Volume 24, Number 6, 2024 ISSN 1562-8353

NONLINEAR DYNAMICS AND SYSTEMS THEORY
An International Journal of Research and Surveys

Volume 24 Number 6 2024

CONTENTS

 ISSN 1562-8353 Printed in Ukraine
To receive contents and abstracts by e-mail, visit our Website at: **http://www.e-ndst.kiev.ua**

NONLINEAR DYNAMICS & SYSTEMS THEORY Volume 24, No. 6, 2024

Nonlinear Dynamics and Systems Theory

An International Journal of Research and Surveys

InforMath Publishing Group
http://www.e-ndst.kiev.ua

Nonlinear Dynamics and Systems Theory

An International Journal of Research and Surveys

INSTRUCTIONS FOR CONTRIBUTORS

(1) General. Nonlinear Dynamics and Systems Theory (ND&ST) is an international journal devoted to publishing peer-refereed, high quality, original papers, brief notes and review articles focusing on nonlinear dynamics and systems theory and their practical applications in engineering, physical and life sciences. Submission of a manuscript is a representation that the submission has been approved by all of the authors and by the institution where the work was carried out. It also represents that the manuscript has not been previously published, has not been copyrighted, is not being submitted for publication elsewhere, and that the authors have agreed that the copyright in the article shall be assigned exclusively to InforMath Publishing Group by signing a transfer of copyright form. Before submission, the authors should visit the website:

http://www.e-ndst.kiev.ua

for information on the preparation of accepted manuscripts. Please download the archive Sample_NDST.zip containing example of article file (you can edit only the file Samplefilename.tex).

(2) Manuscript and Correspondence. Manuscripts should be in English and must meet common standards of usage and grammar. To submit a paper, send by e-mail a file in PDF format directly to

Professor A.A. Martynyuk, Institute of Mechanics,
Nesterov str.3, 03057, Kiev-57, Ukraine
e-mail: journalndst@gmail.com

or to one of the Regional Editors or to a member of the Editorial Board. Final version of the manuscript must typeset using LaTex program which is prepared in accordance with the style file of the Journal. Manuscript texts should contain the title of the article, name(s) of the author(s) and complete affiliations. Each article requires an abstract not exceeding 150 words. Formulas and citations should not be included in the abstract. AMS subject classifications and key words must be included in all accepted papers. Each article requires a running head (abbreviated form of the title) of no more than 30 characters. The sizes for regular papers, survey articles, brief notes, letters to editors and book reviews are: (i) 10-14 pages for regular papers, (ii) up to 24 pages for survey articles, and (iii) 2-3 pages for brief notes, letters to the editor and book reviews.

(3) Tables, Graphs and Illustrations. Each figure must be of a quality suitable for direct reproduction and must include a caption. Drawings should include all relevant details and should be drawn professionally in black ink on plain white drawing paper. In addition to a hard copy of the artwork, it is necessary to attach the electronic file of the artwork (preferably in PCX format).

(4) References. Each entry must be cited in the text by author(s) and number or by number alone. All references should be listed in their alphabetic order. Use please the following style:

Journal: [1] H. Poincare, Title of the article. *Title of the Journal* **volume** (issue) (year) pages. [Language]

Book: [2] A.M. Lyapunov, *Title of the Book.* Name of the Publishers, Town, year.

Proceeding: [3] R. Bellman, Title of the article. In: *Title of the Book.* (Eds.). Name of the Publishers, Town, year, pages. [Language]

(5) Proofs and Sample Copy. Proofs sent to authors should be returned to the Editorial Office with corrections within three days after receipt. The corresponding author will receive a sample copy of the issue of the Journal for which his/her paper is published.

(6) Editorial Policy. Every submission will undergo a stringent peer review process. An editor will be assigned to handle the review process of the paper. He/she will secure at least two reviewers' reports. The decision on acceptance, rejection or acceptance subject to revision will be made based on these reviewers' reports and the editor's own reading of the paper.

NONLINEAR DYNAMICS AND SYSTEMS THEORY
An International Journal of Research and Surveys
Published by InforMath Publishing Group since 2001

Volume 24 Number 6 2024

CONTENTS

Founded by A.A. Martynyuk in 2001.
Registered in Ukraine Number: KB No 5267 / 04.07.2001.

NONLINEAR DYNAMICS AND SYSTEMS THEORY

An International Journal of Research and Surveys

Impact Factor from SCOPUS for 2023: SJR – 0.298, SNIP – 1.032 and CiteScore – 1.6

Nonlinear Dynamics and Systems Theory (ISSN 1562–8353 (Print), ISSN 1813–7385 (Online)) is an international journal published under the auspices of the S.P. Timoshenko Institute of Mechanics of National Academy of Sciences of Ukraine and Curtin University of Technology (Perth, Australia). It aims to publish high quality original scientific papers and surveys in areas of nonlinear dynamics and systems theory and their real world applications.

AIMS AND SCOPE

Nonlinear Dynamics and Systems Theory is a multidisciplinary journal. It publishes papers focusing on proofs of important theorems as well as papers presenting new ideas and new theory, conjectures, numerical algorithms and physical experiments in areas related to nonlinear dynamics and systems theory. Papers that deal with theoretical aspects of nonlinear dynamics and/or systems theory should contain significant mathematical results with an indication of their possible applications. Papers that emphasize applications should contain new mathematical models of real world phenomena and/or description of engineering problems. They should include rigorous analysis of data used and results obtained. Papers that integrate and interrelate ideas and methods of nonlinear dynamics and systems theory will be particularly welcomed.

PUBLICATION AND SUBSCRIPTION INFORMATION

Nonlinear Dynamics and Systems Theory will have 6 issues in 2025, printed in hard copy (ISSN 1562–8353) and available online (ISSN 1813–7385), by InforMath Publishing Group, Nesterov str., 3, Institute of Mechanics, Kiev, MSP 680, Ukraine, 03057. Subscription prices are available upon request from the Publisher, EBSCO Information Services (`mailto:journals@ebsco.com`), or website of the Journal: `http://e-ndst.kiev.ua`. Subscriptions are accepted on a calendar year basis. Issues are sent by airmail to all countries of the world. Claims for missing issues should be made within six months of the date of dispatch.

ABSTRACTING AND INDEXING SERVICES

Papers published in this journal are indexed or abstracted in: Mathematical Reviews / MathSciNet, Zentralblatt MATH / Mathematics Abstracts, PASCAL database (INIST–CNRS) and SCOPUS.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 537–548

Using a 2-D Discrete Chaotic Map to Create a Safe Data in Symmetric Systems

Salah Adoui * and Brahim Benzeghli

Department of Mathematics, University of Batna 2, Batna, Algeria

Received: May 5, 2024; Revised: November 24, 2024

Abstract: In this work, we focus on the utility of chaotic systems of dimension 2 to generate symmetric keys which will be used to encrypt and decrypt data. Non-linear dynamical 2-D systems with chaotic logistic maps have properties that give us the means to hide data to be shared [10]. The two most important properties that are very useful in this work are: the non-linearity that gives a significant complexity to our keys, and the sensibility to the initial conditions that radically transforms our systems as soon as there is a minimal change [1].

Keywords: *matrices; Zeraoulia-Sprott maps; logistic maps; chaos; cryptography; xor operation.*

Mathematics Subject Classification (2010): 70K55, 70K75, 93-00.

1 Introduction

The use of chaotic maps of dimension 2 can be an interesting approach for the encryption of text. 2-D chaotic maps such as the Henon map or the standard map have 2-dimensional dynamic chaos properties [3]. Zeraoulia and Sprott [6] have proposed a new chaotic map of dimension 2,

$$\forall n \in \mathbb{N}; \quad \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} \dfrac{-ax_n}{1+y_n^2} \\ x_n + by_n \end{pmatrix},$$

where $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ are given initial terms, that has the same properties, moreover, these maps can be used to generate complex pseudo-random sequences serving as encryption keys.

* Corresponding author: mailto:s.adoui@univ-batna2.dz

The chaotic transformations of these maps allow to introduce a strong confusion and diffusion in the encryption process. This blurs the links between the clear and encrypted texts and makes statistical analysis more difficult.

Zeraoulia and Sprott map parameters (such as the initial x_0 and y_0 terms, control parameters a and b) [6], [7] can be used to generate unique and unpredictable encryption keys. The sensitivity to the initial conditions ensures a strong uniqueness of the keys. The unpredictability and complexity of these maps make encryption more resistant to brute force attacks etc.

The implementation of 2-D chaotic maps in encryption algorithms requires additional calculations, but increases computing power and facilitates their integration. The use of 2-D chaotic maps for text encryption can provide enhanced security due to the complexity and unpredictability introduced by 2-dimensional chaos. However, it is important to design and optimize the implementation to achieve a balance between security and performance.

In this work, we used the chaotic maps for the creation of a key in matrix form, whose components are the successive terms of a logistic map. This key will be used in symmetric encryption [8], [9] and decryption of our data (text, digital image [11], etc.) using the logical operation xor.

2 The 2-D Rational Discrete Chaotic Map

In [6], [7], a 2-D discrete rationale map is given by

$$\forall n \in \mathbb{N};\quad \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{0.1 + x_n^2} - a y_n \\ \dfrac{1}{0.1 + y_n^2} + b x_n \end{pmatrix}, \textit{ where } \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \textit{ are given initial terms,} \tag{1}$$

a and b are parameters, the 2-D chaotic map (1) is more complicated than the 1-D one. In their papers [6, 7], Zeraoulia and Sprot have proposed a new 2-D chaotic map given by

$$\forall n \in \mathbb{N};\quad \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} \dfrac{-a x_n}{1 + y_n^2} \\ x_n + b y_n \end{pmatrix}, \textit{ where } \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \textit{ are given initial terms,} \tag{2}$$

and a and b are bifurcation parameters. This map is algebraically simple but with more complications, it produces several new chaotic attractors. It leads, according to the values of a and b, to a convergent sequence, a continuation subject to oscillations or a chaotic sequel [6], [7]. The following four cases are considered, with the associated properties proved:

- $|a| < 1$, $|b| < 1$: Global asymptotic stability;
- $|a| < 1$, $|b| > 1$: Existence of unbounded solutions;
- $|a| > 1$, $|b| < 1$: Localization of non-trivial global attractor;
- $|a| > 1$, $|b| > 1$: Existence of unbounded solutions.

And there are some regular and chaotic regions when

$$\begin{cases} (a,b) \in [2,4] \times [0,1] \\ (a,b) \in [3,4] \times [-1,0]. \end{cases} \quad \text{or}$$

3 Construction of a Matrix $\mathfrak{M}$ Using Terms of 2-D Rational Discrete Chaotic Map

Consider the square matrix $\mathfrak{M} = (\mathfrak{m}_{ij})$ generated by the Zeraoulia-Sprott terms. And we obtain the following similar *Toeplidz* matrix $\mathfrak{M}$ [5]:

$$\mathfrak{M}_n = \begin{pmatrix} \frac{x_1+y_1}{2} & x_2 & x_3 & x_4 & \cdots & \cdots & \cdots & x_n \\ y_2 & \frac{x_2+y_2}{2} & x_{n+1} & x_{n+2} & \cdots & \cdots & \cdots & x_{2n-3} \\ y_3 & y_{n+1} & \frac{x_3+y_3}{2} & x_{2n-2} & \cdots & \cdots & \cdots & x_{3n-6} \\ y_4 & y_{n+2} & y_{2n-2} & \frac{x_4+y_4}{2} & \cdots & \cdots & \cdots & x_{4n-9} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & y_{\frac{n^2-n}{2}} \\ y_n & y_{2n-3} & y_{3n-6} & y_{4n-9} & \cdots & \cdots & y_{\frac{n^2-n}{2}} & \frac{x_n+y_n}{2} \end{pmatrix}. \quad (3)$$

We can also write

$$\mathfrak{M} = (m_{ij}) = \begin{cases} m_{ii} = \frac{x_i+y_i}{2} & if \quad i=j, \quad \forall i \in \{1,2,\cdots,n\}, \\ m_{ij} = x_i & if \quad 1<i<j, \\ m_{ij} = y_i & if \quad i>j>1. \end{cases}$$

Encryption data using the matrix $\mathfrak{M}$

We recall that the BB84 protocol (proposed by Charles Bennett and Gilles Brassard in 1984) is a quantum key distribution protocol (QKD) that guarantees the security of the key generation process against attacks. The protocol uses the principles of quantum mechanics to establish a secure key between two interlocutors while detecting the presence of an attempted espionage [2].

4 Application

In our work, the matrix operation we will use is the logical *xor* between two matrices applied component by component. We recall that the *xor* operation is often used in cryptography, especially in flood encryption. This is how it works [4]:

1. **Plain text:** It is the original text that is being protected.

2. **Encryption flow:** It is a sequence of random bits generated using a secure cryptographic algorithm. This bit stream is used as the encryption key.

3. *xor* **operation:** It is a logical operation to compare each bit of the plain text with the corresponding bit of the encryption stream, bit by bit. The result of this *xor* operation gives the encrypted text.

The principle of the *xor* operation is as given in Table 1.

A	1	1	0	0
B	1	0	1	0
$A \oplus B$	0	1	1	0

Table 1: The *xor* operation.

Thus, by applying the *xor* operation between the plain text and the encryption flow, the encrypted text is obtained. Decryption works similarly: the *xor* operation is applied again between the encrypted text and the same encryption stream used for encryption. This makes it possible to find the original plain text. The advantage of the *xor* operation is that it is reversible and very fast to calculate. This is why it is often used in flood encryption algorithms such as One-Time Pad encryption.

Encryption and decryption of a text

In this section we give an application for encrypting and decrypting of a text using some cryptographic techniques.

Let $\mathcal{T}$ be a text to be encrypted, we start by converting it to binary and putting it in a square matrix $\mathcal{T}_c$ of order n.

Let L be the number of characters of the text (letters, symbols, numbers, etc,). We fix the size of the matrix by

$$n = \begin{cases} \sqrt{L} & if \quad \sqrt{L} \in \mathbb{N}, \\ E(\sqrt{L}+1) & if \quad \sqrt{L} \notin \mathbb{N}. \end{cases} \tag{4}$$

Then, we propose the encoding dictionary, see Table 2.

Character	A	B	C	D	E	F	G	H	I	J	K	L	M
Encoded character	01	02	03	04	05	06	07	08	09	10	11	12	13

Character	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Encoded character	14	15	16	17	18	19	20	21	22	23	24	25	26

Character	?	Space	,	.	!	;	:	'
Encoded character	27	28	29	30	31	32	33	34

Table 2: The proposed encoding dictionary.

4.1 Examples

Case 1: (when $n = \sqrt{L} \in \mathbb{N}$).

Let us have the text $\mathcal{T}$= "HELLO BRAHIM, I AM SALAH." We encode it using the previous dictionary, we obtain

$$\begin{aligned} \mathcal{T} &= \text{HELLO BRAHIM, I AM SALAH.} \\ &= 08051212142802170108091229280928011228180112010830. \end{aligned}$$

In this case, we have 25 characters, so $L = 25$ and the matrix size will be $n = 5$.

