–(10) by x^i and summing over i, we get:

$$\left(\tfrac{x-\overline{\gamma}}{x}\right)\chi_0(x,z) = \overline{\gamma}\overline{s}v\chi_{1,1}(z)\left(R(x) - r_0\right) + \overline{\gamma}v\chi_{2,1}(z)\left(R(x) - r_0\right) + \overline{\gamma}\chi_{3,1}(z)\left(R(x) - r_0\right) - \gamma\xi_{0,0}\left(R(x) - r_0\right) - \overline{\gamma}\xi_{0,1}(z) \tag{11}$$

Put $x = 1$ in above Equation.

$$\gamma\chi_0(1,z) = \overline{\gamma}\overline{s}v\chi_{1,1}(z)(1 - r_0) + \overline{\gamma}v\chi_{2,1}(z)(1 - r_0) + \overline{\gamma}\chi_{3,1}(z)(1 - r_0) - \gamma(1 - r_0)\xi_{0,0} - \overline{\gamma}\chi_{0,1}(z) \tag{12}$$

Let $A(z) = \left(\overline{\gamma} + \gamma\overline{\eta} + z\gamma\eta\overline{\beta}\right)$. Then we have:

$$\left(\tfrac{x-A(z)}{x}\right)\chi_1(x,z) = \left(\tfrac{z-r_0}{z}\right)\gamma B_1(x)\xi_{0,0} + \tfrac{\overline{\gamma}}{z}B_1(x)\chi_{0,1}(z) + \gamma B_1(x)\chi_0(1,z) + z\gamma\eta\beta B_1(x)\chi_1(1,z) + \left(\tfrac{\gamma z+\overline{\gamma}r_0}{z}\overline{s}vB_1(x) - z\gamma\eta\beta B_1(x) - A(z)\right)\chi_{1,1}(z) + \left(\tfrac{\gamma z+\overline{\gamma}r_0}{z}\right)B_1(x)\left(v\chi_{2,1}(z) + \chi_{3,1}(z)\right) \tag{13}$$

$$\left(\frac{x - A(z)}{x}\right)\chi_2(x,z) = sA(z)B_2(x)\chi_{1,1}(z) - A(z)\chi_{2,1}(z) + z\gamma\eta\beta B_2(x)\left(\chi_2(1,z) - \chi_{2,1}(z)\right) \tag{14}$$

$$\left(\frac{x - A(z)}{x}\right)\chi_3(x,z) = \overline{sv}A(z)B_3(x)\chi_{1,1}(z) + \overline{v}A(z)B_3(x)\chi_{2,1}(z) - A(z)\chi_{3,1}(z) \tag{15}$$

Putting $x = \overline{\gamma}$ in Eq. (11), we obtain:

$$\gamma\xi_{0,0}\left(R\left(\overline{\gamma}\right)-r_0\right)=\overline{\gamma}\overline{s}v\left(R\left(\overline{\gamma}\right)-r_0\right)\chi_{1,1}(z)+\overline{\gamma}v\left(R\left(\overline{\gamma}\right)-r_0\right)\chi_{2,1}(z)$$
$$+\overline{\gamma}\left(R\left(\overline{\gamma}\right)-r_0\right)\chi_{3,1}(z)-\overline{\gamma}\chi_{0,1}(z) \tag{16}$$

Now, we put $x=A(z)$ in Eqs. (13)–(15).

$$-\left(\tfrac{z-r_0}{z}\right)\gamma B_1\left(A(z)\right)\xi_{0,0}=\tfrac{\overline{\gamma}}{z}B_1\left(A(z)\right)\chi_{0,1}(z)+\gamma B_1\left(A(z)\right)\chi_0\left(1,z\right)$$
$$+z\gamma\eta\beta B_1\left(A(z)\right)\chi_1\left(1,z\right)$$
$$+\left(\tfrac{\gamma z+\overline{\gamma}r_0}{z}\overline{s}vB_1\left(A(z)\right)-z\gamma\eta\beta B_1\left(A(z)\right)-A(z)\right)\chi_{1,1}(z)$$
$$+\left(\tfrac{\gamma z+\overline{\gamma}r_0}{z}\right)B_1\left(A(z)\right)\left(v\chi_{2,1}(z)+\chi_{3,1}(z)\right) \tag{17}$$

$$\chi_{2,1}(z)=\frac{sA(z)B_2\left(A(z)\right)}{\lambda(z)}\chi_{1,1}(z)+\frac{z\gamma\eta\beta B_2\left(A(z)\right)}{\lambda(z)}\chi_2\left(1,z\right) \tag{18}$$

where $\lambda(z)=A(z)+z\gamma\eta\beta B_2(A(z))$

$$\chi_{3,1}(z)=\overline{s}\overline{v}B_3\left(A(z)\right)\chi_{1,1}(z)+\overline{v}B_3\left(A(z)\right)\chi_{2,1}(z) \tag{18a}$$

Using Eqs. (14) and (18) we can write:

$$\chi_{2,1}(z)=\frac{s\lambda_1(z)A(z)B_2\left(A(z)\right)}{\lambda_2(z)}\left(\frac{1-z\overline{\beta}}{1-z}\right)\chi_{1,1}(z) \tag{19}$$

Put above equation in Eq. (18a).

$$\chi_{3,1}(z)=\overline{s}B_3\left(A(z)\right)\left[\overline{s}+\frac{s\lambda_1(z)A(z)B_2\left(A(z)\right)}{\lambda_2(z)}\left(\frac{1-z\overline{\beta}}{1-z}\right)\right]\chi_{1,1}(z) \tag{19a}$$

where $\lambda_1(z)=1-A(z)-z\gamma\eta\beta$
$\lambda_2(z)=\lambda(z)\lambda_1(z)+z\gamma\eta\beta B_2\left(A(z)\right)\left(A(z)+z\gamma\eta\beta\right)$

Performing some algebraic calculations in the above equations, we have the following two equations:

$$\gamma r_0 B_1\left(A(z)\right)\left(\tfrac{1-z\overline{\beta}}{z}\right)\xi_{0,0}=\overline{\gamma}B_1\left(A(z)\right)\left(\tfrac{1-z\overline{\beta}}{z}\right)\chi_{0,1}(z)$$
$$+\left[B_1\left(A(z)\right)G(z)\left(\tfrac{1-z\overline{\beta}}{1-z}\right)\left(\tfrac{z+\overline{\gamma}r_0(1-z)}{z}\right)-\tfrac{z\beta B_1(A(z))}{1-z}-A(z)\right]\chi_{1,1}(z) \tag{20}$$

$$\gamma\xi_{0,0}\left(R\left(\overline{\gamma}\right)-r_0\right)=-\overline{\gamma}\chi_{0,1}(z)+\overline{\gamma}G(z)\left(R\left(\overline{\gamma}\right)-r_0\right)\chi_{1,1}(z) \tag{20a}$$

where $G(z) = \left[\overline{s} + \frac{s\lambda_1(z)A(z)B_2(A(z))}{\lambda_2(z)}\left(\frac{1-z\overline{\beta}}{1-z}\right)\right](v + \overline{v}B_3(A(z)))$

Solving Eqs. (20) and (20a), we get:

$$\chi_{0,1}(z) = \frac{z\gamma\left(R\left(\overline{\gamma}\right) - r_0\right)\mathrm{Nr1}(z)}{\mathrm{Dr}(z)}\xi_{0,0} \tag{21}$$

$$\chi_{1,1}(z) = \frac{\mathrm{Nr}(z)}{\mathrm{Dr}(z)}\xi_{0,0} \tag{22}$$

where $\mathrm{Nr1}(z) = z\beta B_1(A(z)) + (1-z)A(z) - G(z)\left(1 - z\overline{\beta}\right)B_1(A(z))$, $\mathrm{Nr}(z) = \gamma B_1(A(z))\left(1 - z\overline{\beta}\right)(1-z)R(\overline{\gamma})$, and $\mathrm{Dr}(z) = B_1(A(z))G(z)\left(1 - z\overline{\beta}\right)(z + \overline{\gamma}R(\overline{\gamma})(1-z)) - z^2\beta B_1(A(z)) - z(1-z)A(z)$

From Eqs. (19) and (19a), we have:

$$\chi_{2,1}(z) = \left[\frac{s\lambda_1(z)A(z)B_2(A(z))}{\lambda_2(z)}\left(\frac{1-z\overline{\beta}}{1-z}\right)\right]\frac{\mathrm{Nr}(z)}{\mathrm{Dr}(z)}\xi_{0,0} \tag{23}$$

$$\chi_{3,1}(z) = \overline{v}B_3(A(z))\left[\overline{s} + \frac{s\lambda_1(z)A(z)B_2(A(z))}{\lambda_2(z)}\left(\frac{1-z\overline{\beta}}{1-z}\right)\right]\frac{\mathrm{Nr}(z)}{\mathrm{Dr}(z)}\xi_{0,0} \tag{24}$$

Theorem 1 The Markov chain $\{Z_n: n \geq 1\}$ has stationary distribution with probability generating function given by:

$$\chi_0(x,z) = \frac{\gamma xz\left(R(x) - R(\overline{\gamma})\right)\mathrm{Nr1}(z)}{(x-\overline{\gamma})\,\mathrm{Dr}(z)}\xi_{0,0};$$

$$\chi_1(x,z) = \frac{\gamma x\left(1 - z\overline{\beta}\right)(1-z)R(\overline{\gamma})A(z)\left(B_1(x) - B_1(A(z))\right)}{(x - A(z))\,\mathrm{Dr}(z)}\xi_{0,0}$$

$$\chi_2(x,z) = \frac{xsA(z)\mathrm{Nr2}(z)\mathrm{Nr}(z)}{(x - A(z))\,\mathrm{Dr}(z)}\xi_{0,0}$$

$$\chi_3(x,z) = \frac{\left(B_3(x) - B_3(A(z))\right)}{(x - A(z))}\frac{\overline{v}xA(z)\left[\overline{s}\lambda_2(z) + s\gamma\eta A(z)B_2(A(z))\left(1 - z\overline{\beta}\right)\right]\mathrm{Nr}(z)}{\lambda_2(z)\mathrm{Dr}(z)}\xi_{0,0}$$

where, $\mathrm{Nr2}(z)=\left\{B_2(x)\left(\frac{1-z\overline{\beta}}{1-z}\right)(1+z\gamma\eta\beta)-s\gamma\eta\frac{B_2(A(z))}{\lambda_2(z)}\left(1-z\overline{\beta}\right)A(z)\left(1+\frac{z\beta B_2(x)}{1-z}\right)\right\}$,

$$\rho_2=\eta\beta B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)+\eta\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)$$
$$-\eta\beta B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\left(\overline{s}+s\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\left(v+\overline{v}B_3\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right),$$

$\xi_{0,0}=\frac{\rho_2}{R(\overline{\gamma})\rho_1}$, and $\rho_1=$

$$\left\{\begin{array}{l}\eta\left(\beta\left(B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)+\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\\-\beta B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\left[\eta\left(\overline{s}+s\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\right.\\\left.\left(v+\overline{v}B_3\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)+s\overline{s}\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right]\\-2\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\left[\left(1-B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\right.\\+\overline{v}B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\left(\overline{s}+s\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\\\left.\left(1-B_3\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\right]\end{array}\right\}$$

Proof Using Eqs. (11)–(15) and (21)–(24) together, we get the values of $\chi_0(x, z)$, $\chi_1(x, z)$, $\chi_2(x, z)$ and $\chi_3(x, z)$. By applying the normalizing condition $\xi_{0,0}+\chi_0(1,1)+\chi_1(1,1)+\chi_2(1,1)+\chi_3(1,1)=1$, we obtain the desired value of $\xi_{0,0}$. This completes the proof of Theorem 1.

Corollary 1 The probability generating functions of the number of customers present in the orbit according to different states of the server are:

Idle state: $\chi_0(1,z)=\frac{z(1-R(\overline{\gamma}))\mathrm{Nr1}(z)}{\mathrm{Dr}(z)}\xi_{0,0}$

Busy with essential service: $\chi_1(1,z)=\frac{(1-z)R(\overline{\gamma})A(z)(1-B_1(A(z)))}{\eta\mathrm{Dr}(z)}\xi_{0,0}$

Busy with second optional service:

$\chi_2(1,z)=\frac{sA(z)\left(1-z\overline{\beta}\right)\left[\lambda_2(z)-\gamma\eta sB_2(A(z))\left(1-z\overline{\beta}\right)(z\gamma\eta\beta+A(z))\right]\mathrm{Nr}(z)}{(1-A(z))\mathrm{Dr}(z)\lambda_2(z)(1-z)}\xi_{0,0}$

Vacation state: $\chi_3(1,z)=\frac{(1-B_3(A(z)))\overline{v}A(z)(\overline{s}\lambda_2(z)+s\gamma\eta A(z)B_2(A(z))\left(1-z\overline{\beta}\right)\mathrm{Nr}(z)}{(1-A(z))\lambda_2(z)\mathrm{Dr}(z)}\xi_{0,0}$

4 Performance Measures

In the following section, we have provided the probabilities of the server in different states along with queue size and system size. This will help the user to predict an approximate waiting time and to create a hustle free system.

- The system is empty with probability:

$$\xi_{0,0}=\frac{\rho_2}{R\left(\overline{\gamma}\right)\rho_1}$$

- The server is in idle state with probability:

$$\chi_0(1,1)=\left(R\left(\overline{\gamma}\right)-1\right)\frac{\rho_2}{R\left(\overline{\gamma}\right)\rho_1}$$

- The probability that server is busy with first essential service is given by:

$$\chi_1(1,1) = \frac{-2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\left(1 - B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\right)}{\rho_1}$$

- The probability that server is busy with optional service is given by:

$$\chi_2(1,1) = \frac{-\beta s\overline{s}\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)}{\rho_1}$$

- The probability that the server is on vacation is found to be:

$$\chi_3(1,1) = \frac{-2\overline{v}\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\left(1 - B_3\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\right)}{\rho_1}$$

- The average number of customers in the retrial orbit is given by:

$$L_o = \left(\chi_0'(1,z) + \chi_1'\left(1,z\right) + \chi_2'\left(1,z\right) + \chi_3'\left(1,z\right)\right)_{z=1}$$

where $\chi_0'(1,z)\big|_{z=1} = (R(\overline{\gamma}) - 1)\frac{\rho_2}{R(\overline{\gamma})\rho_1}\left\{\frac{\text{num1}''\text{dem1}' - \text{num1}'\text{dem1}''}{2\text{dem1}'''}\right\}$, $\chi_1'(1,z)\big|_{z=1} = \frac{\rho_2}{\eta R(\overline{\gamma})\rho_1}\left\{\frac{\text{num2}''\text{dem1}' - \text{num2}'\text{dem1}''}{2\text{dem1}'''}\right\}$,
$\chi_2'(1,z)\big|_{z=1} = \frac{s\rho_2}{R(\overline{\gamma})\rho_1}\left\{\frac{\text{num3}'\text{dem2}' - \text{num3}'\text{dem2}''}{2\text{dem2}'''}\right\}$, and $\chi_3'(1,z)\big|_{z=1} = \frac{\overline{v}\rho_2}{R(\overline{\gamma})\rho_1}\left\{\frac{\text{num4}'\text{dem2}' - \text{num4}'\text{dem2}''}{2\text{dem2}'''}\right\}$

- The average number of customers in the system is given by:

$$L_s = L_o + \chi_1(1,1) + \chi_2(1,1)$$

(All the necessary symbols are given in Appendix).

5 Numerical Results

This section provides some of the very useful results obtained via performing a programme in MATLAB software. These are basically the effect of various parameters on the average queue length. We have also calculated the values of different probabilities given in Sect. 4. The default parameters taken are $\gamma = 0.95$, $\beta = 0.9$, $\eta = 0.02$, $s = 0.05$, $\overline{v} = 0.1$ and $r = 0.85$. Further, we have assumed that retrial time, service times (FES and SOS) and vacation time follow geometric distribution. The respective distribution functions are taken as $R(x) = \frac{r}{1-\overline{r}x}$; $B_1(x) = B_2(x) = \frac{7x}{10-3x}$ and $B_3(x) = \frac{v}{(1-\overline{v}x)}$. The values of the probabilities thus obtained are $\xi_{0,0} = 0.2464$, $\chi_0(1,1) = 0.0354$, $\chi_1(1,1) = 0.5053$, $\chi_2(1,1) = 0.2346$

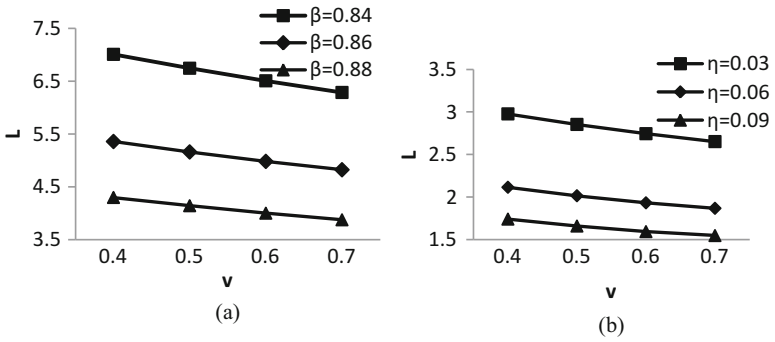


Fig. 1 (**a**) L vs. v with varying β. (**b**) L vs. v with varying η

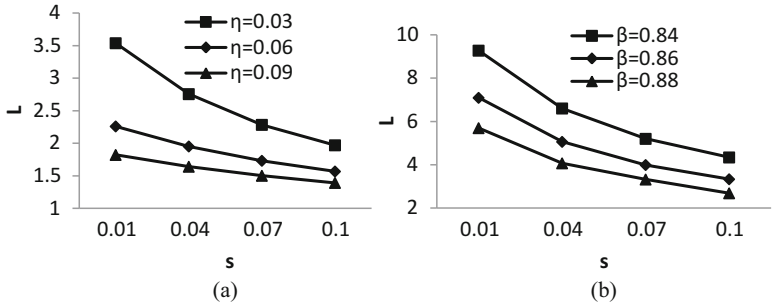


Fig. 2 (**a**) L vs. s with varying η. (**b**) L vs. s with varying β

and $\chi_3(1,1) = 0.2712$. The average queue length is 3.0997. Moreover, we have represented the results in the following four figures:

In Fig. 1a, b, the trend of average queue length, 'L', is observed against v by varying the impatient (η) and priority (β) parameter, respectively. It can be seen that 'L' decreases linearly if the probability of going on vacation ($\overline{v}$) is decreasing which is as expected. The average queue size further decreases while we increase η and β. Then, in Fig. 2a, b, average queue size 'L' is studied against the probability of going on optional service s by varying η and β, respectively. We observe that the average queue length first decreases sharply then gradually it is almost attaining a constant value with an increase in η and β. Thus, we conclude that priority and impatient behaviour can change the average queue size and therefore we can model a better system by varying these parameters.

6 Conclusion

We have worked on Geo/G/1 retrial queue with second optional service plus preferred and impatient customers. Firstly, probability generating function method is applied to analyze the underlying Markov's process. We have then provided the various useful results including steady state probabilities and system size. Few numerical examples performed on MATLAB software are also given which will help the system designers to work more efficiently. Our model is applicable in digital and communication field wherein the data is transferred in discrete format. The efficiency of the model can be increased by including the concepts like bulk arrival or state-dependent server. There is also a scope of doing cost optimization of the considered model to develop a cost-effective system under techno-economic constraints.

A.1 Appendix

$$\lambda = \left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) + \gamma\eta\beta B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$\lambda' = \gamma\eta\overline{\beta} + \gamma^2\eta^2\beta\overline{\beta}B_2'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) + \gamma\eta\beta B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$\lambda_1' = -\gamma\eta$$

$$\lambda_2' = -\gamma\eta\lambda + \gamma\eta\beta B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)(1+\gamma\eta) + \gamma^2\eta^2\beta\overline{\beta}B_2'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$\begin{aligned}\lambda_2'' = {} & 2\lambda'\lambda'_1 + 2\gamma^2\eta^2\beta\overline{\beta}B_2'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)(1+\gamma\eta) \\ & + 2\gamma^2\eta^2\beta B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) + \gamma^3\eta^3\beta\overline{\beta}^2B_2''\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\end{aligned}$$

$$N_1' = -\overline{s}\gamma\eta\beta B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$N_1'' = -2\overline{s}\lambda_2'$$

$$N_1''' = -3\overline{s}\lambda_2''$$

$$N_2' = -\gamma\eta\beta s\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$\begin{aligned}N_2'' = {} & -2s\gamma^2\eta^2\beta\overline{\beta}B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) - 2s\gamma^2\eta^2\beta\overline{\beta}\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) \\ & + 2s\gamma\eta\overline{\beta}\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\end{aligned}$$

$$\begin{aligned}N_2''' = {} & 6s\gamma^2\eta^2\beta\overline{\beta}^2B_2'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\lambda_1' - 6s\gamma\eta\beta\overline{\beta}^2B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\lambda_1' \\ & - 6s\gamma\eta\beta\overline{\beta}^2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\lambda_1' + s\gamma^2\eta^2\beta\overline{\beta}^2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right) \\ & B_2''\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\lambda_1'\end{aligned}$$

$$N_3 = v + \overline{v}B_3\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$N_3' = \overline{v}\gamma\eta\overline{\beta}B_3'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$N_3'' = \overline{v}\gamma^2\eta^2\overline{\beta}^2B_3''\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$\varsigma' = \left(N_1' + N_2'\right)N_3$$

$$\varsigma'' = 2N_1'N_3' + 2N_1''N_3 + 2N_2'N_3' + 2N_2''N_3$$

$$\varsigma''' = 3N_1'N_3'' + 3N_1''N_3' + N_1'''N_3 + 3N_2'N_3'' + 3N_2''N_3' + N_2'''N_3$$

$$E' = -\gamma\eta\beta^2B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$\begin{aligned}E'' = {} & -2\gamma\eta\beta^2B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\\ & -2\gamma^2\eta^2\beta^2\overline{\beta}B_1'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\\ & -2\beta B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\lambda_2'\end{aligned}$$

$$F' = -\gamma^2\eta\beta\overline{\beta}B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$F'' = -3\gamma\eta\overline{\beta}\lambda_2' - \gamma^2\eta^2\beta\overline{\beta}B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)$$

$$H_1' = -\beta B_2\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'$$

$$H_2' = \beta B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'$$

$$H_1'' = 2\overline{\beta}B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\varsigma' - 2\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) - \overline{\beta}B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\varsigma''$$

$$\begin{aligned}H_2'' = {} & 2\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma' + \beta B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma''\\ & -2\overline{\beta}B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma' + 2\beta B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'\left(1 - \overline{\gamma}R\left(\overline{\gamma}\right)\right)\end{aligned}$$

$$\begin{aligned}H_2''' = {} & 3\gamma^2\eta^2\overline{\beta}^2\beta B_1''\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma' + 3\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma''\\ & -6\gamma\eta\overline{\beta}^2B_1'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'\\ & +6\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'\left(1 - \overline{\gamma}R\left(\overline{\gamma}\right)\right) + \beta B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'''\\ & -3\overline{\beta}B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma''\\ & +3\beta B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma''\left(1 - \overline{\gamma}R\left(\overline{\gamma}\right)\right) - 5\overline{\beta}B_1\left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\varsigma'\\ & \left(1 - \overline{\gamma}R\left(\overline{\gamma}\right)\right)\end{aligned}$$

$$E_1' = -E'$$

$$E_1'' = -2E' - E''$$

$$F_1' = -F'$$

$$F_1'' = -2F' - F''$$

$$E_1''' = -6\beta B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2 - 12\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2 - 12\beta B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2' - 3\left(\gamma\eta\overline{\beta}\right)^2\beta B_1''\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2 - 6\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2' - 3\beta B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2''$$

$$F_1''' = -7\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2' - 6\gamma\eta\overline{\beta}\lambda_2 - 6\gamma\eta\overline{\beta}\lambda_2' - 2\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\lambda_2''$$

$$W_1' = -\gamma\beta R\left(\overline{\gamma}\right)B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)$$

$$W_1'' = \gamma R\left(\overline{\gamma}\right)\left(-2\gamma\eta\overline{\beta}\beta B_1'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) + 2\overline{\beta}B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)$$

$$W_2' = -\overline{\beta}\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) + \gamma\eta\overline{\beta}\beta$$

$$W_3' = \lambda_2' - s\gamma\eta\left\{\gamma\eta\overline{\beta}\beta B_2'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) - \overline{\beta}B_2\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) + \gamma\eta\beta B_2\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right\}$$

$$W_4' = -\gamma\eta\overline{\beta}B_3'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)$$

$$W_5' = \overline{s}\lambda_2' + s\left\{\gamma\eta\overline{\beta}\beta B_2\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) - \left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\overline{\beta}B_2\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right) + \gamma\eta\overline{\beta}\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\beta B_2'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right.$$

$$\mathrm{num1}' = E' + F' + H_1'$$

$$\mathrm{num1}'' = E'' + F'' + H_1''$$

$$\mathrm{dem1}' = H_2' + E_1' + F_1'$$

$$\mathrm{dem1}'' = H_2'' + E_1'' + F_1''$$

$$\mathrm{dem1}''' = H_2''' + E_1''' + F_1'''$$

$$\mathrm{num2}' = -R\left(\overline{\gamma}\right)\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\left(1 - B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)$$

$$\mathrm{num2}'' = -2R\left(\overline{\gamma}\right)\gamma\eta\overline{\beta}\left(1 - B_1\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right) + 2\gamma\eta\overline{\beta}R\left(\overline{\gamma}\right)\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)B_1'\left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)$$

$$\mathrm{num3}' = W_1'W_2W_3$$

$$\mathrm{num3}'' = W_1''W_2W_3 + 2W_1'W_2'W_3 + 2W_1'W_2W_3'$$

$$\mathrm{num4}' = W_1'W_4W_5$$

$$\mathrm{num4}' = W_1''W_4W_5 + 2W_1'W_4'W_5 + 2W_1'W_4W_5'$$

$$\mathrm{dem2}' = \left(1 - \left(\overline{\gamma}+\gamma\overline{\eta}+\gamma\eta\overline{\beta}\right)\right)\mathrm{dem1}'$$

$$\text{dem2}'' = -2\gamma\eta\overline{\beta}\text{dem1}' + \left(1 - \left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\right)\text{dem1}''$$

$$\text{dem2}''' = -3\gamma\eta\overline{\beta}\text{dem1}'' + \left(1 - \left(\overline{\gamma} + \gamma\overline{\eta} + \gamma\eta\overline{\beta}\right)\right)\text{dem1}'''$$

References

1. T. Yang, H. Li, On steady state queue size distribution of the discrete-time Geo/G/1 queue with repeated customers. Queue. Syst. **21**, 199–215 (1995)
2. I. Atencia, P. Moreno, A discrete time retrial queue with general retrial times. Queue. Syst. **48**, 5–21 (2004)
3. J. Wang, Q. Zhao, Discrete-time Geo/G/1 retrial queue with general retrial times and starting failures. Math. Comput. Model. **45**, 853–863 (2007)
4. J. Wang, P. Zhang, A discrete-time retrial queue with negative customers and unreliable server. Comput. Ind. Eng. **56**(4), 1216–1222 (2009)
5. S. Upadhyaya, Performance analysis of a discrete-time Geo/G/1 retrial queue under J-vacation policy. Int. J. Ind. Syst. Eng. **29**(3), 369–388 (2018)
6. A.K. Hassan, S. Rabia, F. Taboly, A discrete-time Geo/G/1 retrial queue with general retrial times and balking customers. J. Korean Stat. Soc. **37**, 335–348 (2008)
7. S. Rabia, An improved truncation technique to analyse a Geo/PH/1 retrial queue with impatient customers. Comput. Oper. Res. **46**, 69–77 (2014)
8. M. Jain, S. Agarwal, A discrete-time GeoX/G/1 retrial queueing system with starting failure and optional service. Int. J. Operat. Res. **8**(4), 428–457 (2010)
9. F. Zhang, Z. Zhu, A discrete-time unreliable Geo/G/1 retrial queue with balking customers, second optional service and general retrial times. Math. Probl. Eng. **2013**, 1–13 (2013)
10. Y. Chen, L. Cai, C. Wei, A discrete-time Geo/G/1 retrial queue with balking customer, second optional service, Bernoulli vacation and general retrial time, in *Fuzzy Systems and Operations Research and Management*, (Springer International Publishing, Cham, 2016), pp. 255–266

On the Product and Ratio of Pareto and Maxwell Random Variables

Noura Obeid and Seifedine Kadry

Abstract The distribution of product and ratio of random variables is widely used in many areas of biological and physical sciences, econometric, classification, ranking, and selection and has been extensively studied by many researchers. In this chapter, the analytical distributions of the product XY and ratio X/Y are derived when X and Y are Pareto and Maxwell random variables, respectively, distributed independently of each other.

Keywords Product Distribution · Ratio Distribution · Pareto Distribution · Maxwell Distribution · cumulative distribution function · probability density function · Moment of order r · variance · Survival function · Hazard function.

1 Introduction

Engineering, physics, economics, order statistics, classification, ranking, selection, number theory, genetics, biology, medicine, hydrology, and psychology, all these applied problems depend on the distribution of product and ratio of random variables [15, 25].

As an example of the use of the product of random variables in physics, Sornette [26] mentions that

"...To mimic system size limitation, Takayasu, Sato, and Takayasu introduced a threshold x_c ...and found a stretched exponential truncating the power-law pdf beyond x_c . Frisch and Sornette recently developed a theory of extreme deviations generalizing the central limit theorem which, when applied to multiplication of random variables, predicts the generic presence of stretched exponential pdfs. The

N. Obeid · S. Kadry (✉)
Department of Mathematics and Computer Science, Faculty of Science, Beirut Arab University, Beirut, Lebanon
e-mail: s.kadry@bau.edu.lb

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_26

problem thus boils down to determining the tail of the pdf for a product of random variables . . . "

Several authors have studied that the product distributions for independent random variables come from the same family or different families; see [13] for t and Rayleigh families, [9] for Pareto and Kumaraswamy families, [18] for the t and Bessel families, and [6] for the independent generalized gamma-ratio family.

Examples of the use of the ratio of random variables include Mendelian inheritance ratios in genetics, mass-to-energy ratios in nuclear physics, target-to-control precipitation in meteorology, and inventory ratios in economics.

Several authors have studied that the ratio distributions for independent random variables come from the same family or different families. For the historical review, see [11, 12] for the normal family, [22] for Student's t family, [3] for the Weibull family, [8] for the noncentral chi-squared family, [23] for the gamma family, [20] for the beta family, [16] for the logistic family, [17] for the Frechet family, [1] for the inverted gamma family, [14] for the Laplace family, [21] for the generalized-F family, [27] for the hypoexponential family, [25] for the gamma and Rayleigh families, and [10] for gamma and exponential families.

In this chapter, the analytical probability distributions of XY and X/Y are derived, where X and Y are two independent Pareto and Maxwell distributions, respectively, with probability density functions (pdfs)

$$f_X(x) = \frac{ca^c}{x^{c+1}} \tag{1.1}$$

$$f_Y(y) = \sqrt{\frac{2}{\pi}} \frac{y^2 e^{\frac{-y^2}{2b^2}}}{b^3}, \tag{1.2}$$

respectively, for $a \leq x < \infty$, $a > 0$, $c > 0$, $0 < y < \infty$, and $b > 0$.