We set the order $n = 5$ and create a matrix using the code obtained, we get

$$\mathcal{T}_d = \begin{pmatrix} 08 & 05 & 12 & 12 & 14 \\ 28 & 02 & 17 & 01 & 08 \\ 09 & 12 & 29 & 28 & 09 \\ 28 & 01 & 12 & 28 & 18 \\ 01 & 12 & 01 & 08 & 30 \end{pmatrix}.$$

The matrix $\mathcal{T}_d$ has a matrix $\mathcal{T}_b$ whose elements are the binary conversions of each component of the matrix $\mathcal{T}_d$. So

$$\mathcal{T}_b = \begin{pmatrix} 01000 & 00101 & 01100 & 01100 & 01110 \\ 11100 & 00010 & 10001 & 00001 & 01000 \\ 01001 & 01100 & 11101 & 11100 & 01001 \\ 11100 & 00001 & 01100 & 11100 & 10010 \\ 00001 & 01100 & 00001 & 01000 & 11110 \end{pmatrix}.$$

The encryption key

Using the *Zeraoulia-Sprott* map:

- Let x_0, y_0, $2 < a < 4$ and $0 < b < 1$ be the four parameters exchanged between the interlocutors through the quantum channel using the $BB84$ protocol.
- After we have exchanged the four previous parameters, we introduce the Zeraoulia-Sprott map defined in (1) to calculate the $l = \frac{n^2-n}{2}$ first terms of the sequences $\{x_1, \cdots, x_l\}$ and $\{y_1, \cdots, y_l\}$.

In this case, we fix the parameters $x_0 = 2$, $y_0 = -3$, $a = 3$ and $b = 0.5$, we get the terms shown in Table 3.

i	1	2	3	4	5	6	7	8
x_i	-0,6000	1,4400	-3,8485	4,4402	-1,1743	0,3905	-1,1068	2,6322
y_i	0,5000	-0,3500	1,2650	-3,2160	2,8322	0,2417	0,5113	-0,8511
$\frac{x_i+y_i}{2}$	-0,0500	0,5450	-1,2917	0,6121	0,8289	X	X	X

i	9	10	11
x_i	-4,5791	2,3405	-0,5368
y_i	2,2066	-3,4757	0,6027

Table 3: 11 first terms of the Zeraoulia-Sprott sequence and 5 first means.

Using these terms, we construct a matrix of order $n = 5$:

$$\mathfrak{M}_5 = \begin{pmatrix} \frac{x_1+y_1}{2} & x_2 & x_3 & x_4 & x_5 \\ y_2 & \frac{x_2+y_2}{2} & x_6 & x_7 & x_8 \\ y_3 & y_6 & \frac{x_3+y_3}{2} & x_9 & x_{10} \\ y_4 & y_7 & y_9 & \frac{x_4+y_4}{2} & x_{11} \\ y_5 & y_8 & y_{10} & y_{11} & \frac{x_5+y_5}{2} \end{pmatrix}$$

$$= \begin{pmatrix} -0,0500 & 1,4400 & -3,8485 & 4,4402 & -1,1743 \\ -0,3500 & 0,5450 & 0,3905 & -1,1068 & 2,6322 \\ 1,2650 & 0,2417 & -1,2917 & -4,5791 & 2,3405 \\ -3,2160 & 0,5113 & 2,2066 & 0,6121 & -0,5368 \\ 2,8322 & -0,8511 & -3,4757 & 0,6027 & 0,8289 \end{pmatrix}.$$

As the basis of the binaries is $\{0,1\}$, then the class of a negative real number converted to binary is the same as that of the same number taken without sign (positive). For this, we can take all the components of the matrix in absolute values. So we get a new matrix with same results,

$$\mathfrak{M}_5^+ = \begin{pmatrix} 0,0500 & 1,4400 & 3,8485 & 4,4402 & 1,1743 \\ 0,3500 & 0,5450 & 0,3905 & 1,1068 & 2,6322 \\ 1,2650 & 0,2417 & 1,2917 & 4,5791 & 2,3405 \\ 3,2160 & 0,5113 & 2,2066 & 0,6121 & 0,5368 \\ 2,8322 & 0,8511 & 3,4757 & 0,6027 & 0,8289 \end{pmatrix}.$$

To be able to work with natural integer components, we must get rid of commas, for this, we propose to multiply the components of the matrix $\mathfrak{M}_5^+$ by a power k of 10 chosen according to our needs, then we take only the integer part of each component after multiplication. In our example, we can take $k = 3$, therefore, all components of $\mathfrak{M}_5^+$ must be multiplied by 10^3. We get a new matrix

$$E(\mathfrak{M}_5^+) = \begin{pmatrix} 50 & 1440 & 3848 & 4440 & 1174 \\ 350 & 545 & 390 & 1106 & 2632 \\ 1265 & 241 & 1291 & 4579 & 2340 \\ 3216 & 511 & 2206 & 612 & 536 \\ 2832 & 851 & 3475 & 602 & 828 \end{pmatrix}.$$

We convert the components of the matrix $E(\mathfrak{M}_5^+)$ into binary and we obtain a matrix K which will be the common key for encryption and decryption,

$$K = \begin{pmatrix} 110010 & 10110100000 & 111100001000 & 1000101011000 & 10010010110 \\ 101011110 & 1000100001 & 110000110 & 10001010010 & 101001001000 \\ 10011110001 & 11110001 & 10100001011 & 1000111100011 & 100100100100 \\ 110010010000 & 111111111 & 100010011110 & 1001100100 & 1000011000 \\ 101100010000 & 1101010011 & 110110010011 & 1001011010 & 1100111100 \end{pmatrix}.$$

Encryption

To encrypt the text $\mathcal{T}$, we use the formula $\mathcal{T}_c = \mathcal{T}_b \oplus K$. So

$$\mathcal{T}_c = \begin{pmatrix} 111010 & 10110100101 & 111100010100 & 1000101100100 & 10010100100 \\ 101111010 & 1000100011 & 110010111 & 10001001101 & 101001010000 \\ 10011111010 & 11111101 & 10100101000 & 1000111111111 & 100100101101 \\ 110010101100 & 1000000000 & 100010101010 & 1010000000 & 1000101010 \\ 101100010001 & 1101011111 & 110110010100 & 1001100010 & 1101011010 \end{pmatrix}.$$

$\mathcal{T}_c$ will be sent to the receiver.

Decryption

The recipient receives $\mathcal{T}_c$ and decrypts it to obtain the initial matrix using the formula

$$\mathcal{T}_c \oplus K = \mathcal{T}_b.$$

Proof.

$$\begin{aligned} \mathcal{T}_c \oplus K &= \mathcal{T}_b \oplus K \oplus K \\ &= \mathcal{T}_b \oplus O_5 \quad (O_5 \ \textit{is the null matrix of order 5}) \\ &= \mathcal{T}_b. \end{aligned}$$

After this, the receiver converts the binary matrix $\mathcal{T}_b$ to the decimal matrix $\mathcal{T}_d$, then he uses the dictionary, see Table 2, to obtain the initial text " $\mathcal{T}$ = HELLO BRAHIM, I AM SALAH."

Case 2: ($\sqrt{L} \notin \mathbb{N}$).

In the case when $\sqrt{L}$ is not a perfect square, we put all the obtained codes in order from left to right and down, and we put zeros in the remaining places.

Let us take the text $\mathcal{T}$= "HELLO BRAHIM, I AM SALAH'S FRIEND." We encode it using the previous dictionary, we obtain

$$\begin{aligned} \mathcal{T} &= \text{HELLO BRAHIM, I AM SALAH'S FRIEND.} \\ &= 08051212142802170108091229280928011228180112010834192806180905140430. \end{aligned}$$

In this case, we have 34 characters, so $L = 34$ and the matrix size will be $n = E(\sqrt{34} + 1) = 6$.

We set the order $n = 6$ and create a matrix using the code obtained, we get

$$\mathcal{T}_d = \begin{pmatrix} 08 & 05 & 12 & 12 & 14 & 28 \\ 02 & 17 & 01 & 08 & 09 & 12 \\ 29 & 28 & 09 & 28 & 01 & 12 \\ 28 & 18 & 01 & 12 & 01 & 08 \\ 34 & 19 & 28 & 06 & 18 & 09 \\ 05 & 14 & 04 & 30 & 00 & 00 \end{pmatrix}.$$

The matrix $\mathcal{T}_d$ has a matrix $\mathcal{T}_b$ whose elements are the binary conversions of each component of the matrix $\mathcal{T}_d$. So

$$\mathcal{T}_b = \begin{pmatrix} 01000 & 00101 & 01100 & 01100 & 01110 & 11100 \\ 00010 & 10001 & 00001 & 01000 & 01001 & 01100 \\ 11101 & 11100 & 01001 & 11100 & 00001 & 01100 \\ 11100 & 10010 & 00001 & 01100 & 00001 & 01000 \\ 100010 & 10011 & 11100 & 110 & 10010 & 1001 \\ 101 & 1110 & 100 & 11110 & 0000 & 0000 \end{pmatrix}.$$

The encryption key

Using the *Zeraoulia-Sprott* map:

- Let x_0, y_0, $2 < a < 4$ and $0 < b < 1$ be the four parameters exchanged between the interlocutors through the quantum channel using the $BB84$ protocol.
- After we have exchanged the four previous parameters, we introduce the Zeraoulia-Sprott map defined in (1) to calculate the $l = \frac{n^2-n}{2}$ first terms of the sequences $\{x_1, \cdots, x_l\}$ and $\{y_1, \cdots, y_l\}$.

The matrix size will be 6, so we calculate the first 16 terms of each of the two sequences whose first 11 terms will be the same as those calculated in the first case, see Table 3. We get

i	$\cdots$	6	$\cdots$	11	12	13	14	15	16
x_i	$\cdots$	0,3905	$\cdots$	-0,5368	1,1813	-3,3577	4,7266	-1,5780	0,3952
y_i	$\cdots$	0,2417	$\cdots$	0,6027	-0,2354	1,0636	-2,8259	3,3136	0,0788
$\frac{x_i+y_i}{2}$	$\cdots$	0,3161	$\cdots$	X	X	X	X	X	X

Using these terms, we construct a matrix of order $n = 6$:

$$\mathfrak{M}_6 = \begin{pmatrix} \frac{x_1+y_1}{2} & x_2 & x_3 & x_4 & x_5 & x_6 \\ y_2 & \frac{x_2+y_2}{2} & x_7 & x_8 & x_9 & x_{10} \\ y_3 & y_7 & \frac{x_3+y_3}{2} & x_{11} & x_{12} & x_{13} \\ y_4 & y_8 & y_{11} & \frac{x_4+y_4}{2} & x_{14} & x_{15} \\ y_5 & y_9 & y_{12} & y_{14} & \frac{x_5+y_5}{2} & x_{16} \\ y_6 & y_{10} & y_{13} & y_{15} & y_{16} & \frac{x_6+y_6}{2} \end{pmatrix}$$

$$= \begin{pmatrix} -0,0500 & 1,4400 & -3,8485 & 4,4402 & -1,1743 & 0,3905 \\ -0,3500 & 0,5450 & -1,1068 & 2,6322 & -4,5791 & 2,3405 \\ 1,2650 & 0,5113 & -1,2917 & -0,5368 & 1,1813 & -3,3577 \\ -3,2160 & -0,8511 & 0,6027 & 0,6121 & 4,7266 & -1,5780 \\ 2,8322 & 2,2066 & -0,2354 & -2,8259 & 0,8289 & 0,3952 \\ 0,2417 & -3,4757 & 1,0636 & 3,3136 & 0,0788 & 0,3161 \end{pmatrix}.$$

Then we get a new positive matrix with same results

$$\mathfrak{M}_6^+ = \begin{pmatrix} 0,0500 & 1,4400 & 3,8485 & 4,4402 & 1,1743 & 0,3905 \\ 0,3500 & 0,5450 & 1,1068 & 2,6322 & 4,5791 & 2,3405 \\ 1,2650 & 0,5113 & 1,2917 & 0,5368 & 1,1813 & 3,3577 \\ 3,2160 & 0,8511 & 0,6027 & 0,6121 & 4,7266 & 1,5780 \\ 2,8322 & 2,2066 & 0,2354 & 2,8259 & 0,8289 & 0,3952 \\ 0,2417 & 3,4757 & 1,0636 & 3,3136 & 0,0788 & 0,3161 \end{pmatrix}.$$

We follow the same steps we did in the first case, we get the integer matrix

$$E(\mathfrak{M}_6^+) = \begin{pmatrix} 50 & 1440 & 3848 & 4440 & 1174 & 390 \\ 350 & 545 & 1106 & 2632 & 4579 & 2340 \\ 1265 & 511 & 1291 & 536 & 1181 & 3357 \\ 3216 & 851 & 602 & 612 & 4726 & 1578 \\ 2832 & 2206 & 235 & 2825 & 828 & 395 \\ 241 & 3475 & 1063 & 3313 & 78 & 316 \end{pmatrix}.$$

We convert the components of the matrix $E(\mathfrak{M}_5^+)$ into binary and we obtain a matrix

$K = [K^{'}|K^{''}]$, where

$$K^{'} = \begin{pmatrix} 110010 & 10110100000 & 111100001000 \\ 101011110 & 1000100001 & 10001010010 \\ 10011110001 & 111111111 & 10100001011 \\ 110010010000 & 1101010011 & 1001011010 \\ 101100010000 & 100010011110 & 11101011 \\ 11110001 & 110110010011 & 10000100111 \end{pmatrix}$$

$$K^{''} = \begin{pmatrix} 1000101011000 & 10010010110 & 110000110 \\ 101001001000 & 1000111100011 & 100100100100 \\ 1000011000 & 10010011101 & 110100011101 \\ 1001100100 & 1001001110110 & 11000101010 \\ 101100001001 & 1100111100 & 110001011 \\ 110011110001 & 1001110 & 100111100 \end{pmatrix}.$$

This matrix K will be the common key for encryption and decryption.

Encryption

To encrypt the text $\mathcal{T}$, we use the formula $\mathcal{T}_c = \mathcal{T}_b \oplus K = [\mathcal{T}_c^{'}|\mathcal{T}_c^{''}]$, where

$$\mathcal{T}_c^{'} = \begin{pmatrix} 111010 & 10110100101 & 111100010100 \\ 101100000 & 1000110010 & 10001010011 \\ 10100001110 & 1000011011 & 101000010100 \\ 110010101100 & 1101100101 & 1001011011 \\ 101100110010 & 100010110001 & 100000111 \\ 11110110 & 110110100001 & 10000101011 \end{pmatrix},$$

$$\mathcal{T}_c^{''} = \begin{pmatrix} 1000101100100 & 10010100100 & 110100010 \\ 101001010000 & 10000111101100 & 100100110000 \\ 1000110100 & 10010011110 & 110100101001 \\ 1001110000 & 1001001110111 & 11000110010 \\ 101100001111 & 1101001110 & 110010100 \\ 110100001111 & 1001110 & 100111100 \end{pmatrix}.$$

$\mathcal{T}_c$ will be sent to the receiver.

Decryption

The recipient receives $\mathcal{T}_c$ and decrypts it to obtain the initial matrix using the formula

$$\mathcal{T}_c \oplus K = \mathcal{T}_b.$$

Proof.

$$\begin{aligned} \mathcal{T}_c \oplus K &= \mathcal{T}_b \oplus K \oplus K \\ &= \mathcal{T}_b \oplus O_6 \quad (O_6 \ \textit{ is the null matrix of order 6}) \\ &= \mathcal{T}_b. \end{aligned}$$

After this, the receiver converts the binary matrix $\mathcal{T}_b$ to the decimal matrix $\mathcal{T}_d$, then he uses the dictionary, see Table 2, to obtain the initial text

$\mathcal{T}$ = " HELLO BRAHIM, I AM SALAH'S FRIEND."

Generalisation

In general, to encrypt a text of length L, we follow the same steps as in the previous example.

Step 1: The text

(a) We choose a dictionary to code the text $\mathcal{T}$ (In our example, we have used the dictionary given in Table 2).

(b) Let $\mathcal{T}_d$ be the matrix of order n. The order n is calculated according to the formula (4). And the components of $\mathcal{T}_d$ are the obtained codes put in order from left to right and down, and we put zeros in the remaining places.

(c) We convert the components of the matrix $\mathcal{T}_d$ into binaries, and we obtain a matrix $\mathcal{T}_b$.

Step 2: The encryption key

(a) Let us choose the chaotic sequence of *Zeraoulia* of dimension 2 (see formula (2)). Then we fix the first terms (x_0, y_0) and two parameters a and b such that the chaos is assured [6], [7], the calculus of the first $\dfrac{n^2-n}{2}$ terms of the vector sequence $(x_i, y_i)_{i\in\{1,\cdots,\frac{n^2-n}{2}\}}$ gives us the components of $\mathfrak{M}_n$, the matrix of order n in the form given in the expression (3).

(b) We take all the components of the matrix $\mathfrak{M}_n$ in absolute values to get the matrix $\mathfrak{M}_n^+$.

(c) We multiply the components of the matrix $\mathfrak{M}_n^+$ by a power k of 10 chosen according to our needs, then we take only the integer part of each component after multiplication, then we create a new matrix denoted $E(\mathfrak{M}_n^+)$ whose components are the integer parts of the components of $\mathfrak{M}_n^+$.

(d) We convert the components of the matrix $E(\mathfrak{M}_n^+)$ into binary and we obtain a matrix K which will be the common key for encryption and decryption.

Step 3: Encryption and Decryption

(a) To encrypt the encoded text $\mathcal{T}_b$, we use the formula

$$\mathcal{T}_c = \mathcal{T}_b \oplus K.$$

(b) To decrypt $\mathcal{T}_c$ for obtaining the initial matrix, we use the formula

$$\mathcal{T}_c \oplus K = \mathcal{T}_b.$$

(c) The binary matrix $\mathcal{T}_b$ will be converted to a decimal matrix which will decoded to a text using the initial dictionary, see Table 2, to obtain the initial text " $\mathcal{T}$ ".

5 Performance and Security Analysis

To study the efficacy of our text encryption, we test its security. The proposed method should resist several types of attacks because its symmetric keys used during the encryption and decryption must be transmitted through an unsecured channel.

Cryptanalysis

To determine the key, it is necessary to use techniques more secured and compliant against attacks, these techniques are called the key exchange protocols. In our system, we detail how we can obtain a secret key using the properties of matrices for encrypting and decrypting text and sensibility of chaotic maps to initial conditions. The question is: Can we ensure the security of this encryption? For this, to raise the security levels of our system, we have introduced chaotic logistic maps in cryptography.

For the implementation of the proposed scheme, we choose the size of text $127^2 < L \leq 128^2$. The proposed scheme key $\mathcal{K}$ is none-deterministic because the interlocutors use an arbitrary matrix ($\mathcal{K}$) for getting a common secret key $\mathcal{K}$.

We use the proposed key generation method with x_i, y_i; $i \in \{0, \cdots, 8128\}$; ($x_i, y_i$ are the components of the key $\mathcal{K}$). This provides 10^{k+1} possible cases to obtain one component of the key $\mathcal{K}$. So it provides $(10^{k+1})^{n^2} = 10^{(k+1)n^2}$ possible cases to obtain the key K.

For $n = 128$ and $k = 14$, we get $(10^{15})^{128^2} = 10^{(15)16384} = 10^{245760} > (2^3)^{245760} = 2^{737280}$.

We have also 10^{20} possible cases for getting the term x_0 (with 20 decimal digits after the comma in the set of 10 numbers $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$) if we use the 2-D rational discrete chaotic map that contains 2 parameters a, b and the initial terms (x_0, y_0), so this provides

$$(10^{20})^4 = 10^{80} > 2^{260}.$$

The key space is wide enough for a brute force attack or exhaustive attack is not possible.

6 Concluding Remarks

We know that chaos can be exploited in encryption algorithms to improve the security and robustness of encryption systems. That is why we have included a chaotic system of dimension 2 with the following advantages:

- Chaotic encryption systems use non-linear dynamic systems to generate complex pseudo-random sequences that serve as encryption keys.
- Sensitivity to initial conditions of chaotic systems makes encryption very difficult to break and ensures a strong uniqueness of the generated keys.
- Chaotic properties can be used to generate complex and unpredictable encryption keys from initial parameters.
- Chaotic transformations can be incorporated into the dissemination and confusion stages of encryption algorithms to further blur the links between the plain and encrypted texts.
- This makes encryption more resistant to statistical analysis attacks.
- Unpredictability and sensitivity to chaotic system parameters can be used to enhance the security of encrypted communication protocols.

- This makes them more resistant to brute force attacks and model analysis attacks.

The judicious use of chaos in encryption algorithms leads to safer and more robust encryption systems against various cryptographic attacks.

In our work, we used a chaotic system of dimension two, which already has the venture to keep the two most important options: the non-linearity of the system and its sencibility to the initial conditions, in addition to that, it has increased the security level of the shared key to 2^{260}, which far exceeds the known threshold.

References

[1] A. Sambas, S. Vaidyanathan, S. F. Al-Azzawi, M. K. M. Nawawi, M. A. Mohamed, Z. A. Zakaria, S. S. Abas and M. Mamat. Modelling and MultiSim Simulation of a New Hyperchaos System with No Equilibrium Point. *Nonlinear Dynamics and Systems Theory* **23** (4) (2023) 422–433.

[2] S. Adoui, B. Benzeghli and L. Noui. Sharing Keys Using Circular Matrices and Logistic Maps Through Quantum Channel. *Advances in Mathematics: Scientific Journal* CCS **12** (2022) 1361–1378.

[3] M. Ausloos. *The Logistic Map and the Route to Chaos: From the Beginnings to Modern Applications*. Springer, 2006.

[4] G. Boole. *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*. Cambridge (London: Macmillan, Barclay and Macmillan George Bell), 2006.

[5] A. Böttcher and B. Silbermann. *Introduction to Large Truncated Toeplitz Matrices*. Springer, New York, 2012.

[6] Z. Elhadj and J. C. Sprott. On the Dynamics of a New Simple 2-d Rational Discrete Mapping. *Int. J. Bifurc. Chaos* **21** (2011) 155–160.

[7] G. Chen, E.V. Kudryashova, N. V. Kuznetsov and G. A. Leonov. Dynamics of the Zeraoulia-Sprott map revisited. *International Journal of Bifurcation and Chaos* **26** (07) (2016) 1650126.

[8] J. Lu, X. Wu, J. Lû and K. Kang. A new discrete chaotic system with rational fraction and its dynamical behaviors. *Chaos Solit. Fract.* **22** (2004) 311–319.

[9] Marzieh Azadi and Hossein Jafari. Lie Symmetry Reductions of a Coupled Kdv System of Fractional Order. *Nonlinear Dynamics and Systems Theory* **18** (1) (2018) 22–28.

[10] N. Djafri, T. Hamaizia and F. Derouiche. Boundedness and Dynamics of a Modified Discrete Chaotic System with Rational Fraction. *Nonlinear Dynamics and Systems Theory* **21** (1) (2021) 68–75.

[11] W. S. Mada Sanjaya, Akhmad Roziqin, Agung Wijaya Temiesela, M. Fauzi Badru Zaman, Aria Dewa Wibiksana and Dyah Anggraeni. Moore-Spiegel Chaotic Encryption for Digital Images and Voices. *Nonlinear Dynamics and Systems Theory* **23** (5) (2023) 445–460.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 549–560

The Advection-Diffusion-Reaction Equation: A Numerical Approach Using a Combination of Approximation Techniques

N. Arar [1], Z. Laouar [2,3]* and A. Hioual [4]

[1] *Mathematics and Decision Sciences Laboratory (LAMASD), Department of Mathematics, University Constantine 1, Frères Mentouri, Algeria.*
[2] *Laboratory of Applied Mathematics and Didactics, ENS El Katiba Assia Djebar, Constantine, Algeria.*
[3] *University Center Abdelhafid Boussouf, Mila, Algeria.*
[4] *Laboratory of Dynamical Systems and Control, University of Oum El Bouaghi, Oum El Bouaghi 04000, Algeria.*

Received: April 19, 2024; Revised: November 3, 2024

Abstract: In this paper, we propose a numerical technique for solving the advection-diffusion-reaction equation. The presented approach is based on coupling two numerical methods to address the problem posed with the Robin boundary conditions perturbed with a small parameter ε, in terms of spatial and temporal variables. We start by employing a Galerkin method for the spatial descretization, using a compact basis of the Legendre polynomials to derive a system of ordinary differential equations. This system is then solved using a Crank-Nicolson scheme, with the temporal domain uniformly discretized. The obtained numerical results demonstrate the effectiveness of the proposed numerical method and the convergence of the approximate solution to the analytic solution of the classical problem with the homogeneous Dirichlet boundary conditions when ε approaches zero. This makes it particularly useful for approximating the solutions of such problems of partial differential equations appearing in reaction-diffusion systems, where the explicit solution is unknown under various types of boundary conditions.

Keywords: *advection-diffusion-reaction equation; Galerkin method; Legendre polynomials; Robin boundary conditions; Crank-Nicolson scheme.*

Mathematics Subject Classification (2010): 65N35, 65N30, 65N22, 33C45, 70K99, 93C20.

* Corresponding author: mailto:z.laouar@centre-univ-mila.dz

1 Introduction

Partial differential equations (PDEs) represent a mathematical tool that connect different functions and their derivatives. These functions reflect, generally, some physical quantities like heat and waves [1]. Until now, their study is still very active, gathering the resolution of systems, daily problems and the development of mathematics. The problem for the majority of PDEs appears in the calculation of the analytic solution, which is generally impossible. For this reason, mathematicians headed to other tools such as numerical resolution to approximate the solution of such problems in an effective way in terms of time and results.

In this paper, we focus on the numerical study of the advection-diffusion-reaction equation. The latter brings together three important processes: advection, diffusion and reaction. It is formulated as follows:

$$\frac{\partial C}{\partial t} = D_0 \frac{\partial^2 C}{\partial x^2} - V_0 \frac{\partial C}{\partial x} - K_0 C + f(x,t),$$

where D_0 is the diffusion coefficient, V_0 is the convective velocity, K_0 is the reaction constant and $f(x,t)$ is a scalar function often called the source term. It models, according to the problem, a heat source, chemical reaction, injection/production wells, etc.

This equation occurs in several scientific disciplines such as biology, astrophysics, and industrial and environmental issues. It models many phenomena, for example energy transfer, mass transfer and also the transfer of the heat through a permeable medium and the transport of a chemical or biological pollutant through an underground aquifer system [2–4]. Generally, it describes the phenomenon of the distribution of some quantities in space and time. The advection-diffusion-reaction equation can be viewed as a special case of a reaction-diffusion system (systems that describe the dynamics of chemical concentrations reacting and diffusing in space), where the process of advection is also taken into account. In other words, this equation is an extension of reaction-diffusion equations to include the effects of advection [5].

Moving to numerical solution, different methods have been introduced and show hight accuracy to approximate the desired solution, we can cite: finite differences, finite volumes and finite elements. The principle of these methods is the same for all the numerical methods "*searching for discrete numerical values that approach the exact solution*" [6].

Another branch of numerical methods recently appeared are the spectral methods developed by D. Gottieb and S. Orszag in 1970, based on the use of a finite expansion of certain eigenfunctions obtained from the Sturm-Liouville problem [7–10]. This development gives a high level of precision which is superior to the other mentioned methods, thus, it requires a small number of grid points to get the desired precision [11]. Another advantage of these methods is that they are less intensive in terms of time and memory compared to finite elements, but they become less precise if we consider problems with complex geometry. These approaches are applicable for the resolution of different problems such as the resolution of ODEs, linear and nonlinear PDEs and eigenvalue problems [10, 12–14].

In this work, we develop an efficient numerical method basing on a coupling of spectral methods and finite differences schemes. The model problem is posed with perturbed boundary conditions of Robin type, so that we can apply a Legendre-Galerkin approach according to the spatial variable and a Cranck-Nicolson scheme according to the temporal one. The way in which the conditions are perturbed makes it possible to compare the obtained approximation and the exact solution of the same problem with the Dirichlet boundary conditions.

The structure of this paper is as follows. In Section 2, we remind some preliminaries and essential tools required for the elaboration of the presented study. Next, the model problem is presented in Section 3 with the adaptative variational formulation, the existence and uniqueness of the solution are also proved. In Section 4, we outline the principle of the presented approach and we study its convergence and give an estimation of the error of approximation. Then, in the same section, the implementation of the proposed technique and the coupling with the finite differences scheme are exposed to obtain the final system to solve. Section 5 addresses the proof of the efficiency of the algorithm via different numerical examples by showing the convergence of the approximation to the analytic solution of the classic problem with the homogeneous Dirichlet boundary conditions, when ε reaches zero.

2 Preliminaries

Let $I = (-1, 1)$. We define

$$L^2(I) = \{v\,;\ \ v \text{ is mesurable on } I \ \text{ and } \ \|v\| < +\infty\}.$$

The scalar product is $\langle u, v\rangle = (u, v)_{L^2} = \int_I u(x)v(x)\,\mathrm{d}x$, and the norm is defined by $\|v\|_{L^2} = (v, v)^{\frac{1}{2}}_{L^2}$.