2 Distribution of the Product X Y

Theorem 2.1 Suppose X and Y are independent and distributed according to (1.1) and (1.2), respectively. Then, the cumulative distribution function (*cdf*) of $Z = XY$ can be expressed as

$$F_Z(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ 1 - \frac{(ab)^c}{z^c}\sqrt{\frac{2}{\pi}} 2^{\frac{c+1}{2}} \Gamma(\frac{c+3}{2}) & \\ -\sqrt{\frac{2}{\pi}} \frac{z e^{-z^2/(2a^2b^2)}}{ab} - \frac{1}{\sqrt{\pi}} \Gamma(1/2, z^2/(2a^2b^2)) & \text{if } z > 0. \\ + 2^{\frac{c+1}{2}} \frac{(ab)^c}{z^c}\sqrt{\frac{2}{\pi}} \Gamma(\frac{c+3}{2}, z^2/(2a^2b^2)) & \end{cases} \tag{2.1}$$

Proof The cumulative distribution function of X (1.1) is $F_X(x) = 1-(\frac{a}{x})^c$. Thence, the cumulative distribution function (cdf) of XY can be written as

$$F_Z(z) = \int_0^{z/a} (1 - \left(\frac{ay}{z}\right)^c) f_Y(y))dy. \tag{2.2}$$

We can write $F_Z(z)$ as

$$\begin{aligned} F_Z(z) = &\int_0^{+\infty} \left(1 - \left(\frac{ay}{z}\right)^c\right) f_Y(y)dy \\ &- \int_{z/a}^{+\infty} \left(1 - \left(\frac{ay}{z}\right)^c\right) f_Y(y)dy. \end{aligned} \tag{2.3}$$

Let $I_1 = \int_0^{+\infty} (1 - (\frac{ay}{z})^c) f_Y(y)$, and $I_2 = \int_{z/a}^{+\infty} (1 - (\frac{ay}{z}))^c f_Y(y)dy$

$$F_Z(z) = I_1 - I_2$$

Calculus of I_1

$$\begin{aligned} I_1 &= \int_0^{+\infty} \left(1 - \left(\frac{ay}{z}\right)^c\right) f_Y(y) \\ &= \int_0^{+\infty} f_Y(y)dy - \int_0^{+\infty} \frac{a^c y^c}{z^c} f_Y(y)dy \\ &= 1 - \frac{a^c}{z^c b^3}\sqrt{\frac{2}{\pi}} \int_0^{+\infty} y^{c+2} e^{-y^2/(2b^2)} dy. \end{aligned}$$

Let $u = \frac{y^2}{2b^2}$. Then, we get

$$\begin{aligned} I_1 &= 1 - \frac{a^c}{z^c b^3}\sqrt{\frac{2}{\pi}} 2^{\frac{c+1}{2}} b^{c+3} \int_0^{+\infty} u^{\frac{c+1}{2}} e^{-u} du \\ &= 1 - \frac{a^c b^c}{z^c}\sqrt{\frac{2}{\pi}} 2^{\frac{c+1}{2}} \Gamma(\frac{c+3}{2}). \end{aligned}$$

Calculus of I_2

$$\begin{aligned} I_2 &= \int_{z/a}^{+\infty} \left(1 - \left(\frac{ay}{z}\right)^c\right) f_Y(y)dy \\ &= \int_{z/a}^{+\infty} f_Y(y)dy - \int_{z/a}^{+\infty} \frac{(ay)^c}{z^c} f_Y(y)dy. \end{aligned}$$

Let $I = \int_{z/a}^{+\infty} f_Y(y)dy$ and $I_4 = \int_{z/a}^{+\infty} \frac{(ay)^c}{z^c} f_Y(y)dy$

$$I_2 = I - I_4.$$

Substituting $u = \frac{y^2}{2b^2}$ and using $\Gamma(1/2, x) = \sqrt{\pi}\,\mathrm{erfc}(\sqrt{x})$ for $x > 0$ in I, we get

$$\begin{aligned} I &= \sqrt{\frac{2}{\pi}} \frac{1}{b^3} \int_{z/a}^{+\infty} y^2 e^{-y^2/(2b^2)} dy \\ &= \frac{2}{\sqrt{\pi}} \int_{z^2/(2a^2b^2)}^{\infty} u^{1/2} e^{-u} du \\ &= \frac{2}{\sqrt{\pi}} \left[\frac{z e^{\frac{-z^2}{2a^2b^2}}}{ab\sqrt{2}} + \frac{\sqrt{\pi}}{2} \mathrm{erfc}\left(\frac{z}{\sqrt{2}ab} \right) \right] \\ &= \sqrt{\frac{2}{\pi}} \frac{z e^{\frac{-z^2}{2a^2b^2}}}{ab} + \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}, \frac{z^2}{2a^2b^2} \right). \end{aligned}$$

If we substitute $u = \frac{y^2}{2b^2}$ in I_4, we get

$$I_4 = \frac{a^c b^c}{z^c} 2^{\frac{c+1}{2}} \sqrt{\frac{2}{\pi}} \int_{z^2/(2a^2b^2)}^{\infty} u^{\frac{c+1}{2}} e^{-u} du = 2^{\frac{c+1}{2}} \frac{a^c b^c}{z^c} \sqrt{\frac{2}{\pi}} \Gamma\left(\frac{c+3}{2}, \frac{z^2}{2a^2b^2} \right).$$

Then,

$$I_2 = \sqrt{\frac{2}{\pi}} \frac{z e^{\frac{-z^2}{2a^2b^2}}}{ab} + \frac{1}{\sqrt{\pi}} \Gamma(\frac{1}{2}, \frac{z^2}{2a^2b^2}) - 2^{\frac{c+1}{2}} \frac{a^c b^c}{z^c} \sqrt{\frac{2}{\pi}} \Gamma\left(\frac{c+3}{2}, \frac{z^2}{2a^2b^2} \right).$$

And finally, we obtain for $z > 0$

$$\begin{aligned} F_Z(z) &= I_1 - I_2 \\ &= 1 - \frac{(ab)^c}{z^c} \sqrt{\frac{2}{\pi}} 2^{\frac{c+1}{2}} \Gamma\left(\frac{c+3}{2} \right) \\ &\quad - \sqrt{\frac{2}{\pi}} \frac{z e^{-z^2/(2a^2b^2)}}{ab} - \frac{1}{\sqrt{\pi}} \Gamma\left(1/2, z^2/(2a^2b^2) \right) \\ &\quad + 2^{\frac{c+1}{2}} \frac{(ab)^c}{z^c} \sqrt{\frac{2}{\pi}} \Gamma\left(\frac{c+3}{2}, z^2/(2a^2b^2) \right). \end{aligned}$$

□

Corollary 2.2 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the probability density function of $Z = XY$, when $z > 0$, can be written as

$$f_Z(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \frac{c2^{\frac{c+1}{2}}(ab)^c}{z^{c+1}}\sqrt{\frac{2}{\pi}}\left[\Gamma(\frac{c+3}{2}) - \Gamma(\frac{c+3}{2}, z^2/2(ab)^2)\right] & \text{if } z > 0. \end{cases} \tag{2.4}$$

Proof The probability density function $f_Z(z)$ follows by differentiation using

$$\frac{d}{dz}\left[ze^{\frac{-z^2}{2a^2b^2}}\right] = e^{\frac{-z^2}{2a^2b^2}}\left(1 - \frac{z^2}{a^2b^2}\right)$$

$$\frac{d}{dz}\left[\Gamma\left(1/2, \frac{z^2}{2a^2b^2}\right)\right] = \frac{-\sqrt{2}e^{\frac{-z^2}{2a^2b^2}}}{ab}$$

$$\frac{d}{dz}\left[\Gamma\left(\frac{c+3}{2}, \frac{z^2}{2a^2b^2}\right)\right] = \frac{-z^{c+2}e^{\frac{-z^2}{2a^2b^2}}}{2^{\frac{c+1}{2}}(ab)^{c+3}}$$

$$\frac{d}{dz}\left[\frac{\Gamma\left(\frac{c+3}{2}, \frac{z^2}{2a^2b^2}\right)}{z^c}\right] = \frac{-z^2e^{\frac{-z^2}{2a^2b^2}}}{2^{\frac{c+1}{2}}(ab)^{c+3}} - \frac{c\Gamma\left(\frac{c+3}{2}, \frac{z^2}{2a^2b^2}\right)}{z^{c+1}}.$$

□

Corollary 2.3 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the moment of order r of $Z = XY$, when $c > r$, can be written as

$$\begin{aligned} E[Z^r] &= c(ab)^c 2^{\frac{c+1}{2}}\sqrt{\frac{2}{\pi}}\frac{\alpha^{r-c}}{(c-r)}\Gamma\left(\frac{c+3}{2}\right) \\ &\quad - c(ab)^c\alpha^{r-c}\sqrt{\frac{2}{\pi}}\frac{2^{\frac{c+1}{2}}}{(c-r)}\Gamma\left(\frac{c+3}{2}, \frac{\alpha^2}{2a^2b^2}\right) \\ &\quad - c(ab)^r\sqrt{\frac{2}{\pi}}\frac{2^{\frac{r+1}{2}}}{(r-c)}\Gamma\left(\frac{r+3}{2}, \frac{\alpha^2}{2a^2b^2}\right). \end{aligned}$$

Proof

$$
\begin{aligned}
E[Z^r] &= \int_0^{+\infty} z^r f_Z(z) \\
&= \int_\alpha^{+\infty} \frac{c2^{\frac{c+1}{2}}(ab)^c}{z^{c+1-r}}\sqrt{\frac{2}{\pi}}\Gamma\left(\frac{c+3}{2}\right)dz \\
&\quad - c2^{\frac{c+1}{2}}(ab)^c\sqrt{\frac{2}{\pi}}\int_\alpha^{+\infty}\frac{\Gamma\left(\frac{c+3}{2},\frac{z^2}{2a^2b^2}\right)}{z^{c+1-r}}dz \\
&= c2^{\frac{c+1}{2}}(ab)^c\sqrt{\frac{2}{\pi}}\frac{\alpha^{r-c}}{(c-r)}\Gamma\left(\frac{c+3}{2}\right) \\
&\quad - c2^{\frac{c+1}{2}}(ab)^c\sqrt{\frac{2}{\pi}}\int_\alpha^{+\infty}\frac{\Gamma\left(\frac{c+3}{2},\frac{z^2}{2a^2b^2}\right)}{z^{c+1-r}}dz.
\end{aligned}
$$

Let

$$
I' = \int_\alpha^{+\infty}\frac{\Gamma\left(\frac{c+3}{2},\frac{z^2}{2a^2b^2}\right)}{z^{c+1-r}}dz,
$$

and substituting $u = \frac{z^2}{2a^2b^2}$ in I', we obtain

$$
I' = \frac{1}{2^{\frac{c+2-r}{2}}(ab)^{c-r}}\int_{\frac{\alpha^2}{2a^2b^2}}^{\infty}\frac{\Gamma(\frac{c+3}{2},u)}{u^{\frac{c+2-r}{2}}}du.
$$

Integration by parts implies

$$
\begin{aligned}
I' = \frac{1}{2^{\frac{c+2-r}{2}}(ab)^{c-r}}\Bigg[&\frac{2^{1-\frac{r-c}{2}}\alpha^{r-c}}{(ab)^{r-c}(c-r)}\Gamma\left(\frac{c+3}{2},\frac{\alpha^2}{2a^2b^2}\right) \\
&+\frac{2}{(r-c)}\Gamma\left(\frac{r+1}{2}+1,\frac{\alpha^2}{2a^2b^2}\right)\Bigg].
\end{aligned}
$$

And finally, we obtain

$$
\begin{aligned}
E[Z^r] &= c(ab)^c 2^{\frac{c+1}{2}}\sqrt{\frac{2}{\pi}}\frac{\alpha^{r-c}}{(c-r)}\Gamma\left(\frac{c+3}{2}\right) \\
&\quad - c(ab)^c\alpha^{r-c}\sqrt{\frac{2}{\pi}}\frac{2^{\frac{c+1}{2}}}{(c-r)}\Gamma\left(\frac{c+3}{2},\frac{\alpha^2}{2a^2b^2}\right) \\
&\quad - c(ab)^r\sqrt{\frac{2}{\pi}}\frac{2^{\frac{r+1}{2}}}{(r-c)}\Gamma\left(\frac{r+3}{2},\frac{\alpha^2}{2a^2b^2}\right).
\end{aligned}
$$

□

Corollary 2.4 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the expected value of $Z = XY$, when $c > 1$ and $r = 1$, can be written as

$$\begin{aligned} E[Z] = c(ab)^c 2^{\frac{c+1}{2}} \sqrt{\frac{2}{\pi}} \frac{\alpha^{1-c}}{(c-1)} \Gamma\left(\frac{c+3}{2}\right) \\ - c(ab)^c \alpha^{1-c} \sqrt{\frac{2}{\pi}} \frac{2^{\frac{c+1}{2}}}{(c-1)} \Gamma\left(\frac{c+3}{2}, \frac{\alpha^2}{2a^2b^2}\right) \\ - c(ab) \sqrt{\frac{2}{\pi}} \frac{2}{(1-c)} \Gamma\left(2, \frac{\alpha^2}{2a^2b^2}\right). \end{aligned}$$

Corollary 2.5 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the variance of $Z = XY$, when $c > 2$, can be written as

$$\begin{aligned} \sigma^2 = c(ab)^c 2^{\frac{c+1}{2}} \sqrt{\frac{2}{\pi}} \frac{\alpha^{2-c}}{(c-2)} \Gamma\left(\frac{c+3}{2}\right) \\ - c(ab)^c \alpha^{2-c} \sqrt{\frac{2}{\pi}} \frac{2^{\frac{c+1}{2}}}{(c-2)} \Gamma\left(\frac{c+3}{2}, \frac{\alpha^2}{2a^2b^2}\right) \\ - c(ab)^2 \sqrt{\frac{2}{\pi}} \frac{2^{\frac{3}{2}}}{(2-c)} \Gamma\left(\frac{5}{2}, \frac{\alpha^2}{2a^2b^2}\right) \\ - \left[c(ab)^c 2^{\frac{c+1}{2}} \sqrt{\frac{2}{\pi}} \frac{\alpha^{1-c}}{(c-1)} \Gamma\left(\frac{c+3}{2}\right) \right. \\ - c(ab)^c \alpha^{1-c} \sqrt{\frac{2}{\pi}} \frac{2^{\frac{c+1}{2}}}{(c-1)} \Gamma\left(\frac{c+3}{2}, \frac{\alpha^2}{2a^2b^2}\right) \\ \left. - c(ab) \sqrt{\frac{2}{\pi}} \frac{2}{(1-c)} \Gamma\left(2, \frac{\alpha^2}{2a^2b^2}\right) \right]^2. \end{aligned}$$

Proof By definition of variance,

$$\sigma^2 = E[Z^2] - (E[Z])^2.$$

□

Corollary 2.6 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the survival function of $Z = XY$ can be written as

$$S_Z(z) = \begin{cases} 1 & \text{if } z \le 0 \\ \dfrac{(ab)^c}{z^c}\sqrt{\dfrac{2}{\pi}}2^{\frac{c+1}{2}}\Gamma\left(\dfrac{c+3}{2}\right) & \\ +\sqrt{\dfrac{2}{\pi}}\dfrac{ze^{-z^2/(2a^2b^2)}}{ab} + \dfrac{1}{\sqrt{\pi}}\Gamma\left(1/2, z^2/(2a^2b^2)\right) & \text{if } z > 0. \\ -2^{\frac{c+1}{2}}\dfrac{(ab)^c}{z^c}\sqrt{\dfrac{2}{\pi}}\Gamma\left(\dfrac{c+3}{2}, z^2/(2a^2b^2)\right) & \end{cases} \tag{2.5}$$

Corollary 2.7 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the Hazard function of $Z = XY$ for $z > 0$ can be written as

$$h_Z(z) = \begin{cases} 0 & \text{if } z \le 0 \\ \dfrac{\frac{c2^{\frac{c+1}{2}}(ab)^c}{z^{c+1}}\sqrt{\frac{2}{\pi}}\left[\Gamma\left(\frac{c+3}{2}\right) - \Gamma\left(\frac{c+3}{2}, \frac{z^2}{2(ab)^2}\right)\right]}{\frac{(ab)^c}{z^c}\sqrt{\frac{2}{\pi}}2^{\frac{c+1}{2}}\Gamma\left(\frac{c+3}{2}\right) + \sqrt{\frac{2}{\pi}}\frac{ze^{-\frac{z^2}{2(ab)^2}}}{ab} + \frac{1}{\sqrt{\pi}}\Gamma\left(1/2, \frac{z^2}{2(ab)^2}\right) - 2^{\frac{c+1}{2}}\frac{(ab)^c}{z^c}\sqrt{\frac{2}{\pi}}\Gamma\left(\frac{c+3}{2}, \frac{z^2}{2(ab)^2}\right)} & \text{if } z > 0. \end{cases} \tag{2.6}$$

3 Distribution of the Ratio X/Y

Theorem 3.1 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the cumulative distribution function of $Z = X/Y$ can be written as (Fig. 1)

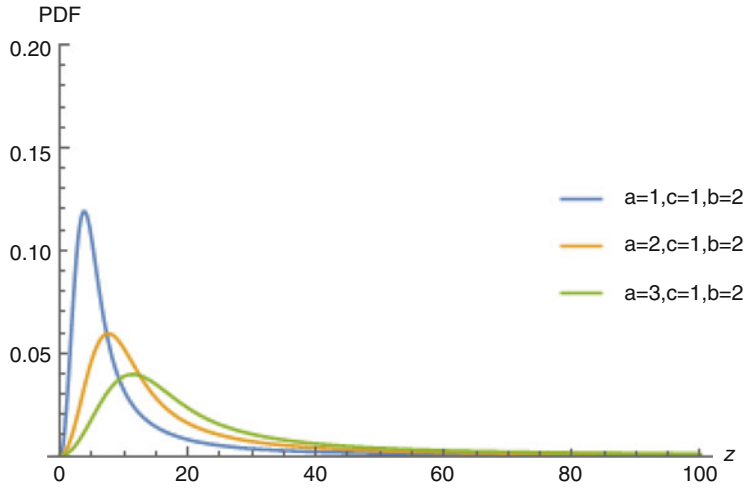


Fig. 1 Plots of the pdf (2.4) for $a = 1, 2, 3$, $c = 1$, and $b = 2$

$$F_Z(z) = \begin{cases} 0 & \text{if } z \le 0 \\ \frac{\sqrt{2}a}{\sqrt{\pi}bz} e^{-\frac{a^2}{2b^2z^2}} + \frac{1}{\sqrt{\pi}}\Gamma\left(1/2, \frac{a^2}{2b^2z^2}\right) - \frac{a^c}{z^c b^c \sqrt{\pi} 2^{c/2-1}}\Gamma\left(\frac{3-c}{2}, \frac{a^2}{2b^2z^2}\right) & \text{if } z > 0, c < 3 \\ \frac{\sqrt{2}a}{\sqrt{\pi}bz} e^{-\frac{a^2}{2b^2z^2}} + \frac{1}{\sqrt{\pi}}\Gamma\left(1/2, \frac{a^2}{2b^2z^2}\right) - \frac{a^3}{z^3 b^3 \sqrt{2\pi}} E_{\frac{c-1}{2}}\left(\frac{a^2}{2b^2z^2}\right) & \text{if } z > 0, (\frac{3}{2} - \frac{c}{2}) \in \mathbb{Z}^-. \end{cases} \tag{3.1}$$

Proof The cumulative distribution function of X (1.1) is $F_X(x) = 1-(\frac{a}{x})^c$. Thence, the cumulative distribution function (cdf) of X/Y can be written as

$$\begin{aligned} F_Z(z) &= Pr(\frac{X}{Y} \le z) \\ &= \int_0^\infty F_X(zy) f_Y(y) dy \\ &= \int_{a/z}^\infty \left[1 - \left(\frac{a}{yz}\right)^c\right] f_Y(y) dy \\ &= \int_{a/z}^\infty f_Y(y) dy - \int_{a/z}^\infty \frac{a^c}{y^c z^c} f_Y(y) dy. \end{aligned} \tag{3.2}$$

Let $I_1 = \int_{a/z}^\infty f_Y(y)dy$, and $I_2 = \int_{a/z}^\infty \frac{a^c}{y^c z^c} f_Y(y)dy$

$$F_Z(z) = I_1 - I_2.$$

Calculus of I_1

$$I_1 = \int_{a/z}^\infty \sqrt{\frac{2}{\pi}} \frac{y^2}{b^3} e^{-\frac{y^2}{2b^2}} dy.$$

Let $u = \frac{y^2}{2b^2}$, then

$$\begin{aligned} I_1 &= \frac{2}{\sqrt{\pi}} \int_{\frac{a^2}{2z^2b^2}}^\infty \sqrt{u} e^{-u} du \\ &= \frac{2}{\sqrt{\pi}} \left[\frac{a e^{-\frac{a^2}{2b^2z^2}}}{bz\sqrt{2}} + \frac{\sqrt{\pi}}{2} \operatorname{erfc}\left(\sqrt{\frac{a^2}{2b^2z^2}}\right)\right] \\ &= \frac{2}{\sqrt{\pi}} \left[\frac{a e^{-\frac{a^2}{2b^2z^2}}}{bz\sqrt{2}} + \frac{1}{2}\Gamma\left(1/2, \frac{a^2}{2b^2z^2}\right)\right] \\ &= \frac{\sqrt{2}a e^{-\frac{a^2}{2b^2z^2}}}{\sqrt{\pi}bz} + \frac{1}{\sqrt{\pi}}\Gamma\left(1/2, \frac{a^2}{2b^2z^2}\right). \end{aligned}$$

Calculus of I_2

$$\begin{aligned} I_2 &= \int_{a/z}^{\infty} \frac{a^c}{y^c z^c} f_Y(y) dy \\ &= \int_{a/z}^{\infty} \frac{a^c}{(yz)^c} \sqrt{\frac{2}{\pi}} \frac{y^2}{b^3} e^{-\frac{y^2}{2b^2}} dy \\ &= \frac{a^c}{z^c b^3} \sqrt{\frac{2}{\pi}} \int_{a/z}^{\infty} \frac{e^{-\frac{y^2}{2b^2}}}{y^{c-2}} dy. \end{aligned}$$

Substituting $u = \frac{y^2}{2b^2}$, we obtain

$$\begin{aligned} I_2 &= \frac{a^c}{z^c b^c \sqrt{\pi} 2^{c/2-1}} \int_{\frac{a^2}{2b^2 z^2}}^{\infty} u^{-c/2+1/2} e^{-u} du \\ &= \frac{a^c}{z^c b^c \sqrt{\pi} 2^{c/2-1}} \Gamma\left(\frac{3-c}{2}, \frac{a^2}{2b^2 z^2}\right). \end{aligned}$$

Finally, for $z > 0$, we get

$$\begin{aligned} F_Z(z) &= I_1 - I_2 \\ &= \frac{\sqrt{2}a}{\sqrt{\pi} bz} e^{-\frac{a^2}{2b^2 z^2}} + \frac{1}{\sqrt{\pi}} \Gamma\left(1/2, \frac{a^2}{2b^2 z^2}\right) \\ &\quad - \frac{a^c}{z^c b^c \sqrt{\pi} 2^{c/2-1}} \Gamma\left(\frac{3-c}{2}, \frac{a^2}{2b^2 z^2}\right). \end{aligned}$$

□

For $(\frac{3}{2} - \frac{c}{2}) \in \mathbb{Z}^-$, using **Lemma 4**, we have

$$\Gamma(\frac{3}{2} - \frac{c}{2}, a^2/2z^2b^2) = \left[\frac{a^2}{2z^2b^2}\right]^{\frac{3}{2}-\frac{c}{2}} E_{\frac{c-1}{2}}\left(\frac{a^2}{2z^2b^2}\right),$$

and

$$F_Z(z) = \frac{\sqrt{2}a}{\sqrt{\pi} bz} e^{-\frac{a^2}{2b^2 z^2}} + \frac{1}{\sqrt{\pi}} \Gamma\left(1/2, \frac{a^2}{2b^2 z^2}\right) - \frac{a^3}{z^3 b^3 \sqrt{2\pi}} E_{\frac{c-1}{2}}\left(\frac{a^2}{2b^2 z^2}\right). \tag{3.3}$$

Corollary 3.2 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the probability density function of $Z = X/Y$, for $c < 3$, can be written as

$$f_Z(z)=\begin{cases}0 & \text{if } z\le 0\\ \frac{a^3e^{-\frac{a^2}{2b^2z^2}}}{\sqrt{\pi}b^3z^4}(\sqrt{2}-1)+\frac{ca^c\Gamma(\frac{3-c}{2},\frac{a^2}{2b^2z^2})}{b^c\sqrt{\pi}2^{\frac{c}{2}-1}z^{c+1}} & \text{if } z>0, c<3\\ \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3z^4}e^{-\frac{a^2}{2b^2z^2}}-\frac{a^5}{b^5z^6\sqrt{2\pi}}E_{\frac{c-3}{2}}(\frac{a^2}{2b^2z^2})+\frac{3a^3}{b^3\sqrt{2\pi}z^4}E_{\frac{c-1}{2}}(\frac{a^2}{2b^2z^2}) & \text{if } z>0, (\frac{3}{2}-\frac{c}{2})\in\mathbb{Z}^-.\end{cases} \tag{3.4}$$

Proof The probability density function $f_Z(z)$ in (3.5) follows by differentiation using

$$\frac{d}{dz}\left(\frac{e^{-\frac{a^2}{2b^2z^2}}}{z}\right)=\frac{a^2e^{-\frac{a^2}{2b^2z^2}}}{b^2z^4}-\frac{e^{-\frac{a^2}{2b^2z^2}}}{z^2}$$

$$\frac{d}{dz}\left(\Gamma(1/2,\frac{a^2}{2b^2z^2})\right)=\frac{a\sqrt{2}e^{-\frac{a^2}{2b^2z^2}}}{bz^2}$$

$$\frac{d}{dz}\left(\frac{\Gamma(\frac{3-c}{2},\frac{a^2}{2b^2z^2})}{z^c}\right)=\frac{a^{3-c}e^{-\frac{a^2}{2b^2z^2}}}{2^{1-\frac{c}{2}}b^{3-c}z^4}-\frac{c\Gamma(\frac{3-c}{2},\frac{a^2}{2b^2z^2})}{z^{c+1}}$$

$$\frac{d}{dz}\left(\frac{E_{\frac{c-1}{2}}(\frac{a^2}{2b^2z^2})}{z^3}\right)=E_{\frac{c-3}{2}}(\frac{a^2}{2b^2z^2})\frac{a^2}{b^2z^6}-\frac{3}{z^4}E_{\frac{c-1}{2}}(\frac{a^2}{2b^2z^2}).$$

□

Corollary 3.3 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the moment of order r of $Z=X/Y$, for $\alpha>0$, can be written as

$$E[Z^r]=\begin{cases}\left[\frac{(\sqrt{2}-1)a^r}{\sqrt{\pi}b^r2^{\frac{r-1}{2}}}+\frac{a^{r-c}}{(r-c)2^{\frac{r-c}{2}}b^{r-c}}\right]\left[\Gamma(\frac{3-r}{2})-\Gamma(\frac{3-r}{2},\frac{a^2}{2b^2\alpha^2})\right]\\ +\frac{ca^c\alpha^{r-c}}{b^c\sqrt{\pi}2^{\frac{c}{2}-1}(c-r)}\Gamma(\frac{3-c}{2},\frac{a^2}{2b^2\alpha^2}) & \text{if } c<3\\ \sqrt{\frac{2}{\pi}}\frac{a^r}{b^r2^{\frac{r-1}{2}}}\left[\Gamma(\frac{3-r}{2})-\Gamma(\frac{3-r}{2},\frac{a^2}{2b^2\alpha^2})\right]\\ -\frac{a^5}{b^5\sqrt{2\pi}}\left[\sum_{i=0}^{\frac{c-3}{2}}(\frac{a^2}{b^2})^i\frac{E_{\frac{c-3}{2}-i}(\frac{a^2}{2b^2\alpha^2})\alpha^{-(2i+5)+r}}{\prod_{j=2}^{i+2}\left[(2j+1)-r\right]}\right] & \text{if } (\frac{3}{2}-\frac{c}{2})\in\mathbb{Z}^-, r<3.\\ +\frac{3a^3}{b^3\sqrt{2\pi}}\left[\sum_{i=0}^{\frac{c-1}{2}}(\frac{a^2}{b^2})^i\frac{E_{\frac{c-1}{2}-i}(\frac{a^2}{2b^2\alpha^2})\alpha^{-(2i+3)+r}}{\prod_{j=1}^{i+1}\left[(2j+1)-r\right]}\right]\end{cases} \tag{3.5}$$

Proof For $c < 3$,

$$\begin{aligned} E[Z^r] &= \int_{-\infty}^{+\infty} z^r f_Z(z)dz \\ &= \int_{\alpha}^{+\infty} \frac{a^3(\sqrt{2}-1)}{\sqrt{\pi}b^3} z^{r-4} e^{-\frac{a^2}{2b^2z^2}} dz \\ &\quad + \int_{\alpha}^{+\infty} \frac{ca^c z^{r-c-1}}{b^c\sqrt{\pi}2^{c/2-1}} \Gamma(\frac{3-c}{2}, \frac{a^2}{2b^2z^2})dz. \end{aligned} \tag{3.6}$$

Let $I_1 = \int_{\alpha}^{+\infty} \frac{a^3(\sqrt{2}-1)}{\sqrt{\pi}b^3} z^{r-4} e^{-\frac{a^2}{2b^2z^2}} dz$ and $I_2 = \int_{\alpha}^{+\infty} \frac{ca^c z^{r-c-1}}{b^c\sqrt{\pi}2^{c/2-1}} \Gamma(\frac{3-c}{2}, \frac{a^2}{2b^2z^2})dz$

$$E[Z^r] = I_1 + I_2.$$

Calculus of I_1
Let $u = \frac{a^2}{2b^2z^2}$, we obtain

$$\begin{aligned} I_1 &= \int_{\alpha}^{+\infty} \frac{a^3(\sqrt{2}-1)}{\sqrt{\pi}b^3} z^{r-4} e^{-\frac{a^2}{2b^2z^2}} dz \\ &= \frac{(\sqrt{2}-1)a^r}{\sqrt{\pi}b^r 2^{\frac{r-1}{2}}} \int_0^{\frac{a^2}{2b^2\alpha^2}} u^{2-\frac{r+3}{2}} e^{-u} du \\ &= \frac{(\sqrt{2}-1)a^r}{\sqrt{\pi}b^r 2^{\frac{r-1}{2}}} \left[\Gamma(\frac{3-r}{2}) - \Gamma(\frac{3-r}{2}, \frac{a^2}{2b^2\alpha^2}) \right]. \end{aligned}$$

Calculus of I_2
Integration by parts implies

$$\begin{aligned} I_2 &= \int_{\alpha}^{+\infty} \frac{ca^c z^{r-c-1}}{b^c\sqrt{\pi}2^{c/2-1}} \Gamma\left(\frac{3-c}{2}, \frac{a^2}{2b^2z^2}\right) dz \\ &= -\frac{ca^c}{b^c\sqrt{\pi}2^{c/2-1}} \frac{\alpha^{r-c}}{(r-c)} \Gamma\left(\frac{3-c}{2}, \frac{a^2}{2b^2\alpha^2}\right) \\ &\quad - \frac{a^{3-c}}{(r-c)2^{\frac{-c+1}{2}}b^{-c+3}} \int_{\alpha}^{\infty} z^{r-4} e^{-a^2/2b^2z^2} dz. \end{aligned}$$

Substituting $u = \frac{a^2}{2b^2z^2}$ in the integral above, we get

$$I_2 = \frac{ca^c}{b^c\sqrt{\pi}2^{c/2-1}}\frac{\alpha^{r-c}}{(c-r)}\Gamma\left(\frac{3-c}{2}, \frac{a^2}{2b^2\alpha^2}\right) - \frac{a^{r-c}}{(r-c)2^{\frac{r-c}{2}}b^{r-c}}\left[\Gamma\left(\frac{3-r}{2}, \frac{a^2}{2b^2\alpha^2}\right) + \Gamma\left(\frac{3-r}{2}\right)\right].$$

And finally, we get $E[Z^r] = I_1 + I_2$.
For $(\frac{3}{2} - \frac{c}{2}) \in \mathbb{Z}^-$,

$$E[Z^r] = \int_\alpha^\infty z^r f_Z(z)dz = \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3}\int_\alpha^\infty z^{-4+r}e^{-\frac{a^2}{2b^2z^2}}dz - \frac{a^5}{b^5\sqrt{2\pi}}\int_\alpha^\infty z^{-6+4}E_{\frac{c-3}{2}}\left(\frac{a^2}{2b^2z^2}\right)dz + \frac{3a^3}{b^3\sqrt{2\pi}}\int_\alpha^\infty z^{-4+r}E_{\frac{c-1}{2}}\left(\frac{a^2}{2b^2z^2}\right)dz.$$

Let $I_1 = \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3}\int_\alpha^\infty z^{-4+r}e^{-\frac{a^2}{2b^2z^2}}dz$, $I_2 = \frac{a^5}{b^5\sqrt{2\pi}}\int_\alpha^\infty z^{-6+r}E_{\frac{c-3}{2}}\left(\frac{a^2}{2b^2z^2}\right)dz$,

and $I_3 = \frac{3a^3}{b^3\sqrt{2\pi}}\int_\alpha^\infty z^{-4+r}E_{\frac{c-1}{2}}\left(\frac{a^2}{2b^2z^2}\right)dz$. Then,

$$E[Z^r] = I_1 - I_2 + I_3.$$

Calculus of I_1

$$I_1 = \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3}\int_\alpha^\infty z^{-4+r}e^{-\frac{a^2}{2b^2z^2}}dz.$$

Substituting $u = \frac{a^2}{2b^2z^2}$, we get

$$I_1 = \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3}\int_\alpha^\infty z^{-4+r}e^{-\frac{a^2}{2b^2z^2}}dz = \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3}\left[\frac{a^{r-3}}{b^{r-3}2^{\frac{r-1}{2}}}\int_0^{\frac{a^2}{2b^2\alpha^2}} u^{\frac{1-r}{2}}e^{-u}du\right] = \sqrt{\frac{2}{\pi}}\frac{a^3}{b^3}\left[\frac{a^{r-3}}{b^{r-3}2^{\frac{r-1}{2}}}\left[\Gamma\left(\frac{3-r}{2}\right) - \Gamma\left(\frac{3-r}{2}, \frac{a^2}{2b^2\alpha^2}\right)\right]\right] = \sqrt{\frac{2}{\pi}}\frac{a^r}{b^r2^{\frac{r-1}{2}}}\left[\Gamma\left(\frac{3-r}{2}\right) - \Gamma\left(\frac{3-r}{2}, \frac{a^2}{2b^2\alpha^2}\right)\right].$$

Calculus of I_2

$$I_2 = \frac{a^5}{b^5\sqrt{2\pi}} \int_\alpha^\infty z^{-6+r} E_{\frac{c-3}{2}} \left(\frac{a^2}{2b^2z^2}\right) dz.$$

Integration by parts implies

$$I_2 = \frac{a^5}{b^5\sqrt{2\pi}} \left[\frac{\alpha^{-5+r}}{5-r} E_{\frac{c-3}{2}} \left(\frac{a^2}{2b^2\alpha^2}\right) + \int_\alpha^\infty \frac{z^{-5+r}}{5-r} E_{\frac{c-3}{2}-1} \left(\frac{a^2}{2b^2z^2}\right) \frac{a^2}{b^2z^3} dz\right].$$