For every positive m, we define the Sobolev space by

$$H^m(I) = \left\{v\,;\ \ \frac{\partial^k v}{\partial x^k} \in L^2(I), \quad 0 \le k \le m\right\},$$

and the standard semi-norm and norm are $|v|_{L^2} = \left\|\frac{\partial v}{\partial x}\right\|_{L^2}$; $\|v\|_{H^m} = \sum_{k=0}^{m} \left\|\frac{\partial^k v}{\partial x^k}\right\|_{L^2}$.
Let $(H^1(-1, 1))^*$ be the dual space of $H^1(-1, 1)$ with a norm defined by

$$\|g\|_{H^{1*}} = \sup_{\substack{v\in H^1(I)\\ v\neq 0}} \frac{\langle g, v\rangle}{\|v\|_{H^1}}.$$

Thus, and since $\|v\|_{L^2} \le C\|v\|_{H^1}$ for all $v \in H^1(I)$, we can write

$$\|g\|_{H^{1*}} \le C\|g\|_{L^2}. \tag{1}$$

We denote by $\mathscr{P}_N$ the space of polynomials of degree that is less than or equal to N.
Let $L_n(x)$; $x \in I$ be the standard Legendre polynomial of degree n. The family of Legendre polynomials $L_k(x)_{k\in\mathbb{N}}$ constitutes a Hilbert basis of $L^2(I)$ and they are solutions of the following differential Legendre equation:

$$\left(1 - x^2\right) L_n^{''}(x) - 2xL_n^{'}(x) + n\,(n + 1)\,L_n(x) = 0, \quad n \ge 0.$$

The polynomial $L_n(x)$ is of degree n for all $n \in \mathbb{N}$, and the coefficient of its highest degree term is $\dfrac{(2n)!}{2^n(n!)^2}$. They satisfy

$$\forall n \neq m \in \mathbb{N}, \qquad \int_{-1}^{1} L_n(x)L_m(x)\,\dot{\mathrm{x}} = 0 \quad \text{ and } \quad \int_{-1}^{1} L_n^2(x)\,\dot{\mathrm{x}} = \frac{2}{2n+1}.$$

The Legendre polynomials satisfy the following recurrence relations [9]:

$$L_n(1) = 1 \quad \text{and} \quad L_n(-x) = (-1)^n L_n(x) \Longrightarrow L_n(-1) = (-1)^n,$$
$$(n+1)\, L_{n+1}(x) - (2n+1)\, x L_n(x) + n L_{n-1}(x) = 0, \quad n \geq 1,$$
$$(2n+1) L_n(x) = L'_{n+1}(x) - L'_{n-1}(x), \quad n \geq 1.$$

3 The Model Problem

We consider the advection-diffusion-reaction problem with mixed Robin-type boundary conditions disturbed with a small parameter ε,

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,x) + \alpha \dfrac{\partial u}{\partial x}(t,x) - \beta \dfrac{\partial^2 u}{\partial x^2}(t,x) + \lambda u(t,x) = f(t,x);\ -1 < x < 1,\ t > 0, \\ u(t,-1) - \varepsilon \dfrac{\partial u}{\partial x}(t,-1) = 0; \\ u(t,1) + \varepsilon \dfrac{\partial u}{\partial x}(t,1) = 0, \end{cases} \tag{2}$$

where $\varepsilon \in]0,1]$, and $u(0,x) = u_0 = g(x)$ is a given initial condition.
In this study, we focus on the case where the reaction, advection and diffusion coefficients are scalars. Let $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}_+^*$ and $\lambda \in \mathbb{R}$.

Multiplying the equation of problem (2) by v which depends only on x, and integrating by parts on I, we obtain

$$\int_{-1}^{1} \frac{\partial u}{\partial t} v(x)\,\mathrm{d}x + \beta \int_{-1}^{1} \frac{\partial u}{\partial x}\frac{dv}{dx}\,\mathrm{d}x + \alpha \int_{-1}^{1} \frac{\partial u}{\partial x}(t,x) v(x)\,\mathrm{d}x + \lambda \int_{-1}^{1} u(t,x)v(x)\,\mathrm{d}x$$
$$=$$
$$\int_{-1}^{1} f(t,x)v(x)\,\mathrm{d}x + \beta \left[\frac{\partial u}{\partial x} v(x)\right]_{-1}^{1}.$$

The boundary conditions give

$$\frac{\partial u}{\partial x}(-1) = \frac{u(-1)}{\varepsilon}, \quad \frac{\partial u}{\partial x}(1) = -\frac{u(1)}{\varepsilon}.$$

Hence the weak formulation of the problem (2) is

$$\begin{cases} \text{Find } \ u(t) \in H^1(I) \ \text{ such that} \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\langle u(t), v\rangle + a(u(t), v) = \langle f(t), v\rangle, \end{cases} \tag{3}$$

with the initial condition $u(0) = g(x)$ and where

$$\begin{aligned} a(u(t), v) =& \beta \int_{-1}^{1} \frac{\partial u}{\partial x}(t,x)\frac{dv}{dx}(x)\,\mathrm{d}x + \alpha \int_{-1}^{1} \frac{\partial u}{\partial x}(t,x)v(x)\,\mathrm{d}x + \lambda \int_{-1}^{1} u(t,x)v(x)\,\mathrm{d}x \\ &+ \frac{\beta}{\varepsilon}(u(1)v(1) + u(-1)v(-1)). \end{aligned} \tag{4}$$

Theorem 3.1 *Let $T > 0$ be a final time, $g \in L^2(I)$ be an initial data and $a(.,.)$ be the bilinear form given in (4). The following problem has a unique solution $u \in L^2(]0,T[;H^1(I)) \cap C([0,T];L^2(I))$:*

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\langle u(t), v\rangle + a(u(t), v) = \langle f(t), v\rangle, \quad \forall v \in H^1(I), \quad 0 < t < T, \\ u(t=0) = g(x). \end{cases} \tag{5}$$

In addition, we have

$$\frac{1}{2}\int_{-1}^{1} u^2(x,t)\,\mathrm{d}x + \beta\int_0^t\int_{-1}^{1}\left|\frac{\partial u}{\partial x}(s,x)\right|^2 \mathrm{d}x\,\mathrm{d}s + \lambda\int_0^t\int_{-1}^{1} u^2(s,x)\,\mathrm{d}x\,\mathrm{d}s$$
$$+\frac{\alpha}{2}\int_0^t (u^2(s,1) - u^2(s,-1))\,\mathrm{d}s + \frac{\beta}{\varepsilon}\int_0^t (u^2(s,1) + u^2(s,-1))\,\mathrm{d}s$$
$$=$$
$$\frac{1}{2}\int_{-1}^{1} u^2(0,x)\,\dot{x} + \int_0^t\int_{-1}^{1} f(s,x)u(s,x)\,\mathrm{d}x\,\mathrm{d}s.$$

This leads to the following energy estimate:

$$\|u\|^2_{C([0,T];L^2(I))} + m\|u\|^2_{L^2(]0,T[;H^1(I))} \le C\left(\|u_0\|^2_{L^2} + \|f\|^2_{L^2([0,T];L^2(I))}\right). \tag{6}$$

Proof. It is clear that $a(.,.)$ is a symmetric bilinear form. So, we prove the continuity and coercivity to ensure the existence and uniqueness.

The continuity is ensured by using the Cauchy-Schwarz inequality and the fact that

$$\mid u(\pm 1) \mid \le \sup_{x\in[-1,1]} \mid u(x) \mid \; = \|u\|_{L^\infty}.$$

So, $\exists \delta > 0,\ \delta = \theta + |\alpha| + \dfrac{2\beta}{\varepsilon}$ such that

$$\forall u(t), v \in H^1(I) \qquad \mid a(u(t), v) \mid \le \delta \|u(t)\|_{H^1}\|v\|_{H^1}$$

for $\theta = \max\limits_{\lambda\in\mathbb{R},\ \beta\in\mathbb{R}_+^*} \{\beta, |\lambda|\}$.

The coercivity is also ensured. In fact, we have

$$a(u(t), u(t)) \ge \beta \left\|\frac{\partial u}{\partial x}(t)\right\|^2_{L^2} + \frac{\alpha}{2}\left(u^2(t,1) - u^2(t,-1)\right) + \lambda\|u(t)\|^2_{L^2}.$$

So

- If $\alpha \ge 0$ and $\lambda \ge 0$, we have $M = \underset{\lambda\in\mathbb{R},\ \beta\in\mathbb{R}_+^*}{min} \{\beta, \lambda\} > 0$ such that

$$\forall u(t) \in H^1(I), \quad a\left(u(t), u(t)\right) \ge M\|u(t)\|^2_{H^1}.$$

- If $\alpha \ge 0$ and $\lambda \le 0$, we have $M = \beta > 0$ and $\eta = \beta - \lambda$ such that

$$a(u(t), u(t)) + \eta\|u(t)\|^2_{L^2} \ge M\|u(t)\|^2_{H^1}, \quad \forall u(t) \in H^1(I).$$

- If $\alpha < 0$ and $\lambda \geq 0$, we have $\eta = -\frac{\alpha}{2}$ and $M = \underset{\lambda\in\mathbb{R},\, \beta\in\mathbb{R}_+^*}{min}\{\beta, \lambda\} > 0$ such that

$$a(u(t), u(t)) + \eta\|u(t)\|_{L^2}^2 \geq M\|u(t)\|_{H^1}^2, \quad \forall u(t) \in H^1(I).$$

- If $\alpha < 0$ and $\lambda \leq 0$, we have $\eta = \beta - \frac{\alpha}{2} - \lambda$ and $M = \beta$ such that

$$a(u(t), u(t)) + \eta\|u(t)\|_{L^2}^2 \geq M\|u(t)\|_{H^1}^2, \quad \forall u(t) \in H^1(I).$$

Hence the existence and uniqueness of the solution of the problem (5) is proved.
For the second equality, by integrating the formula (3) on [0,t] and for all $t \in [0, T]$, we obtain the energy equality (6).

Using the previous energy equality and posing $\sigma = \underset{\beta\in\mathbb{R}_+^*,\lambda\in\mathbb{R}}{min}\{\beta, \lambda\}$, we have

$$\frac{1}{2}\|u(t)\|_{L^2}^2 + \sigma\int_0^t \|u(s)\|_{H^1}^2\,\mathrm{d}s + \frac{\alpha}{2}\int_0^t (u^2(s,1) - u^2(s,-1))\,\mathrm{d}s$$
$$\leq \frac{1}{2}\|u_0\|_{L^2}^2 + \int_0^t \|f(t)\|_{L^2}\|u(s)\|_{L^2}\,\mathrm{d}s + \frac{2\beta}{\varepsilon}\int_0^t \|u\|_{H^1}^2\,\mathrm{d}s.$$

And from Young's algebraic inequality, there exists $k > 0$ such that

$$\|u(t)\|_{L^2}^2 + 2\left(\sigma - k - \frac{2\beta}{\varepsilon}\right)\int_0^t \|u(s)\|_{H^1}^2\,\mathrm{d}s + \alpha\int_0^t (u^2(s,1) - u^2(s,-1))\,\mathrm{d}s$$
$$\leq \|u_0\|_{L^2}^2 + \frac{1}{2k}\int_0^t \|f(s)\|_{L^2}^2\,\mathrm{d}s.$$

By a simple calculation, we show

$$\|u(t)\|_{L^2}^2 + 2\left(\sigma - k - \frac{2\beta}{\varepsilon} - \frac{\alpha}{2}\right)\int_0^t \|u(s)\|_{H^1}^2\,\mathrm{d}s \leq \|u_0\|_{L^2}^2 + \frac{1}{2k}\int_0^t \|f(s)\|_{L^2}\,\mathrm{d}s.$$

And the desired energy estimate (6) is obtained for $m = 2\left(\sigma - k - \frac{2\beta}{\varepsilon} + \frac{\alpha}{2}\right)$ and $C = \max_{k\geq 0}\left\{1, \frac{1}{2k}\right\}$.

4 Legendre-Galerkin Approximation for the Advection-Diffusion-Reaction Equation

Let N be a positive integer, we consider

$$\mathscr{L}_N = [\{L_0, L_1, \ldots, L_N\}].$$

We define the finite dimensional space V_N included in the space $H^1(I)$ by

$$V_N = \{v \in \mathscr{L}_N \;\; such\ that \;\; v(-1) - \varepsilon v'(-1) = 0 \;\;\; and \;\;\; v(1) + \varepsilon\, v'(1) = 0\}.$$

Then the Legendre spectral scheme for (3) is

$$\left\{\begin{array}{c} \text{Find } \; u_N(t) \in V_N \; \text{ such that} \\ \frac{d}{dt}\langle u_N(t), v_N\rangle + a(u_N(t), v_N) = \langle f(t), v_N\rangle; \quad \forall\, v_N \in V_N. \end{array}\right. \tag{7}$$

At this stage, we choose as the basis functions for V_N a family of polynomials constructed from the orthogonal Legendre polynomials given as

$$\varphi_k(x) = L_k(x) + a_k L_{k+1}(x) + b_k L_{k+2}(x); \qquad k = 0, 1, 2, \dots, \tag{8}$$

where a_k and b_k are the coefficients that can be determined once φ_k verifies the boundary conditions of the problem (2). Then, for $k \geq 0$ and $\varepsilon > 0$, we obtain

$$a_k = 0; \qquad b_k = -\frac{1 + \dfrac{\varepsilon}{2} k(k+1)}{1 + \dfrac{\varepsilon}{2}(k+2)(k+3)}.$$

Since the elements of the polynomial family $\{\varphi_k\}_k$ are linearly independent, we have $V_N = [\{\varphi_0, \varphi_1, \dots, \varphi_{N-2}\}]$, and the desired approximation can be written as

$$u_N(t, x) = \sum_{k=0}^{N-2} u_k(t)\varphi_k(x). \tag{9}$$

4.1 Convergence and error estimation

Theorem 4.1 *Let u_N be the solution of the problem* (7). *There exists a constant C which depends on M and does not depend on N so that for any $t > 0$, we have*

$$\|u_N(t)\|_{L^2}^2 + M \int_0^t \|u_N(s)\|_{L^2}^2 \,\mathrm{d}s \leq \|u_N(0)\|_{L^2}^2 + C \int_0^t \|f(s)\|_{L^2}^2 \,\mathrm{d}s. \tag{10}$$

Proof. By taking $v = u_N(t)$, in the problem (7), we obtain for every $t > 0$,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_N(t)\|_{L^2}^2 + a(u_N(t), u_N(t)) = \langle f(t), u_N(t)\rangle_{L^2}.$$

From the coercivity of the bilinear form $a(.,.)$ and by making use of Young's algebraic inequality, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u_N(t)\|_{L^2}^2 + M\|u_N(t)\|_{L^2}^2 \leq \frac{1}{M}\|f(t)\|_{L^2}^2. \tag{11}$$

Finally, we integrate the expression (11) for $t \in [0, T]$ to have (10) with $u_N(0) = g(x)$ and the constant $C = \dfrac{1}{M}$ does not depend on N.

Now, to show the convergence of the proposed spectral method, we introduce the following theorem. First, we define the function e by $\mathrm{e}(t) = R_N u(t) - u_N(t)$, where

$$R_N : H^1(I) \longrightarrow V_N \; ; \qquad \|u - R_N u\|_{H^1} \xrightarrow[N \to +\infty]{} 0, \;\; \forall u \in H^1(I).$$

Theorem 4.2 *Let u be the solution of the problem* (5) *and u_N be the solution of the problem (7). Then we have the following error estimation:*

$$\|\mathrm{e}(t)\|^2 + M\int_0^t \|\mathrm{e}(s)\|_{H^1(-1,1)}^2 \,\mathrm{d}s$$
$$\leq \|\mathrm{e}(0)\|^2 + C\int_0^t \left\|\frac{\partial u}{\partial t} - R_N\frac{\partial u}{\partial t}\right\|_{H^{1*}}^2 \mathrm{d}s + \delta\int_0^t \|(u - R_N u)\|_{H^1}^2 \,\mathrm{d}s,$$

where C and δ are two constants not depending on N.

Proof. We can write (cf. Chapter 6 in [9])

$$\frac{1}{2}\frac{d}{dt}\|\mathrm{e}\|^2 + M\|\mathrm{e}\|^2_{H^1} \leq \left|\langle \frac{\partial u}{\partial t} - R_N\frac{\partial u}{\partial t}, \mathrm{e}\rangle + a(u - R_N u, \mathrm{e})\right|. \tag{12}$$

From the continuity of $a(.,.)$ and the formula (1), we have

$$\left|\langle \frac{\partial u}{\partial t} - R_N\frac{\partial u}{\partial t}, \mathrm{e}\rangle + a(u - R_N u, \mathrm{e})\right| \leq \|\mathrm{e}\|_{H^1}\left(C\left\|\frac{\partial u}{\partial t} - R_N\frac{\partial u}{\partial t}\right\|_{H^{1*}} + \delta\|u - R_N u\|_{H^1}\right).$$

By replacing the latter inequality in (12) and integrating for $t > 0$, we obtain the desired estimate.

4.2 Implementation

In order to solve the problem (7), we start by substituting the approximation given in (9) and defined using the spectral basis (8). By taking the test functions v_N as the basis function, the spectral scheme becomes, for all $j = \overline{0, N-2}$,

$$\frac{\mathrm{d}}{\mathrm{dt}}\sum_{k=0}^{N-2} u_k(t)\langle\varphi_k, \varphi_j\rangle + \sum_{k=0}^{N-2} u_k(t)a(\varphi_k, \varphi_j) = \langle f(t), \varphi_j\rangle.$$

So

$$\langle\varphi_k, \varphi_j\rangle = \int_{-1}^{1}\varphi_k(x)\varphi_j(x)\,\mathrm{d}x, \quad \langle f(t), \varphi_j\rangle = \int_{-1}^{1} f(t,x)\varphi_j(x)\,\mathrm{d}x.$$

$$\begin{aligned} a(\varphi_k, \varphi_j) =& \beta\int_{-1}^{1}\varphi_k'(x)\varphi_j'(x)\,\mathrm{d}x + \lambda\int_{-1}^{1}\varphi_k(x)\varphi_j(x)\,\mathrm{d}x + \alpha\int_{-1}^{1}\varphi_k'(x)\varphi_j(x)\,\mathrm{d}x \\ &+ \frac{\beta}{\varepsilon}(\varphi_k(1)\varphi_j(1) + \varphi_k(-1)\varphi_j(-1)). \end{aligned}$$

Then we obtain the matrix form

$$\frac{\mathrm{d}}{\mathrm{dt}}AU(t) + BU(t) = C(t), \tag{13}$$

where $U(t) = (u_0(t), \cdots, u_{N-2}(t))^{\mathrm{T}}$ is the vector of the unknown coefficients and A and B are the $(N-1)\times(N-1)$ matrices defined by

$$A_{kj} = \langle\varphi_k, \varphi_j\rangle, \quad B_{kj} = a(\varphi_k, \varphi_j) \quad \text{and} \quad C(t) = (\langle f(t), \varphi_1\rangle, \cdots, \langle f(t), \varphi_{N-2}\rangle)^{\mathrm{T}}.$$

To solve the obtained system of ordinary differential equations (13), we propose a scheme of Crank-Nicolson. For this, we discretize the domain $[-1,1]$ using a constant step Δx and the time domaine $[0,T]$ is discretized by a step Δt. We denote by U_i^n the value of the solution U at node x_i and at time t_n and we write the scheme as follows:

$$(A + \frac{\Delta t}{2}B)U_i^{n+1} = (A - \frac{\Delta t}{2}B)U_i^n + \frac{\Delta t}{2}(C(t_n) + C(t_{n+1})); \qquad U_i^0 = (g(x))_i, \tag{14}$$

where $(g(x))_i$ is the value of $g(x)$ in each node x_i of the discretization of $[-1,1]$.

5 Numerical Results

In order to test the performance of the described method, we propose three examples on $[-1,1]$, where we solve numerically the equation (2) and show the convergence of the approximate solution, when ε reaches zero, to the analytic solution of the problem with boundary conditions of Dirichlet type.

The figures are obtained for $Nx = 16$ nodes in the domain $[-1,1]$, and 100 nodes in the time domain $[0,T]$. The error of the approximation, in the case where the Robin boundary conditions are considered, is calculated for ε fixed, according to the following formula:

$$error = \|u_N - u_{N+2}\|_\infty; \qquad N = 2, 4, \dots, 2\ell, \dots \tag{15}$$

Example 5.1 We consider the problem posed in (2) with $\alpha = \lambda = 0$, $\beta = 1$, $f(t,x) = 0$, $T = 3$ and the initial condition is given by $u(0,x) = \cos(\frac{\pi}{2}x)$. For the homogeneous Dirichlet boundary conditions, the analytic solution is given by

$$u(t,x) = \exp(-\frac{\pi^2}{4}t)\cos(\frac{\pi}{2}x). \tag{16}$$

In Figure 1, we observe the convergence of the obtained approximate solution, when taking the decreasing values of ε, to the analytic solution (16). The values of ε are taken between 0.08 and 0.0025 when $t = 1.5$ and between 0.25 and 0.04 when $x = 0.2$, to show that its behaviour remains the same all over the domain $[-1,1]$ at different instances of time.

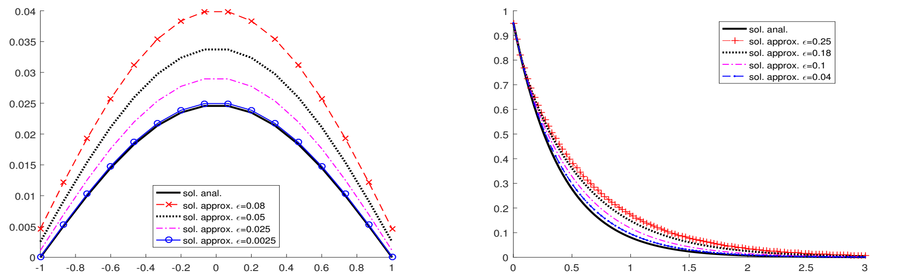


Figure 1: The behavior of the solution of (14) for $t = 1.5$ (left), and for $x = 0.2$ (right), when ε reaches 0 and $N = 6$.

Example 5.2 We consider, in (2), $\alpha = 0$, $\lambda = 0.001$, $\beta = 1$, $T = 3$. The initial condition is given by $u(0,x) = \sin(\pi x)$. For the case of the homogeneous Dirichlet boundary conditions, the analytic solution is given by

$$u(t,x) = \sin(\pi x)\exp(-\lambda t). \tag{17}$$

In Table 1, the error of approximation (15) is calculated for $N = 10$ at different points from $[-1,1]$ using different values of ε. The obtained results show that the numerical solution of (14) converges to the exact solution of the problem with the homogeneous Dirichlet boundary conditions when ε approaches zero. The variations of the error of approximation (15) are given in Figure 2 as a function of N for fixed $\varepsilon = 0.5$. We mention here that the error of approximation is calculated for $N = 42$ and is 8.129410e − 10.

x	$\varepsilon = 0.1$	$\varepsilon = 0.01$	$\varepsilon = 0.001$	$\varepsilon = 0.0001$
-1.000000	$2.847374 \cdot 10^{-1}$	$3.100927 \cdot 10^{-2}$	$3.128715 \cdot 10^{-3}$	$3.131517 \cdot 10^{-4}$
-0.600000	$1.708432 \cdot 10^{-1}$	$1.860373 \cdot 10^{-2}$	$1.874042 \cdot 10^{-3}$	$1.845087 \cdot 10^{-4}$
-0.200000	$5.693988 \cdot 10^{-2}$	$6.194439 \cdot 10^{-3}$	$6.184256 \cdot 10^{-4}$	$5.532692 \cdot 10^{-5}$
0.066665	$1.897638 \cdot 10^{-2}$	$2.061521 \cdot 10^{-3}$	$2.030002 \cdot 10^{-4}$	$1.532227 \cdot 10^{-5}$
0.466667	$1.328897 \cdot 10^{-1}$	$1.448028 \cdot 10^{-2}$	$1.467799 \cdot 10^{-3}$	$1.536447 \cdot 10^{-4}$
0.866667	$2.467852 \cdot 10^{-1}$	$2.688325 \cdot 10^{-2}$	$2.717881 \cdot 10^{-3}$	$2.774053 \cdot 10^{-4}$

Table 1: The error of approximation as a function of x for $N = 10$.

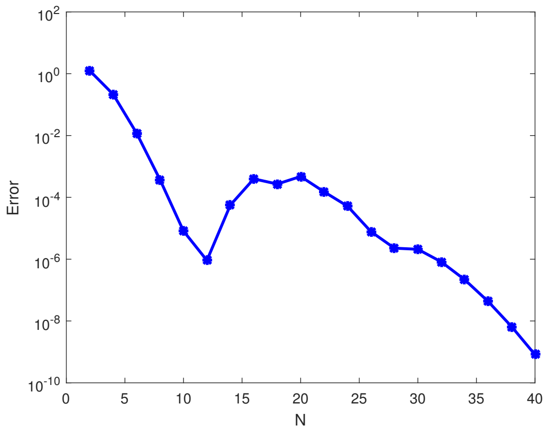


Figure 2: Logarithmic approximation error as a function of N with $\varepsilon = 0.5$.

Example 5.3 We consider, in (2), $\alpha = 0.3$, $\beta = 0.1$, $\lambda = -0.15$, $f(t,x) = 0$ and $T = 1$.The initial condition is given by $u(0,x) = cos(\pi \frac{x}{2})$.

Figure 3 presents, on the left side, the variations of the error as a function of N taking $\varepsilon = 0.1$. We see the decrease in the error curve with the growth of N. Note that the error for $N = 80$ is 2.460133e − 06. The same behavior of the error is depicted in Figure 4 for $\varepsilon = 0.001$. Moreover, on the right side, for the same value of ε and for $N = 18$, we represent the approximate solution of the problem. Finally, in Table 2, we show different values of the approximation error calculated according to the formula (15) and for $\varepsilon = 0.001$.

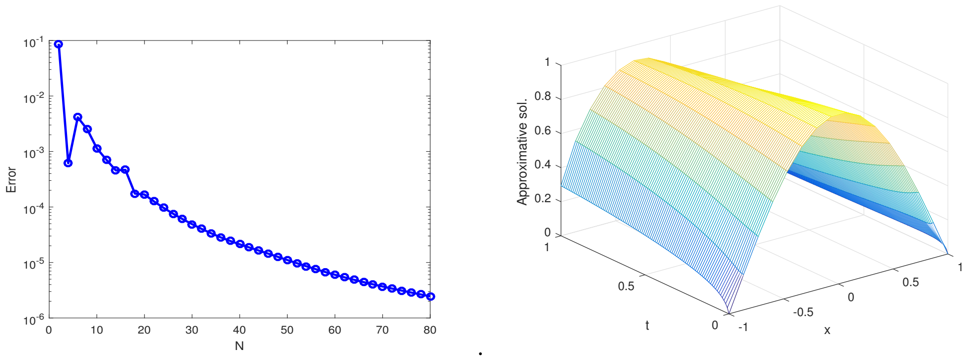


Figure 3: Logarithmic approximation error as a function of N for $\varepsilon = 0.1$ (on the left), and approximate solution of (2) for $\varepsilon = 0.1$ and $N = 18$ (on the right).

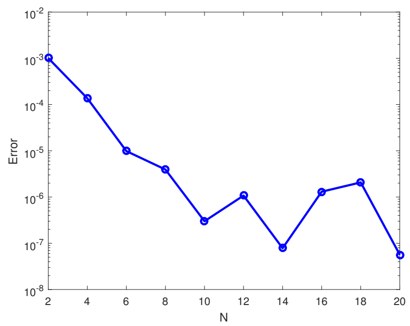


Figure 4: Logarithmic approximation error as a function of N for $\varepsilon = 0.001$.

N	error
2	$1.009365 \cdot 10^{-3}$
4	$1.364608 \cdot 10^{-4}$
6	$9.846206 \cdot 10^{-6}$
8	$3.931209 \cdot 10^{-6}$
10	$2.996409 \cdot 10^{-7}$
12	$1.075871 \cdot 10^{-6}$
14	$7.827772 \cdot 10^{-8}$
16	$1.304010 \cdot 10^{-6}$
18	$2.071257 \cdot 10^{-6}$
20	$5.646181 \cdot 10^{-8}$

Table 2: Error values as a function of N for $\varepsilon = 0.001$.

6 Conclusion

In this work, a combined algorithm of numerical methods is proposed for treating the advection-diffusion-reaction equation posed with the Robin boundary conditions perturbed with a small parameter ε. The technique is based on a Legendre-Galerkin method devoted to the spatial descretization and a Crank-Nicolson scheme to treat the obtained temporal system. The results obtained show high accuracy and good behavior, especially when comparing the approximate solution to the analytical solution of the problem posed with homogeneous boundary conditions allowing us to obtain the solution of the problem treated for different types of boundary conditions. The presented study offers a new accurate technique to approximate the solution of a partial differential equation which can be applied to approximate the solution of reaction-diffusion systems, where the explicit solution is unknown.

Acknowledgment

The authors acknowldge support from Directoriate General for Scientifc Research and Technological Development (DGRSDT), Algeria.

References

[1] H. Rienhard. *Equations aux dérivés partielles.* Dunod, France, 2004.

[2] M. K. Kadalbajoo and A. S. Yadaw. Finite Difference Finite elements and B-Spline Collocation Methods Applied to Two Parameter Singularly Pertubated Boundry Value Problems 1. *Journal of Numerical Analysis, Industrial and Applied Mathematics* **5** (2011) 163–180.

[3] J. D. Logan. *Transport modeling in hydrogeochemical systems.* Interdisciplinary Applied Mathematics, Springer, New York, 2001.

[4] Z. Laouar, N. Arar and A. Talaat. Efficient spectral Legendre Galerkin approach for the advection diffusion equation with constant and variable coefficients under mixed Robin boundary conditions. *Advances in the Theory of Nonlinear Analysis and its Application* **7** (1) (2023) 133−147.

[5] S. A. Kim. Complete analytic solutions for convection-diffusion-reaction source equations without using an inverse Laplace transform. Sci. Rep., **10**: 8040, 2020.

[6] A. Quarteroni and R. Sacco and F. Saleri. *Numerical Mathematics.* Texts in Applied Mathematics. Springer-Verlag, New York, 2007.