Integration by parts of the integral above implies

$$\begin{aligned} I_2 = \frac{a^5}{b^5\sqrt{2\pi}} \Bigg[& \frac{\alpha^{-5+r}}{5-r} E_{\frac{c-3}{2}} \left(\frac{a^2}{2b^2\alpha^2}\right) \\ & + \frac{a^2\alpha^{-7+r}}{(5-r)(7-r)b^2} E_{\frac{c-3}{2}-1} \left(\frac{a^2}{2b^2\alpha^2}\right) \\ & + \frac{a^2a^2}{(5-r)b^2b^2} \int_\alpha^\infty \frac{z^{-10+r}}{7-r} E_{\frac{c-3}{2}-2} \left(\frac{a^2}{2b^2z^2}\right) dz\Bigg]. \end{aligned}$$

By recurrence, we get I_2, where $E_0(u) = \frac{e^{-u}}{u}$.
Same proof for the calculation of I_3. □

Corollary 3.4 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the survival function of $Z = X/Y$ can be written as

$$S_Z(z) = \begin{cases} 1 & \text{if} z \le 0 \\ 1 - \frac{\sqrt{2}a}{\sqrt{\pi}bz} e^{-\frac{a^2}{2b^2z^2}} - \frac{1}{\sqrt{\pi}}\Gamma(1/2, \frac{a^2}{2b^2z^2}) + \frac{a^c}{z^c b^c \sqrt{\pi} 2^{c/2-1}} \Gamma(\frac{3-c}{2}, \frac{a^2}{2b^2z^2}) & \text{if } z > 0, c < 3 \\ 1 - \frac{\sqrt{2}a}{\sqrt{\pi}bz} e^{-\frac{a^2}{2b^2z^2}} - \frac{1}{\sqrt{\pi}}\Gamma(1/2, \frac{a^2}{2b^2z^2}) + \frac{a^3}{z^3b^3\sqrt{2\pi}} E_{\frac{c-1}{2}}(\frac{a^2}{2b^2z^2}) & \text{if } z > 0, (\frac{3}{2} - \frac{c}{2}) \in \mathbb{Z}^-. \end{cases} \tag{3.7}$$

Proof By definition of the survival function,

$$S_Z(z) = 1 - F_Z(z).$$

□

Corollary 3.5 Assume that X and Y are independent Pareto (1.1) and Maxwell (1.2) random variables, respectively. A representation for the Hazard function of $Z = X/Y$ can be written as (Fig. 2)

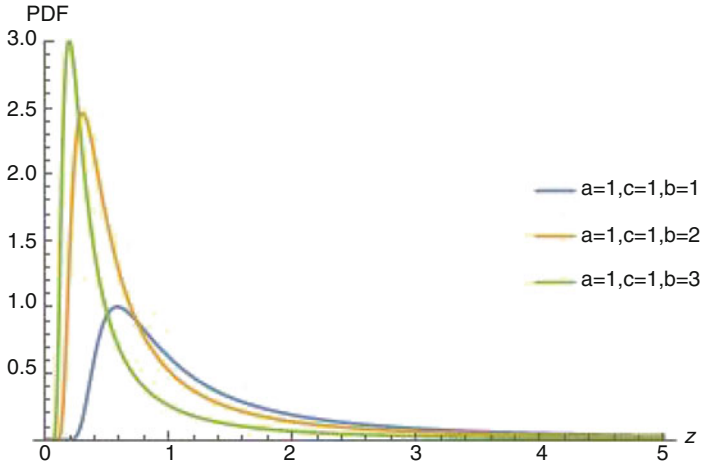


Fig. 2 Plots of the pdf (3.5) for $a = 1$, $c = 1$, and $b = 1, 2, 3$

$$h_Z(z) = \begin{cases} 0 & \text{if} z \le 0 \\ \dfrac{b^c\sqrt{\pi}2^{\frac{c}{2}-1}z^{c+1}a^3e^{-\frac{a^2}{2b^2z^2}}(\sqrt{2}-1)+\sqrt{\pi}b^3z^4ca^c\Gamma(\frac{3-c}{2},\frac{a^2}{2b^2z^2})}{\sqrt{\pi}b^2z^4\left[\sqrt{\pi}b^{c+1}z^{c+1}2^{\frac{c}{2}-1}-b^cz^c2^{\frac{c}{2}-1}(\sqrt{2}ae^{-\frac{a^2}{2b^2z^2}})-b^{c+1}z^{c+1}2^{\frac{c}{2}-1}\Gamma(1/2,\frac{a^2}{2b^2z^2})+zba^c\Gamma(\frac{3-c}{2},\frac{a^2}{2b^2z^2})\right]} & \text{if } z>0, c<3 \\ \dfrac{2b^2a^3z^2e^{-\frac{a^2}{2b^2z^2}}-a^5E_{\frac{c-3}{2}}(\frac{a^2}{2b^2z^2})+3a^3b^2z^2E_{\frac{c-1}{2}}(\frac{a^2}{2b^2z^2})}{z^3b^2\left[\sqrt{2\pi}z^3b^3-2b^2z^2ae^{-\frac{a^2}{2b^2z^2}}-\sqrt{2}z^3b^3\Gamma(1/2,\frac{a^2}{2b^2z^2})+a^3E_{\frac{c-1}{2}}(\frac{a^2}{2b^2z^2})\right]} & \text{if } z>0, (\frac{3}{2}-\frac{c}{2})\in\mathbb{Z}^-. \end{cases} \tag{3.8}$$

Proof By definition of the hazard function,

$$h_Z(z) = \frac{f_Z(z)}{S_Z(z)}. \tag{3.9}$$

□

4 Applications

If x is the random variable describing the amplification of the ith amplifier, then the total amplification $x = x_1x_2 \ldots x_n$ is also a random variable, and it is important to know the distribution of this product. For example, suppose an electric circuit with two amplifiers in series, X_1 is a random variable that follows Pareto distribution with parameter $c = 1$ and $a = 2$, and X_2 is a random variable that follows Maxwell

distribution with parameter $b = 2$, then the total amplification gain is $Z = X_1.X_2$, and by using our result, their pdf is

$$f_Z(z) = \frac{8}{z^2}\sqrt{\frac{2}{\pi}}\left[\Gamma(2) - \Gamma(2, \frac{z^2}{32})\right].$$

Another example involves the distribution of ratio of two independent variables. Let us consider the below PERT network. A PERT chart, sometimes called a PERT diagram, is a project management tool used to schedule, organize, and coordinate tasks within a project. It provides a graphical representation of a project's time line that allows project managers to break down each individual task in the project for analysis (Figs. 3 and 4).

In the above network, we are interesting of the feasibility of starting the series of activities, say A and B, on the same date may be investigated by considering the random variable $Z = \frac{A}{B}$. This idea suggests that through examination of such probabilities as $Pr(\frac{A}{B} > k)$ and $Pr(k^{'} < \frac{A}{B} < k)$, the need for rescheduling A or B may be determined. For instance, if the time to accomplish task A is a random variable that follows Pareto distribution with parameter $c = 1$ and $a = 1$ and task B is a random variable that follows Maxwell distribution with parameter $b = 1$, then by using our result, their pdf is

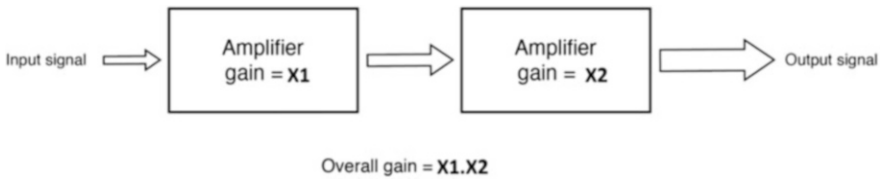


Fig. 3 An electric circuit with two amplifiers in series

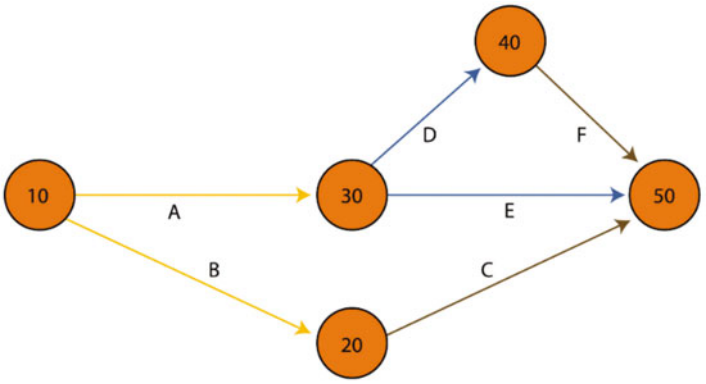


Fig. 4 PERT network

$$f_Z(z) = \frac{(\sqrt{2}-1)e^{-\frac{1}{2z^2}}}{\sqrt{\pi}z^4} + \sqrt{\frac{2}{\pi}}\frac{\Gamma(1,\frac{1}{2z^2})}{z^2}.$$

5 Conclusion

Determining distributions of the functions of random variables is a very crucial task, and this problem has been attracted a number of researchers because there are numerous applications in risk management, finance, economics, science, and many other areas. In this chapter, we have found analytically the expressions of the pdf, the cdf, the moment of order r, the survival function, and the Hazard function, for the product and ratio distributions of Pareto and Maxwell random variables. Our results have been illustrated in two figures of the probability density function (pdf) for the distributions of XY and X/Y.
Finally, we have discussed two examples of engineering applications for the distribution of the product and ratio.

References

1. M.M. Ali, M. Pal, J. Woo, On the ratio of inverted gamma variates. Austrian J. Stat. **36**(2), 153–159 (2007)
2. A. Asgharzadeh, S. Nadarajah, F. Sharafi, Weibull Lindley distributions. Stat. J. **16**(1), 87–113 (2018)
3. A.P. Basu, R.H. Lochner, On the distribution of the ratio of two random variables having generalized life distributions. Technometrics **13**(2), 281–287 (1971)
4. G. Beylkin, L. Monzón, I. Satkauskas, On computing distributions of products of non-negative independent random variables. Appl. Comput. Harmonic Anal. **46**(2), 400–416 (2019)
5. F. Brian, K. Adem, Some results on the gamma function for negative integers. Appl. Math. Inf. Sci. **6**(2), 173–176 (2012)
6. C.A. Coelho, J.T. Mexia, On the distribution of the product and ratio of independent generalized gamma-ratio. Sankhya Ind. J. Stat. **69**(2), 221–255 (2007)
7. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series and Products*, vol. 6 (Academic Press, Cambridge, 2000)
8. D.L. Hawkins, C.-P. Han, Bivariate distributions of some ratios of independent noncentral chi-square random variables. Commun. Stat. Theory Methods **15**(1), 261–277 (1986)
9. L. Idrizi, On the product and ratio of Pareto and Kumaraswamy random variables. Math. Theory Model. **4**(3), 136–146 (2014)
10. L. Joshi, K. Modi, On the distribution of ratio of gamma and three parameter exponentiated exponential random variables. Ind. J. Stat. Appl. **3**(12), 772–783 (2014)
11. P.J. Korhonen, S.C. Narula, The probability distribution of the ratio of the absolute values of two normal variables. J. Stat. Comput. Simul. **33**(3), 173–182 (1989)
12. G. Marsaglia, Ratios of normal variables and ratios of sums of uniform variables. J. Am. Stat. Assoc. **60**(309), 193–204 (1965)
13. K. Modi, L. Joshi, On the distribution of product and ratio of t and Rayleigh random variables. J. Calcutta Math. Soc. **8**(1), 53–60 (2012)

14. S. Nadarajah, The linear combination, product and ratio of Laplace random variables. Statistics **41**(6), 535–545 (2007)
15. S. Nadarajah, D. Choi, Arnold and Strauss's bivariate exponential distribution products and ratios. New Zealand J. Math. **35**, 189–199 (2006)
16. S. Nadarajah, A.K. Gupta, On the ratio of logistic random variables. Comput. Stat. Data Anal. **50**(5), 1206–1219 (2006)
17. S. Nadarajah, S. Kotz, On the ratio of fréchet random variables. Qual. Quant. **40**(5), 861–868 (2006)
18. S. Nadarajah, S. Kotz, On the product and ratio of t and Bessel random variables. Bullet. Inst. Math. Acad. Sin. **2**(1), 55–66 (2007)
19. S. Park, On the distribution functions of ratios involving Gaussian random variables. ETRI J. **32**(6), (2010)
20. T. Pham-Gia, Distributions of the ratios of independent beta variables and applications. Commun. Stat. Theory Methods **29**(12), 2693–2715 (2000)
21. T. Pham-Gia, N. Turkkan, Operations on the generalized-variables and applications. Statistics **36**(3), 195–209 (2002)
22. S.J. Press, The t-ratio distribution. J. Am. Stat. Assoc. **64**(325), 242–252 (1969)
23. S.B. Provost, On the distribution of the ratio of powers of sums of gamma random variables. Pakistan J. Stat. **5**(2), 157–174 (1989)
24. A.P. Prudnikov, Y.A. Brychkov, O.I. Marichev, *Integrals and Series*, vol. 2, no. 3 (Gordon and Breach Science Publishers, Amsterdam, 1986)
25. M. Shakil, B.M.G. Kibria, Exact distribution of the ratio of gamma and Rayleigh random variables. Pakistan J. Stat. Oper. Res. **2**(2), 87–98 (2006)
26. D. Sornette, Multiplicative processes and power laws. Phys. Rev. E **57**, 4811–4813 (1998)
27. K. Therrar, S. Khaled, The exact distribution of the ratio of two independent hypoexponential random variables. Brit. J. Math. Comput. Sci. **4**(18), 2665–2675 (2014)

Performance Analysis of a Discrete-Time Retrial Queue with Bernoulli Feedback, Starting Failure and Single Vacation Policy

Shweta Upadhyaya

Abstract In this study, we consider an unreliable discrete-time Geo/G/1 retrial queue with Bernoulli feedback. During the idle time, server may leave for a vacation of random length according to which at any time instant, when the system becomes empty and no new customer arrives, the server may leave for a vacation of random length and will immediately return from the vacation if at least one customer arrives in the system. The arrival stream is composed of repeated customers, priority customers, and impatient customers. The service time, retrial time, and vacation time are defined by general distribution. The probability generating function (PGF) method and supplementary variable technique (SVT) are employed to derive expressions for system size, orbit size, and other performance measures. A numerical illustration is provided to validate our results with the real-life systems.

Keywords Discrete-time queue · Retrial customers · Priority customers · Impatient customers · Starting failure · Bernoulli feedback · Single vacation · System size

1 Introduction

Discrete-time retrial queueing models are most suitable for the performance evaluation of asynchronous transfer mode (ATM) multiplexers and switches. In these models, the time axis is divided into fixed-length slots and the service of a customer must start and end at slot boundaries. In computer networks, if a packet is lost, the packet may be retransmitted at a later time by a retransmission mechanism such as the TCP (Transmission Control Protocol). In this study, an attempt has been made to provide a remedy for modeling some discrete-time (digital) systems of day-to-

S. Upadhyaya (✉)
Amity Institute of Applied Sciences, Amity University, Noida, UP, India
e-mail: supadhyay@amity.edu

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_27

day life viz. Broadband Integrated Services Digital Network (BISDN) and related computer communication technologies, wherein the models for continuous-time queues fail. As discrete-time systems are more appropriate than their continuous-time equivalent, these have been applied in modeling and solving many congestion problems of real world viz. cellular communication networks in which the service area is divided into cells. Users (mobile stations) in each cell are served by a base station with a limited number of channels. Therefore, only a limited number of users can communicate at the same time.

These category of queues answers many practical problems that arise in these areas and further contribute in the advancement of telecommunication and computer network. In the past years, lot of researchers got interested towards this field and has done worth mentioning work. Atencia and Moreno [1] studied the Geo/G/1 model with discrete parameters and general retrial times and provided the generating functions of the system size as well as the orbit size. Wang and Zhao [2] extensively studied the same model with the condition of starting failures and an optional service. Later on, Jain and Agarwal [3] got motivated from their work and worked on the batch arrival of this model under same conditions. Aboul-Hassan et al. [4] have discussed Geo/G/1 queue with geometric retrial time under discrete scenario and have derived all the interesting performance measures. They have further provided numerical results showcasing the effect of impatience on different parameters. Gao and Liu [5] too examined $\text{Geo}^{X}/G/1$ but their work included the concept of Bernoulli feedback and impatient customers. Moreover, Lan and Tang [6] investigated discrete-time queueing system where server may undergo working breakdowns.

Our study on queueing models with vacation is basically motivated by its abundant applications with the advancement of technology in the area of communication systems and computer networks. The various computers, routers, and switches in such a network may be modeled as individual queues. One of the best real-life examples of different types of vacation is given here. In order to reduce the energy consumption of the mobile cellular network, the base station can be switched off while there is no call in the retrial orbit. During the period that the base station is switched off, the new arrival fresh calls are deposited in the orbit and the base station will seek to serve the calls from the retrial orbit after the channel is switched on. The period when the base station is switched off may be considered as ***vacation*** in queueing terminology. Now a days there is a tremendous increase in application of vacation queues with discrete and retrial phenomenon. Their impact can be found in production or inventory system and cellular network areas. A few impactful work includes that of Gao and Wang [7], Zhang and Zhu [8], and Upadhyaya [9]. Gao and Wang [7] described a batch arrival discrete-time queue with repeated attempts, working vacation, and vacation interruption as well. They have obtained the desired results by applying *supplementary variable method*. Zhang and Zhu [8] combined urgent and normal vacations in a single queueing model and have further explored the relationship between the discrete-time queue and the corresponding continuous-time queue. Upadhyaya [9] used the generating function method and provided performance measures for retrial queue in discrete-time parameter wherein server

can opt at most J number of vacations. Recently, mean queue length of Geo/G/1 discrete retrial system with second optional service and vacation circumstances has been obtained by Gunasekaran and Jeyakumar [10].

This study is motivated by modeling discrete-time Geo/G/1 unreliable retrial queue with priority and impatient customers under Bernoulli feedback and vacation. This work proves to be useful in solving congestion problems in analyzing discrete-time retrial queues with priority in different frameworks. Up to the best of our knowledge, no such type of work has been done on discrete-time retrial queues till now.

The rest of the work done is as follows. Section 2 describes the system by stating requisite assumptions and notations. The system size distribution has been explored in Sect. 3 via probability generating function (PGF) method and supplementary variable technique (SVT). We have explored various useful performance measures of our model in Sect. 4. Numerical results including sensitivity analysis is provided in Sect. 5. Finally, Sect. 6 includes some concluding remarks and future scope of our work for the researchers working in this field.

2 System Description

In this study, we consider a discrete-time retrial queueing system where in server may leave for vacation of random length when the queue becomes empty. Here, ***time*** acts as a discrete random variable (called ***slot***) and arrivals and departure can only occur at boundary epochs of these time slots. Following assumptions have been made to formulate the discrete-time Geo/G/1 retrial queueing model with Bernoulli feedback and starting failure under single vacation policy:

We consider an ***early arrival system*** (EAS) (cf. [11]) according to which departure occurs in the interval (n^-, n) whereas arrivals and retrials occur in the interval (n, n^+), where n^- is the instant immediately before the epoch n and n^+ is the instant immediately after the epoch n. The customers reach the system according to a geometric arrival process with parameter λ $(0 < \lambda < 1)$. During the busy state of the server, the arriving customer has three choices; it enters the orbit with probability $p\overline{\eta} = p\,(1-\eta)$ $(0 < p \leq 1, 0 \leq \eta \leq 1)$ to retry for service (the customer is called repeated customer) or interrupts the customer under service to start his own service with probability $p\eta$ (the customer is called priority customer) or departs from the system with probability $\overline{p} = 1 - p$ without being served (the customer is called impatient customer). During the free time of the server, the customer at the head of the retrial queue initiates its service immediately and the interrupted customer joins the orbit. The interrupted customer may start the service from the beginning. The retrial time begins only after service completion or repair completion or vacation completion. It is assumed that the time between successive retrials follow general distribution $\{\varsigma_i\}_0^\infty$ with generating function $\Phi(x) = \sum_{i=0}^{\infty} \varsigma_i x^i$.

When an arbitrary customer finds that the server is free, it forces the server to activate and start its service immediately with probability "ϑ"; otherwise, when the server is not activated successfully with probability "$\overline{\vartheta} = 1 - \vartheta$" due to some faults, it is sent to repair by the repairman. The server may go for a vacation of random length when the queue becomes empty. During vacation, the server may undergo various maintenance activities such as virus scan, disk cleaning, formatting or simply leaves the system for recreation. The service times, repair times, and vacation times are assumed to be independent and identically distributed with arbitrary distribution $\{g_{1,i}\}_0^{\infty}$, $\{g_{2,i}\}_0^{\infty}$, and $\{v_i\}_0^{\infty}$; generating function $G_1(x) = \sum_{i=0}^{\infty} g_{1,i}x^i$, $G_2(x) = \sum_{i=0}^{\infty} g_{2,i}x^i$, and $W(x) = \sum_{i=0}^{\infty} v_i x^i$ and the rth factorial moments μ_r, γ_r, and θ_r, $r \geq 1$, respectively.

In this study, we have incorporated real-life phenomenon in which after service completion if the customer is not fully satisfied with its service, then it may either join the head of the retrial queue again for another service with probability "ϖ" or depart from the system with complementary probability "$\varpi = 1 - \varpi$," where $0 \leq \varpi < 1$; this queueing situation is called *Bernoulli feedback*. The queue discipline is FCFS (first come first served) and the inter-arrival time, retrial times, service time, repair times, and vacation times are assumed to be mutually independent.

3 Queue Size Distribution

The queueing system under consideration can be described by means of the process $\Gamma_n = (Y_n, \chi_{0,n}, \chi_{1,n}, \chi_{2,n}, \chi_{3,n}, N_n)$ at time epoch n^+ (the instant immediately after time slot n), where Y_n represents the state of the server. $Y_n = 0$, 1, 2, or 3 according to whether the server is idle, busy, breakdown, or under vacation, and N_n represents the number of customers in the retrial queue. If $Y_n = 0$ and $N_n > 0$, $\chi_{0,n}$ represents the remaining retrial time; if $Y_n = 1$, $\chi_{1,n}$ represents the remaining service time; if $Y_n = 2$, $\chi_{2,n}$ represents the remaining repair time; and if $Y_n = 3$, $\chi_{3,n}$ represents the remaining vacation time. Thus $\{X_n, n \geq 0\}$ forms a Markov chain with the following state space:

$$E = \{(0,0)\} \cup \{(j,i,k) : j = 0, 2, i \geq 1, k \geq 1\} \cup \{(j,i,k) : j = 1, 3, i \geq 1, k \geq 0\}$$

We define the following stationary probabilities:

$$\pi_{0,0} = \lim_{n\to\infty} \Pr\{Y_n = 0, N_n = 0\}$$

$$\pi_{0,i,k} = \lim_{n\to\infty} \Pr\{Y_n = 0, \chi_{0,n} = i, N_n = k\}; \qquad i \geq 1, \quad k \geq 1$$

$$\pi_{1,i,k} = \lim_{n\to\infty} \Pr\left\{Y_n = 1, \chi_{1,n} = i, N_n = k\right\}; \qquad i \geq 1, \quad k \geq 0$$

$$\pi_{2,i,k} = \lim_{n\to\infty} \Pr\left\{Y_n = 2, \chi_{2,n} = i, N_n = k\right\}; \qquad i \geq 1, \quad k \geq 1$$

$$\pi_{3,i,k} = \lim_{n\to\infty} \Pr\left\{Y_n = 3, \chi_{3,n} = i, N_n = k\right\}; \qquad i \geq 1, \quad k \geq 0$$

In this work, we put forward our efforts in finding the above stationary probabilities of defined Markov chain. The Kolmogorov equations for the stationary distribution of the system are constructed as follows:

$$\pi_{0,0} = \overline{\lambda}\,\pi_{0,0} + \overline{\lambda}\,v_{1,1,0} \tag{1}$$

$$\pi_{0,i,k} = \overline{\lambda}\,\pi_{0,i+1,k} + \varpi\overline{\lambda}\,\varsigma_i\pi_{1,1,k-1} + \overline{\varpi}\,\overline{\lambda}\,\varsigma_i\pi_{1,1,k} + \overline{\lambda}\,\varsigma_i\pi_{2,1,k} + \overline{\lambda}\,\varsigma_i v_{1,1,k}; \quad i \geq 1, k \geq 1 \tag{2}$$

$$\begin{aligned}\pi_{1,i,k} = {} & \delta_{0k}\lambda\vartheta g_{1,i}\pi_{0.0} + \left(1-\delta_{0,k}\right)\lambda\vartheta g_{1,i}\sum_{j=1}^{\infty}\pi_{0,j,k} + \overline{\lambda}\vartheta g_{1,i}\pi_{0,1,k+1} \\ & + \left(1-\delta_{0,k}\right)\lambda\vartheta\varpi g_{1,i}\pi_{1,1,k-1} + \left(\overline{\varpi}\lambda + \varpi\overline{\lambda}\varsigma_0\right)\vartheta g_{1,i}\pi_{1,1,k} \\ & + \left(\overline{\lambda} + \lambda\,\overline{p}\right)\pi_{1,i+1,k} + \left(1-\delta_{0,k}\right)p\lambda\overline{\eta}\pi_{1,i+1,k-l} + \overline{\varpi}\overline{\lambda}\varsigma_0\vartheta g_{1,i}\pi_{1,1,k+1} \\ & + \left(1-\delta_{0,k}\right)p\eta\lambda g_{1,i}\sum_{j=2}^{\infty}\pi_{1,j,k-1} + (1-\delta_{0k})\,\lambda\vartheta g_{1,i}\pi_{2,1,k} \\ & + \overline{\lambda}\varsigma_0\vartheta g_{1,i}\pi_{2,1,k+1} + \lambda\vartheta g_{1,i}v_{1,k} + \overline{\lambda}\vartheta\varsigma_0 g_{1,i}v_{1,k+1}; \qquad i \geq 1, \quad k \geq 0\end{aligned} \tag{3}$$

$$\begin{aligned}\pi_{2,i,k} = {} & \delta_{1k}\lambda\overline{\vartheta}g_{2,i}\pi_{0.0} + \left(1-\delta_{1,k}\right)\lambda\overline{\vartheta}g_{2,i}\sum_{j=1}^{\infty}\pi_{0,j,k-1} + \overline{\lambda\vartheta}g_{2,i}\pi_{0,1,k} \\ & + \left(1-\delta_{1,k}\right)\varpi\lambda\overline{\vartheta}g_{2,i}\pi_{1,1,k-2} + \overline{\varpi}\overline{\lambda}\varsigma_0\overline{\vartheta}g_{2,i}\pi_{1,1,k} + \left(1-\delta_{1,k}\right)p\eta\lambda g_{2,i} \\ & \times\sum_{j=2}^{\infty}\pi_{2,j,k-1} + (1-\delta_{1k})\,\lambda\overline{\vartheta}g_{2,i}\pi_{2,1,k-1} + \overline{\lambda}\varsigma_0\overline{\vartheta}g_{2,i}\pi_{2,1,k} \\ & + \lambda\overline{\vartheta}g_{2,i}v_{1,k-1} + \overline{\lambda\vartheta}\varsigma_0 g_{2,i}v_{1,k}; \qquad i \geq 1, \quad k \geq 1\end{aligned} \tag{4}$$

$$v_{i,k} = \overline{\lambda} v_{i+1,k} + \left(1 - \delta_{0,k}\right) \lambda v_{i+1,k-1} + \overline{\varpi}\overline{\lambda}\delta_{0,k} v_i \pi_{1,1,0}; \quad i \geq 1, \quad k \geq 0 \tag{5}$$

where $\delta_{ij} = \begin{cases} 1, \text{ if } i = j \\ 0, \text{ otherwise} \end{cases}$ is the Kronecker's symbol.

Let us define the following generating functions and auxiliary generating functions for solving the above equations:

$$\Omega_m(x, z) = \sum_{i=1}^{\infty} \sum_{k=d}^{\infty} \pi_{m,i,k} x^i z^k; \quad m = 0, 2, \quad d = 1; \quad m = 1, \quad d = 0;$$

$$W(x, z) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} v_{i,k} x^i z^k.$$

$$\Omega_{m,i}(z) = \sum_{k=d}^{\infty} \pi_{m,i,k} z^k; \quad m = 0, 2, d = 1; \quad m = 1, d = 0; \quad i \geq 1;$$

$$W_i(z) = \sum_{k=1}^{\infty} v_{i,k} z^k; \quad i \geq 1.$$

The normalizing condition is given by:

$$\pi_{0,0} + \Omega_0(1, 1) + \Omega_1(1, 1) + \Omega_2(1, 1) + \Omega_2(1, 1) = 1 \tag{6}$$

Multiply Eq. (5) by z^k and summing over k and on doing some algebraic manipulations, we result in:

$$W_i(z) = \chi(z) W_{i+1}(z) + \overline{\varpi}\overline{\lambda} v_i \pi_{1,1,0}; \quad i \geq 1 \tag{7}$$

Multiplying Eq. (7) by x^i and then summing over i, we get:

$$\left[\frac{x - \chi(z)}{x}\right] W(x, z) = -\chi(z) W_1(z) + \overline{\varpi}\overline{\lambda} W(x) \pi_{1,1,0}; \quad i \geq 1 \tag{8}$$

Putting $x = \chi(z)$ in Eq. (8), we get:

$$W_1(z) = \frac{\overline{\varpi}\overline{\lambda} W(\chi(z)) \pi_{1,1,0}}{\chi(z)} \tag{9}$$

Use of Eq. (9) in Eq. (8) results in:

$$W(x, z) = \frac{\overline{\varpi}\overline{\lambda} x \left[W(x) - W(\chi(z))\right] \pi_{1,1,0}}{[x - \chi(z)]} \tag{10}$$

Taking derivatives of Eq. (10) with respect to x and letting $x = z = 0$, we get:

$$\pi_{1,1,0} = \frac{\lambda \pi_{0,0}}{\overline{\varpi}\,\overline{\lambda} W\left(\overline{\lambda}\right)} \quad \text{and} \quad v_{1,0} = \frac{\lambda}{\overline{\lambda}} \pi_{0,0} \tag{11}$$

Multiplying Eqs. (1)–(4) by z^k and summing over k and then doing some algebraic manipulations and using results from Eq. (11), we get the auxiliary generating functions for the stationary distribution of the Markov chain $\{\Gamma_n, n \geq 0\}$ when $0 \leq z \leq 1$ as:

$$\begin{aligned} \Omega_{0,i}(z) = {} & \overline{\lambda}\Omega_{0,i+1}(z) + \kappa(z)\overline{\lambda}\varsigma_i \Omega_{1,1}(z) + \overline{\lambda}\varsigma_i \Omega_{2,1}(z) \\ & - \frac{\lambda\varsigma_i\left[\chi(z)\left\{1+W\left(\overline{\lambda}\right)\right\} - \overline{\lambda} W(\chi(z))\right]\pi_{0,0}}{\chi(z) W\left(\overline{\lambda}\right)}; \qquad i \geq 1 \end{aligned} \tag{12}$$

$$\begin{aligned} \Omega_{1,i}(z) = {} & \lambda\vartheta g_{1,i}\Omega_0(1,z) + \left(\overline{\lambda} + \lambda\,\overline{p} + \lambda p\overline{\eta} z\right)\Omega_{1,i+1}(z) + \lambda p\eta g_{1,i} z \Omega_1(1,z) \\ & + \frac{\overline{\lambda}\vartheta g_{1,i}}{z}\Omega_{0,1}(z) + \left[\frac{\kappa(z)\left(\lambda z + \overline{\lambda}\varsigma_0\right)\vartheta}{z} - p\lambda\eta z\right] g_{1,i}\Omega_{1,1}(z) + \frac{\left(\lambda z + \overline{\lambda}\varsigma_0\right)\vartheta g_{1,i}}{z}\Omega_{2,1}(z) \\ & + \frac{\left(\lambda z + \overline{\lambda}\varsigma_0\right)\vartheta g_{1,i}}{z} v_{1,1}(z) - \frac{\overline{\lambda}\varsigma_0 \vartheta g_{1,i}\left[\chi(z)\left\{1+W\left(\overline{\lambda}\right)\right\} - \overline{\lambda} W(\chi(z))\right]\pi_{0,0}}{z\chi(z) W\left(\overline{\lambda}\right)} \end{aligned} \tag{13}$$

$$\begin{aligned} \Omega_{2,i}(z) = {} & \lambda z \overrightarrow{\vartheta} g_{2,i}\Omega_0(1,z) + \gamma(z)\Omega_{2,i+1}(z) + \lambda p\eta z g_{2,i}\Omega_2(1,z) + \overline{\lambda\vartheta} g_{2,i}\Omega_{0,1}(z) \\ & + \kappa(z)\left(\lambda z + \overline{\lambda}\varsigma_0\right)\overline{\vartheta} g_{2,i}\Omega_{1,1}(z) + \left(\lambda z\overline{\vartheta} + \overline{\lambda\vartheta}\varsigma_0 - p\lambda\eta z\right) g_{2,i}\Omega_{2,1}(z) \\ & + \left(\lambda z + \overline{\lambda}\varsigma_0\right)\overline{\vartheta} g_{2,i} v_{1,1}(z) - \frac{\overline{\lambda}\varsigma_0 \overline{\vartheta} g_{2,i}\Gamma(z)\pi_{0,0}}{\chi(z) W\left(\overline{\lambda}\right)}; \qquad i \geq 1 \end{aligned} \tag{14}$$

where $\kappa(z) = \overline{\varpi} + \varpi z$, $\gamma(z) = \overline{\lambda} + \lambda\overline{p} + \lambda p\overline{\eta} z$, $\chi(z) = \overline{\lambda} + \lambda z$, and $\Gamma(z) = \chi(z)\left\{1 + W\left(\overline{\lambda}\right)\right\} - \overline{\lambda} W\left(\chi(z)\right)$.

Now, multiplying Eqs. (12)–(14) by x^i and summing over "i" and then doing some algebraic manipulations, we get the stationary distribution of the Markov chain $\{X_n, n \geq 0\}$ for $0 \leq z \leq 1$:

$$\begin{aligned} \left[\frac{x - \overline{\lambda}}{x}\right]\Omega_0(x,z) = {} & -\overline{\lambda}\Omega_{0,1}(z) + \overline{\lambda}\left[\Phi(x) - \varsigma_0\right]\left[\kappa(z)\Omega_{1,1}(z) + \Omega_{2,1}(z)\right] \\ & - \frac{\lambda\left[\Phi(x) - \varsigma_0\right]\Gamma(z)\pi_{0,0}}{\chi(z) W\left(\overline{\lambda}\right)} \end{aligned} \tag{15}$$

$$\left[\frac{x-\gamma(z)}{x}\right]\Omega_1(x,z) = \frac{\overline{\lambda}\vartheta[1-\overline{\eta}z]G_1(x)\Omega_{0,1}(z)}{z} + \left\{\frac{\vartheta[1-\overline{\eta}z]\left[z+\overline{\lambda}\varsigma_0(1-z)\right]}{z[1-z]}\right\}G_1(x)\Omega_{2,1}(z) + \left[\left\{\frac{\vartheta\kappa(z)[1-\overline{\eta}z]\left[z+\overline{\lambda}\varsigma_0(1-z)\right]}{z[1-z]}\right\}G_1(x) - \frac{\eta\, zG_1(x)}{[1-z]} - \gamma(z)\right]\Omega_{1,1}(z) - \frac{\overline{\lambda}\vartheta[1-\overline{\eta}z][z\{\chi(z)-W(\chi(z))\}+\varsigma_0(1-z)\Gamma(z)]G_1(x)\,\pi_{0,0}}{z[1-z]\chi(z)W(\overline{\lambda})} \tag{16}$$

$$\left[\frac{x-\gamma(z)}{x}\right]\Omega_2(x,z) = \overline{\lambda\vartheta}\,[1-\overline{\eta}z]\,G_2(x)\Omega_{0,1}(z) + \left\{\frac{\overline{\vartheta}\kappa(z)[1-\overline{\eta}z]\left[z+\overline{\lambda}\varsigma_0(1-z)\right]}{[1-z]}\right\}G_2(x)\Omega_{1,1}(z) + \left[\left\{\frac{\overline{\vartheta}[1-\overline{\eta}z]\left[z+\overline{\lambda}\varsigma_0(1-z)\right]}{[1-z]}\right\}G_2(x) - \frac{\eta\, zG_2(x)}{[1-z]} - \gamma(z)\right]\Omega_{2,1}(z) - \frac{\overline{\lambda\vartheta}[1-\overline{\eta}z][z\{\chi(z)-W(\chi(z))\}+\varsigma_0(1-z)\Gamma(z)]G_2(x)\,\pi_{0,0}}{[1-z]\chi(z)W(\overline{\lambda})} \tag{17}$$

Letting $x = \overline{\lambda}$ in Eq. (15) and $x = \gamma(z)$ in Eqs. (16) and (17), we get three simultaneous equations in terms of $\Omega_{0,1}(z)$, $\Omega_{1,1}(z)$, and $\Omega_{2,1}(z)$ as follows:

$$\Rightarrow \overline{\lambda}\left[\Phi\left(\overline{\lambda}\right)-\varsigma_0\right]\left[\kappa(z)\Omega_{1,1}(z)+\Omega_{2,1}(z)\right]-\overline{\lambda}\Omega_{0,1}(z) = \frac{\lambda\left[\Phi\left(\overline{\lambda}\right)-\varsigma_0\right]\Gamma(z)\pi_{0,0}}{\chi(z)W\left(\overline{\lambda}\right)} \tag{18}$$

$$\begin{aligned}\Rightarrow\ & \upsilon\left[1-\overline{\beta}z\right]\left[\overline{\lambda}\zeta_0(1-z)+z\right]G_1(\gamma(z))\left[\kappa(z)\Omega_{1,1}(z)+\Omega_{2,1}(z)\right]\\ & -\left[z(1-z)\gamma(z)+\beta z^2G_1(\gamma(z))\right]\Omega_{1,1}(z)\\ & +\overline{\lambda}\vartheta\,[1-z]\,[1-\overline{\eta}z]\,G_1(\gamma(z))\,\Gamma_{0,1}(z)\\ & = -\frac{\lambda\vartheta[1-\overline{\eta}z][z\{\chi(z)-W(\chi(z))\}+\varsigma_0(1-z)\Gamma(z)]G_1(\gamma(z))\ \pi_{0,0}}{z\chi(z)W(\overline{\lambda})}\end{aligned} \tag{19}$$

$$\begin{aligned}\Rightarrow\ & \overline{\vartheta}\,[1-\overline{\eta}z]\left[\overline{\lambda}\varsigma_0(1-z)+z\right]G_2(\gamma(z))\left[\kappa(z)\Omega_{1,1}(z)+\Omega_{2,1}(z)\right]\\ & -[(1-z)\gamma(z)+\beta zG_2(\gamma(z))]\,\Omega_{2,1}(z)\\ & +\overline{\lambda}\overline{\upsilon}\,[1-z]\,[1-\overline{\eta}z]\,G_2(\gamma(z))\,\Omega_{0,1}(z)\\ & = -\frac{\lambda\overline{\vartheta}[1-\overline{\eta}z][z\{\chi(z)-W(\chi(z))\}+\varsigma_0(1-z)\Gamma(z)]G_1(\gamma(z))\pi_{0,0}}{\chi(z)W(\overline{\lambda})}\end{aligned} \tag{20}$$

Solving Eqs. (18)–(20) for $\Omega_{0,1}(z)$, $\Omega_{1,1}(z)$ and $\Omega_{2,1}(z)$, we obtain:

$$\Omega_{0,1}(z)=\frac{\lambda z\left[\Phi\left(\overline{\lambda}\right)-\varsigma_0\right]\left[\Gamma(z)\Gamma_1(z)\Gamma_2(z)-(1-\overline{\eta}z)\left(\lambda+W\left(\overline{\lambda}\right)\right)\mathrm{B}(z)G_1(\gamma(z))\right]\pi_{00}}{\overline{\lambda}\chi(z)W\left(\overline{\lambda}\right)\Lambda(z)} \tag{21}$$

$$\Omega_{1,1}(z)=\frac{\lambda\vartheta\,(1-\overline{\eta}z)\,\Gamma_2(z)G_1\,(\gamma(z))\,\left[(1-z)\,\Phi\left(\overline{\lambda}\right)\Gamma(z)+z\,\{\gamma(z)-W\,(\gamma(z))\}\right]\pi_{00}}{\chi(z)W\left(\overline{\lambda}\right)\Lambda(z)} \tag{22}$$

$$\Omega_{2,1}(z)=\frac{\lambda\overline{\vartheta}z\Gamma_1(z)G_2\,(\gamma(z))\,\left[(1-z)\,\Phi\left(\overline{\lambda}\right)\Gamma(z)+z\,\{\gamma(z)-W\,(\gamma(z))\}\right]\pi_{00}}{\chi(z)W\left(\overline{\lambda}\right)\Lambda(z)} \tag{23}$$

where $\Gamma_i(z)=(1-z)\gamma(z)+\eta z G_i(\gamma(z)),\quad i=1,2;\ \Lambda(z)=[1-\overline{\eta}z]\left[z+\overline{\lambda}\,(1-z)\,\Phi\left(\overline{\lambda}\right)\right]\mathrm{B}(z)-z\Gamma_1(z)\Gamma_2(z),\ \mathrm{B}(z)=\vartheta G_1\,(\gamma(z))\,\kappa(z)\Gamma_2(z)+\overline{\vartheta}zG_2\,(\gamma(z))\,\Gamma_1(z);$ $\alpha=\overline{\lambda}+\lambda\overline{p}+\lambda p\overline{\eta}$.

Further, substituting the expressions for $\Omega_{0,1}(z)$, $\Omega_{1,1}(z)$ and $\Omega_{2,1}(z)$ in Eqs. (15)–(17), we can obtain the expressions for $\Omega_0(x, z)$, $\Omega_1(x, z)$, and $\Omega_2(x, z)$.

4 Performance Measures

To predict the performance of our developed queueing system, we compute some useful performance measures and then visualize the effects of various critical parameters on these measures. In this section, we obtain various useful performance measures of interests for the developed model as given below:

4.1 Steady State Probabilities

- The probability that the system is empty is given by:

$$\pi_{0,0}=\frac{\sigma_2}{\Phi\left(\overline{\lambda}\right)\sigma_1}$$

  where $\sigma_1=p\{\alpha+\eta G_1(\alpha)\}-\eta^2G_1(\alpha)W(\alpha)-2\alpha\{1-G_1(\alpha)W(\alpha)\}$, $\sigma_2=p[\eta G_1(\alpha)+\alpha\{1-G_1(\alpha)W(\alpha)\}]$.
- The probability that the server is idle is given by:

$$P\,[I]=\Omega_0\,(1,1)=\frac{\left[\Phi\left(\overline{\lambda}\right)-1\right]\sigma_2}{\Phi\left(\overline{\lambda}\right)\sigma_1}$$

- The probability that the server is busy is given by:

$$P\left[B\right] = \Omega_1\left(1,1\right) = \frac{2\alpha\vartheta\left[G_1\left(\alpha\right) - 1\right]}{\sigma_1}$$

- The probability that the server is under repair is given by:

$$P\left[R\right] = \Omega_2\left(1,1\right) = \frac{2\alpha\overline{\vartheta}\left[G_2\left(\alpha\right) - 1\right]}{\sigma_1}$$

- The probability that the server is on vacation is given by:

$$P\left[V\right] = W\left(1,1\right) = \frac{2\alpha G_1\left(\alpha\right)\left[W\left(\alpha\right) - 1\right]}{\sigma_1}$$

- The probability that a customer is lost is given by:

$$P\left[L\right] = \overline{p}\lambda\Omega_1\left(1,1\right) = \frac{2\alpha\overline{p}\lambda\left[G_1\left(\alpha\right) - 1\right]}{\sigma_1}$$

4.2 Average Number and Mean Waiting Time of Customers in the System

The mean number of customers in the orbit (or retrial queue) and the mean waiting time of customers in the queue are denoted by $E[L]$ and $E[W]$, respectively, and are given by:

$$E\left[L\right] = \Omega'(1) = \Omega_0'\left(1,1\right) + \Omega_1'\left(1,1\right) + \Omega_2'\left(1,1\right) + W'\left(1,1\right)$$

$$E\left[WT\right] = \frac{E\left[L\right]}{\lambda}$$

where $W'\left(1,1\right) = \frac{\left[\lambda^2\theta_2 + 2\lambda\overline{\lambda}(\theta_1 - 1)\left\{1 - \Phi\left(\overline{\lambda}\right)\right\}\right]}{2\left[\Phi\left(\overline{\lambda}\right)\left\{\lambda + W\left(\overline{\lambda}\right)\right\} - \lambda(1-\theta_1)\right]}$, $\Omega_0'\left(1,1\right) = \frac{\vartheta\,\Phi\left(\overline{\lambda}\right)\left[\mathrm{Num}_0''\Omega' - \mathrm{Num}_0'\Omega''\right]\pi_{00}}{2(\Omega')^2}$, $\Omega_1'\left(1,1\right) = \frac{\vartheta\,\Phi\left(\overline{\lambda}\right)\left[\mathrm{Num}_1''\Omega' - \mathrm{Num}_1'\Omega''\right]\pi_{00}}{2(\Omega')^2}$, and $\Omega_2'\left(1,1\right) = \frac{\overline{\vartheta}\,\Phi\left(\overline{\lambda}\right)\left[\mathrm{Num}_2''\Omega' - \mathrm{Num}_2'\Omega''\right]\pi_{00}}{2p(\Omega')^2}$.

$$X = \eta b_1\left(\beta\right) b_2\left(\beta\right)$$

$$Y = \eta b_1\left(\beta\right)$$

$$\eta = 1 - p\lambda\beta$$

$$X' = p\lambda\overline{\eta}\eta\left\{G_1'(\beta)G_2(\beta) + G_1(\beta)G_2'(\beta)\right\} + \eta\left(1 + \varpi\vartheta + \overline{\vartheta}\right)G_1(\beta)G_2(\beta) - \beta\left\{\vartheta G_1(\beta) + \overline{\vartheta}G_2(\beta)\right\}$$

$$Y' = -\beta + \eta G_1(\beta) + p\lambda\overline{\eta}\eta G_1'(\beta)$$

$$T_i' = -\beta + \eta G_i(\beta) + p\lambda\overline{\eta}\eta G_i'(\beta); i = 1, 2$$

$$T_i'' = -2p\lambda\overline{\eta} + p\lambda\overline{\eta}\eta G_i'(\beta) + \eta(p\lambda\overline{\eta})^2 G_i''(\beta), i = 1, 2$$

$$x_1 = \vartheta G_1'(\beta)T_2' + \overline{\vartheta}G_2'(\beta)T_1'$$

$$x_2 = \varpi\vartheta G_1'(\beta)G_2(\beta) + \overline{\vartheta}G_1(\beta)G_2'(\beta)$$

$$x_3 = \varpi\vartheta G_1(\beta)T_2' + \overline{\vartheta}G_2(\beta)T_1'$$

$$x_4 = \vartheta G_1(\beta)T_2'' + \overline{\vartheta}G_2(\beta)T_1''$$

$$x_5 = \vartheta G_1''(\beta)G_2(\beta) + \overline{\vartheta}G_1(\beta)G_2''(\beta)$$

$$X'' = 2\left(p\lambda\overline{\eta}x_1 + p\lambda\overline{\eta}\eta x_2 + x_3\right) + x_4 + (p\lambda\overline{\eta})^2\eta x_5$$

$$\mathrm{Num}_0' = \beta\eta\left[\vartheta G_2(\beta) + \overline{\vartheta}G_1(\beta)\right] - \eta\left(1 + \eta\vartheta\overline{\varpi} - \lambda\eta\right)G_1(\beta)G_2(\beta)$$

$$\mathrm{Num}_0'' = \eta X'' - T_2T_1'' - T_1T_2'' - 2\left\{\left(\overline{\eta} - \lambda\eta\right)X'\right\} + \lambda\overline{\eta}X - 2\left\{T_2T_1' + T_1'T_2' + T_2'T_1\right\}$$

$$\mathrm{Num}_1' = -\vartheta\beta\left[1 - G_1(\beta)\right]$$

$$\mathrm{Num}_1'' = 2\vartheta p\lambda\overline{\eta}\left[\beta G_1'(\beta) - \{1 - G_1(\beta)\}\right]$$

$$\mathrm{Num}_2' = -\overline{\vartheta}\beta\left[1 - G_2(\beta)\right]$$

$$\mathrm{Num}_2'' = 2\overline{\upsilon}\left[p\lambda\overline{\eta}\beta G_2'(\beta) - (p\lambda\overline{\eta} + \beta)\left\{1 - G_2(\beta)\right\}\right]$$

$$W = \frac{T_1}{T_2}$$

$$W' = \frac{T_2T_1' - T_1T_2'}{(T_2)^2}$$

$$W'' = \frac{(T_2)^2\left[T_2T_1'' - T_1T_2''\right] - 2T_2T_2'\left[T_2T_1' - T_1T_2'\right]}{(T_2)^4}$$

$$\Omega' = p\Lambda'(1)$$

$$\Omega'' = p\Lambda''(1)$$

$$\Re_1 = \vartheta G_1(\beta) + \overline{\vartheta}G_2(\beta)\, W$$

$$\Re_1' = \frac{\left[\eta^2\left(\varpi\vartheta + \overline{\vartheta}\right)G_1(\beta)\left(G_2(\beta)\right)^2 + p\lambda\eta^2\overline{\eta}G_1'(\beta)\left(G_2(\beta)\right)^2 - \beta\overline{\vartheta}\eta\left(G_2(\beta) - G_1(\beta)\right)G_2(\beta)\right]}{\eta^2\left(G_2(\beta)\right)^2}$$

$$\begin{aligned}\Re_1'' = {} & 2\left[\varpi\vartheta p\lambda\overline{\eta}G_1'(\beta) + \overline{\vartheta}W'(1)G_2(\beta) + \overline{\vartheta}p\lambda\overline{\eta}WG_2'(\beta) + \overline{\vartheta}p\lambda\overline{\eta}W'G'_2(\beta)\right] \\ & + \vartheta(p\lambda\overline{\eta})^2G_1''(\beta) + \overline{\vartheta}\left(p\lambda\overline{\beta}\right)^2WG_2''(\beta) + \overline{\vartheta}W''G_2(\beta)\end{aligned}$$

$$\Lambda''(1) = 2\left[1 - \overline{\lambda}\Phi\left(\overline{\lambda}\right)\right]\left[\eta\Re_1' - \overline{\eta}\Re_1\right] - 2\overline{\eta}\Re_1' + \eta\Re_1'' - T_1'' - 2T_1'$$

5 Numerical Results

In this section, we provide some numerical examples exploring the effects of some sensitive system parameters on mean orbit size. For computation purpose,

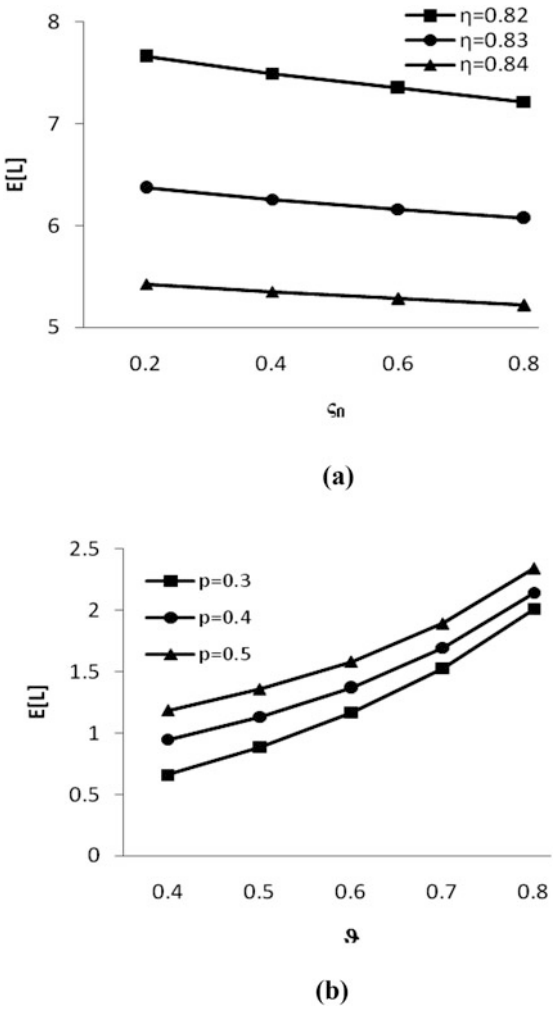


Fig. 1 Effect of (**a**) ς_0 and (**b**) ϑ on mean orbit size for different values of η and p

we assume retrial time distribution, service time distribution, and vacation time distribution to be geometric distributed with generating functions $\Phi(x) = \frac{(1-r)}{1-rx}$, $G_1(x) = \frac{7x}{10-3x}$, and $W(x) = \frac{(1-v)}{1-vx}$, respectively. We coded a program in MATLAB software and plotted Fig. 1a, b showing the trend in mean orbit size vs. ς_0 and ϑ for different values of η and p. The default parameters for these figures are taken as $\lambda = 0.95$, $\eta = 0.9$, $p = 0.02$, $r = 0.15$, and $v = 0.9$.

From these figures, it is clear that mean orbit size tends to decrease linearly by increasing either retrial probability ς_0 or probability η. On the other hand, mean orbit size increases first slowly then sharply and afterwards becomes almost constant with the increase in either the probability of successful activation of server ϑ or balking probability p. This feature matches with many real-life service systems

such as observed generally in ***AIIMS hospital*** in Delhi wherein if high priority patients arrive to the doctor then queue length (number of patients waiting in queue) tends to increase as server is busy in serving those customers first then it starts service of non-priority customers. Moreover, patient calls arrive at telephone in the doctor's room for appointment start accumulating in the virtual queue or buffer if the server or telephone system at doctor's room is activated successfully with high probability ϑ. We observe some queueing situations in the hospital that on seeing a long queue in the hospital, some patients become impatient and may balk with higher probability $(1 - p)$ which in turn decreases the queue length. We also observe that when patients retry for service with more probability, then queue length tends to decrease linearly. Doctor may also leave the system for some urgent work or simply recreation (vacation). Overall, we conclude that we can control the queue length in most of service centers such as hospitals, banks, post offices, super markets, airports by controlling some sensitive system parameters such as ς_0, ϑ, η, and p.

6 Conclusion

In this chapter, we have studied a discrete-time Geo/G/1 retrial queue subject to single vacation with both Bernoulli feedback and starting failure. Many immense works have been done on the vacation policy retrial queueing systems in continuous time but according to our study, this combination has not been done in discrete environment. Here, by using the probability generating function method and supplementary variable technique, we have derived the expressions for some performance measures of the system such viz. long run probabilities, orbit size, and system size. This proposed model validates with the real-life systems and illustrated by some numerical results. In future, one may generalize this model by assuming arrivals in batches and/or by including more realistic feature of optional service along with essential service.

References

1. I. Atencia, P. Moreno, A discrete time retrial queue with general retrial times. Queue. Syst. **48**, 5–21 (2004)
2. J. Wang, Q. Zhao, Discrete-time Geo/G/1 retrial queue with general retrial times and starting failures. Math. Comput. Model. **45**, 853–863 (2007)
3. M. Jain, S. Agarwal, A discrete-time GeoX/G/1 retrial queueing system with starting failure and optional service. Int. J. Operat. Res. **8**(4), 428–457 (2010)
4. A.K. Aboul-Hassan, S.I. Rabia, A.A. Al-Mujahid, A discrete-time *Geo*/*G*/1 retrial queue with starting failures and impatient customers, in *Transactions on Computational Science VII. Lecture Notes in Computer Science*, ed. by M. L. Gavrilova, C. J. K. Tan, vol. 5890, (Springer, Berlin, 2010)
5. S. Gao, Z. Liu, A repairable Geo^{X}/G/1 retrial queue with Bernoulli feedback and impatient customers. Acta. Math. Appl. Sin. **30**(1), 205–222 (2014)

6. S. Lan, Y. Tang, Performance analysis of a discrete-time queue with working breakdowns and searching for the optimum service rate in working breakdown period. J. Syst. Sci. Inf. **5**(2), 176–192 (2017)
7. S. Gao, J. Wang, Discrete-time $Geo^X/G/1$ retrial queue with general retrial times, working vacation and vacation interruption. Qual. Technol. Quantitat. Manag. **10**(4), 495–512 (2013)
8. F. Zhang, Z. Zhu, A discrete-time Geo/G/1 retrial queue with two different types of vacations. Math. Probl. Eng. **2015**, 835158., 12 pages (2015)
9. S. Upadhyaya, Performance analysis of a discrete-time Geo/G/1 retrial queue under J-vacation policy. Int. J. Ind. Syst. Eng. **29**(3), 369–388 (2018)
10. P. Gunasekaran, S. Jeyakumar, An analysis of discrete time Geo/G/1 retrial queue with second optional service with a vacation. In. J. Rec. Technol. Eng. **7**(6S2), 2277–3878 (2019)
11. J.J. Hunter, *Mathematical Techniques of Applied Probability, Discrete-Time Models: Techniques and Applications* (Academic Press, New York, NY, 1983), p. 2

Monofractal and Multifractal Analysis of Indian Agricultural Commodity Prices

Neha Sam, Vidhi Vashishth, and Yukti

Abstract The Indian commodity market is characterized by high volatility. When considering the agro-based commodity market, the prices may sometimes vary on a daily basis and regional basis. For the purpose of our research, we have restricted our region of study to the Indian national capital New Delhi. This paper aims to find out whether commodity markets follow a pattern with respect to prices, and if they do, then whether this could be determined by using basic fractal theory and determination of Hurst exponent. We have followed a suitable algorithm to find the Hurst exponent using statistical methods, specifically linear regression and time series analysis, wherein time is the independent variable and price of the commodity considered is dependent. The reason why time series analysis is chosen is because of the tendency of a time series to regress strongly to its mean. A statistical measure chosen to classify time series is the Hurst exponent. Initially, we have focused on onion prices for the years 2013 to 2017. The data set has been derived from the official website of the Consumer Affairs Department of the Government of India. The daily retail prices for Delhi for the month of June were observed and analyzed. We eventually aim to investigate if the market for onions has a long-term memory and will it be suitable to extend this conclusion to all other agro-based commodities. Our study has been motivated by the Fractal Market Hypothesis (FMH) that analyses the daily randomness of the market. We seek to find out whether the commodity market follows such a pattern provided that external factors remain constant. By external factors, we mean the variations that occur in the market with time, which include the demand, inflation, global price change, changes in the economy, etc. Keeping this in mind, we have attempted a time series analysis, using the monofractal analysis, at the end of which we would be estimating the Hurst exponent. The determination of Hurst exponent will help us to classify the time series as persistent or anti-persistent, i.e., how strong is the tendency of the time series to revert to its long-term mean value. Further, the multifractal analysis has been used to detect small as well as large fluctuations within the time series

N. Sam · V. Vashishth · Yukti (✉)
Department of Mathematics, Jesus and Mary College, University of Delhi, New Delhi, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_28

taken into consideration. This result would thus lead us to understand if prices in the commodity market could be remotely predicted, and what is the strength of the time series to return to its long-term mean value. Hence, this fractal analysis can be used to determine the characteristics of the prices in an agro-based economy.

Keywords Fractal theory · Prices · Monofractal · Pattern · Multifractal

1 Introduction

The concept of memory cannot be ruled out, when one comes across patterns in economic variables. It is defined from the point of view of economic models, backed by continuous time approach (as per the Concept of Dynamic Memory in Economics by Valentina V. Tarasova and Vasily E. Tarasov). With this realization, we attempt to find if prices in the Indian agro-based commodity market have a tendency of exhibiting a long-term memory. The analysis in this paper considers the market for onions in the National capital Delhi region of India [1].

The volatility of the price levels in the Indian commodity market forms a sizable portion of the overall price fluctuations seen in the economy, in the past few years. Because the Indian economy is agro-based, it is highly sensitive to even the smallest of changes in the agro-based commodity market. Interestingly, onions show a high degree of variability in prices, among all agro-based products. Thus, a closer examination of onion prices in the Indian economy may give us an insight of whether the market, in general, can be remotely predicted. Fractal Market Hypothesis (FMH) forms the basis for studying the daily randomness of the market [2–5]. The Hurst exponent approach has been used for quantifying the results obtained.

From among the multiple statistical methods known, we find linear regression and time series analysis the most suitable for the algorithm proposed in this paper. The statistical measure chosen to classify time series is the Hurst exponent. The calculation of the Hurst exponent has been done, considering time as the independent variable and price of the commodity considered (onions) as dependent. The reason for choosing time series analysis is its tendency to regress strongly to its mean [6–9].

Further, in the course of this research, it was observed that the value of the Hurst exponent (H), as found, indicates that the time series for the chosen commodity is a long-range dependent process or persistent. Henceforth, the multifractal detrended fluctuation analysis (MF-DFA) was employed. It involves calculating the q-order Hurst exponent Hq, using the root-mean-square (RMS) approach to observe short-term and long-term fluctuations in the time series. The q-dependent RMS values obtained were converging to the overall RMS value of the time series, which was evident for the time series being multifractal. Another advantage of MF-DFA lies in the ability of the analysis to identify fluctuations even in dynamic time series. The analysis has been extensively used in medical science, and stock markets, all over the world. Being used in agro-based commodity markets is a first-of-its-kind

application. This highlights a promising scope of this analysis for policymaking in the country [10–12].

2 Background

Commodities have been traded in India, since time immemorial. If the trade takes place in bulk, it is greatly influenced by weather conditions. Soft commodities, such as agricultural products, are greatly impacted by external factors. Farming patterns are one such factor, which are largely profit-driven.

Importance of Agriculture in Indian Economy Since independence, the primary sectors (Agriculture and Allied Activities) have contributed a major portion to India's GDP, which bears testimony to the fact that India is an agriculturally driven economy. It is only natural then that emphasis be laid on welfare of the people employed in the primary sector, while drafting national policies. It has been seen time and again that such policy decisions affect the course of political governance in the nation. In such a scenario, if we are able to detect a pattern, then there could be incorporation of exemplary and innovative methods in the process of policy formulation [3].

Fractals and Fractal Market Hypothesis A fractal is a never-ending, self-similar pattern. Earlier researches have shown that financial markets have a fractal-like property. Fractal Market Hypothesis (FMH) is based on the assumption that history repeats itself, and hence finds application in estimating asset prices. The ambit of using FMH could be widened if this research successfully establishes a similar fractal-like pattern in commodity markets too [13, 14].

Finally, in India, besides being a staple diet food, onion is also hoarders' favourite, in times of tight supply and rising prices. India is second only to China in onion production. Due to the reason of it being produced on such a large scale, onion often tends to act as an effective indicator of inflation, through its prices. Another mysterious aspect about this vegetable is that its prices have been continuously rising despite increased production. Hence, our research draws legitimacy from the fact that, to some extent, onion prices are independent of external factors [15].

3 Methodology

3.1 Monofractal Analysis (Hurst Exponent Approach)

The Hurst exponent approach has been used to perform monofractal analysis to find the existence of fractal patterns in the prices of an agricultural commodity. The

Hurst exponent (H) can be used to quantify the character of randomness exhibited in a time series via an autocorrelation measurement [16, 17].

Value of Hurst exponent lying between 0 and 0.5 represents variables that show anti-correlated behaviour.

Value of Hurst exponent equal to 0.5 represents the process that is purely random.

Value of Hurst exponent lying between 0.5 and 1 represents that behaviour is positively correlated and there is persistence of definite patterns.

Autocorrelation function C, given by,

$$C = 2^{(2H-1)-1},$$

, where H is the Hurst exponent, can illustrate the effect of influence by the present on the future. A simple relation $D = 2-H$, where H is the Hurst exponent, gives the fractal dimension of the time series. Dimension for fractals is essentially a statistical quantity of a fractal. Fractal dimension provides an idea of how a fractal is taking up space, if zoomed down to finer and finer scales [6, 18].

3.2 Data Collection

Onion prices for the years 2013–2017 are considered. The data has been extracted from the official website of the Consumer Affairs Department of the Government of India.

The daily retail prices of onion in Delhi for the month of July, August, and September were considered. Observation and suitable analysis were performed on the data [18, 20].

3.3 Method: Determination of Hurst Exponent

Finding the Hurst exponent is the major part of applying monofractal analysis using (R/S) analysis [18, 21, 22].

The algorithm of finding Hurst exponent using R/S analysis is as follows:

1. Split the time series of size M into disjoint subsets of time intervals $D_j (j = 1, \ldots, J)$ of size m.
2. $\bar{y_j}$ denotes the mean of values in each of the subsets. $x_{k,j}, k = 1, \ldots, m$ represents the value to each corresponding time value.
3. The cumulative deviation $\hat{y_{i,j}} (i = 1, \ldots, m)$ is calculated for each D_j.
4. $RangeR_j = max(\hat{y_{i,j}}, i = 1, \ldots, m) - min(\hat{y_{i,j}}, i = 1, \ldots, m)$.
   S_j= standard deviation for each D_j.
   (R/S)(m)= average of R_j/S_j for j=1,. . .,J.

The relation between the Hurst exponent and above calculated values is

$$log(R/S)_m = logc + Hlogm,$$

where c = constant and H = Hurst exponent.

Linear regression is applied to the above equation to obtain the Hurst exponent H [18].

3.4 Monofractal Detrended Fluctuation Analysis (DFA)

Monofractal DFA is performed to analyse monofractal fluctuations in a time series. Monofractal DFA involves the following steps [10]:

1. Convert a noise-like time series into a random walk-like time series.
2. Compute root-mean-square (RMS) variation of a time series including computation of local fluctuation in the time series as RMS of the time series within non-overlapping segments. RMS is defined to be, as the name suggests, the square root of arithmetic mean of squares of values.
3. The amplitudes of the local RMS are summarized into an overall RMS. The overall RMS of the segments with small sample sizes is dominated by the fast fluctuations in the time series, whereas the overall RMS for segments with large sample sizes is dominated by slow fluctuations. The power law relation between the overall RMS for multiple segment sample sizes (i.e., scales) is defined by a monofractal detrended fluctuation analysis (DFA) and is called the Hurst exponent [10].
4. First divide time series into non-overlapping segments of size t. The detrending procedure is done by estimating a polynomial trend $x^n_{w,t}$, within each segment w by least-square fitting and subtracting this trend from the original profile ("detrending"), $X_t(i) = X(i) - x^n_{w,t}(i)$. The degree of the polynomial can be varied in order to eliminate constant ($m = 0$), linear ($m = 1$), etc. The variance of the detrended profile $X_t(i)$ in each segment w yields the mean-square fluctuations, $F^2_{DFAm}(w,t) = 1/t \sum_{i=1}^{t}[X^2_t(i)]$.

   $F^2_{DFAm}(w,t)$ are averaged over all segments w to obtain the mean fluctuations $F_2(t)$. Slope of regression line of log $F_2(t)$ versus log(t) gives an approximation of Hurst exponent [10, 23].

Next, we can compare this Hurst exponent with as calculated using rescaled range analysis or (R/S) analysis in Sect. 3.3.