[7] Akutowicz and J. Dwin. Sur l'approximation spectrale. *Annales scientifiques de l'École Normale Supérieure* **82** (1965) 297–325.

[8] C. I. Christov. A complete orthogonal system of functions in $L_2(-\infty, +\infty)$ space. *SIAM J. Appl. Math.* **42** (1982) 1337–1344.

[9] A. Quarteroni, C. Canuto, C. Hussaini and T. A. Zang. *Spectral Methods, Fundamentals in Single Domains.* Springer-Verlag Berlin Heidelberg, Berlin, 2006.

[10] J. Shen. Efficient Chebyshev-Legendre Galerkin Methods for Elliptic Problems. *Houston Journal of Mathematics* **70** (1996) 233–239.

[11] N. Arar, B. Deghdough, S. Dekkiche, Z. Torch and A. M. Nagy. Numerical Solution of the Burgers' Equation Using Chelyshkov Polynomials. *International Journal of Applied and Computational Mathematics* **10** (1) (2024) 33.

[12] Guo and Y. Ben. *Spectral Methods and Their Application.* World Scientific Publishing Co. Pte. Ltd, Singapore, 1998.

[13] A. Quarteroni, C. Canuto, C. Hussaini and T. A. Zang. *Spectral Methods in Fluid Dynamics.* Springer-Verlag Berlin Heidelberg, Berlin, 1988.

[14] J. Shen, T. Tang and L. Wang. *Spectral Methods, Algorithms, Analysis and Applications.* Springer-Verlag Berlin Heidelberg, Berlin, 2011.

[15] M. A. Mohamed and A. Sambas. Numerical Solution of the Black-Scholes Partial Differential Equation for the Option Pricing Model Using the ADM-Kamal Method. *Nonlinear Dyn. Syst. Theory* **23** (3) (2023) 295–309.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 561–574

Superlinear Problem with Inverse Coefficient for a Time-Fractional Parabolic Equation with Integral Over-Determination Condition

Iqbal M. Batiha [1,2,*], Amal Benguesmia [3], Moataz Alosaimi [4], Taki-Eddine Oussaeif [3], Nidal Anakira [5,6] and Mohammad Odeh [7]

[1] *Department of Mathematics, Al Zaytoonah University of Jordan, Amman 11733, Jordan.*
[2] *Nonlinear Dynamics Research Center (NDRC), Ajman University, Ajman, UAE.*
[3] *Department of Mathematics and Informatics, Oum El Bouaghi University, Oum El Bouaghi, Algeria.*
[4] *Department of Mathematics and Statistics, Taif University, Taif 21944, Saudi Arabia.*
[5] *Faculty of Education and Arts, Sohar University, Sohar 3111, Oman.*
[6] *Jadara Research Center, Jadara University, Irbid 21110, Jordan.*
[7] *Department of Mathematics, Jouf University, Tabarjal, Saudi Arabia.*

Received: February 29, 2024; Revised: November 30, 2024

Abstract: The inverse problem of finding the right-hand side of a nonlinear fractional parabolic equation with an integral over-determination supplementary condition is examined in this study. The functional analysis method, which is based on energy inequality and the density of the range of the operator created by the problem addressed, is used to demonstrate the existence, uniqueness, and continuous dependence on the data of the direct problem. The existence theorem is then obtained from the solution of the given problem, starting with the uniqueness theorem, making the energy inequality method, also known as the method of a priori estimates, a higher character method. The hardest part of this approach is figuring out which functional spaces to use, E and F, and if the inverse problem can be solved uniquely under the right circumstances. The existence and uniqueness of the solution to the inverse problem, which arises frequently in engineering and physics modeling of diverse processes, are established using the fixed point theorem.

Keywords: *fixed point theorem; nonlocal integral condition; inverse nonlinear problem.*

Mathematics Subject Classification (2010): 35R11, 34A12.

* Corresponding author: mailto:i.batiha@zuj.edu.jo

1 Introduction

The purpose of this paper is to examine if it is possible to solve the following two functions: $\{u(x,t), f(t)\}$ fulfilling the fractional parabolic equation that follows:

$$^{C}D_t^{\alpha}u - \Delta u + \beta u + u^3 = f(t)g(x,t),\ x \in \Omega, t \in (0,T), \tag{1}$$

with the initial condition

$$u(x,0) = 0,\ x \in \Omega, \tag{2}$$

the boundary condition

$$u(x,t) = 0,\ (x,t) \in \partial\Omega \times [0,T], \tag{3}$$

and the non-local condition

$$\int_{\Omega} v(x)u(x,t)dx = E(t),\ t \in [0,\ T], \tag{4}$$

where $\partial\Omega$ is taken to be a regular boundary of a bounded domain Ω in $\mathbb{R}^n$. There are two known functions, g and E, and a positive constant, β.

Numerous real-world situations naturally give rise to inverse issues for the heat equations, see [1–3]. The integral condition (4) provides further information about how to solve the inverse problem in this case. The integral condition is a crucial modeling tool in the theory of PDEs in physics and engineering [4–9]. It is important to remember that integral over-determination processes may not always be successful in addressing non-local challenges [10, 11]. Numerous approaches to solving issues brought on by non-local issues have been put out thus far. The type of the involved non-local boundary value dictates the chosen method [12–14]. Numerous authors have studied the inverse parabolic problem and its unique solvability, focusing on conditions of type (4), see for example, [2, 3, 15–18]. The existence and uniqueness of inverse problem solutions for different parabolic equations with unknown source functions have also been the subject of several studies. Reversing problems with a parabolic equation's determination term and over-determination condition were also considered in [19, 20].

Fractional differential equations (FDEs) are created by generating differential equations to any desired order [21–25]. Because fractional differential equations are used to simulate complicated phenomena, they are significant in the fields of engineering, physics, and applied mathematics [26, 27]. Because of this, engineers and scientists have shown a growing interest in them in recent years. Since FDEs have memory and non-local relations in space and time, they can be used to simulate complex phenomena [28–33]. Herein, the tools, which will be used in our investigation, are the energy inequality method and the fixed point theorem. The structure of the energy inequality approach can be summed up as follows:

- First, we write the problem in the form of an operational equation

  $$Lu = F,\ u \in D(L),$$

  where a Banach space E is considered, and the operator L is studied from it to a suitable Hilbert space F.
- The a priori estimate for the operator L is then established.
- Next, we establish the density of this operator's collection of values in space F.

The results of the previous procedure will help us in investigating the existence, uniqueness and continuous dependence of the problem at hand.

2 Functional Spaces

When it comes to inverse coefficient problems for time fractional parabolic equations under integral over-determination conditions, functional spaces are essential tools for delving into intricate mathematical difficulties. Unknown coefficient identification becomes a difficult challenge when studying dynamic systems governed by partial differential equations, which gives rise to these problems. Function spaces are used in this study to give the inverse coefficient superlinear problem a comprehensive framework for analysis and solution. We shall include some definitions and lemmas pertaining to our study in the sections that follow. Let us clarify the conventions and notations we will use:

$$g^*(t) = \int_\Omega g(x,t)\cdot v(x)dx,\ Q = \Omega\times(0,T). \tag{5}$$

- The left Caputo derivative is given by

$$^cD_t^\alpha u := \frac{1}{\Gamma(1-\alpha)}\int_0^t \frac{\partial u(x,\tau)}{\partial\tau}\frac{1}{(t-\tau)^\alpha}d\tau. \tag{6}$$

- The left Riemann-Liouville derivative is given by

$$^RD_t^\alpha u := \frac{1}{\Gamma(1-\alpha)}\frac{\partial}{\partial t}\int_0^t \frac{u(x,\tau)}{(t-\tau)^\alpha}d\tau. \tag{7}$$

- The right Riemann-Liouville derivative is given by

$$^R_tD^\alpha u(x,t) := \frac{1}{\Gamma(1-\alpha)}\frac{\partial}{\partial t}\int_t^T \frac{u(x,\tau)}{(t-\tau)^\alpha}d\tau. \tag{8}$$

Because the Caputo version is easier to handle under homogenous initial conditions, several authors contend that it is more natural. A direct calculation can confirm the link between the two concepts (6) and (7) as follows:

$$^rD_t^\alpha u = {}^cD_t^\alpha u + \frac{u(x,0)}{\Gamma(1-\alpha)t^\alpha}. \tag{9}$$

Definition 2.1 [34] For each real $\alpha > 0$, we define the space $^lH_0^\alpha(I)$ as the closure of $C_0^\infty(I)$ with regard to the norm $\|u\|_{^lH_0^\alpha(I)}$ as follows:

$$\|u\|_{^lH^\alpha(I)} := \left(\|u\|^2_{L^2(I)} + |u|^2_{^lH_0^\alpha(I)}\right)^{\frac{1}{2}}, \tag{10}$$

where

$$|u|^2{}_{^lH^\alpha(I)} = \left\|{}_0^RD_t^\alpha u\right\|_{L^2(I)}.$$

Definition 2.2 For each real $\alpha > 0$, the space $^rH_0^\alpha(I)$ is defined as the closure of $C_0^\infty(I)$ with regard to the norm $\|u\|_{^rH_0^\alpha(I)}$ as follows:

$$\|u\|_{^rH_0^\alpha(I)} := \left(\|u\|^2_{L^2(I)} + |u|^2{}_{^rH_0^\alpha(I)}\right)^{\frac{1}{2}}, \tag{11}$$

where

$$|u|^2_{^rH_0^\alpha(I)} := \left\|{}_t^R\partial_T^\alpha u\right\|^2_{L^2(I)}.$$

Lemma 2.1 *[8, 34] If $u \in {}^{l}H^{\alpha}(I)$ and $v \in C_0^{\infty}(I)$ for any real $\alpha \in \mathbb{R}_+$, then*

$$\left({}^{R}D_t^{\alpha}u(t), v(t)\right)_{L^2(I)} = \left(u(t), {}_t^{R}D^{\alpha}v(t)\right)_{L^2(I)}.$$

Lemma 2.2 *[8, 34] For $0 < \alpha < 2, \alpha \neq 1, u \in H_0^{\frac{\alpha}{2}}(I)$, we have*

$${}^{R}D_t^{\alpha}u(t) = {}^{R}D_t^{\frac{\alpha}{2}}\, {}^{R}D_t^{\frac{\alpha}{2}}u(t).$$

Lemma 2.3 *[8, 34] For $\alpha \in \mathbb{R}_+$, the semi-norms $|.|_{{}^{l}H^{\alpha}(I)}, |.|_{{}^{r}H^{\alpha}(I)}$ and $|.|_{{}^{c}H^{\alpha}(I)}$ are equivalent, for which $\alpha \neq n + \frac{1}{2}$. Thus, we have*

$$|u|_{{}^{l}H^{\alpha}(I)} \cong |.|_{{}^{r}H^{\alpha}(I)} \cong |.|_{{}^{c}H^{\alpha}(I)}.$$

Lemma 2.4 *The space ${}^{l}H_0^{\alpha}(I)$ is complete for every real $\alpha > 0$ with respect to the norm (10).*

Definition 2.3 The space of square functions, in the Bochner sense, integrated with the scalar product is represented by $L_2\left(0, T, L_2(0, d)\right)$, and it is given by

$$(u, w)_{L_2(0,T,L_2(0,d))} = \int_0^T (u, w)_{L_2(0,d)}dt. \tag{12}$$

3 The Direct Fractional Parabolic Problem's Solvability

One of the fundamental aspects of a more general study of inverse coefficient super-linear problems related to time fractional parabolic equations under integral over-determination conditions is the solvability of the direct fractional parabolic problem. Understanding the forward dynamics regulated by partial differential equations is essential for comprehending the behavior of the underlying systems, and this is achieved through the study of direct fractional parabolic problems.

3.1 Problem setting

In the rectangular domain $Q = (0,\ d) \times (0,\ T)$, where $d, T < \infty$ and $0 < \alpha < 1$, we will examine the existence and uniqueness of solution $u = u(x, t)$ to the following fractional parabolic problem:

$$\begin{cases} {}^{c}D_t^{\alpha}u - \left(\frac{\partial^2 u(x,t)}{\partial x^2}\right) + \beta u + u^3 = \tilde{f}(x, t) \ \text{ in } Q, \\ u(x, 0) = 0,\ \forall x \in (0, d), \\ u(0, t) = u(d, t) = 0,\ \forall t \in (0, T), \end{cases}$$

whose fractional parabolic equation is nonlinear and provided as follows:

$$\mathcal{L}u = {}^{c}D_t^{\alpha}u - \frac{\partial^2 u}{\partial x^2} + \beta u + u^3 = \tilde{f}$$

with the initial condition

$$\ell u = u(x, 0) = 0,\ \forall x \in (0, d)$$

and

$$u(0, t) = u(d, t) = 0,\ \forall t \in (0, T),$$

where $\tilde{f}$ is a known function and $b \in \mathbb{R}_*^+$.

Within this segment, we exhibit the existence and uniqueness of the solution for the problem (1)-(3) as a resolution of the subsequent operator equation

$$Lu = \mathcal{F}, \tag{13}$$

for which $L = (\mathcal{L}, \ell)$, and the domain of definition $D(L) = B$ that can be outlined as

$$D(L) = \left\{ u \backslash u \in L^2(Q) \cap L^4(Q), {}^cD_t^\alpha u, \frac{\partial u}{\partial x} \in L^2(Q) \right\}.$$

The operator L is defined in the space between B and F, where B is the Banach space containing all functions $u(x,t)$ with a finite norm of the form

$$\|u\|_B^2 = \left\| {}^cD_t^{\frac{\alpha}{2}} u \right\|_{L^2(Q)}^2 + \left\| \frac{du}{dx} \right\|_{L^2(Q)}^2 + \|u\|_{L^2(Q)}^2 + \|u\|_{L^4(Q)}^4,$$

and F is the Hilbert space consisting of all Fourier elements $(f, 0)$ such that the norm $L^2(Q)$ is finite.

Theorem 3.1 *For each function $u \in B$, we have the inequality*

$$\|u\|_B \leq C \|Lu\|_{L^2(Q)}, \tag{14}$$

where C is a positive constant independent of u.