3.5 Multifractal Analysis

A method to observe and investigate the multifractal spectrum of the time series is q-order fluctuation analysis or detrended fluctuation analysis. It was originally

introduced by Peng et al. [24] and is used to reliably detect long-range (auto-) correlations.

In a multifractal time series, local fluctuation will be of extreme large magnitude for segments w within the time periods of large fluctuations and extreme small magnitude for segments w within the time periods of small fluctuations. Consequently, the multifractal time series are not normally distributed, and all q-order statistical moments should to be considered [10].

Therefore, q-order Hurst exponent is calculated for different values of q.

Algorithm [25] :

For time series $Y = y_i{}_{i=1}^{n}$ and n the length of the data:

1. calculate the $Z = z_i{}_{i=1}^{n}$,

$$z_i = \sum_{m=1}^{i} (y_m - \bar{y})$$

, where $\bar{y}$ is the mean value of Y;
2. divide the profile time series Z into non-overlapping segments of equal length r;
3. calculate the trend of each segment, and the detrended time series $x_t(i)$ can be obtained by the following equation:

$$x_t(i) = w_t(i) - p_t(i), (1 < i < r)$$

, where $w_t(i)$: segment time series and $p_t(i)$: trend time series, at each segment t;
4. determine the variance by the following:

$$F^2(r,t) = \frac{1}{r} \sum_{i=1}^{r} x_t^2(i)$$

.

Then, the q-order function fluctuation is given by

$$\begin{cases} F_q(r) = \sqrt[p]{\dfrac{1}{2I_r} \displaystyle\sum_{i=1}^{2I_r} F^2(r,t)^{\frac{q}{2}}}, \; if \; q \neq 0 \end{cases}$$

$$\begin{cases} F_q(r) = \exp\{\dfrac{1}{4I_r} \displaystyle\sum_{i=1}^{2I_r} ln(F^2(r,t))\}, \; if \; q = 0 \end{cases}$$

Plot F_q for different values of q [25].

4 Results

4.1 Data

Figure 1 is the graphical representation of daily retail prices of onions in New Delhi, India quantified in Rs. per kg, and the data has been recorded from July to September for each year from 2013 to 2017.

Here, the time period is scaled along the X-axis, and the price per kg of onions in the corresponding time period in INR is marked along the Y-axis. The graph has been plotted on R software.

4.2 Results Derived from the Hurst Exponent Approach

Table 1 displays the values of autocorrelation function, fractal dimension, and Hurst exponent values for each year in Columns 1, 2, and 3, respectively. The value of the

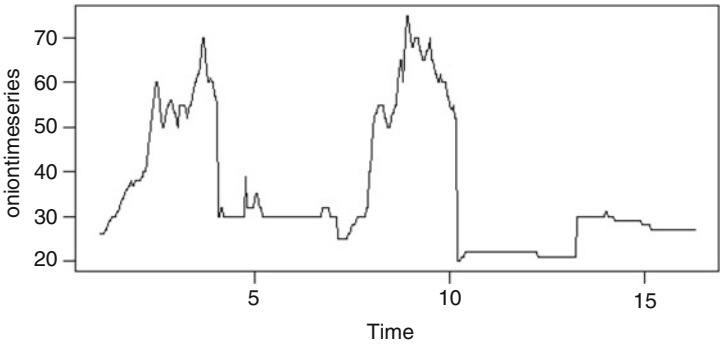


Fig. 1 Graphical representation of daily retail prices of onions in New Delhi in Rs. per kg from 2013–17. (Source: Data used to plot the graph has been taken from the website of the Consumer Affairs Department of the Government of India, consumeraffairs.nic.in)

Table 1 Values of autocorrelation function, fractal dimension, and Hurst exponent for 2013–17

Year	$C=2^{2H-1}-1$ (Autocorrelation function)	D=2-H (Dimension)	Value of Hurst exponent(H)
2013	0.5360	1.1904	0.8096
2014	0.0039	1.4972	0.5028
2015	0.1769	1.3825	0.6175
2016	0.0397	1.4719	0.5281
2017	0.5195	1.1982	0.8018

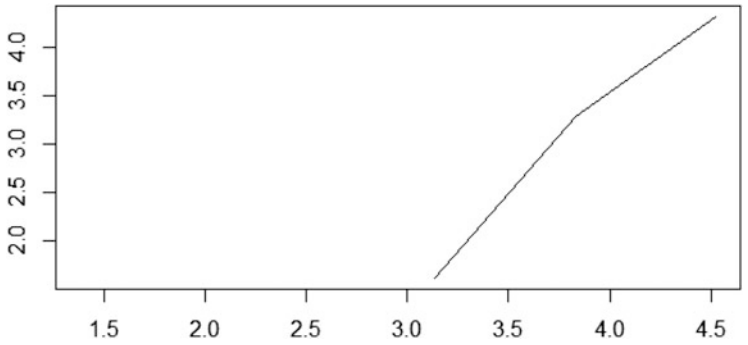


Fig. 2 V-statistic v/s log(n) graph for 2013. . (Source: Authors' calculations)

autocorrelation function describes the influence of present (present price variations in this case) on the future. Fractal dimension values reflect how completely a fractal appears to fill up space as one zooms down to finer and finer scales. It is notable from Table I that fractal dimensions are fractional in nature unlike dimensions of shapes in classical geometry.

The Hurst exponent values represented in Column 3 have been calculated on R software. Clearly, the Hurst exponent for each year from 2013 to 2017 is greater than 0.5; therefore, the variations are not completely random and can be predicted in the short run. This also implies that the variations show fractal characteristics.

4.3 Test to Establish the Persistence of the Time Series

In order to test the stability of the Hurst exponent, the V_n versus log(n) graph is plotted. V statistic is given by

$$V_n = \frac{(R/S)_n}{\sqrt{n}}.$$

Figures 2, 3, 4, 5, and 6 represent V_n versus log(n) graphs for the years 2013–2017.

The graphs clearly represent that the V_n versus log(n) graphs for all the years are upward sloping, and it was claimed that the process is persistent, and thus, stability of the Hurst exponent is established [18].

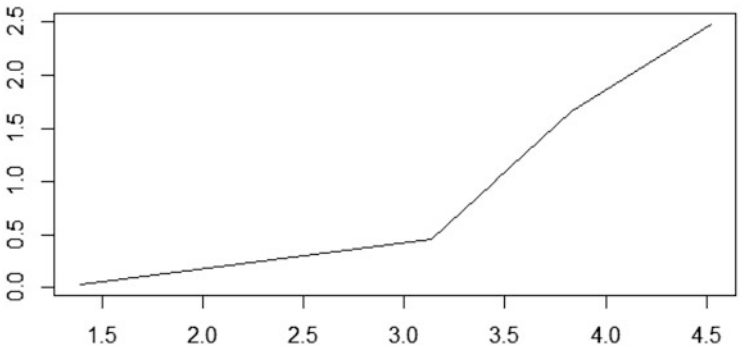


Fig. 3 V-statistic v/s log(n) graph for 2014. (Source: Authors' calculations)

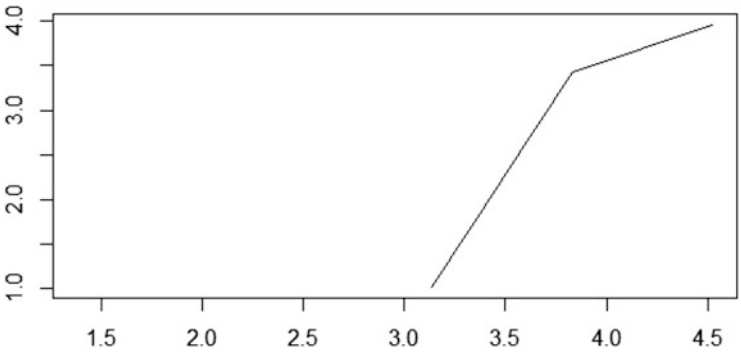


Fig. 4 V-statistic v/s log(n) graph for 2015. (Source: Authors' calculations)

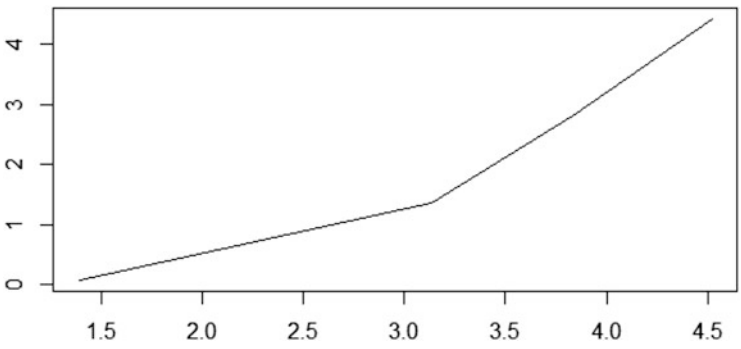


Fig. 5 V-statistic v/s log(n) graph for 2016. (Source: Authors' calculations)

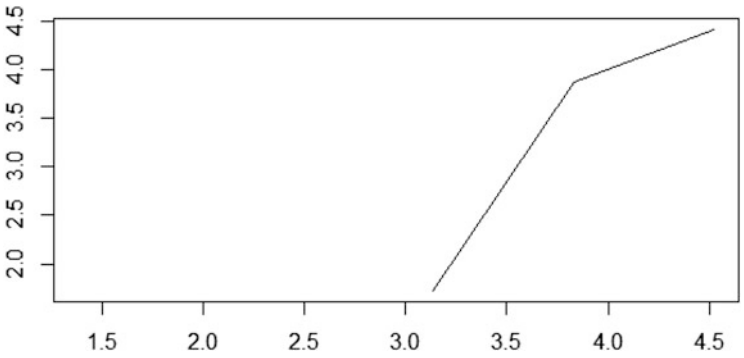


Fig. 6 V-statistic v/s log(n) graph for 2017. (Source: Authors' calculations)

4.4 Calculation of Hurst Exponent for Compiled Data from 2013 to 2017

The Hurst exponent is now calculated for the entire time series using a suitable algorithm on R software. The value of the Hurst exponent thus calculated comes out to be **0.7137**.

4.5 Monofractal Detrended Fluctuation Analysis (DFA)

In order to perform the analysis, MATLAB (computer programming language) has been used. A stepwise process shall be employed to graphically analyze variations and fluctuations in the time series corresponding to the data. Note that the data taken here shall be compiled data for all years from 2013 to 2017. The compiled data is converted into a 460×1 array.

4.5.1 Random Walk-Like Time Series

As a preliminary step for DFA, the time series should be first converted into random walk-like time series using a suitable code on MATLAB. The graph thus obtained is different from Fig. 1, thus displaying that the time series has been converted into a random walk. This is illustrated in Fig. 7.

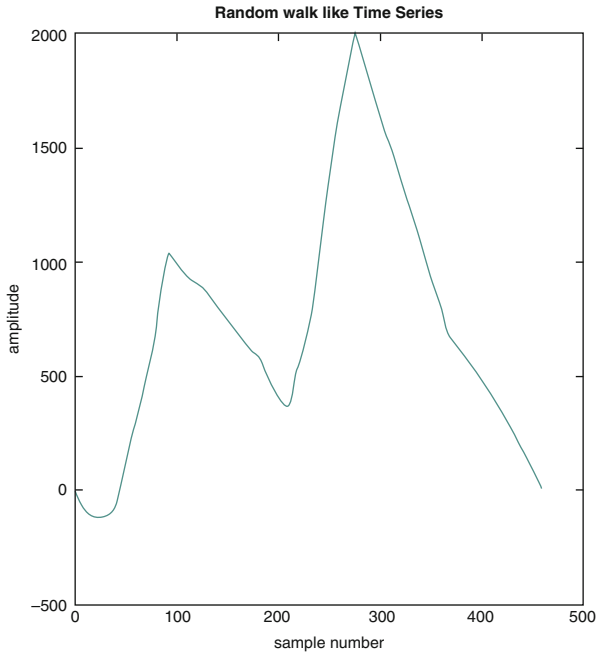


Fig. 7

4.5.2 Local Root-Mean-Square Variation of the Time Series (RMS)

The root mean square of a time series is the square root of the mean of the squared data entries. Using MATLAB, it was determined that the value of RMS for the time series is 36.6833.

4.5.3 Graphical Representation of the Overall Root Mean Square of Time Series

The continually changing fluctuation would influence the overall RMS. Using a suitable algorithm in MATLAB, the overall RMS denoted by F is plotted such that it is represented in log coordinates. Figure 8 illustrates the overall RMS where the X-axis represents logged values of the overall RMS and the Y-axis represents segments of the sample size.

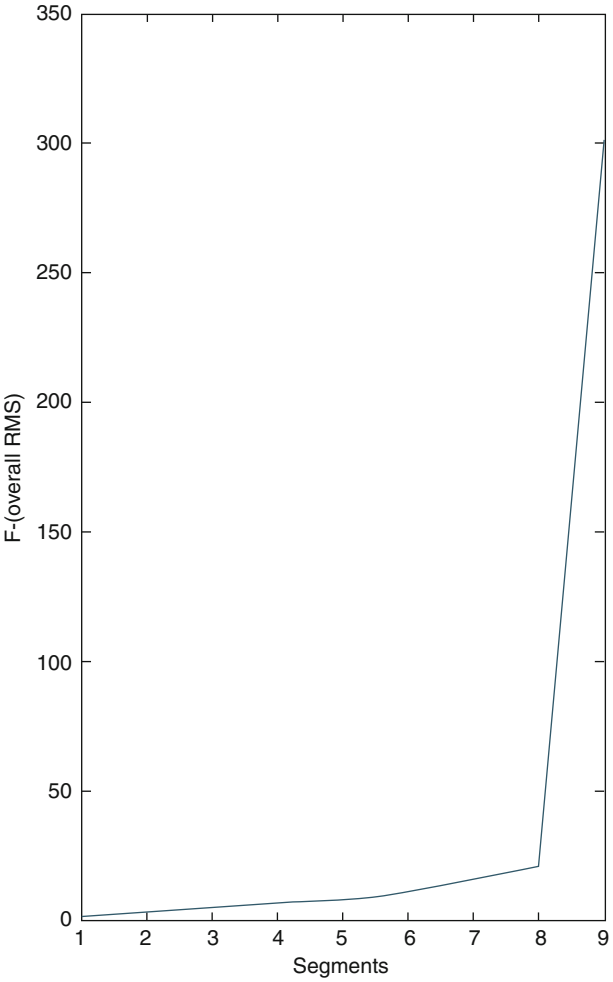
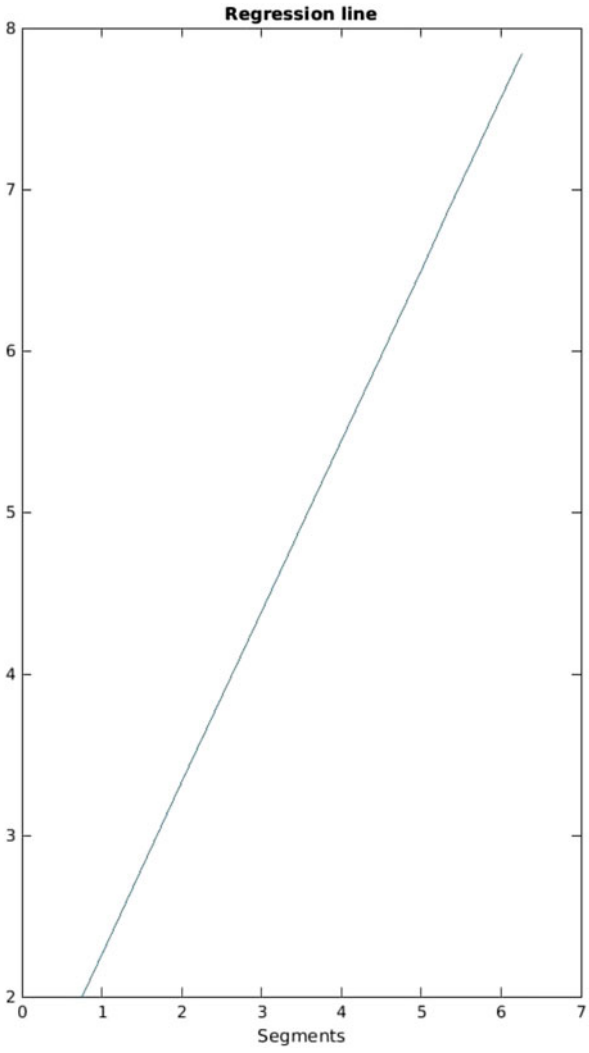


Fig. 8

4.5.4 Graphical Representation of the Regression Line and Verification of Hurst Exponent

As the concluding step in the detrended fluctuation analysis, the best fit regression line is plotted. The slope of the regression line is the Hurst exponent. Figure 9 represents the regression line. The slope of the line as calculated on MATLAB is equal to **0.7509**.

Fig. 9



4.6 Multifractal Detrended Fluctuation Analysis (MFDFA)

MFDFA is performed to determine whether the time series displays multifractal-like characteristics. In order to establish this, the q-order Hurst exponents are defined as the slopes of the regression line, and the computation is done using looping commands on MATLAB. The graph thus obtained is given in Fig. 10. It can be observed that the regression lines thus plotted with $q = -10, 0, 10$ converge.

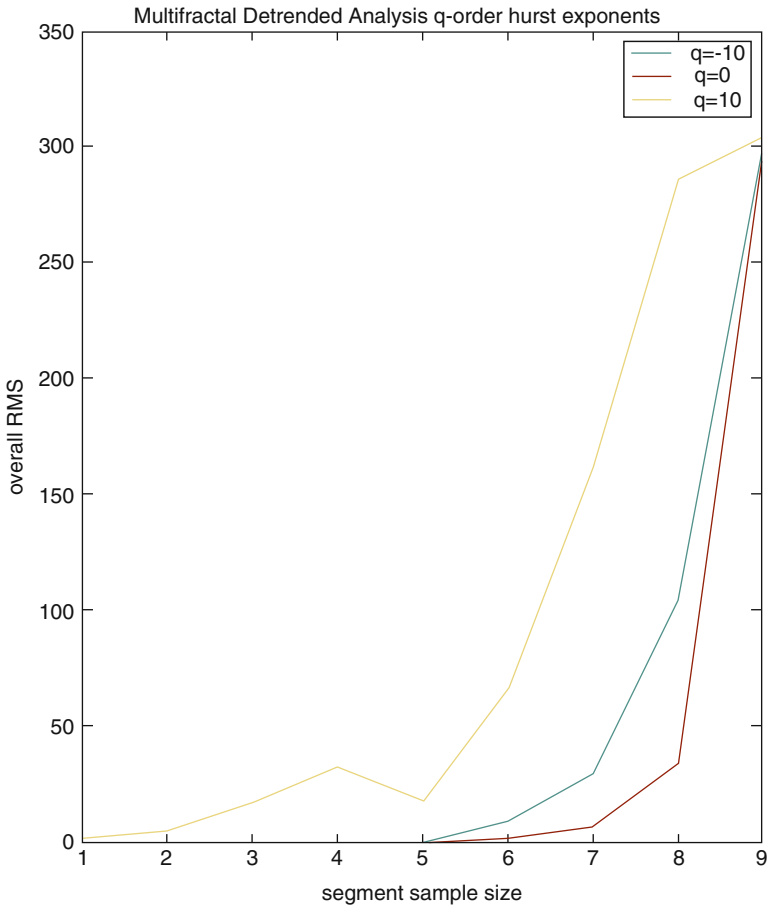


Fig. 10

5 Analysis

First and foremost, the Hurst exponent value derived from the Hurst exponent approach for the compiled data for the years 2013–2017 is found to be approximately equal to the value computed by the detrended fluctuation analysis approach. It can also be inferred from Fig. 7 that the time series that has been converted into a random walk-like structure has distinct periods with large and small fluctuations, which is a characteristic of multifractals. Additionally from Fig. 10, it can be observed that the slopes of the regression lines are q-dependent and the q-order Hurst exponent are the slopes, and the lines thus plotted are seen to converge at larger segments, which is another characteristic of time series exhibiting multifractal patterns.

Conclusions

The Hurst exponent values calculated by both the Hurst exponent approach and the detrended fluctuation analysis make it evident that the prices of onions from 2013–2017 exhibit fractal characteristics. Therefore, the retail onion prices series is monofractal. By the multifractal detrended fluctuation analysis results mentioned in Sect. 4.5 and the analysis in Sect. 5, it can be inferred that the onion retail price series also displays certain multifractal-like characteristics.

Hence, we can conclude that the time series of the onion prices is persistent in nature displaying monofractal and multifractal-like characteristics. Since the study has been restricted to a single commodity, this result may or may not be generalized to the other agro-based commodities of the Indian market. The scope of this research is extensive as fractal-like properties exhibited by a time series imply it could be remotely predicted. Multifractal-like properties displayed by the time series imply promising results towards future prediction of price fluctuations in the market, and hence, knowledge of these fluctuations could benefit the policymakers in drafting policies that can prepare the economy to tackle major price variations.

Acknowledgments We are thankful to Dr. Anu Saxena and Department of Mathematics, Jesus and Mary College for the interesting opportunity. We would also like to extend our gratitude to our dear mentor Dr. Monica Rani for guiding and supporting us throughout.

References

1. V.V. Tarasova, V.E. Tarasov, Concept of dynamic memory in economics, vol . 55 (2017). arXiv.1712.09088
2. B.B. Mandelbrot, J. Bus. **36**, (1963)
3. W.L. Andrew, Econometrica **59**, (1991)
4. A. Chatrath, B. Adrangi, K.K. Dhanda, Agric. Econ. **2**(27), (2002)
5. E. Peters (ed.), *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics* (Wiley, New York, 1994)
6. S. Mansukhani, *Analytics Magazine by INFORMS* (The Institute for Operations Research and the Management Sciences, Catonsville, 2012)
7. A.Y. Schumann, J.W. Kantelhardt, Phys. A **390**, (2011)
8. F.M. Siokis, Phys. A **395**, (2014)
9. B.B. Mandelbrot, J. Bus. **40**, (1967)
10. E.A.F. Ihlen, Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. Front. Physiol. (2012). https://doi.org/10.3389/fphys.2012.00141
11. A.P. Geoffrey, Int. J. Forecast. **1**(16), (1994)
12. C.D. Scott, R.E. Smalley, J. Nanosci. Nanotechnol. **3**(75), (2003)
13. L. Kristoufek, Fractal Markets Hypothesis and the Global Financial crisis: Scaling , investment horizons and liquidity. Adv. Complex Syst. **15**, 1250065 (2012)
14. K. Yin, H. Zhang, W. Zhang, Q. Wei, Romanian J. Econ. Forecast. **3**(16), (2013)
15. S. Ahluwalia, *Onion Prices and Indian Politics* (Observers Research Foundation (ORF), Delhi, 2015)

16. I. Pilgrim, R.P. Taylor, *Fractal Analysis of Time-Series Data Sets: Methods and Challenges* **309** (2018). https://doi.org/10.5772/intechopen.81958
17. H.E. Hurst, Trans. Am. Soc. Civil Eng. **116**, (1951)
18. Y. Wang, X. Su, X. Zhan, Fractal analysis of the agricultural products prices time series. Int. J. u-e Serv. Sci. Technol. **8**(10), 395–404 (2015)
19. Daily—Retail and Wholesale Pricing—Price Monitoring Cell—Consumer Affairs Website of India—Ministry of Consumer Affairs, Food and Public Distribution—Government of India. https://fcainfoweb.nic.in/reports/report_menu_web.aspx
20. Daily Price Arrival Market Bulletin—Price and Arrivals Statistics—Statistic and Market Info—National Horticulture Board—Ministry of Agriculture and Farmers Welfare—Government of India. http://nhb.gov.in/
21. B.B. Mandelbrot, J.R. Wallis, Water Resour. Res. **5**(5), (1969)
22. B.B. Mandelbrot, Probab. Theory Relat. Fields **4**(31), (1975)
23. J.W. Kantelhardt, *Fractal and Multifractal Time Series* (Springer, Berlin, 2008)
24. C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L.Goldberger, Mosaic organization of DNA nucleotides. Phys. Rev. E **49**, 1685 (1994)
25. X. Zhang, G. Zhang, L. Qiu, B. Zhang, Y. Sun, Z. Gui, Q. Zhang, A modified multifractal detrended fluctuation analysis (MFDFA) approach for multifractal analysis of precipitation in Dongting Lake Basin, China (2019)
26. W. Taylor, D.W. Bunn, Manage. Sci. **2**(145), (1999)

Empirical Orthogonal Function Analysis of Subdivisional Rainfall over India

K. C. Tripathi and M. L. Sharma

Abstract The Indian Summer Monsoon Rainfall (ISMR) that takes place during May, June, July, and August each year is a factor that contributes significantly to the socio-economical growth of the Indian subcontinent. ISMR accounts for about 70–75% of the annual rainfall over the region. However, the distribution of the precipitation over the spatial domain is not uniform, and there may be simultaneous flood and draught. The spatial and temporal distribution of the precipitation can lead to a better forecasting model with lesser number of unknown parameters. In the present study, the 142-year monthly data set of 19 subdivisions of India from the Indian Institute of Tropical Meteorology is analyzed to decode the precipitation signals and redistribute the dimensions based on variance and co-variance matrices. This is known as Empirical Orthogonal Functions Analysis. It is observed that the entire data set of 19 dimensions can be redistributed in 3 dimensions with a relatively less information being lost. The five eigenvectors of the co-variance matrix are discussed. The paper is concluded with discussion on the employment of intelligent system algorithms for the extraction of further lower dimensions in the data so as to further reduction in the data.

Keywords Subdivisional rainfall · Correlation · Empirical orthogonal functions · Variance · Pattern recognition

1 Introduction

The Asian Monsoon system at macro- and microlevels is a matter of interest among meteorologists [1]. It remains a vital component of the global monsoon system. The monsoon affects a large part of the world population and is predominantly responsible for the precipitation over the globe [2]. Indian Summer Monsoon

K. C. Tripathi (✉) · M. L. Sharma
Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India
e-mail: kctripathi@mait.ac.in; mlsharma@mait.ac.in

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_29

Rainfall (ISMR), the total precipitation occurring during June, July, August, and September (JJAS), accounts for about 75–80% of the annual rainfall over the Indian subcontinent. At the regional level, the western and central India receives more than 90% of the total precipitation during the JJAS period, while the southern and north western India receives about 50% of the total precipitation during the JJAS period [3]. It is claimed that more than 50% of the total earth population is affected by the Asian monsoon [4]. A small delay in the arrival of Monsoon may bring catastrophe in the entire subcontinent. Further, the droughts and floods associated with the Indian monsoon have a significant effect on the socio-political aspects of India [5–8]. Hence, the analysis of variation of the ISMR is important not only for scientific understanding but also for a better socio-economical development of the subcontinent. There are two factors about the ISMR that is of significant interest: the interannual variability and the intraseasonal variability [9]. Researchers have always been interested in analyzing the variability of the ISMR. These variations result from global climatic conditions as well as local weather effects. The distribution of the spatial and temporal precipitation over India is highly unpredictable, and hence, this remains a topic of study in all times. Further, the variability of the rainfall is reflected at three scales: the All India scale, the regional scale, and the subdivisional scale. In the present study, interest lies in the redistribution of the subdivisional rainfall. The All India and the regional scales are macrolevel manifestation of this microlevel distribution. The temporal variances, i.e., the interannual and intraseasonal variations, have also been kept out of scope of the study so as to avoid digressing from the topic of interest. The Empirical Orthogonal Function Analysis or the Principal Components Analysis [10] refers to the reduction in dimension of a data set by rotating the original space of the data observations so as to redistribute the variances along orthogonal axes. It is a mathematical tool that is more commonly used in pattern recognition, but of late it has been freely used by meteorologists across the globe for the analysis of meteorological data such as precipitation [11–14] and sea surface temperature [15]–[16]. The present study aims to study the correlations in the amount of precipitation recorded at substation level across the Indian climatic region. The information contained in the hidden dimensions can then be used for better model development for rainfall forecasting at the microlevel.

2 Methodology

2.1 Data

We have used the data set of the Indian Institute of Tropical Meteorology (IITM) obtained from the official website http://www.tropmet.res.in/Data. The data set comprises monthly rainfall with a resolution of upto 1 decimal in mm. It is a 142 year all-India rainfall as well rainfall of 30 subdivisions of India during

the period 1871–2012. We have taken into consideration 19 of the subdivisions based on regions and homogeneity. These subdivisions are: Assam and Meghalaya, Gangetic West Bengal, Jharkhand, Bihar, East UP, West UP, Haryana, Punjab, West Rajasthan, East Rajasthan, West MP, East MP, Gujarat, Saurashtra, Madhya Maharashtra, Chhattisgarh, Coastal Andhra Pradesh, Tamil Nadu, and Kerala. Figure 1 shows the map of India and the substations therein.

2.2 *Empirical Orthogonal Functions (EOFs)*

A meteorological data set usually consists of a large number of attributes. The original space of attributes of the data set makes the "attribute space" or the original space of the data. Each attribute may be considered as an axis in terms of the coordinate system. A brief description of EOF analysis is presented here. Consider

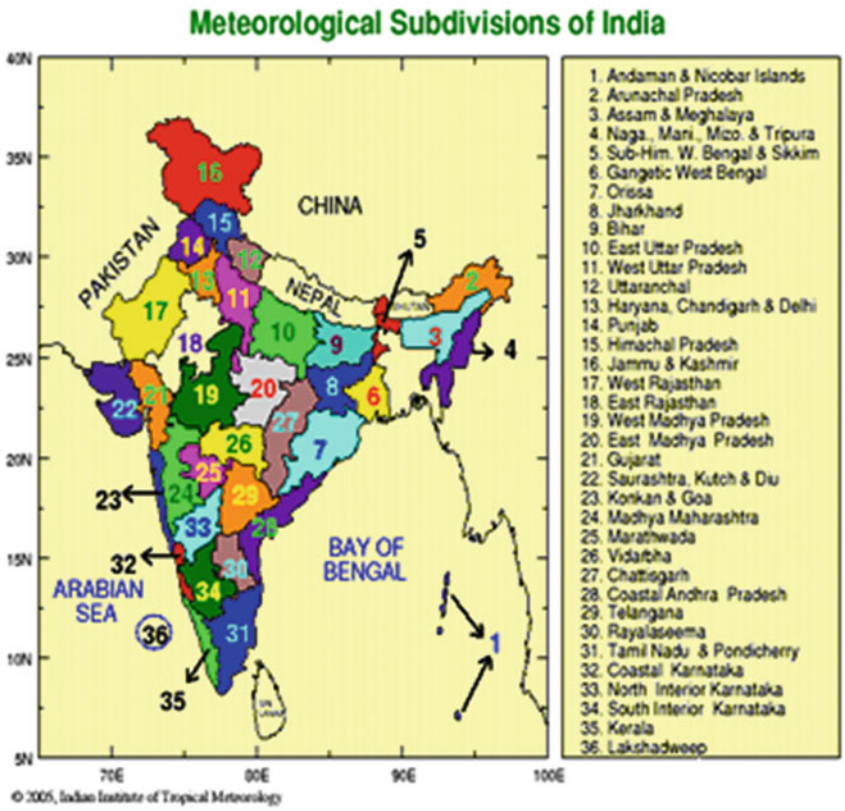


Fig. 1 Meteorological subdivisions of India (Indian Institute of Tropical Meteorology, Pune)

a data set **D** in n dimensions of the Euclidian space. Each observation can then be considered as an n-dimensional vector or a single point in the n-dimensional space. The objective of EOF analysis is to find a new set of orthogonal axes such that when the original cloud of n-dimensional data is manifested in the new space, called the "feature space," the variances are redistributed so that a few axes retain the maximum variance of the data. The total variance of the data in the feature space is the same as the total variance of the data in the original attribute space.
Let R^n be the original space of **D** and S^n be the transformed (feature space). Then, EOF analysis is the discovery of a function:

$$f : R^n \rightarrow S^n. \tag{1}$$

If the total variance (standard deviation) of **D** in R^n is $\phi(\mathbf{D})$, then the total variance of **D** in S^n is also $\phi(\mathbf{D})$. The variance is written as a scalar here to emphasis the fact that it is the total variance and not the attributewise variance. We shall call **D** in S^n as $\mathbf{D_S}$. It can be shown that the optimum choice for S^n is the system in which the axes are the eigenvectors of the co-variance (correlation) matrix of **D** [10]. Let $\{e_i | 1 \leq i \leq n\}$ be the set of eigenvectors of the correlation matrix of **D**. Then, each e_i is an n-dimensional vector in R^n. Each e_i is mathematically viewed as a column vector of n dimensions. We have

$$S^n = \{e_i | 1 \leq i \leq n\}. \tag{2}$$

As pointed out above, we have $\phi(\mathbf{D}) = \phi(\mathbf{D_S})$ but with two differences: (i) the features (axes) in S^n are uncorrelated and (ii) the variances of e_i are such that a few e_i explain, or account for, maximum amount of $\mathbf{D_S}$. This enables us to discard those dimensions in S^n, or those e_i, which explain insignificant ratio of $\mathbf{D_S}$, thus reducing the dimension of the data. It can be shown that the amount of variance explained by a vector e_r of S^n is the eigenvalue of e_r. Thus, those eigenvectors from S^n are retained that explain maximum variance and the rest are discarded. We thus have the reduced feature space given by

$$U^n = \{e_i | 1 \leq i \leq m : m < n\}. \tag{3}$$

The set U^n is the set of EOFs. The analysis can be continued by representing the data **D** in U^n, called $\mathbf{D_U}$. The method incorporates two steps, i.e., (i) transform the set **D** from R^n to S^n calling it $\mathbf{D_S}$ and (ii) retain the subset $\mathbf{D_U}$ of $\mathbf{D_S}$ in U^n. Step (i) is achieved by first evaluating S^n as discussed and projecting **D** on S^n. U^n is then obtained by discarding those e_i that explain insignificant proportion of$\mathbf{D_S}$. $\mathbf{D_U}$ is the subset of $\mathbf{D_S}$ along U^n.