Proof. We now employ the function $Mu = u(x,t)$ together with the scalar product in $L^2(Q)$ of (1), where $Q = (0,d)x(0,T)$. Consequently, we can have

$$\begin{aligned} \int_Q \mathcal{L}u \cdot Mu \, dxdt &= \int_Q {}^cD_t^\alpha u(x,t) \cdot u(x,t) \, dxdt - \int_Q \left(\frac{\partial^2 u(x,t)}{\partial x^2} \right) \cdot u(x,t) \, dxdt \\ &+ b \int_Q u^2(x,t) \, dxdt + \int u^4(x,t) \, dxdt \\ &= \int_Q f(x,t) u(x,t) \, dxdt. \end{aligned} \tag{15}$$

Due to $u(x,0) = 0$, and by using Lemmas 2.1, 2.2 and 2.3, we get

$$\begin{aligned} \int_Q {}^cD_t^\alpha u(x,t) \cdot u(x,t) dxdt &= \left({}^cD_t^\alpha u(x,t), u(x,t)\right)_{L^2(Q)} \\ &= \left({}^RD_t^{\frac{\alpha}{2}} \, {}^RD_t^{\frac{\alpha}{2}} u(x,t), u(x,t) \right)_{L^2(Q)} \\ &= \left({}^RD_t^{\frac{\alpha}{2}} u(x,t), {}_t^RD^{\frac{\alpha}{2}} u(x,t) \right)_{L^2(Q)} \\ &= |u|^2{}_{{}^cH^\alpha(Q)} \cong |u|^2{}_{{}^lH^\alpha(Q)} = \left\| {}^CD_t^{\frac{\alpha}{2}} u \right\|_{L^2(Q)}^2. \end{aligned}$$

With the use of the relationship ($|ab| \leq \frac{\varepsilon a^2}{2} + \frac{b^2}{2\varepsilon}$) coupled with the integral by parts, we obtain

$$\|^C D_t^{\frac{\alpha}{2}} u\|^2_{L^2(Q)} + \|\frac{du}{dx}\|^2_{L^2(Q)} + (b - \frac{\varepsilon}{2})\|u\|^2_{L^2(Q)} + \|u\|^4_{L^4(Q)} \leq \frac{1}{2\varepsilon}\|f\|^2_{L^2(Q)}.$$

So, for $\varepsilon \leq 2b$, we can have

$$\|^c D_t^{\frac{\alpha}{2}} u\|^2_{L^2(Q)} + \|\frac{du}{dx}\|^2_{L^2(Q)} + \|u\|^2_{L^2(Q)} + \|u\|^4_{L^4(Q)} \leq c\|f\|^2_{L^2(Q)}$$

with

$$c = \frac{1}{2\varepsilon \min\left(1, b - \frac{\varepsilon}{2}\right)}.$$

Consequently, we get

$$\|u\|_B \leq C\|Lu\|_{L^2(Q)},$$

where $C = \sqrt{c}$.

Proposition 3.1 *There is a closure for the operator L from B to F.*

Proof. Consider $(u_n)_{n\in\mathbb{N}} \subset D(L)$ is a sequence in which $u_{n\to 0}$ in B, and $Lu_{n\to\mathcal{F}}$ in F. Herein, we should show $f \equiv 0$. To this end, we notice that in B, the convergence of u_n to 0 causes

$$u_{n\to 0} \ \text{ in } \ (C_0^\infty(Q))'. \tag{16}$$

Given the continuity of the fractional derivative, the continuity distribution of the function u^2, and 162 derivation of the first order of $(C_0^\infty(Q))'$ in $(C_0^\infty(Q))'$ as a special case of the fractional derivative, the relationship (16) involves

$$\mathcal{L}u_{n\to 0} \ \text{ in } \ (C_0^\infty(Q))'. \tag{17}$$

Furthermore, in $L^2(Q)$, the convergence of Lv_n to f yields

$$\mathcal{L}v_{n\to f} \ \text{ in } \ (C_0^\infty(Q))'. \tag{18}$$

Due to the limit in $(C_0^\infty(Q))'$ is unique, we may infer from (17) and (18) that $f \equiv 0$. Therefore, the operator L is closeable.

We will define $D(\bar{L})$ as the domain of definition of $\bar{L}$ and let $\bar{L}$ be the closure of L in the material that follows.

Definition 3.1 Problem (1)–(3) has a strong solution, which is the operator equation

$$\bar{L}u = \mathcal{F}.$$

Furthermore, we may expand the previous estimate to a strong solution, meaning we would get the estimate

$$\|u\|_B \leq C\|\bar{L}u\|_F, \ \forall u \in D(\bar{L}). \tag{19}$$

Corollary 3.1 *Problem (1)–(3) has a unique strong solution that is constantly dependent on $f \in F$.*

Corollary 3.2 *The closure of $R(L)$ and the range $R(\bar{L})$ of the operator $\bar{L}$ in F are equal, i.e.,*

$$R(\bar{L}) = \overline{R(L)}.$$

Proof. First, if the solution exists, we will show that it is unique. For this purpose, we assume that u_1 and u_2 are two solutions such that $\eta = u_1 - u_2$. So, η will satisfy

$$\begin{cases} {}^cD_t^\alpha \eta(x,t) - \left(\frac{\partial^2 \eta(x,t)}{\partial x^2}\right) + b\eta(x,t) + u_1^3 - u_2^3 = 0, \quad \text{in } Q, \\ \eta(x,0) = 0, \ \forall x \in (0,d), \\ \eta(x,t) = 0, \ \forall (x,t) \in \partial\Omega \times (0,T) \\ . \end{cases},$$

for which

$$ {}^cD_t^\alpha \eta(x,t) - \left(\frac{\partial^2 \eta(x,t)}{\partial x^2}\right) + b\eta(x,t) + u_1^3 - u_2^3 = 0, \quad \text{in } Q. \tag{20}$$

Using the scalar product of (20) and η in $L^2(\Omega)$, we obtain

$$\int_\Omega {}^cD_t^\alpha \eta(x,t) \cdot \eta(x,t)dx - \int_\Omega \left(\frac{\partial^2 \eta(x,t)}{\partial x^2}\right) \cdot \eta(x,t)\,dx + b\int_\Omega \eta^2(x,t)\,dx \\ + \int_\Omega (u_1^3 - u_2^3)(u_1 - u_2)\,dx = 0.$$

Due to $\eta(x,0) = 0$, with the use of Lemmas 2.1, 2.2 and 2.3 together with integrating by parts, we obtain

$$\|{}^cD_t^{\frac{\alpha}{2}} \eta\|^2_{L^2(\Omega)} + \|\frac{d\eta}{dx}\|^2_{L^2(\Omega)} + \|\eta\|^2_{L^2(\Omega)} + \int_\Omega (u_1^3 - u_2^3)(u_1 - u_2)dx = 0. \tag{21}$$

The last item on the left-hand side of equation (21) is positive since λ^3 is a monotone function in λ (on $\Omega = (0,d)$), and this leads to the following conclusion from equation (20):

$$\|\eta\|^2_{L^2(\Omega)} \leq 0,$$

which implies $u_1 = u_2$ for all $t\epsilon(0,T)$. We will now go back and illustrate the result we discuss. To this end, we let $z \in \overline{R(l)}$. Thus, there exists a sequence $(z_n)_{n\in\mathbb{N}}$ in $R(L)$, for which $\lim_n z_n = z$. So, just as $(z_n)_{n\in\mathbb{N}}$ in $R(L)$, $\exists\,(u_n)_{n\in\mathbb{N}}$ in $D(L)$, for which $Lu_n = z_n$. Assume that $\mathcal{E}, n \geq n_0$, and $m, m' \in \mathbb{N}$, $m \geq m'$, for which u_m and $u_{m\prime}$ satisfy

$$Lu_m = f \text{ and } Lu_{m'} = f.$$

We put $y = u_m - u_m\prime$, then y satisfies

$$\begin{cases} {}^cD_t^\alpha y(x,t) - \left(\frac{\partial^2 y(x,t)}{\partial x^2}\right) + by(x,t) + u_m^3 - u_{m\prime}^3 = 0, \quad \text{in } Q, \\ y(x,0) = 0, \ \forall x \in (0,d), \\ y(x,t) = 0, \ \forall (x,t) \in \partial\Omega \times (0,T) \end{cases}.$$

Using the same method we employed to demonstrate the solution's uniqueness, we can now obtain $y = 0$. This suggests that for every $t \in (0,\ T)$, we obtain

$$0 \leq \|u_m - u_{m'}\| \leq 0,$$

i.e.,

$$\forall\varepsilon \geq 0,\ \exists n_0 \in \mathbb{N},\ \forall m, m' \geq n_0,\ \|u_m - u_{m\prime}\| \leq \varepsilon.$$

Therefore, since E is a Banach space and $(u_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, there exists $u \in E$ such that $\lim_n u_n = u$. With the use of the definition of $\bar{L}$ ($\lim_n u_n = u$ in

E; given that $\lim_n u_n = \lim_n z_n = u$, $\lim_n \bar{L}u_n = z$ since $\bar{L}$ is closed, implying that $\bar{L}u = z$), we find that the function u satisfies $u \in D(\bar{L})$, for which $\bar{L}u = z$. Thus, we have $z \in R(\bar{L})$, and so we obtain

$$\overline{R(L)} \subset R(\bar{L}).$$

We also conclude that it is closed since $R(\bar{L})$ is Banach. It remains to show that this is not the case. For this purpose, we assume that $z \in R(\bar{L})$. Then, given the elements of the set $R(\bar{L})$, there exists a sequence of $(z_n)_n$ in F such that

$$\lim_n z_n = z.$$

Therefore, a matching sequence $(u_n)_{n\in\mathbb{N}}$ exists such that

$$\lim_n \bar{L}u_n = z_n.$$

However, we have a Cauchy sequence in F, which is $(u_n)_{n\in\mathbb{N}}$. Thus, $u \in E$ exists such that

$$\lim_n u_n = u \quad \text{in } E.$$

Consequently, we have $\lim_n \bar{L}u_n = z$. As a consequence, $z \in \overline{R(L)}$, and then we obtain

$$\overline{R(L)} = R(\bar{L}).$$

4 Existence of Solution

In order to prove that the solution exists, we show that for every $u \in B$ and for any arbitrary $\mathcal{F} = (f, 0) \in F$, $R(L)$ is dense in F.

Theorem 4.1 *The problem (1)-(3) has a solution.*

Proof. The definition of F's scalar product is

$$(Lv, W)_F = \int_Q \mathcal{L}v \cdot w dx dt, \tag{22}$$

where $W = (w, 0)$ in $D(L)$. Set $w \in (R(L))^\perp$, and the result is

$$\int_Q {}^cD_t^\alpha u(x,t)w(x,t)dxdt - \int_Q \left(\frac{\partial^2 u(x,t)}{\partial x^2}\right) w(x,t)dxdt + b\int_Q u(x,t)w(x,t)dxdt \\ + \int_Q u^3(x,t)\cdot w(x,t)dxdt = 0.$$

Letting $w = u$ yields

$$\int_Q {}^cD_t^\alpha u(x,t)\cdot u(x,t)\,dx\,dt - \int_Q \left(\frac{\partial^2 u(x,t)}{\partial x^2}\right)\cdot u(x,t)\,dx\,dt + b\int_Q u^2(x,t)\,dx\,dt \\ + \int_Q u^4(x,t)\,dx\,dt = 0.$$

After accounting for the condition of u and integrating by parts each term of (4), we get

$$\left\| {}^{c}D_t^{\frac{\alpha}{2}} u \right\|_{L^2(Q)}^2 + \left\| \frac{du}{dx} \right\|_{L^2(Q)}^2 + b\|u\|_{L^2(Q)}^2 + \|u\|_{L^4(Q)}^4 = 0.$$

So, we get

$$\left\| {}^{c}D_t^{\frac{\alpha}{2}} u \right\|_{L^2(Q)}^2 + b\|u\|_{L^2(Q)}^2 + \|u\|_{L^4(Q)}^4 = -\left\| \frac{du}{dx} \right\|_{L^2(Q)}^2 \leq 0.$$

Then, we have

$$\|u\|_{L^2(Q)}^2 \leq 0.$$

Consequently, $u = 0$ in Q, providing $w = 0$ within Q, and this completes the proof.

5 Solvability of the Main Problem

We assume that the functions that show up in the problem's data are quantifiable and meet the following conditions:

$$\left\{ \begin{array}{c} g \in C\left((0,T), L^2(\Omega)\right), v \in W_2^1(\Omega) \cap L^4(\Omega), E \in W_2^2(0,T), \\ \|g(x,t)\| \leq m, |g^*(t)| \geq r > 0, \text{ for } r \in \mathbb{R},\ (x,t) \in Q \end{array} \right. .$$

The following linear operator provides the relationship between f and u:

$$A : L^2(0,T) \to L^2(0,) \tag{23}$$

such that

$$(\mathrm{Af}(t)) = \frac{1}{g^*} \left\{ \int_\Omega \frac{du}{dx}\frac{dv}{dx} dx + \int_\Omega u^3(x,t) \cdot v(x) dx \right\}. \tag{24}$$

Consequently, for the function f over $L^2(0,T)$, the previous relationship between f and u may be expressed as a second-order linear equation. In other words, we have

$$f = Af + W, \tag{25}$$

where

$$W = \frac{D_t^{\alpha}{}^{\alpha} + \beta E}{g^*}, \tag{26}$$

and $E(0) = 0$.

Theorem 5.1 *Presume that the condition (H) is validated by the data functions (1)-(4) of the inverse problem. Then we have the equivalent of the following statement:*

1. *If the inverse problem (1)-(4) can be solved, then equation (25) can be solved as well.*

2. *The inverse problem (1)-(4) has a solution if equation (25) has a solution and the compatibility requirement $E(0) = 0$ holds.*

Proof. Assume that problem (1)-(4) can be solved. We denote its solution as $\{u, f\}$ in this instance. Now, after integrating the outcomes over Ω and multiplying both sides of (1) by v, the following is obtained:

$$\begin{aligned} {}^cD_t^\alpha \int_\Omega u(x,t)\cdot v(x)dx + \int_\Omega \frac{du}{dx}\frac{dv}{dx}dx + \beta \int_\Omega u(x,t)\cdot v(x)dx \\ + \int_\Omega u^3(x,t)\cdot v(x)dx = f(t)g^*(t). \end{aligned} \tag{27}$$

By applying (4) and (24), we obtain

$$f = Af + \frac{\beta E + {}^cD_t^\alpha E}{g^*}.$$

It is still necessary to demonstrate that u fulfills the integral over-determination condition (4). The function u is subject to the following relation by equation (27):

$${}^cD_t^\alpha E + \int_\Omega \frac{du}{dx}\frac{dv}{dx}dx + \beta E + \int_\Omega u^3(x,t)\cdot v(x)dx = f(t)g^*(t). \tag{28}$$

Equation (27) is subtracted from equation (28) to obtain

$${}^cD_t^\alpha \int_\Omega u(x,t)\cdot v(x)dx + \beta \int_\Omega u(x,t)\cdot v(x)dx = {}^cD_t^\alpha E + \beta E. \tag{29}$$

We determine that u meets the integral condition (4) by integrating the preceding equation and accounting for the compatibility constraint $E(0) = 0$. Consequently, we may infer that the solution to the inverse problem (1)-(4) is $\{u, f\}$.