3 Results and Discussion

The data set is the monthly rainfall data of 142 years arranged in 12 columns. Total 19 such data sets are considered. The time series of each substation is obtained. The length of each of the 19 time series is $142 \times 12 = 1704$. Arranging the time series of 19 substations, each comprising 1704 points, as a 1704×19 matrix gives **D**. Figure 2 shows the contour plot of the correlation matrix of **D**. The numbers indicate a significantly high level of correlation. The eigenvectors of the co-variance matrix of **D** and the corresponding eigenvalues were calculated. The eigenvalues thus calculated are shown in Fig. 3 in ascending order. The percentage of correlation of **D**, i.e., percentage of $\phi(\mathbf{D})$, explained by an eigenvector e_i is

$$PVE(e_j) = \frac{EV(e_j)}{\sum_j EV(e_j)}, \tag{4}$$

where

$PVE(e_j)$: percentage variance explained by eigenvector e_j and
$EV(e_j)$: eigenvalue of the eigenvector e_j.

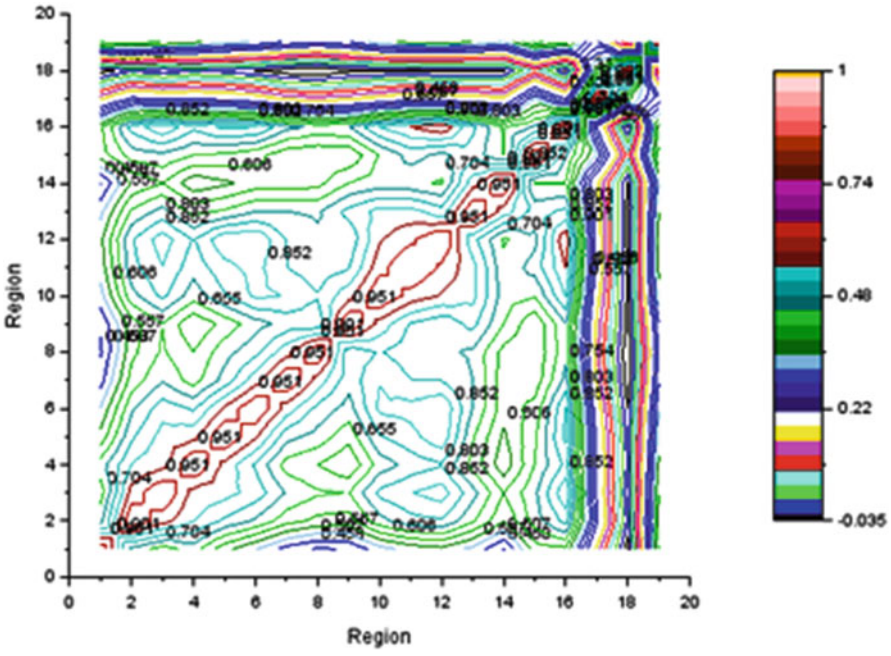


Fig. 2 Contour plot of the correlation matrix of time series, **D**, of rainfall recorded at 19 substations

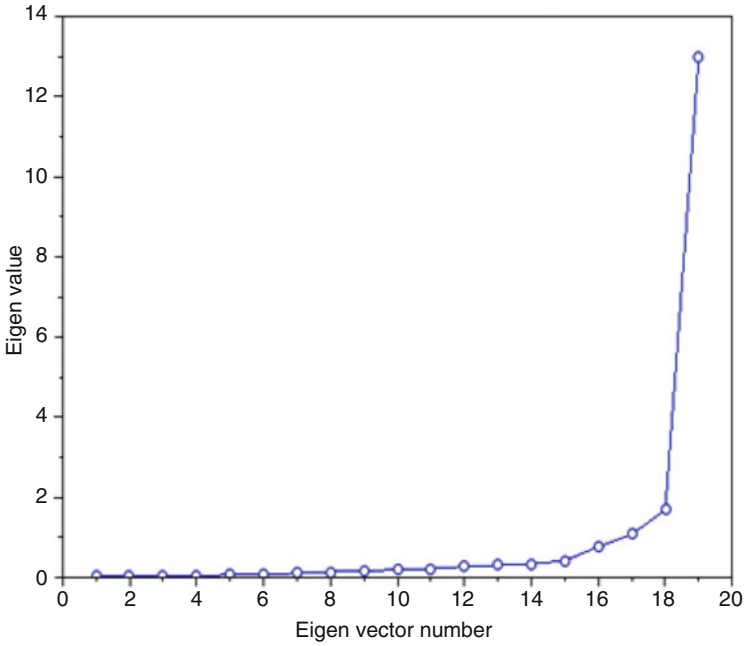


Fig. 3 Eigenvalues of the 19 eigenvectors that make S^n

It can be seen that the last 4 eigenvalues are significantly higher than the remaining ones. The percentage of variance explained by 19th eigenvector is 68.3 %, by 18th vector is 8.9%, by 17th vector is 5.7%, and by 16th vector is 4.1%. The percentage of variance explained by eigenvectors corresponding to 4 largest eigenvalues is 87%. This is a significantly higher percentage. This simply means that the information contained in the observations of 19 subdivisions can be condensed down in 4 sets of observations. The transformed space, S^n, or the eigenspace is embedded in the original space R^n as discussed; hence, all the unit vectors in U^n are represented in terms of axes of R^n.

The eigenvectors corresponding to 4 largest eigenvalues make U^n and have been plotted in Fig. 4. The x-axis represents the subdivision no in R^n, and the y-axis represents the coefficient of each subdivision in U^n. The graph is partitioned into two regions: (1) subdivisions of maximum contribution and (2) subdivisions of minimum contributions. The maximum coefficients (absolute value) are circled. It can be seen that stations 14–19 are the major contributors to U^n, the reduced or the EOF space. These subdivisions are: Saurashtra, Madhya Maharashtra, Chhattisgarh, Coastal Andhra Pradesh, Tamil Nadu, and Kerala. These are predominantly the regions in the south. The patterns in these regions can thus be said to be the major contributors in the overall variation of rainfall in the discussed 19 subdivisions.

Although not desired and not of interest, the eigenvectors corresponding to 4 smallest eigenvalues have been plotted in Fig. 5. As in the previous case, the

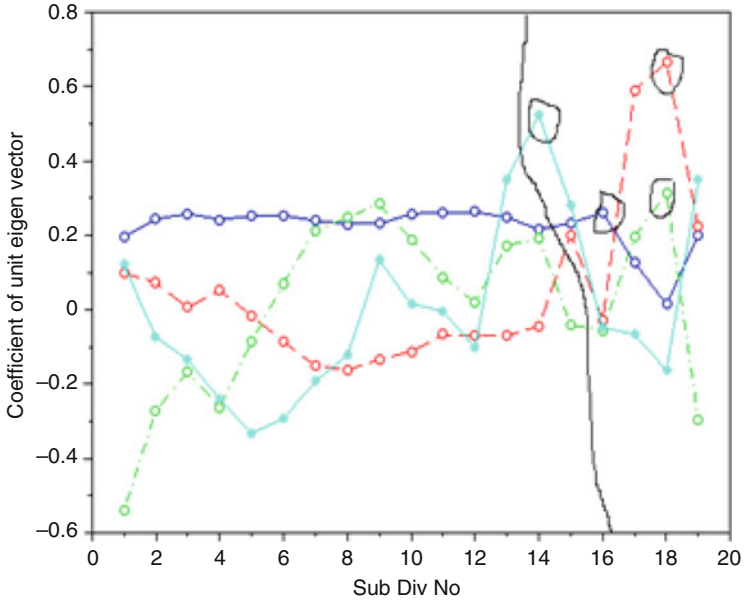


Fig. 4 Eigenvectors corresponding to 4 largest eigenvalues that make U^n

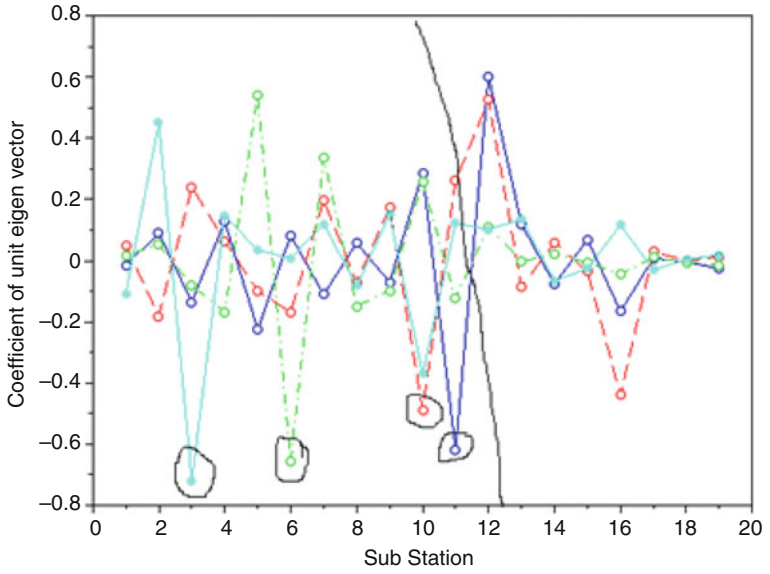


Fig. 5 Eigenvectors corresponding to 4 smallest eigenvalues of U^n

maximum contributors (absolute values) to these unit vectors are highlighted. These subdivisions are the ones numbered 12 or less. These are: Assam and Meghalaya,

Gangetic West Bengal, Jharkhand, Bihar, East UP, West UP, Haryana, Punjab, West Rajasthan, East Rajasthan, West MP, and East MP. It can be said that these regions have comparatively less impact on the overall pattern of the rainfall distributions.

4 Conclusion and Future Scope

Empirical Orthogonal Function Analysis of the 19 subdivisions of a total of 30 of homogenous Indian rainfall was done. The 19 subdivisions made the original data space for the analysis. The eigenvectors of the correlation matrix were obtained and their eigenvalues were analyzed. It was observed that 4 leading eigenvectors (vectors corresponding to 4 largest eigenvalues) or the EOFs account for about 87% of the total variance. This is a significant number and tells us that enormous information is contained in these 4 leading EOFs. The EOFs are the dimensions embedded in the original data space, and hence, the coefficients of the EOFs in the original data space give us information about the contribution of each subdivision in making of those EOFs. It was revealed that the major subdivisions that contributed to these EOFs are the Saurashtra, Madhya Maharashtra, Chhattisgarh, Coastal Andhra Pradesh, Tamil Nadu, and Kerala. These are predominantly the regions in the south. Thus, the rainfall pattern in these regions determines the EOFs that account for the maximum variation in the overall data set. It would be interesting to observe how these EOFs varied over interannual and intraseasonal scales. For this, the EOFs have to be worked out in span of 30 years as this is the duration taken by the IMD to define one climate period. This would make about 5 such analyses. Further, the EOFs during the Indian Summer Monsoon Rainfall (ISMR) period may vary considerably with climatic periods. Such studies shall have to be undertaken in future in order to better model the subdivisional level rain patterns over the Indian region. The fact that 4 leading EOFs account for about 87% of the total variance in the precipitation of 19 subdivisions also compels us to look forward to design prediction models based on these EOFs rather than the 19 subdivisions considered in isolation. In particular, statistical models such as the artificial neural networks are kept in sight.

References

1. J. Shukla, Interannual variability of monsoons, in *Monsoons*, ed. by J.S. Fein, P.L. Stephens (Wiley, Hoboken, 1987), pp. 399–464
2. B. Wang, Q. Ding, Changes in global monsoon precipitation over the past 56 years. Geophys. Res. Lett. **33**, L06711 (2006). https://doi.org/10.1029/2005GL025347
3. K.C. Tripathi, S. Rai, A.C. Pandey, I.M.L. Das, *Southern Indian Ocean SST Indices as Early Predictors of Indian Summer Monsoon Rainfall*, vol. 37 (CSIR, Delhi, 2008), pp. 70–76
4. P.J. Webster, V.O. Magana, T.N. Palmer, J. Shukla, R.A. Tomas, M. Yanai, T. Yasunari, Monsoons: processes, predictability and the prospects for prediction. J. Geophys. Res. **103**, 14451–14510 (1998)

5. B. Parthasarthy, A.A. Munot, D.R. Kothawale, Regression model for estimation of food grains production from summer monsoon rainfall. Agric. Forest Meteorol. **42**, 167–2 (1988)
6. S. Gadgil, Monsoon-ocean coupling. Curr. Sci. **78**, 309–323 (2000)
7. S. Gadgil, P.N. Vinayachandran, P.A. Francis, Droughts of the Indian summer monsoon: role of clouds over the Indian ocean. Curr. Sci. **85**, 1713–1719 (2003)
8. P.K. Xavier, B.N. Goswami, Analog method for realtime forecasting of summer monsoon subseasonal variability. Month. Weather Rev. **135**, 4149–4160 (2007)
9. V. Krishnamurthy, J. Shukla, Intraseasonal and Interannual Variability of Rainfall over India. J. Climate **13**, 4366–4377 (2000)
10. C.M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, New Delhi, 1995), pp. 140–148, 203, 267–268, 372
11. C.V. Singh, Empirical orthogonal function (EOF) analysis of monsoon rainfall and satellite-observed outgoing long-wave radiation for Indian monsoon: a comparative study. Meteorol. Atmos. Phys. **85**(4), 227–234 (2004)
12. S.V. Singh, R.S. Kriplani, Application of extended empirical orthogonal function analysis to interrelationships and sequential evolution of Monsoon fields. Month. Weather Rev. **114**, 1603–1611 (1986)
13. C.V. Singh, Principal components of monsoon rainfall in normal, flood and drought years over India. Int. J. Climatol. **19**, 639–652 (1999)
14. C. Saxena, K.C. Tripathi, P.N. Hrisheekesha, Autoassociative neural network for nonlinear principal component analysis of some atmospheric parameters, in *Proceedings of IEEE International Conference on Research and Development Prospects on Engineering and Technology, EGS Pillay Engineering College, Nagapattinam* (2013)
15. X. Chen, J.M. Wallace, Ka-Kit Tung, Pairwise-rotated EOFs of global SST. J. Clim. **30**, 5473–5489 (2017). https://doi.org/10.1175/JCLI-D-16-0786.1
16. C.H. Nnamchi, F.K. Noel, S. Keenlyside, R. Farneti, Analogous seasonal evolution of the South Atlantic SST dipole indices. Atmos. Sci. Lett. **18**, 396–402 (2017). https://doi.org/10.1002/asl.781

Forecast of Flow Characteristics in Time-Dependent Artery Having Mild Stenosis

A. K. Singh and S. P. Pandey

Abstract In this part, we have considered the blood stream, however, time-subordinate supply route with gentle stenosis. The impact of time on protection from stream ($\bar{\lambda}$), volumetric stream rate(Q), pivotal speed and shear pressure is demonstrated scientifically and graphically. Articulations for dimensionless release variable and dimensionless shear pressure variable are gotten. We have additionally thought about the trademark speed for projection. Basic estimation of Reynolds number at which partition happens has been found under this thought.

1 Introduction

The vehicle of liquids in funnels, cylinders and diverts is significant in numerous organic and biomedical frameworks, especially, in the human cardiovascular frameworks. Normally development of greasy material, for example, calcium on their inward dividers, is known as blood vessel stenosis. The affidavit of atherosclerotic plaque relies upon the geometry of the veins. The most well-known areas to deal stenosis are the arches, intersections and bifurcations of huge and medium class. It is imperative to think about the bio-liquid dynamical parts of the human cardiovascular framework, which have increased more consideration in the ongoing decades regarding the determination and the genesis of atherosclerosis.

Numerous scientists have considered the blood stream in stenosed supply routes of various geometries. An investigation of the influence of a pivotally symmetric time-subordinate development of mellow steno-sister in the lumen of a cylinder whose cross segment is consistent through which a Newtonian liquid is streaming relentlessly has been portrayed by Young [1]. Sahu et al. [2] presented a mathematical analysis of the influence of a mild stenosis on blood flow (couple stress fluid) characteristics (Fig. 1). Srinivasacharya and Srikanth [3] investigated

A. K. Singh (✉) · S. P. Pandey
RBS Engineering Technical Campus, Bichpuri, Agra, India

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_30

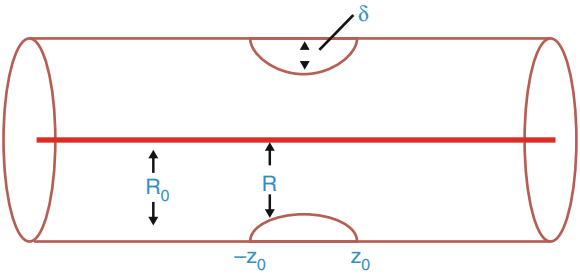


Fig. 1 Geometry of mild stenosis

the consistent spilling impact on the pulsatile nature of couple pressure liquid. Expansion of the speed and the stream rate altogether for a little addition in the couple pressure parameter has been portrayed by Adhikary and Misra [4]. The insecure laminar incompressible progression of Eyring–Powell liquid between two parallel permeable plates with variable suction or infusion speed with the thought of couples tresses and a uniform attractive field has been talked about by Rana and Khan [5]. Reddy et al. [6] added to the mathematical model for couple pressure liquid course through the stenotic annular locale and examined that the impedance has been expanding with progression in the stature and length of stenosis. Hayat et al. [7] studied the Hall consequences for the peristaltic movement of couple pressure liquid in a slanted asymmetric channel with warmth and mass exchange. Adesanya and Makinde [8] researched the impact of couple pressure liquid stream on the enduring dainty stream down warmed slanted plate and examined the impact of couple pressure parameter to chop down the stream velocity and temperature appropriation. Prakash et al. [9] reported that the size of the stenoses diminishes the volumetric stream rate and expands the divider shear worry just as impedance. Prakash and Makinde [10] observed that the impedance is diminished because of the attractive field impact, when patients experienced thermal radiation therapy. Tiwari and Chauhan [11] discussed the effect of plasma layer thickness, varying viscosity, yield stress, permeability and viscosity ratio parameter on the flow variables. Bhatti et al. [12] proposed slip impacts and endoscopy investigation on blood stream of molecule liquid suspension. They explored that weight rise diminishes because of the impact of molecule volume division and friction powers likewise decrease because of the effect of molecule volume part.

2 Mathematical Formulation

The time pace of the sweep of corridor $R(z)$ can be characterized as

$$R = R_0 - \tau\beta_0(1 - e^{-\frac{t}{\tau}})\left(1 + \cos\frac{\pi z}{z_0}\right); -z_0 \le z \le z_0 R = R_0; |z| > z_0 \quad (1)$$

$$\frac{\partial p}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right) = 0 \quad (2)$$

The fundamental condition of movement in barrel-shaped polar directions

$$\frac{\partial p}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right) = 0 \quad (3)$$

$$\frac{\partial p}{\partial r} = 0 \quad (4)$$

Equation (2) can be written as

$$-G = \frac{\mu}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right) \quad (5)$$

where $G = -\frac{\partial p}{\partial z}$. No slip conditions on the stenosis surface are

$$w = 0\, at\, r = R(z) - z_0 \le z \le z_0 \quad (6)$$

$$w = 0\, at\, r = R_0\, |z| \ge z_0 \quad (7)$$

Integrating Eq. (5), one obtains

$$r\frac{\partial w}{\partial r} = -G\frac{r^2}{2\mu} + c_1 \quad (8)$$

Since

$$\frac{\partial w}{\partial r} = 0 \quad (9)$$

on the axis implies that $c_1 = 0$

$$\Rightarrow r\frac{\partial w}{\partial r} = -G\frac{r^2}{2\mu} \quad (10)$$

Integrating (8) and using (6), we get

$$w = -\frac{G}{4\mu}\left(r^2 - R^2\right) \quad (11)$$

Volumetric flow rate through the artery is

$$Q = \int_0^R 2\pi r w dr = \frac{\pi G}{8\mu} R^4 \tag{12}$$

From Eq. (9),

$$G(z) = -\frac{\partial p}{\partial z} = \frac{8\mu Q}{\pi R^4} \tag{13}$$

Integrating (11) along length of the artery and $p = p_1$ at $z = -L$ and $p = p_2$ at $z = L$, we obtain

$$\Delta p = \frac{8\mu Q}{\pi} \int_{-L}^{L} \frac{1}{R^4} dz = \frac{8\mu Q}{\pi R_0^4} \int_{-L}^{L} \frac{1}{\left(\frac{R}{R_0}\right)^4} dz \tag{14}$$

Resistance to flow can be defined as

$$\lambda = \frac{\Delta p}{Q} = \frac{8\mu}{\pi R_0^4} \int_{-L}^{L} \frac{1}{\left(\frac{R}{R_0}\right)^4} dz \tag{15}$$

$$\lambda = \frac{\Delta p}{Q} = \frac{8\mu}{\pi R_0^4} \left[\int_{-L}^{-z_0} \frac{1}{(R/R_0)^4} dz + \int_{-z_0}^{z_0} \frac{1}{(R/R_0)^4} dz + \int_{z_0}^{L} \frac{1}{(R/R_0)^4} dz \right] \tag{16}$$

$$\lambda = \frac{\Delta p}{Q} = \frac{16\mu}{\pi R_0^4} \left[L - z_0 + \int_0^{z_0} \frac{1}{(R/R_0)^4} dz \right] \tag{17}$$

In the normal condition,

$$\lambda_N = \frac{16\mu L}{\pi R_0^4} \tag{18}$$

Resistance to flow ratio can be written as

$$\bar{\lambda} = \frac{\lambda}{\lambda_N} = \left[1 - \frac{z_0}{L} + \frac{1}{L} \int_0^{z_0} \frac{1}{(R/R_0)^4} dz \right] \tag{19}$$

where R/R_0 can be taken from Eq. (1). Wall shear stress (τ_w) is given by the relation:

$$\tau_w = -\frac{R}{2} \frac{\partial p}{\partial z} \tag{20}$$

Using Eq. (11) in Eq. (18), we get

$$\tau_w = \frac{4\mu Q}{\pi R^3} \tag{21}$$

In normal situation,

$$\tau_N = \frac{4\mu Q}{\pi R_0^3} \tag{22}$$

Wall shear stress ratio is given as

$$\bar{\tau}_w = \frac{\tau_w}{\tau_N} = \left(\frac{R}{R_0}\right)^{-3} \tag{23}$$

The vein having the stenosis supplies blood to a specific vascular bed, and it is assumed that the all out weight drop over the course and the vascular bed $(p_1 - p_3)$ are consistent. The all out weight drop can be communicated as

$$p_1 - p_3 = p_1 - p_2 + p_2 - p_3$$

$$\Rightarrow \left(\frac{p_1 - p_3}{Q}\right) = \left(\frac{p_1 - p_2}{Q}\right) + \left(\frac{p_2 - p_3}{Q}\right)$$

$$\Rightarrow \lambda_{13} = \lambda_{12} + \lambda_{23} \tag{24}$$

It is assumed that

$$\lambda_{23} = M(\lambda_{12})_p \tag{25}$$

where $(\lambda_{12})_p$is the resistance to flow of the artery supplying the vascular bed based on Poiseuille flow and M is a constant.
From Eqs. (22) and (23), the expression for dimensionless discharge parameter is given by

$$\frac{(\lambda_{12})_p Q}{p_1 - p_3} = \left[\frac{\lambda_{12}}{(\lambda_{12})_p} + M\right]^{-1} \tag{26}$$

Suppose a stenosis be specified for maximum stenosis height $\frac{\delta_m}{R_0} = 0.2$.
Then,

$$\frac{\delta}{R_0} = 0.2(1 - e^{-t/\tau}) \tag{27}$$

The expression for the maximum wall shear stress in the stenosis can be written as

$$\frac{(\lambda_{12})_p \pi R^3 \tau_w}{4\mu(p_1 - p_3)} = \left[\frac{\lambda_{12}}{(\lambda_{12})_p} + M\right]^{-1} \tag{28}$$

$$\Rightarrow \frac{(\lambda_{12})_p \pi R_0^3 \tau_w}{4\mu(p_1 - p_3)} = \left[\frac{\lambda_{12}}{(\lambda_{12})_p} + M\right]^{-1} \left(\frac{R}{R_0}\right)^{-3} \tag{29}$$

Prediction of Separation

The previous analysis is based on the condition that viscous forces are much larger than inertial forces. Inertial impacts are because of the convective speeding-up terms in the Navier–Stokes condition. Clearly as the size of stenosis increments or Reynolds number builds, the significance of inertial terms cannot be ignored. Two significant impacts because of inertial powers are (i) lower weight at limited area of stenosis because of Bernoulli impact and (ii) separation. The improvement of stenosis happens because of division.

The previous analysis is based on the condition that viscous forces are much larger than inertial forces. Inertial effects are due to the convective acceleration terms in the Navier–Stokes equation. It is obvious that as the size of stenosis increases or Reynolds number increases, the importance of inertial terms cannot be neglected. Two important effects due to inertial forces are (i) lower pressure at narrowed section of stenosis due to Bernoulli effect and (ii) separation. The development of stenosis occurs due to separation.

Characteristic velocity for the bulge is assumed to be given by the equation:

$$v_\delta = 2U\left\{1 - \left(1 - \frac{\delta}{R_0}\right)^2\right\} \tag{30}$$

As $\frac{\delta}{R_0} << 1$,

$$\Rightarrow v_\delta \cong 4U\left(\frac{\delta}{R_0}\right) \tag{31}$$

The Reynolds number is

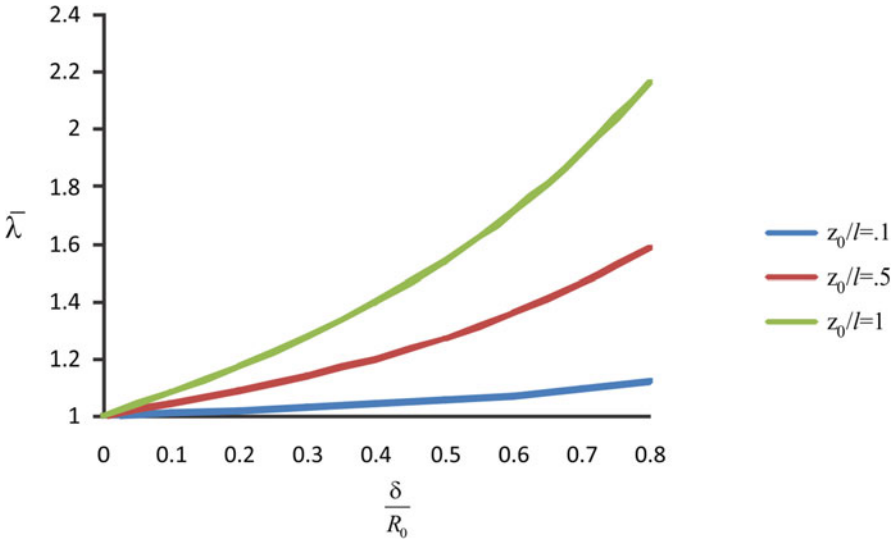
$$(R_e)_\delta = \frac{\delta v_\delta}{v} = 4\left(\frac{\delta}{R_0}\right)^2 \frac{UR_0}{v} \tag{32}$$

In Eq. (29), it is assumed that when Reynolds number reaches some critical value R_{crit}, separation will occur. Thus the condition for separation is

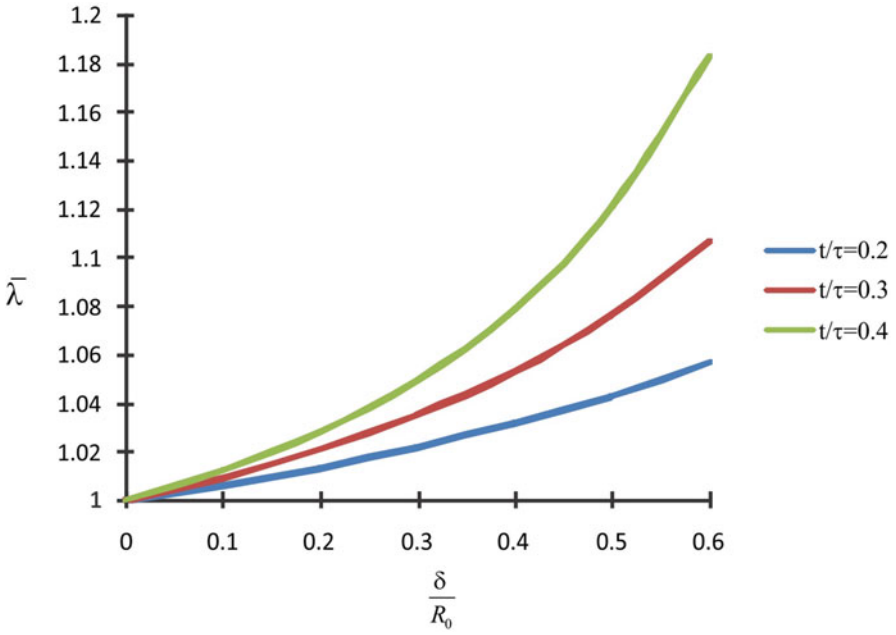
$$\frac{R_{crit}}{4} = \left(\frac{\delta}{R_0}\right)^2 R_0 \tag{33}$$

3 Results and Discussion

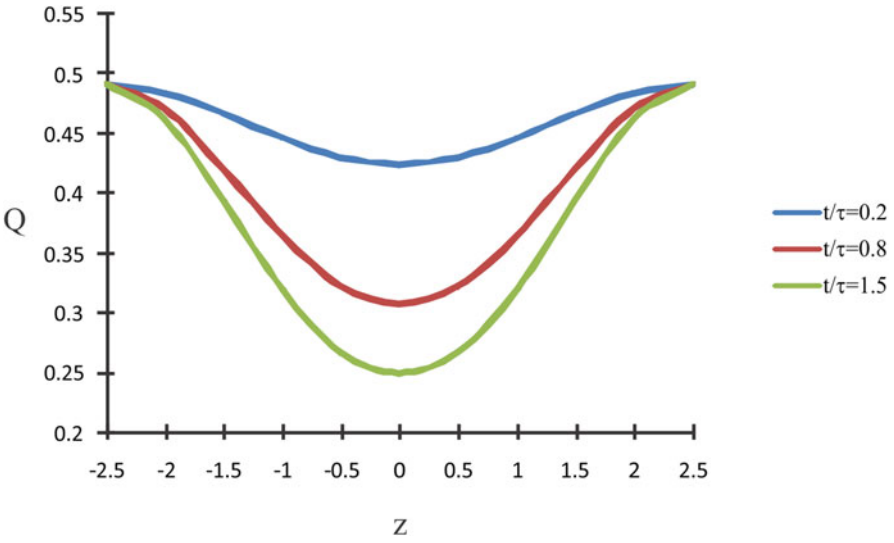
The impact of stenosis stature on protection from stream has appeared in charts (Graph 1 and Graph 2). The outcomes demonstrate that protection from stream increments as stenosis stature increments. The bend marked $z_0/l = 1$ shows transparently the impact of the stenosis on the protection from stream. The point to be noticed is that for $\delta/R_0 = .1$, the protection from stream increments for a perpetual-width tube by about 25%. The bend named $z_0/l = .1$ is explaining the way that if the protection from stream over a long section of vein is considered, the impact of the stenosis is little until a specific estimation of δ/R_0 is surpassed. Past this basic estimation of δ/R_0, the nearness of the stenosis quickly ends up huge. It ought to be stressed that for a gentle stenosis, the adjustment in the genuine weight at a point in the supply route because of the stenosis will in any case be little in contrast with the mean blood vessel weight. Diagram (Graph 2) delineates that this protection from stream increments as the dimensionless time t/τ increments. The variety of volumetric stream rate (Q) with hub separation (z) for various estimations of dimensionless time t/τ is introduced in chart (Graph 3). As the time t/τ expands, the stream rate diminishes. The variety of hub speed profile with the spiral arrangement have appeared in diagram (Graph 4). It is seen that the pivotal speed accomplishing the most extreme extents at the hub ($r = 0$) and least at the limit ($r = R$). The variety in divider shear worry all through the pivotal separation with time has appeared in chart (Graph 5). It is observed that wall shear stress steeply increases in the upstream from its approached value to the peak value at the throat, and decreases in the downstream of the throat. Further those dividers' shear pressure increments as time increments. Restricting conditions for δ/R_0 and Reynolds number (R_e) for different estimations of R_{crit} are introduced in chart (Graph 6). The inexact idea of this examination ought to be perceived and worth got from chart (Graph 6) must be utilized as unpleasant assessments for anticipating partition. It is evident from the chart that, in any event, for mellow stenosis, detachment may happen at a generally little Reynolds number. For instance, for $\delta/R_0 = .1$, the constraining estimation of the corridor Reynolds number (R_e) is around 125, which depends on $R_{crit} = 5$. There will be no division anticipated for the focuses falling beneath the bend (Graph 6). For $z_0/l = 1$ and $M = 10$, the variety in the release parameter and the most extreme divider shearing worry in stenosis have been plotted versus dimensionless time in charts (Graph 7 and Graph 8) separately. It is seen that release is diminishing gradually with the time despite the fact that shear pressure increments quickly. For $t/\tau = 1$, the discharge has decreased by about 2% and the shear has increased approximately 50%.



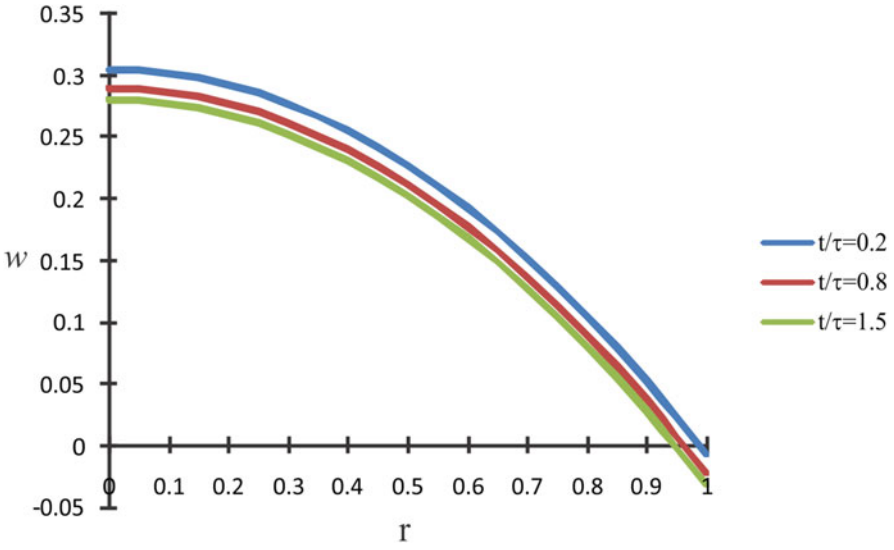
Graph 1 Variation in resistance to flow with stenosis height for different values of z_0/l



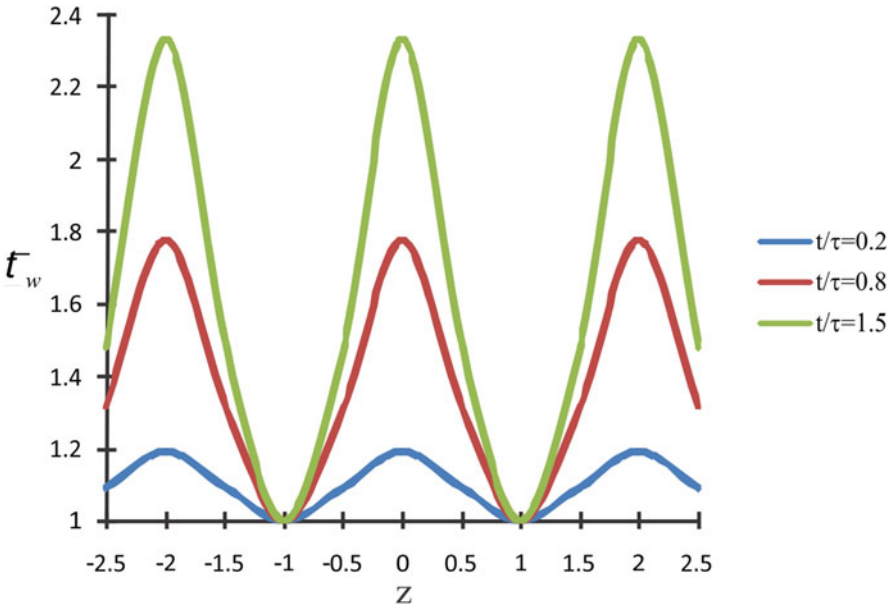
Graph 2 Variation in resistance to flow with stenosis height for different values of time



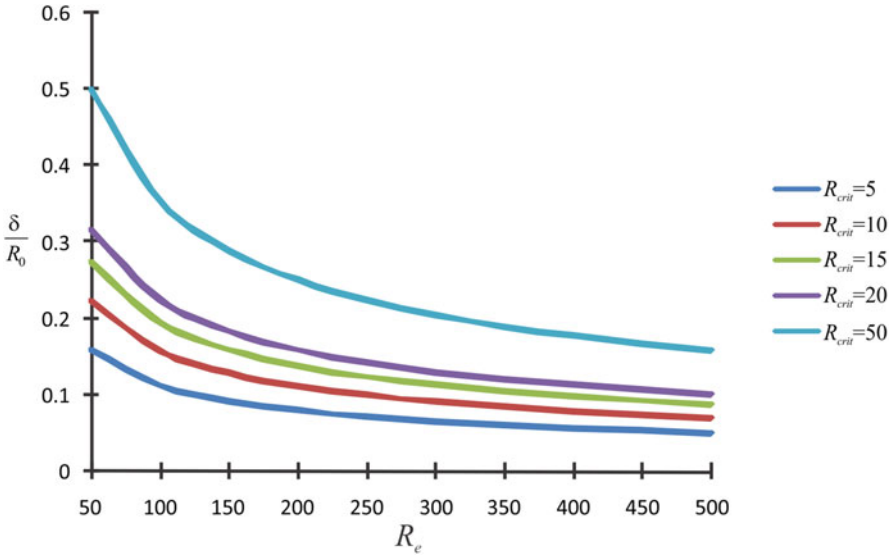
Graph 3 Variation in volumetric flow rate with axial distance for different time



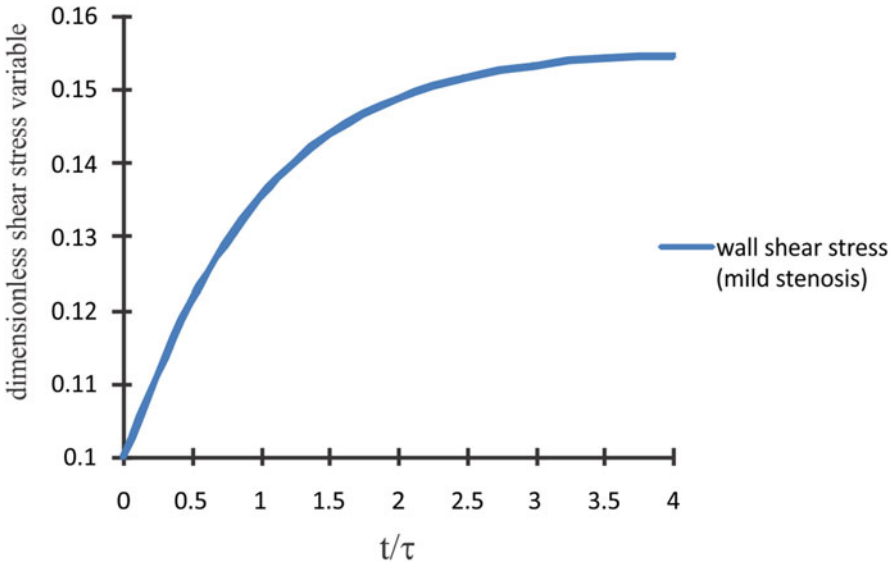
Graph 4 Variation in axial velocity with radial distance for different time



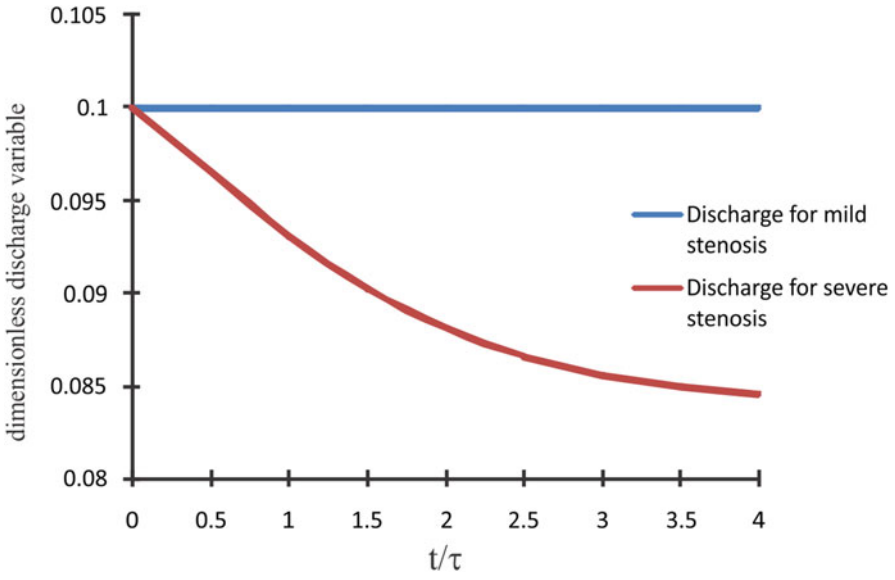
Graph 5 Variation in shear stress with axial distance for different time



Graph 6 Variation in stenosis height with Reynolds number for different values of R_{crit}



Graph 7 Variation in dimensionless discharge variable with time



Graph 8 Variation in dimensionless shear stress with time

4 Closing Comments

The arrangement of a stenosis in a vein may cause numerous serious issues like atherosclerosis and upset the normal capacity of circulatory framework. Expectation of the stream normal for such sort of issue is troublesome, and different rearranging propositions are required to build up a tractable model. In this chapter, Flow of blood through a time dependent artery with axially symmetric stenosis is considered. A surmised arrangement with time subordinate geometric design for a mellow stenosis is obtained.

References

1. D.F. Young, Effect of a time-dependent stenosis on flow through a tube. J. Manuf. Sci. Eng. **90**, 248–54 (1968)
2. M.K. Sahu, S.K. Sharma, A.K. Agrawal, Study of arterial blood flow in stenosed vessel using non-Newtonian couple stress fluid model. Int. J. Dynam. Fluids **6**(2), 248–257 (2010)
3. D. Srinivasacharya, D. Srikanth, Steady streaming effect on the flow of a couple stress fluid through a constricted annulus. ArchMech **64**(2), 137–152 (2012)
4. S.D. Adhikary, J.C. Misra, Pulsating flow of a couple stress fluid in a channel with permeable walls. Forsch Ingenieurwes **77**, 49–57 (2013)
5. M.A. Rana, N. Khan, Effects of couple stresses and variable suction/injection on the unsteady MHD flow of an Eyring Powell fluid between two parallel porous plates. Life Sci. J. **11**(4s), 105–112 (2014)
6. J.V.R. Reddy, D. Srikanth, S.K. Murthy, Mathematical modelling of couple stresses on fluid flow in constricted tapered artery in presence of slip velocity-effects of catheter. Appl. Math. Mech. **35**(8), 947–58 (2014)
7. T. Hayat, M. Iqbal, H. Yasmin, F. Alsaadi, Hall effects on peristaltic flow of couple stress fluid in an inclined asymmetric channel. Int. J. Biomath. **7**(5), 1450047 (2014)
8. S.O. Adesanya, O.D. Makinde, Irreversibility analysis in a couple stress film flow along an inclined heated plate with adiabatic free surface. Physica **432**, 222–229 (2015)
9. Om. Prakash, O.D. Makinde, S.P. Singh, N. Jain, Effects of stenoses on non-Newtonian flow of blood in blood vessels. Int. J. Biomath. **8**(1), 1550010 (2015). http://doi.org/10.1142/S1793524515500102, 13p.
10. J. Prakashand, O.D. Makinde, Radiative heat transfer to blood flow through a stenotic artery in the presence of magnetic field. Lat. Am. Appl. Res. **41**(3), 273–277 (2011)
11. A. Tiwari, S.S. Chauhan, Effect of varying viscosity on two-fluid model of pulsatile blood flow through porous blood vessels: A comparative study. Microvascular Research **123**, 99–110 (2019)
12. M.M. Bhatti, A. Zeeshan, N. Ijaj, Slip effects and endoscopy analysis on blood flow of particle-fluid suspension induced by peristaltic wave. J. Mol. Liquids **218**, 240–245 (2016)

Applicability of Measure of Noncompactness for the Boundary Value Problems in ℓ_p Spaces

Tanweer Jalal and Ishfaq Ahmad Malik

Abstract In this paper, we prove the existence of solution for the boundary value problem for an infinite system of second-order differential equations in ℓ_p space of the form:

$$\frac{\mathrm{d}^2 v_j}{\mathrm{d}t^2} + v_j = f_j(t, v(t)),$$

where $v_j(0) = v_j(T) = 0, t \in [0, T]$, $v(t) = \left(v_j(t)\right)_{j=1}^{\infty}$, and $j = 1, 2, \ldots$.

By applying the concept of measures of non-compactness this boundary value problem is first changed into an equivalent system of integral equations, then the result is obtained for the system of integral equations by using Darbo's fixed point theorem. The result is applied to an example to illustrate the concept.

1 Introduction and Preliminaries

Measures of noncompactness are very important concept widely used in fixed point theory, differential equations, functional equations, integral and integro-differential equations, etc. The fixed point arguments have been used in the study of existence of solutions to functional equations, for instance, the Banach contraction [1, 14, 29, 30] and Schauder's fixed point theorem [12, 15, 16, 18]. These theorems cannot be used in case the compacity and the Lipschitz condition are not satisfied. The measure of noncompactness argument appears as most convenient and useful in such cases.

It was Kuratowski [17], who introduced the idea of measure of noncompactness, and then Banaś and Goebel [5] gave an axiomatic approach to it, in general, a Banach space. Darbo [9] first used the idea of measure of noncompactness to come up with a fixed point theorem for condensing operators, which generalized

T. Jalal (✉) · I. A. Malik
National Institute of Technology, Srinagar, India
e-mail: tjalal@nitsri.net; ishfaq_2phd15@nitsri.net

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_31

the classical Schauder's fixed point theorem and a special type of the Banach contraction principle. In [2–4, 6–8, 20, 21, 25, 26, 28], an infinite system of differential equations has been studied using the idea of measure of noncompactness in a different Banach space.

Let $(X, \|\cdot\|)$ be a Banach space, for any $E \subset X$, $\bar{E}$ denotes closure of E and $conv(E)$ denotes the closed convex hull of E. We denote the family of non-empty bounded subsets of X by $\mathfrak{M}_X$ and family of non-empty and relatively compact subsets of X by $\mathfrak{N}_X$. Let $\mathbb{N}$ denote the set of natural numbers and $\mathbb{R}$ the set of real numbers for $\mathbb{R}_+ = [0, \infty)$, and the axiomatic definition of measure of noncompactness is defined below.

Definition 1 ([7]) A mapping $\chi : \mathfrak{M}_X \to \mathbb{R}_+$ is said to be the measure of noncompactness in X if the following conditions hold:

(i) The family $\text{Ker}\chi = \{E \in \mathfrak{M}_X : \chi(E) = 0\}$ is non-empty and $\text{Ker}\chi \subset \mathfrak{N}_X$;
(ii) $E_1 \subset E_2 \Rightarrow \chi(E_1) \leq \chi(E_2)$;
(iii) $\chi(\bar{E}) = \chi(E)$;
(iv) $\chi(convE) = \chi(E)$;
(v) $\chi[\lambda E_1 + (1-\lambda)E_2] \leq \lambda\chi(E_1) + (1-\lambda)\chi(E_2)$ for $0 \leq \lambda \leq 1$;
(vi) If (E_n) is a sequence of closed sets from $\mathfrak{M}_X$ such that $E_{n+1} \subset E_n$ and $\lim_{n\to\infty} \chi(E_n) = 0$, then the intersection set $E_\infty = \bigcap_{n=1}^{\infty} E_n$ is non-empty.

Further properties of Hausdorff measure of noncompactness χ can be found in [5, 7].

The following Darbo's fixed point theorem will be utilized in our further consideration.

Lemma 1 ([9]) *Let E be a non-empty, bounded, closed and convex subset of Banach space X, with χ as measure of noncompactness, and let $T : E \to E$ be a continuous mapping. Assume that there exists a constant $k \in [0, 1)$ such that $\chi(T(E)) \leq k\chi(E)$ for any non-empty subset E of X. Then, T has a fixed point in the set E.*

The idea of equicontinuous sets is defined as follows:

Definition 2 (Equicontinuous) Let (Ω_1, d) and (Ω_2, d) be two metric spaces and $\mathcal{T}$ the family of functions from Ω_1 to Ω_2. The family $\mathcal{T}$ is equicontinuous at a point $m_0 \in \Omega_1$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that $d(f(m), f(m_0)) < \epsilon$ for all $f \in \mathcal{T}$ and all $m \in \Omega_1$ such that $d(m, m_0) < \delta$. The family is pointwise equicontinuous if it is equicontinuous at each point of Ω_1.

For fixed p , $p \geq 1$, we denote by ℓ_p the Banach sequence space with $\|\cdot\|_p$ norm defined as

$$\|x\|_p = \|(x_n)\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{\frac{1}{p}}$$

for $x = (x_n) \in \ell_p$. In order to apply Lemma 1 in a given Banach space X, we need a formula expressing the measure of noncompactness by a simple formula. Such formulas are known only in a few spaces [5, 7].

For the Banach sequence space $\left(\ell_p, \|\cdot\|_p\right)$, Hausdorff measure of noncompactness is given by

$$\chi(E) = \lim_{n\to\infty} \left\{ \sup_{(e_k)\in E} \left(\sum_{k\geq n} |e_k|^p \right)^{\frac{1}{p}} \right\}, \tag{1}$$

where $E \in \mathfrak{M}_{\ell_p}$. The above formulas will be used in the sequel of the chapter.

In [3, 8, 27], an infinite system of second-order differential equations of the form

$$\frac{\mathrm{d}^2 u_i}{\mathrm{d}t^2} = -f_i(t, u_1, u_2, \ldots) \ , \qquad u_i(0) = u_i(T) = 0 \ , \ i \in \mathbb{N} \ , \ t \in [0, T],$$

is studied in different Banach spaces.

In this chapter, our consideration is

$$\frac{\mathrm{d}^2 v_j}{\mathrm{d}t^2} + v_j = f_j(t, v(t)), \tag{2}$$

where $t \in [0, T]$, $v(t) = \left(v_j(t)\right)_{j=1}^{\infty}$ and $j = 1, 2, \ldots$, in ℓ_p Banach space.

The above system will be studied together with the boundary problem

$$v_j(0) = v_j(T) = 0. \tag{3}$$

The solution is investigated using the infinite system of integral equations and Green's function [13]. Such systems appear in the study of theory of neural sets, theory of branching process and theory of dissociation of polymers [10, 11, 23].

In this chapter, we find the conditions under which the system given in (2) under the boundary condition (3) has solution in the Banach sequence space ℓ_p, and to do so, we define an equivalent infinite system of integral equations. The result is supported by an example.

2 Main Results

By $C(I, \mathbb{R})$ we denote the space of continuously differentiable functions on $I = [a, b]$ and by $C^2(I, \mathbb{R})$ the space of twice continuously differentiable functions on $I = [a, b]$. A function $v \in C^2(I, \mathbb{R})$ is a solution of (2) if and only if v is a solution of the infinite system of integral equations

$$v_j(t) = \int_0^T R(s,t) f_j(s, v(s)) ds \ , \quad t \in I, \tag{4}$$

where Green's function $R(s,t)$ defined on the square I^2 as

$$R(s,t) = \begin{cases} \frac{\sin(t)\sin(T-s)}{\sin(T)} : 0 \le s < t \le T, \\ \frac{\sin(s)\sin(T-t)}{\sin(T)} : 0 \le t < s \le T. \end{cases} \tag{5}$$

This function satisfies the inequality

$$R(s,t) \le \frac{1}{2} \tan(0.5T) \tag{6}$$

for all $(t,s) \in I^2$.

From (4) and (5), we obtain

$$v_j(t) = \int_0^t \frac{\sin(t)\sin(T-s)}{\sin(T)} f_j(s, v(s)) ds + \int_t^T \frac{\sin(s)\sin(T-t)}{\sin(T)} f_j(s, v(s)) ds.$$

Differentiation gives

$$\frac{dv_j}{dt} = \int_0^t \frac{\cos(t)\sin(T-s)}{\sin(T)} f_j(s, v(s)) ds + \int_t^T \frac{-\sin(s)\cos(T-t)}{\sin(T)} f_j(s, v(s)) ds.$$

Again, differentiating gives

$$\begin{aligned} \frac{d^2 v_j}{dt^2} &= \int_0^t \frac{-\sin(t)\sin(T-s)}{\sin(T)} f_j(s, v(s)) ds + \frac{\cos(t)\sin(T-t)}{\sin(T)} f_j(t, v(t)) \\ &\quad + \int_t^T \frac{-\sin(s)\sin(T-t)}{\sin(T)} f_j(s, v(s)) ds + \frac{\sin(t)\cos(T-t)}{\sin(T)} f_j(t, v(t)) \\ &= -\int_0^T R(s,t) f_j(s, v(s)) ds \\ &\quad + \frac{1}{\sin(T)} [\sin(t)\cos(T-t) + \cos(t)\sin(T-t)] f_j(t, v(t)) \\ &= -v_j(t) + f_j(t, v(t)). \end{aligned}$$

Thus, $v_j(t)$ given in (4) satisfies (2). Hence, finding the existence of solution for the system (2) with boundary conditions (3) is equivalent to finding the existence of solution for the infinite system of integral equations (4).

Remark 1 If X is a Banach space and χ_X denotes its Hausdorff measure of noncompactness, then the Hausdorff measure of noncompactness of a subset E of

$C(I, X)$, the Banach space of continuous functions is given by [5, 19, 22, 24]

$$\chi(E) = \sup\{\chi_X(X(t)) : t \in I\}.$$

where E is equicontinuous on the interval $I = [0, T]$,

In order to find the condition under which the system (4) has a solution in ℓ_p, we need the following assumptions:

($\mathbf{A}_1$) The functions f_j are real-valued, defined on the set $I \times \mathbb{R}^\infty$, $(j = 1, 2, 3, \ldots)$.

($\mathbf{A}_2$) An operator f defined on the space $I \times \ell_p$ as

$$(t, v) \mapsto (fv)(t) = \left(f_j(t, v)\right) = (f_1(t, v), f_2(t, v), f_3(t, v), \ldots)$$

transforms the space $I \times \ell_p$ into ℓ_p.

The class of all functions $\{(fv)(t)\}_{t\in I}$ is equicontinuous at each point of the space ℓ_p. That is, for each $v \in \ell_p$, fixed arbitrarily and given $\epsilon > 0$, there exists $\delta > 0$ such that whenever $\|u - v\|_p < \delta$

$$\|(fu)(t) - (fv)(t)\|_p < \epsilon. \tag{7}$$

($\mathbf{A}_3$) For each fixed $t \in I$, $v = \left(v_j\right) \in \ell_p$, the following inequality holds:

$$\left|f_j(t, v(t))\right|^p \le g_j(t) + h_j(t)|v_j|^p \qquad n \in \mathbb{N}, \tag{8}$$

where $h_j(t)$ and $g_j(t)$ are real-valued continuous functions on I. The function g_j $(j = 1, 2, \ldots)$ is continuous on I, and the function series $\sum_{b\ge1} g_b(t)$ is uniformly convergent. Also, the function sequence $\left(h_j(t)\right)_{j\in\mathbb{N}}$ is equibounded on I.

To prove the general result, we set the following constants:

$$g(t) = \sum_{j=1}^{\infty} g_j(t),$$

$$G = \max\{g(t) : t \in I\},$$

$$H = \sup\left\{h_j(t) : t \in I, j \in \mathbb{N}\right\}.$$

Theorem 1 *Under the assumptions* $(\boldsymbol{A}_1) - (\boldsymbol{A}_3)$*, with* $(HT)^{\frac{1}{p}} \tan(0.5T) < 2$*,* $T \neq (2n-1)\pi$*,* $n = 1, 2, \ldots$*, the infinite system of integral equations* (4) *has at least one solution* $v(t) = (v_j(t))$ *in* ℓ_p *space* $p \ge 1$*, for each fixed* $t \in I$*.*

Proof We consider the space $C(I, \ell_p)$ of all continuous functions on $I = [0, T]$ with supremum norm given as

$$\|v\| = \sup_{t\in I} \left\{ \|v(t)\|_p \right\}.$$

Define the operator $\mathcal{F}$ on the space $C(I, \ell_p)$ by

$$\begin{aligned} (\mathcal{F}v)(t) &= \left((\mathcal{F}v)_j(t)\right) \\ &= \left(\int_0^T R(s,t) f_j(s, v(s)) ds \right) \\ &= \left(\int_0^T R(s,t) f_1(s, v(s)) ds, \int_0^T R(s,t) f_2(s, v(s)) ds, \dots \right). \end{aligned} \tag{9}$$

The operator $\mathcal{F}$ as defined in (9) transforms the space $C(I, \ell_p)$ into itself, which we will show. Fix $v = v(t) = \left(v_j(t)\right)$ in $C(I, \ell_p)$; then, for arbitrary $t \in I$ using assumption $(\mathbf{A}_3)$, inequality (6) and Hölder's inequality, we have

$$\begin{aligned} \left(\|(\mathcal{F}v)(t)\|_p\right)^p &= \sum_{j=1}^{\infty} \left| R(s,t) f_j(s, v(s)) ds \right|^p \\ &\le \sum_{j=1}^{\infty} \left\{ \int_0^T |R(s,t)|^p |f_j(s, v(s))|^p ds \right\} \left(\int_0^T ds \right)^{\frac{p}{q}} \\ &\le (T)^{\frac{p}{q}} \sum_{j=1}^{\infty} \left\{ \int_0^T |R(s,t)|^p \left[g_j(s) + h_j(s) |v_j(s)|^p \right] ds \right\} \\ &\le \left(\frac{1}{2} \tan(0.5T) \right)^p (T)^{\frac{p}{q}} \sum_{j=1}^{\infty} \left[\int_0^T g_j(s) ds + \int_0^T h_j(s) |v_j(s)|^p ds \right]. \end{aligned}$$

Now, using Lebesgue's dominated convergence theorem, we get

$$\begin{aligned} \left(\|(\mathcal{F}(v)(t)\|_p\right)^p &\le \left(\frac{T^{\frac{1}{q}}}{2} \tan(0.5T) \right)^p \left(\int_0^T g(s) ds + H \int_0^T \sum_{j=1}^{\infty} |v_j(s)|^p ds \right) \\ &\le \left(\frac{T^{\frac{1}{q}}}{2} \tan(0.5T) \right)^p \left(GT + HT \left(\|v\|_p \right)^p \right) \\ &= \left(\frac{T}{2} \tan(0.5T) \right)^p \left(G + H \left(\|v\|_p \right)^p \right). \end{aligned}$$

Therefore,

$$\left(\|(\mathcal{F}(v)(t)\|_p\right)^p \leq \left(\frac{T}{2}\tan(0.5T)\right)^p \left(G + H\left(\|v\|_p\right)^p\right). \tag{10}$$

Hence, $\mathcal{F}v$ is bounded on the interval I. Thus, $\mathcal{F}$ transforms the space $C(I, \ell_p)$ into itself. From (10), we get

$$\|(\mathcal{F}(v)(t)\|_p \leq \frac{T}{2}\tan(0.5T)\left(G + H\left(\|v\|_p\right)^p\right)^{\frac{1}{p}}. \tag{11}$$

Now, using (4) and following the procedure as above, we get

$$\begin{aligned}
\left(\|v\|_p\right)^p &\leq \left(\frac{T}{2}\tan(0.5T)\right)^p \left(G + H\left(\|v\|_p\right)^p\right) \\
\Rightarrow \left(\|v\|_p\right)^p &\leq \frac{G\left(T\tan(0.5T)\right)^p}{2^p - H(T\tan(0.5T))^p} \\
\Rightarrow \|v\|_p &\leq \frac{G^{\frac{1}{p}}\left(T\tan(0.5T)\right)}{[2^p - H(T\tan(0.5T))^p]^{\frac{1}{p}}}.
\end{aligned}$$

Thus, the positive number

$$r = \frac{G^{\frac{1}{p}}\left(T\tan(0.5T)\right)}{[2^p - H(T\tan(0.5T))^p]^{\frac{1}{p}}}$$

is the optimal solution of the inequality $\frac{T}{2}\tan(0.5T)\left(G + HR^p\right)^{\frac{1}{p}} \leq R$.

Hence, by (11), the operator $\mathcal{F}$ transforms the ball $B_r \subset C(I, \ell_p)$ into itself.

Further, we show that $\mathcal{F}$ is continuous on B_r. Let $\epsilon > 0$ be arbitrarily fixed and $v = (v(t)) \in B_r$ be any arbitrarily fixed function, and then if $u = (u(t)) \in B_r$ is any function such that $\|u - v\| < \epsilon$, then for any $t \in I$, we have

$$\begin{aligned}
\left(\|(\mathcal{F}u)(t) - (\mathcal{F}v)(t)\|_p\right)^p &= \sum_{j=1}^{\infty}\left|\int_0^T R(s,t)\left[f_j(s,u(s)) - f_j(s,v(s))\right]ds\right|^p \\
&\leq \sum_{j=1}^{\infty}\int_0^T |R(s,t)|^p \left|f_j(s,u(s)) - f_j(s,v(s))\right|^p ds \left(\int_0^T ds\right)^{\frac{p}{q}} \\
&\leq (T)^{\frac{p}{q}} \sum_{j=1}^{\infty}\int_0^T |R(s,t)|^p \left|f_j(s,u(s)) - f_j(s,v(s))\right|^p ds.
\end{aligned}$$

Now, by using (6) and the assumption $(\mathbf{A}_2)$ of equicontinuity, we get

$$
\begin{aligned}
&\left(\|(\mathcal{F}u)(t)-(\mathcal{F}v)(t)\|_p\right)^p \\
&\leq (T)^{\frac{p}{q}}\left(\frac{1}{2}\tan(0.5T)\right)^p \sum_{j=1}^{\infty}\int_0^T \left|f_j(s,u(s))-f_j(s,v(s))\right|^p ds \\
&= \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \lim_{m\to\infty}\sum_{j=1}^{m}\int_0^T \left|f_j(s,u(s))-f_j(s,v(s))\right|^p ds \\
&= \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \lim_{m\to\infty}\int_0^T \left(\sum_{j=1}^{m}\left|f_j(s,u(s))-f_j(s,v(s))\right|^p\right) ds.
\end{aligned}
\tag{12}
$$

Define the function $\delta(\epsilon)$ as

$$
\delta(\epsilon)=\sup\left\{|f_j(s,u(s))-f_j(s,v(s))| : u,v\in\ell_p, \|u-v\|\leq\epsilon t\in I, j\in\mathbb{N}\right\}.
$$

Then, clearly $\delta(\epsilon)\to 0$ as $\epsilon\to 0$ since the family $\{(fv)(t) : t\in I\}$ is equicontinuous at every point $v\in\ell_p$.

Therefore, by (12) and using Lebesgue's dominant convergence theorem, we have

$$
\begin{aligned}
\left(\|(\mathcal{F}u)(t)-(\mathcal{F}v)(t)\|_p\right)^p &\leq \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \int_0^T [\delta(\epsilon)]^p ds \\
&= \left(\frac{T}{2}\tan(0.5T)\right)^p [\delta(\epsilon)]^p .
\end{aligned}
$$

This implies that the operator $\mathcal{F}$ is continuous on the ball B_r, since $T\neq(2n-1)\pi$, $n=1,2,\ldots$.

Since $R(s,t)$ as defined in (5) is uniformly continuous on I^2, and so by definition of operator $\mathcal{F}$, it is easy to show that $\{\mathcal{F}u : u\in B_r\}$ is equicontinuous on I. Let $B_{r_1}=conv(\mathcal{F}B_r)$, then $B_{r_1}\subset B_r$ and the functions from the set B_{r_1} are equicontinuous on I.

Let $E\subset B_{r_1}$, then E is equicontinuous on I. If $v\in E$ is a function, then for arbitrarily fixed $t\in I$, we have by assumption $(\mathbf{A}_3)$

$$
\begin{aligned}
\sum_{j=b}^{\infty}\left|(\mathcal{F}v)_j(t)\right|^p &= \sum_{j=b}^{\infty}\left|\int_0^T R(s,t)f_j(s,v(s))ds\right|^p \\
&\leq \sum_{j=b}^{\infty}\left(\int_0^T |R(s,t)||f_j(s,v(s))|\right)^p
\end{aligned}
$$

Using Hölder's inequality and (6), we get

$$\sum_{j=b}^{\infty}\left|(\mathcal{F}v)_j(t)\right|^p \le \sum_{j=b}^{\infty}\left(\int_0^T |R(s,t)|^p |f_j(s,v(s))|^p ds\right)\left(\int_0^T ds\right)^{\frac{p}{q}}$$

$$\le T^{\frac{p}{q}}\left(\frac{1}{2}\tan(0.5T)\right)^p \sum_{j=b}^{\infty}\left(\int_0^T |f_j(s,v(s))|^p ds\right).$$

Using Lebesgue's dominant convergence theorem and the assumption $(\mathbf{A}_3)$ gives

$$\sum_{j=b}^{\infty}\left|(\mathcal{F}v)_j(t)\right|^p$$

$$\le \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \int_0^T \left\{\sum_{j=b}^{\infty}[g_j(s)+h_j(s)|v_j(s)|^p]\right\} ds$$

$$= \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \left\{\int_0^T\left(\sum_{j=b}^{\infty} g_j(s)\right) ds + \int_0^T\left(\sum_{j=b}^{\infty} h_j(s)|v_j(s)|^p\right) ds\right\}$$

$$\le \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \left\{\int_0^T\left(\sum_{j=b}^{\infty} g_j(s)\right) ds + H\int_0^T \sum_{j=b}^{\infty}|v_j(s)|^p ds\right\}.$$

Taking supremum over all $v \in E$, we obtain

$$\sup_{v\in E}\sum_{j=b}^{\infty}\left|(\mathcal{F}v)_j(t)\right|^p$$

$$\le \left(\frac{T^{\frac{1}{q}}}{2}\tan(0.5T)\right)^p \left\{\int_0^T\left(\sum_{j=b}^{\infty} g_j(s)\right) ds + H\sup_{v\in E}\int_0^T \sum_{j=b}^{\infty}|v_j(s)|^p ds\right\}.$$

Using the definition of Hausdorff measure of noncompactness in ℓ_p space and noting that E is the set of equicontinuous functions on I, then by using Remark 1 we get

$$(\chi(\mathcal{F}E))^p \le HT\left(\frac{1}{2}\tan(0.5T)\right)^p (\chi(E))^p$$

$$\Rightarrow \chi(\mathcal{F}E) \le (HT)^{\frac{1}{p}}\left(\frac{1}{2}\tan(0.5T)\right)\chi(E).$$

Therefore, if $(HT)^{\frac{1}{p}}\left(\frac{1}{2}\tan(0.5T)\right) < 1$, that is, $(HT)^{\frac{1}{p}}\tan(0.5T) < 2$, then by Lemma 1, the operator $\mathcal{F}$ on the set B_{r_1} has a fixed point, which completes the proof of the theorem. □

Now, the system of integral equations (4) is equivalent to the boundary value problem (2), and we conclude that the infinite system of second-order differential equations (2) satisfying the boundary conditions (3) has at least one solution $v(t) = (v_1(t), v_2(t), \ldots) \in \ell_p$ such that $v_j(t) \in C^2(I, \ell_p)$, $(j = 1, 2, \ldots)$ for any $t \in I$, if the assumptions of Theorem 1 are satisfied.

Note The value of T is chosen such that the condition $(HT)^{\frac{1}{p}}\tan(0.5T) < 2$ is satisfied.

The above result is illustrated by the following example.

Example 1 Consider the infinite system of second-order differential equations in ℓ_2

$$\frac{\mathrm{d}^2 v_n}{\mathrm{d}t^2} + v_n = \frac{t3^{-nt}}{n} + \sum_{b=n}^{\infty} \frac{\cos t}{(1+2n)\sqrt{(b-1)!}} \cdot \frac{v_b(t)[1-(b-n)v_b(t)]}{(b-n+1)}, \tag{13}$$

for $n = 1, 2, \ldots$.

Solution Comparing (13) with (2), we have

$$f_n(t, v) = \frac{t3^{-nt}}{n} + \sum_{b=n}^{\infty} \frac{\cos t}{(1+2n)\sqrt{(b-1)!}} \cdot \frac{v_b(t)[1-(b-n)v_b(t)]}{(b-n+1)}. \tag{14}$$

Assumption $(\mathbf{A}_1)$ of Theorem 1 is clearly satisfied. We now show that assumption $(\mathbf{A}_2)$ of Theorem 1 is also satisfied, that is,

$$|f_n(t, v)|^2 \le g_n(t) + h_n(t)|v_n|^2. \tag{15}$$

Using the Cauchy–Schwarz inequality and Eq. (13), we have

$$\begin{aligned}
&|f_n(t, v)|^2 \\
&= \left| \frac{t3^{-nt}}{n} + \sum_{b=n}^{\infty} \frac{\cos t}{(1+2n)\sqrt{(b-1)!}} \cdot \frac{v_b(t)[1-(b-n)v_b(t)]}{(b-n+1)} \right|^2 \\
&\le 2\left\{ \frac{t^2 3^{-2nt}}{n^2} + \left\{ \sum_{b=n}^{\infty} \frac{\cos t}{(1+2n)\sqrt{(b-1)!}} \cdot \frac{v_b(t)[1-(b-n)v_b(t)]}{(b-n+1)} \right] \right\}^2 \\
&\le 2\frac{t^2 3^{-2nt}}{n^2} + 2\left(\sum_{b=n}^{\infty} \frac{\cos^2 t}{(1+2n)^2(b-1)!} \right) \cdot \sum_{b=n}^{\infty} \left(\frac{v_b(t)[1-(b-n)v_b(t)]}{(b-n+1)} \right)^2 .
\end{aligned}$$

Now, using the fact that $\frac{1-\alpha\beta}{\beta} \leq \frac{1}{(2\beta)^2}$ for any real $\alpha, \beta, \beta \neq 0$, we have

$$|f_n(t,v)|^2 \leq 2\frac{t^2 3^{-2nt}}{n^2} + 2\frac{\cos^2 t}{(1+2n)^2} \times e \times \left(v_n^2 + \sum_{b=n+1}^{\infty} \frac{v_b(t)[1-(b-n)v_b(t)]}{(b-n+1)} \right)$$

$$\leq 2\frac{t^2 3^{-2nt}}{n^2} + 2\frac{e[\cos^2 t]}{(1+2n)^2}(v_n^2) + 2\frac{e[\cos^2 t]}{(1+2n)^2} \times \sum_{b=n+1}^{\infty} \left(\frac{1}{2(b-n)} \right)^2$$

$$\leq 2\frac{t^2 3^{-2nt}}{n^2} + \frac{1}{2}\frac{e[\cos^2 t]}{(1+2n)^2} \times \frac{\pi^2}{6} + 2\frac{e[\cos^2 t]}{(1+2n)^2}(v_n^2).$$

Hence, by taking

$$g_n(t) = 2\frac{t^2 3^{-2nt}}{n^2} + \frac{\pi^2}{12}\frac{e[\cos^2 t]}{(1+2n)^2} \quad , \quad h_n(t) = 2\frac{e[\cos^2 t]}{(1+2n)^2},$$

it is clear that $g_n(t)$ and $h_n(t)$ are real-valued continuous functions on I. Also,

$$|g_n(t)| \leq 2\frac{T^2}{n^2} + \frac{\pi^2}{12}\frac{e}{(1+2n)^2}$$

$$\leq \left(2T^2 + \frac{\pi^2 e}{12} \right) \frac{1}{n^2}$$

for all $t \in I$. Thus, by the Weierstrass test for uniform convergence of the function series, we see that $\sum_{b\geq 1} g_b(t)$ is uniformly convergent on I.

Furthermore, we have

$$|h_j(t)| \leq \frac{2e}{(1+2n)^2}$$

for all $t \in I$.

Thus, the function sequence $\left(h_j(t)\right)$ is equibounded on I. Thus, (14) is satisfied, and hence the assumption $(\mathbf{A}_3)$ is satisfied.

Also,

$$G = \sup \left\{ \sum_{b\geq 1} g_b(t) : t \in I \right\} = \left(2T^2 + \frac{\pi^2 e}{12} \right) \frac{\pi^2}{6},$$

and

$$H = \sup\left\{h_j(t) : t \in I\right\} = \frac{2e}{9}.$$

The assumption $(\mathbf{A}_2)$ is also satisfied as for fixed $t \in T$ and $(v_j(t)) = (v_1(t), v_2(t), \ldots) \in \ell_2$, we have

$$\sum_{j=1}^{\infty} \left|f_j(t, v)\right|^2 = \sum_{j=1}^{\infty} g_j(t) + \sum_{j=1}^{\infty} h_j(t)|v_j(t)|^2$$

$$\leq G + H \sum_{j=1}^{\infty} |v_j(t)|^2.$$

Hence, the operator $f = (f_j)$ transforms the space (I, ℓ_2) into ℓ_2.

Also, for $\epsilon > 0$ and $u = (u_j)$, $v = (v_j)$ in ℓ_2 with $\|u - v\|_2 < \epsilon$, we have

$$\left(\|(fu)(t) - (fv)(t)\|_2\right)^2 = \sum_{n=1}^{\infty} |f_n(t, u(t)) - f_n(t, v(t))|^2$$

$$= \sum_{n=1}^{\infty} \left\{ \left| \sum_{b=n}^{\infty} \frac{(\cos t)u_b(t)[1 - (b-n)u_b(t)]}{(1+2n)(b-n+1)\sqrt{(b-1)!}} - \frac{(\cos t)v_b(t)[1 - (b-n)v_b(t)]}{(1+2n)(b-n+1)\sqrt{(b-1)!}} \right|^2 \right\}$$

$$\leq \sum_{n=1}^{\infty} \left\{ \left(\frac{1}{(1+2n)^2}\right) \left| \sum_{b=n}^{\infty} \frac{u_b(t)[1 - (b-n)u_b(t)] - v_b(t)[1 - (b-n)v_b(t)]}{(b-n+1)\sqrt{(b-1)!}} \right|^2 \right\}$$

$$\leq \sum_{n=1}^{\infty} \left\{ \left(\frac{1}{(1+2n)^2}\right) \left[\sum_{b=n}^{\infty} \left| \frac{(u_b(t) - v_b(t))[1 - (b-n)(u_b(t) + v_b(t))]}{\sqrt{(b-1)!}(b-n+1)} \right| \right]^2 \right\}$$

Using Holder's inequality, we get [31]

$$\left(\|(fu)(t) - (fv)(t)\|_2\right)^2$$

$$\leq \sum_{n=1}^{\infty} \left\{ \frac{1}{(1+2n)^2} \left(\sum_{b=n}^{\infty} \frac{1}{(b-1)!} \right) \left[\sum_{b=n}^{\infty} \left| \frac{(u_b(t) - v_b(t))[1 - (b-n)(u_b(t) + v_b(t))]}{(b-n+1)} \right|^2 \right] \right\}$$

$$\leq e \sum_{n=1}^{\infty} \left\{ \frac{1}{(1+2n)^2} \left[\sum_{b=n}^{\infty} |u_b(t) - v_b(t)|^2 \left| \frac{1 - (b-n)(u_b(t) + v_b(t))}{(b-n+1)} \right|^2 \right] \right\}$$

$$\leq e \sum_{n=1}^{\infty} \left\{ \frac{1}{(1+2n)^2} \left[\sum_{b=n}^{\infty} |u_b(t) - v_b(t)|^2 \right] \right\}$$

$$< e\epsilon^2 \sum_{n=1}^{\infty} \left\{ \frac{1}{(1+2n)^2} \right\}$$

$$\leq e \left(\frac{\pi^2}{8} \right) \epsilon^2.$$

Thus, for any $t \in I$, we have

$$\| (fu)(t) - (fv)(t) \|_2 < \frac{\pi \epsilon \sqrt{e}}{2\sqrt{2}}.$$

Therefore, the family $\{(fv)(t) : t \in I\}$ is equicontinuous.

Finally, we see that the condition $(HT)^{\frac{1}{p}} \tan(0.5T) < 2$ is satisfied for all $T \leq 2$.

So, by Theorem 1, there exists at least one solution to given infinite system of differential equations (13) in $C(I, \ell_2)$.

References

1. R.P. Agarwal, B. Mouffak, S. Hamani, A survey on existence results for boundary value problems of nonlinear fractional differential equations and inclusions. Acta Appl. Math. **109**(3), 973–1033 (2010)
2. A. Aghajani, M. Mursaleen, A. Shole Haghighi, Fixed point theorems for Meir-Keeler condensing operators via measure of noncompactness. Acta Math. Sci. **35**(3), 552–566 (2015)
3. A. Aghajani, E. Pourhadi, Application of measure of noncompactness to ℓ_1-solvability of infinite systems of second order differential equations. Bull. Belg. Math. Soc. Simon Stevin **22**(1), 105–118 (2015)
4. J. Banaś, Applications of measures of weak noncompactness and some classes of operators in the theory of functional equations in the Lebesgue space. Nonlin. Anal. T.M.A. **30**(6), 3283–3293 (1997)
5. J. Banaś, K. Goebel, *Measures of Noncompactness in Banach Spaces: Lecture Notes in Pure and Applied Mathematics* (Marcel Dekker, New York and Basel, 1980)
6. J. Banaś, M. Lecko, Solvability of infinite systems of differential equations in Banach sequence spaces. J. Comput. Appl. Math. **137**(2), 363–375 (2001)
7. J. Banaś, M. Mursaleen, *Sequence Spaces and Measures of Noncompactness with Applications to Differential and Integral Equations* (Springer, 2014)
8. J. Banaś, M. Mursaleen, S.M.H. Rizvi, Existence of solutions to a boundary-value problem for an infinite system of differential equations. Electron. J. Differ. Equ. **262**, 1–12 (2017)
9. G. Darbo, Punti uniti in trasformazioni a codominio non compatto. Rend. Sem. Mat. Univ. Padova **24**, 84–92 (1955)
10. K. Deimling, *Nonlinear Functional Analysis* (Courier Corporation, 2010)
11. K. Deimling, *Ordinary Differential Equations in Banach Spaces* (Springer, 2006)
12. B.C. Dhage, D. O'Regan, A fixed point theorem in Banach algebras with applications to nonlinear integral equation. Funct. Differ. Equ. **7**(4), 259–267 (2000)
13. G.D. Duffy, *Green's Function with Applications* (Chapman and Hall/CRC, London, 2004)
14. M. Jleli, B. Samet, Existence of positive solutions to a coupled system of fractional differential equations. Math. Method Appl. Sci. **6**(38), 1014–1031 (2015)
15. S. Karlin, L. Nirenberg, On a theorem of P. Nowosad. J. Math. Anal. Appl. **17**(1), 61–67 (1967)

16. J. Klamka, Schauder's fixed-point theorem in nonlinear controllability problems. Control Cybernet. **29**, 153–165 (2000)
17. C. Kuratowski, Sur les espaces complets. Fund. Math. **1**(15), 301–309 (1930)
18. Z. Liu, M.K. Shin, Applications of Schauder's Fixed-point theorem with respect to iterated functional equations. Appl. Math. Lett. **14**(8), 955–962 (2001)
19. I.A. Malik, T. Jalal, Measures of noncompactness in $\left(\bar{N}_{\Delta}^{q}\right)$ summable difference sequence spaces. Filomat **32**(15), 5459–5470 (2018)
20. I.A. Malik, T. Jalal, Application of measure of noncompactness to infinite systems of differential equations in ℓ_p spaces. Rend. Circ. Mat. Palermo (2), 1–12 (2019). https://doi.org/10.1007/s12215-019-00411-6
21. I.A. Malik, T. Jalal, Boundary value problem for an infinite system of second order differential equations in ℓ_p spaces. Math. Bohem. (Published online June 20, 2019)
22. I.A. Malik, T. Jalal, Measures of noncompactness in $\bar{N}(p, q)$ summable sequence spaces. Operators Matrices **13**(4), 1191–1205 (2019). https://doi.org/10.7153/oam-2019-13-79
23. I.A. Malik, T. Jalal, Infinite system of integral equations in two variables of Hammerstein type in c_0 and ℓ_1 spaces. Filomat **33**(11), 3441–3455 (2019)
24. E. Malkowsky, V. Rakočević, An introduction into the theory of sequence spaces and measures of noncompactness. Matematički institut SANU (2000)
25. M. Mursaleen, Application of measure of noncompactness to infinite system of differential equations. Can. Math. Bull. **56**(2), 388–394 (2013)
26. M. Mursaleen, S.A. Mohiuddine, Applications of measures of noncompactness to the infinite system of differential equations in ℓ_p spaces. Nonlinear Anal. **75**(4), 2111–2115 (2012)
27. M. Mursaleen, S.M.H. Rizvi, Solvability of infinite systems of second order differential equations in c_0 and ℓ_1 by Meir-Keeler condensing operators. Proc. Am. Math. Soc. **144**(10), 4279–4289 (2016)
28. M. Mursaleen, S.M.H. Rizvi, B. Samet, Solvability of a class of boundary value problems in the space of convergent sequences. Appl. Anal. **97**(11), 1829–1845 (2018)
29. V. Muresan, Volterra integral equations with iterations of linear modification of the argument. Novi. Sad. J. Math. **33**(2), 1–10 (2003)
30. I.M. Olaru, An integral equation via weakly Picard operators. Fixed Point Theory **11**(1), 97–106 (2010)
31. W. Rudin, *Real and Complex Analysis* (Tata McGraw-Hill Edu., 2006)

Differential Equations Involving Theta Functions and h-Functions

H. C. Vidya and B. Ashwath Rao

Abstract Ramanujan in his notebook recorded elegant continued fraction identities and mentioned some of the appealing formulas involving it. The purpose of this chapter is to acquire the connection among the continued fraction of order 12 with h-functions. In this chapter, we additionally construct certain beautiful differential identities containing h-functions by utilizing explicit relations recorded by Shaun Cooper.

1 Introduction

The continued fraction of order 12 was established by M. S. M. Naika et al. [1] as a special case of fascinating continued fraction identity noted by Ramanujan in his second notebook [2, p.74]. Shaun Cooper [3] in his book recorded some basic properties of h-functions and also established relations involving Eisenstein series of various levels and h-functions. Furthermore, they have established the relation among cubic continued fraction and Rogers–Ramanujan's continued fraction with h-functions. Recently, B. C. Berndt et al. [4] formed certain differential equations to prove the identities of orders 14 and 35 in Section 8, 9 and 10. They have systematically derived several new differential equations for eta function quotients in Section 10. Recently, H. C. Vidya and B. R. Srivatsa Kumar [5] constructed certain differential equations involving theta functions.

This chapter involves identities that relate continued fraction of order 12 with h-functions. Furthermore, we construct certain differential equations involving theta functions and h-functions, which are achieved by adopting some of the Eisenstein series relations recorded by S. Cooper. In Sect. 3, we express continued fraction

H. C. Vidya · B. A. Rao (✉)
National Institute of Technology, Puducherry, India
e-mail: tjalal@nitsri.net; ishfaq_2phd15@nitsri.net

V. K. Singh et al. (eds.), *Recent Trends in Mathematical Modeling and High Performance Computing*, Trends in Mathematics,
https://doi.org/10.1007/978-3-030-68281-1_32

of order 12 in terms of h-functions. In Sect. 4, we construct certain differential equations involving h-functions. Section 2 is dedicated to record some preliminary results.

2 Preliminaries

For $|q| < 1$, the h-function is defined by

$$h = h(q) = q \prod_{k=1}^{\infty} \frac{(1-q^{12k-1})(1-q^{12k-11})}{(1-q^{12k-5})(1-q^{12k-7})}.$$

For any complex a and q with $|q| < 1$, the q-series is defined by

$$(a;q)_\infty := \prod_{n=0}^{\infty}(1-aq^n).$$

For $|ab| < 1$, Ramanujan's general theta function [6, p.35] is given by

$$f(a,b) := \sum_{n=-\infty}^{\infty} a^{n(n+1)/2} b^{n(n-1)/2} = (-a,-b,ab;ab)_\infty.$$

The special cases of theta functions are

$$f(-q) := f(-q,-q^2) = \sum_{n=-\infty}^{\infty} (-1)^n q^{n(3n-1)/2} = (q;q)_\infty,$$

$$\varphi(q) := f(q,q) = \sum_{n=-\infty}^{\infty} q^{n^2} = (-q;q^2)^2{}_\infty (q^2;q^2)_\infty = \frac{(-q;-q)_\infty}{(q;-q)_\infty}.$$

Ramanujan's cubic continued fraction $G(q)$ is defined as

$$G(q) := \frac{q^{1/3} f(-q,-q^5)}{f(-q^3,-q^3)} = \frac{q^{1/3}}{1} \underset{+}{} \frac{q+q^2}{1} \underset{+}{} \frac{q^2+q^4}{1} \underset{+}{} \frac{q^3+q^6}{1} \underset{+\ldots}{}.$$

The continued fraction of order 12 is defined by

$$U(q) := \frac{qf(-q,-q^{11})}{f(-q^5,-q^7)} = \frac{q(1-q)}{(1-q^3)} \underset{+}{} \frac{q^3(1-q^2)(1-q^4)}{(1-q^3)(1+q^6)} \underset{+}{} \frac{q^3(1-q^8)(1-q^{10})}{(1-q^3)(1+q^{12})} \underset{+\ldots}{}.$$

The cubic continued fraction in terms of h-function as recorded by S. Cooper [3] is

$$G^3(q) = h\frac{(1-h)^2}{(1+h^2)^2}. \tag{1}$$

The weight two modular form y_{12} in terms of h-function is defined by

$$y_{12} = q\frac{d}{dq}logh = 1 - \sum_{s=1}^{\infty}\chi_{12}(s)\frac{sq^s}{1-q^s},$$

where

$$\chi_{12}(s) = \begin{cases} 1 & \text{if } s = 1 \text{or} 11 (\text{mod} 12), \\ -1 & \text{if } s = 5 \text{or} 7 (\text{mod} 12), \\ 0 & \text{otherwise.} \end{cases}$$

The Ramanujan-type Eisenstein series is defined by

$$P(q) := 1 - 24\sum_{k=1}^{\infty}\frac{kq^k}{1-q^k}.$$

Lemma *The following relation among cubic continued fraction and theta functions hold:*

$$8G^3(q) = 1 - \frac{\varphi^4(-q)}{\varphi^4(-q^3)}. \tag{2}$$

Proof For a proof, see Chapter 20 [6, p. 345]. □

Lemma *The following identity holds:*

$$\frac{\varphi(q)}{\varphi(q^3)} = \frac{1+U(q)}{1-U(q)}. \tag{3}$$

Proof For a proof, see [7]. □

Lemma *We have*

$$G(q) + G(-q) + 2G^2(-q)G^2(q) = 0. \tag{4}$$

Proof For a proof, see [8]. □

Lemma ([3]) *The following equality holds:*

$$\frac{1}{24}(P(q)-9P(q^3)-4P(q^4)+36P(q^{12}))=\frac{(1-h^2)}{(1+h^2)}y_{12}, \tag{5}$$

$$\frac{1}{24}(4P(q^2)-16P(q^4)-12P(q^6)+48P(q^{12}))=\frac{(1-h^2)}{(1-h+h^2)}y_{12}, \tag{6}$$

$$\frac{1}{24}(P(q)+3P(q^3)+8P(q^4)+12P(q^6)-24P(q^{12}))=h\frac{dy_{12}}{dh}, \tag{7}$$

$$\frac{1}{24}(-3P(q)-4P(q^4)-12P(q^6)+4P(q^2)+3P(q^3)+36P(q^{12})) = \frac{(1-h^2)}{(1-4h+h^2)}y_{12}, \tag{8}$$

$$\frac{1}{24}(-P(q)+6P(q^2)+9P(q^3)-8P(q^4)-54P(q^6)+72P(q^{12})) = \frac{(1-h^2)}{(1-2h+h^2)}y_{12}, \tag{9}$$

$$\frac{1}{24}(3P(q)-14P(q^2)-3P(q^3)+8P(q^4)+6P(q^6)+24P(q^{12})) = \frac{(1-h^2)}{(1+2h+h^2)}y_{12}. \tag{10}$$

Proof *For a proof, see [3].* □

3 Expression of Continued Fraction of Order 12 in Terms of *h*-Function

In his book, S. Cooper [3] documented an expression involving cubic continued fraction with h-functions. Using these quadratic transformation formulas, interestingly, we are able to discover relations including continued fraction of order 12 with h-functions.

Theorem 1.1 *We have*

$$(i)\ \ U(-q)=\frac{\sqrt{1-4h+h^2}-\sqrt{1+h^2}}{\sqrt{1-4h+h^2}+\sqrt{1+h^2}},$$

$$(ii)\ \ U(q) = \frac{(1+h^2)\left(1-(1-h)\sqrt{1-6h+h^2}\right)^3 - 8h^2(1-h)^4 + 1}{(1+h^2)\left(1-(1-h)\sqrt{1-6h+h^2}\right)^3 - 8h^2(1-h)^4 - 1}.$$

Proof (i) Replacing q with $-q$ in (3) and raising to the power 4, we see that

$$\frac{\varphi^4(-q)}{\varphi^4(-q^3)} = \left(\frac{1+U(-q)}{1-U(-q)}\right)^4.$$

On comparing the above equation with (2) and then equating the resulting expression with (1), we deduce

$$8h\frac{(1-h)^2}{(1+h^2)^2} = \left(\frac{1+U(-q)}{1-U(-q)}\right)^4.$$

Rearranging the above equation for $U(-q)$ using maple, we obtain (i) and (ii). Solving (4) for $G(-q)$, we deduce

$$G(-q) = \frac{-1 \pm \sqrt{1-8G^3(q)}}{4G^2(q)}.$$

Using (1), on the right of the above equation, we note that the first factor becomes

$$G(-q) = \frac{(1+h^2)^{1/3}[1-(1-h)\sqrt{1-6h+h^2}]}{4h^{2/3}(1-h)^{4/3}},$$

and the second factor becomes

$$G(-q) = \frac{(1+h^2)^{1/3}[1+(1-h)\sqrt{1-6h+h^2}]}{4h^{2/3}(1-h)^{4/3}}.$$

Now, applying L'Hospital's rule, the first factor tends to zero in some neighbourhood of $q = 0$, and the second factor does not vanish. Thus, by analytic continuation in $|q| < 1$, we have

$$G(-q) = \frac{(1+h^2)^{1/3}[1-(1-h)\sqrt{1-6h+h^2}]}{4h^{2/3}(1-h)^{4/3}}.$$

Replacing q with $-q$ in (2), comparing with (3) and further using the above identity, we arrive at (ii). □

4 Construction of Differential Equations

S. Cooper [3] supplied certain identities involving Eisenstein series of numerous levels with h-functions. We have framed differential equations with the aid of these identities that contain theta functions and h-functions.

Theorem 1.2 *If*

$$v = q\frac{f(-q)f^3(-q^{12})}{f^3(-q^3)f(-q^4)},$$

then

$$\frac{q}{v}\frac{dv}{dq} - \frac{(1-h^2)}{(1+h^2)}y_{12} = 0.$$

Proof Employing the definition of theta function, we achieve

$$v = q\frac{(q;q)_\infty (q^{12};q^{12})_\infty^3}{(q^3;q^3)_\infty^3 (q^4;q^4)_\infty}.$$

Logarithmically differentiating v with respect to q and then simplifying, we deduce that

$$\frac{1}{v}\frac{dv}{dq} = \frac{1}{q} - \frac{1}{q}\left[\sum_{s=1}^{\infty}\frac{kq^s}{1-q^s} - 3\sum_{s=1}^{\infty}\frac{3sq^{3s}}{1-q^{3s}} - \sum_{s=1}^{\infty}\frac{4sq^{4s}}{1-q^{4s}} + 3\sum_{s=1}^{\infty}\frac{12sq^{12s}}{1-q^{12s}}\right].$$

Expressing the above sum in terms of the known Eisenstein series, we arrive at

$$\frac{q}{v}\frac{dv}{dq} = \frac{1}{24}[P(q) - 9P(q^3) - 4P(q^4) + 36P(q^{12})].$$

Employing (5) in the above equation, we get the required result. □

Theorem 1.3 *If*

$$v = q\frac{f^2(-q^2)f^4(-q^{12})}{f^4(-q^4)f^2(-q^6)},$$

then

$$\frac{q}{v}\frac{dv}{dq} - \frac{(1-h^2)}{(1-h+h^2)}y_{12} = 0.$$

Proof Utilizing the definition of theta function and further logarithmically differentiating v with respect to q, we deduce

$$\frac{1}{v}\frac{dv}{dq}=\frac{1}{q}-\frac{1}{q}\left[2\sum_{s=1}^{\infty}\frac{2sq^{2s}}{1-q^{2s}}-4\sum_{s=1}^{\infty}\frac{4sq^{4s}}{1-q^{4s}}-2\sum_{s=1}^{\infty}\frac{6sq^{6s}}{1-q^{6s}}+4\sum_{n=1}^{\infty}\frac{12sq^{12s}}{1-q^{12s}}\right].$$

Using the definition of Eisenstein series, we obtain

$$\frac{q}{v}\frac{dv}{dq}=\frac{1}{24}[4P(q^2)-16P(q^4)-12P(q^6)+48P(q^{12})].$$

Now, upon using (6), we obtain the desired result. □

Theorem 1.4 *If*

$$v=\frac{f(-q)f(-q^3)f^2(-q^4)f^2(-q^6)}{f^2(-q^{12})},$$

then

$$\frac{q}{v}\frac{dv}{dq}-h\frac{dy_{12}}{dh}=0.$$

Proof Utilizing the definition of theta function and further logarithmically differentiating v with respect to q, we arrive at

$$\frac{1}{v}\frac{dv}{dq}=-\frac{1}{q}\left[\sum_{s=1}^{\infty}\frac{sq^{2s}}{1-q^{2s}}+\sum_{s=1}^{\infty}\frac{3sq^{3s}}{1-q^{3s}}+2\sum_{s=1}^{\infty}\frac{4sq^{4s}}{1-q^{4s}}+2\sum_{s=1}^{\infty}\frac{6sq^{6s}}{1-q^{6s}}-2\sum_{s=1}^{\infty}\frac{12sq^{12s}}{1-q^{12s}}\right].$$

Furthermore, upon using the definition of Eisenstein series, we obtain

$$\frac{q}{v}\frac{dv}{dq}=\frac{1}{24}[P(q)+3P(q^3)+8P(q^4)+12P(q^6)-24P(q^{12})].$$

Using (7), we arrive at the desired result. □

Theorem 1.5 *If*

$$v=q\frac{f^3(-q)f(-q^4)f^2(-q^6)}{f^2(-q^2)f(-q^3)f^3(-q^{12})},$$

then

$$\frac{q}{v}\frac{dv}{dq}+\frac{(1-h^2)}{(1-4h+h^2)}y_{12}=0.$$

Proof Applying the definition of theta function and then logarithmically differentiating v, we arrive at

$$\frac{1}{v}\frac{dv}{dq} = \frac{1}{q} - \frac{1}{q}\left[-3\sum_{s=1}^{\infty}\frac{sq^s}{1-q^s} + 2\sum_{s=1}^{\infty}\frac{2sq^{2s}}{1-q^{2s}} + \sum_{s=1}^{\infty}\frac{3sq^{3s}}{1-q^{3s}} - \sum_{s=1}^{\infty}\frac{4sq^{4s}}{1-q^{4s}} - 2\sum_{s=1}^{\infty}\frac{6sq^{6s}}{1-q^{6s}} + 3\sum_{s=1}^{\infty}\frac{12sq^{12s}}{1-q^{12s}}\right].$$

Using the definition of Eisenstein series, we obtain

$$\frac{q}{v}\frac{dv}{dq} = -\frac{1}{24}\left[-3P(q) + 4P(q^2) + 3P(q^3) - 4P(q^4) - 12P(q^6) + 36P(q^{12})\right].$$

Using (8), we get the required result. □

Theorem 1.6 *If*

$$v = \frac{1}{q}\frac{f(-q)f^2(-q^4)f^9(-q^6)}{f^3(-q^2)f^3(-q^3)f^6(-q^{12})},$$

then

$$\frac{q}{v}\frac{dv}{dq} + \frac{(1-h^2)}{(1-2h+h^2)}y_{12} = 0.$$

Proof Using the definition of theta function, logarithmically differentiating v and then simplifying, we deduce

$$\frac{1}{v}\frac{dv}{dq} = -\frac{1}{q} + \frac{1}{q}\left[-\sum_{s=1}^{\infty}\frac{sq^s}{1-q^s} + 3\sum_{s=1}^{\infty}\frac{2sq^{2s}}{1-q^{2s}} + 3\sum_{s=1}^{\infty}\frac{3sq^{3s}}{1-q^{3s}} - 2\sum_{n=1}^{\infty}\frac{4sq^{4s}}{1-q^{4s}} - 9\sum_{s=1}^{\infty}\frac{6sq^{6s}}{1-q^{6s}} + 6\sum_{s=1}^{\infty}\frac{12sq^{12s}}{1-q^{12s}}\right].$$

Furthermore, using the definition of Eisenstein series, we obtain

$$\frac{q}{v}\frac{dv}{dq} = \frac{1}{24}[P(q) - 6P(q^2) - 9P(q^3) + 8P(q^4) + 4P(q^6) - 72P(q^{12})].$$

Using (9), we obtain the required result. □

Theorem 1.7 *If*

$$v = q\frac{f^3(-q)f^2(-q^4)f(-q^6)f^2(-q^{12})}{f^7(-q^2)f(-q^3)},$$

then

$$\frac{q}{v}\frac{dv}{dq} - \frac{(1-h^2)}{(1+2h+h^2)}y_{12} = 0.$$

Proof Applying the definition of theta function, logarithmically differentiating v and then simplifying, we deduce

$$\frac{1}{v}\frac{dv}{dq} = \frac{1}{q} + \frac{1}{q}\Bigg[-3\sum_{s=1}^{\infty}\frac{sq^s}{1-q^s} + 7\sum_{s=1}^{\infty}\frac{2sq^{2s}}{1-q^{2s}} + \sum_{s=1}^{\infty}\frac{3sq^{3s}}{1-q^{3s}} - 2\sum_{s=1}^{\infty}\frac{4sq^{4s}}{1-q^{4s}} - \sum_{s=1}^{\infty}\frac{6sq^{6s}}{1-q^{6s}} - 2\sum_{s=1}^{\infty}\frac{12sq^{12s}}{1-q^{12s}}\Bigg].$$

Furthermore, using the definition of Eisenstein series and relation (10), we obtain the required result. □

References

1. M.S.M. Naika, B.N. Dharmendra, K. Shivashankara, A continued fraction of order twelve. Centr. Eur. J. Math. **6**(3), 393–404 (2008)
2. S. Ramanujan, *Notebooks (2 Volumes)* (Bombay, 1957)
3. S. Cooper, *Ramanujan's Theta Functions* (Springer, 2017)
4. B.C. Berndt, H.H. Chan, S.S. Haung, Incomplete elliptic integrals in Ramanujan's lost notebook. Contemp. Math. **254**, 79–124 (2000)
5. H.C. Vidya, B.R. Srivatsa Kumar, Some studies on Eisenstein series and its applications. Notes Number Theory Discrete Math. **25**(4), 30–43 (2019)
6. B.C. Berndt, *Ramanujan's Notebooks*, Part III (Springer, New York, 1991)
7. K.R. Vasuki, A.A.K. Abdulrawf, G. Sharath, C. Sathish Kumar, On a continued fraction of order 12. Ukr. Math. J. **62**(12), 1866–1878 (2011)
8. H.H. Chan, On Ramanujan's cubic continued fraction. Acta Arith. **73**, 343–345 (1995)