Lemma 5.1 *If (H) is true, then A is a contracting operator in $L^2(0,T)$ for some positive δ.*

Proof. Based on (24), the following estimate can be inferred

$$|Af(t)|^2 \le \frac{2}{r^2}\left[\left\|\frac{du}{dx}\right\|^2_{L^2(\Omega)}\left\|\frac{dv}{dx}\right\|^2_{L^2(\Omega)} + \|u\|^6_{L^4(\Omega)}\|v\|^2_{L^4(\Omega)}\right].$$

Now, we suppose $\|u\|^2_{L^\infty(0,T,L^4(\Omega))} = \Upsilon \ge 0$. Then we obtain

$$|Af(t)|^2 \le \frac{2}{r^2}\left[\left\|\frac{du}{dx}\right\|^2_{L^2(\Omega)}\left\|\frac{dv}{dx}\right\|^2_{L^2(\Omega)} + \Upsilon\|u\|^4_{L^4(\Omega)}\|v\|^2_{L^4(\Omega)}\right].$$

Now, integrating the previous inequality over $(0,\ T)$ yields

$$\begin{aligned} &\int_0^T |Af(t)|^2 dt \\ &\quad\le \frac{2}{r^2}\max\left(\left\|\frac{dv}{dx}\right\|^2_{L^2(\Omega)}, \gamma\|v\|^2_{L^4(\Omega)}\right)\left[\int_0^T \left\|\frac{du}{dx}\right\|^2_{L^2(\Omega)} dt + \int_0^T \|u\|^4_{L^4(\Omega)}\right) dt\right]. \end{aligned} \tag{30}$$

Consequently, we get

$$\|Af\|_{L^2(0,T)} \leq K \left[\int_0^T \left\| \frac{du}{dx} \right\|^2_{L^2(\Omega)} dt + \int_0^T \|u\|^4_{L^4(\Omega)} \right) dt \right]^{\frac{1}{2}},$$

for which

$$K = \sqrt{\frac{2}{r^2} \max \left(\left\| \frac{dv}{dx} \right\|^2_{L^2(\Omega)}, Y \|v\|^2_{L^4(\Omega)} \right)}.$$

After removing a few terms and applying the a priori estimate, we now have

$$\| \frac{du}{dx} \|^2_{L^2(Q)} + \|u\|^4_{L^4(Q)} \leq C \|f\|^2_{L^2(Q)}.$$

Thus, we have

$$\|Af\|_{L^2(0,T)} \leq \delta \|f\|_{L^2(0,T)}, \tag{31}$$

where $\delta = K\sqrt{C}$. The previous relation indicates that there exists a positive δ such that $\delta \leq 1$. Hence, the operator A is a contracting mapping on $L^2(0,T)$, as shown by inequality (31).

Theorem 5.2 *If the compatibility condition and assumption* (H) *are met, then there is only one solution* $\{u, f\}$ *to the inverse problems (1)-(4).*

Proof. It is evident that there is only one solution f for equation (25) in $L^2(0,T)$. It is established by Lemma 2.3 that there is a solution to the inverse problem (1)-(4). We still need to prove that this approach is unique. However, suppose that the inverse problem under consideration has two distinct solutions, $\{u_1, f_1\}$ and $\{u_2, f_2\}$. Now, the theorem on the uniqueness of the solution of the main direct problem (1)-(3) produces $z_1 = z_2$ if the linear operator A contracts on $L^2(0,\ T)$ from Lemma 5.1, resulting in $f_1 = f_2$.

Corollary 5.1 *The solution* f *to equation (25) is continuously dependent on the data* W*, under the presumptions of Theorem 5.1.*

Proof. Let us assume two data sets that meet the conditions of Theorem 5.1: ω and v. For each set of data, ω and v, let f and g represent the solutions to equation (25), respectively. Now, based on (25), we can have

$$f = Af + v, \qquad g = Ag + \omega.$$

In this regard, it is necessary to compute $f - g$. When utilizing (31), it is evident that

$$\|f - g\|_{L^2(0,T)} = \|(Af + v) - (Ag + \omega)\|_{L^2(0,T)} \leq \delta \|f - g\|_{L^2(0,T)} + \|v - \omega\|_{L^2(0,T)}.$$

Consequently, we get

$$\|f - g\|_{L^2(0,T)} \leq \frac{1}{1-\delta} \|v - \omega\|_{L^2(0,T)}.$$

6 Discussion

When the initial condition is homogeneous, the inverse problem of finding the right-hand side of a nonlinear fractional parabolic equation with an integral over-determination condition has been examined. Theoretical analysis has been conducted for this inverse problem. This study has established the conditions for the problem's existence, uniqueness, and continuous dependence on data. The work done in this paper can therefore be continued from a variety of intriguing angles in numerical analysis, particularly with regard to creating efficient numerical techniques that are compliant with integrative type non-local conditions and considering how to solve the same problem but with incompatible initial conditions.

References

[1] M. Ivanchov. *Inverse Problems for Equations of Parabolic Type.* VNTL Publishers, Ukraine, 2003.

[2] O. Taki-Eddine and B. Abdelfatah. An inverse coefficient problem for a parabolic equation under nonlocal boundary and integral overdetermination conditions. *International Journal of Differential Equations and Applications* **2** (3) (2014) 38–43.

[3] T. E. Oussaeif and A. Bouziani. Inverse problem of a hyperbolic equation with an integral overdetermination condition. *Electronic Journal of Differential Equations* **2016** (138) (2016) 1–7.

[4] J. R. Cannon, Y. Lin and S. Wang. Determination of a control parameter in a parabolic partial differential equation. *ANZIAM Journal* **33** (2) (1991) 149–163.

[5] J. R. Cannon, Y. Lin and S. Wang. Determination of source parameter in a parabolic equations. *Meccanica* **27** (2) (1992) 85–94.

[6] M. I. Ivanchov and N. V. Pabyrivska. Simultaneous determination of two coefficients of a parabolic equation in the case of nonlocal and integral conditions. *Ukrainian Mathematical Journal* **53** (5) (2001) 674–684.

[7] A. G. Fatullayev, N. Gasilov and I. Yusubov. Simultaneous determination of unknown coefficients in a parabolic equation. *Applicable Analysis* **87** (1) (2008) 1167–1177.

[8] M. I. Ismailov and F. Kanca. Inverse problem of finding the time-dependent coefficient of heat equation from integral overdetermination condition data. *arXiv* **20** (4) (2010) 463–476.

[9] M. I. Ismailov and F. Kanca. An inverse coefficient problem for a parabolic equation in the case of nonlocal boundary and overdetermination conditions. *Mathematical Methods in the Applied Sciences* **34** (6) (2011) 692–702.

[10] N. I. Ionkin. The solution of a certain boundary value problem of the theory of heat conduction with a nonclassical boundary condition. *Differentsial'nye Uravneniya* **13** (2) (1977) 294–304.

[11] V. L. Kamynin and I. A. Vasin. Asymptotic behaviour of the solutions of inverse problems for parabolic equations with irregular coefficient. *Sbornik: Mathematics* **188** (3) (1997) 371–387.

[12] A. F. Jameel, A. H. Shather, N. R. Anakira, A. K. Alomari and A. Saaban. Comparison for the approximate solution of the second-order fuzzy nonlinear differential equation with fuzzy initial conditions. *Mathematics and Statistics* **8** (5) (2020) 527–534.

[13] A. F. Jameel, N. Anakira, A. K. Alomari, I. Hashim and S. Momani. A new approximation method for solving fuzzy heat equations. *Journal of Computational and Theoretical Nanoscience* **13** (11) (2016) 7825–7832.

[14] S. A. Altaie, N. Anakira, A. Jameel, O. Ababneh, A. Qazza and A. K. Alomari. Homotopy analysis method analytical scheme for developing a solution to partial differential equations in fuzzy environment. *Fractal and Fractional,* **6** (8) (2022) 419.

[15] A. I. Prilepko and D. G. Orlovsky. Determination of the parameter of an evolution equation and inverse problems of mathematical physics. *Differentsial'nye Uravneniya* **21** (4) (1985) 694–701.

[16] T. E. Oussaeif and A. Bouziani. Existence and uniqueness of solutions to parabolic fractional differential equations with integral conditions. *Electronic Journal of Differential Equations* **2014** (179) (2014) 1–10.

[17] A. Lopushans'kyi, H. Lopushans'ka and V. Rapita. Inverse problem in the space of generalized functions. *Ukrainian Mathematical Journal* **68** (2) (2016) 269–282.

[18] V. L. Kamynin. *Elliptic and Parabolic Problems.* CRC Press, Boca Raton, Florida, 2020.

[19] A. Bouziani. Solution forte d'un problem mixte avec condition non locales pour une classe d'equations hyperbolique. *Bulletin de la Classe des sciences* **8** (1) (1997) 53–70.

[20] J. H. He. Some applications of nonlinear fractional differential equations and their approximations. *Bulletin of Science, Technology and Society* **15** (2) (1999) 86–90.

[21] I. M. Batiha, Z. Chebana, T. E. Oussaeif, A. Ouannas and I. H. Jebril. On a weak solution of a fractional-order temporal equation. *Mathematics and Statistics* **10** (5) (2022) 1116–1120.

[22] N. Anakira, Z. Chebana, T. E. Oussaeif, I. M. Batiha and A. Ouannas. A study of a weak solution of a diffusion problem for a temporal fractional differential equation. *Nonlinear Functional Analysis and Applications* **27** (3) (2022) 679–689.

[23] I. M. Batiha, A. Ouannas, R. Albadarneh, A. A. Al-Nana and S. Momani. Existence and uniqueness of solutions for generalized Sturm–Liouville and Langevin equations via Caputo–Hadamard fractional-order operator. *Engineering Computations* **39** (7) (2022) 2581–2603.

[24] A. Lamamri, I. Jebril, Z. Dahmani, A. Anber, M. Rakah and S. Alkhazaleh. Fractional calculus in beam deflection: analyzing nonlinear systems with Caputo and conformable derivatives. *AIMS Mathematics* **9** (8) (2024) 21609–21627.

[25] I. M. Batiha, Z. Chebana, T. E. Oussaeif, A. Ouannas, S. Alshorm and A. Zraiqat. Solvability and dynamics of superlinear reaction diffusion problem with integral condition. *IAENG International Journal of Applied Mathematics* **53** (1) (2023) 113–121.

[26] I. M. Batiha, S. A. Njadat, R. M. Batyha, A. Zraiqat, A. Dababneh and S. Momani. Design fractional-order PID controllers for single-joint robot arm model. *International Journal of Advances in Soft Computing and its Applications* **14** (2) (2022) 96–114.

[27] I. M. Batiha, J. Oudetallah, A. Ouannas, A. A. Al-Nana and I. H. Jebril. Tuning the fractional-order PID-controller for blood glucose level of diabetic patients. *International Journal of Advances in Soft Computing and its Applications* **13** (2) (2021) 1–10.

[28] K. B Oldham and J. Spanier. *The Fractional Calculus Theory and Applications of Differentiation and Integration to Arbitrary Order.* Academic Press, New York, 1974.

[29] J. H. He. Nonlinear oscillation with fractional derivative and its application. *International Conference on Vibrating Engineering* **98** (1) (1998) 288–291.

[30] J. H. He. Approximate analytical solution for seepage flow with fractional derivatives in porous media. *Computer Methods in Applied Mechanics and Engineering* **167** (1-2) (1998) 57–68.

[31] A. I. Prilepko, D. G. Orlovsky and I. A. Vasin. *Methods for Solving Inverse Problems in Mathematical Physics.* CRC Press, Boca Raton, Florida, 2000.

[32] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports* **339** (1) (2000) 1–7.

[33] A. A. Kilbas. *Theory and Application of Fractional Differential Equations*. Elsevier, Amsterdam, 2006.

[34] X. Li and C. Xu. Existence and uniqueness of the weak solution of the space-time fractional diffusion equation and a spectral method approximation. *Communications in Computational Physics* **8** (5) (2010) 1016–1051.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 575–581

On Stability and Convergence of a Fractional Convection Reaction-Diffusion Model

OumKeltoum Benhamouda [1], Mohamed El-Hadi Smakdji [2], Ammar Derbazi [3], Allaoua Boudjedour [2*], Mohamed Dalah [2], Khaled Zennir [4] and Abdelwahab Zarour [2]

[1] *Faculté des Sciences et de Technologie, Université USTHB, Alger, Algeria.*
[2] *University Constantine 1 "FM", FSE, Department of Mathematics, MAM Laboratory.*
[3] *Université Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj, Algeria.*
[4] *Department of Mathematics, College of Science, Qassim University, Saudi Arabia.*

Received: February 4, 2024; Revised: November 17, 2024

Abstract: In this paper, we study the one-dimensional space fractional convection-diffusion problem by using a finite difference method. First, we give the mathematical model of our first initial boundary value problem. In the second step, we develop the discretization of the mathematical model and the development of the scheme for the fractional order type linear diffusion equation. For this scheme, the stability as well as convergence are studied via the Fourier method. At the end, the solutions of some numerical examples are discussed and represented graphically using Matlab. Finally, error analysis shows that the algorithm is convergent.

Keywords: *finite difference schemes; fractional derivative; Caputo fractional derivative; stability; convergence.*

Mathematics Subject Classification (2010): N65M06, 65M12, 35R11, 65L12.

1 Introduction

In this study, we consider the one-dimensional space fractional convection-diffusion problem of Caputo type of order $0 < \alpha < 1$, which is used in the modeling of chemical convection-diffusion. Several techniques for numerical resolution of this type of equation have been studied by several authors [1] - [5]. In most of these techniques, either the solutions of the integrer order differential equation versions of the given problem or the fractional differential equations with initial conditions and boundary conditions are used.

* Corresponding author: mailto:aboudjedour@yahoo.com

The rest of this paper is organized as follows. The following section describes the mathematical model. Sections 2 and 3 introduce model equations and discretization of the mathematical model and development of the scheme. Stability of the approximate scheme is illustrated and described in Section 4. Section 5 describes the convergence of the approximate scheme. Finally, in Section 6, two applications of this technique are given to solve a one-dimensional space fractional convection-diffusion model, numerically.

2 Mathematical Model

We establish a novel mathematical model consisting of a one-dimensional space fractional convection-diffusion problem defined in $\Omega = [0; L]$ and $0 < \alpha \leq 1$ by

$$\frac{\partial u(x,t)}{\partial t} = -c(x)\frac{\partial u(x,t)}{\partial x} + d(x)\frac{\partial^{\alpha} u(x,t)}{\partial x^{\alpha}} + p(x,t), \qquad (x,t) \in \Omega\times]0,\mathrm{T}[, \tag{1}$$

with the boundary conditions

$$u(0,t) = u(L,t) = 0, \qquad x \in \partial\Omega, \tag{2}$$

and the initial condition

$$u(x,0) = f_0(x), \qquad x \in \Omega. \tag{3}$$

3 Discretization of the Mathematical Model and Development of the Scheme

The present study deals with the discretization of the mathematical model which describes the one-dimensional space fractional convection-diffusion problem. First, we discretise the domain [L, R]. We define

$$x_i = x_0 + ih, \text{ and } t_j = t_0 + jk, \quad \forall i = 0,1,\ldots,M \text{ and } \forall j = 0,1,\ldots,N, \tag{4}$$

k represents the time step size and h represents the space step length.

Let us assume that

$$u(x_i,t_j) = u_i^j, \quad p(x_i,t_j) = p_i^j, \quad c(x_i) = c_i, \quad d(x_i) = d_i, \quad f_0(x_i) = f_{0,i}. \tag{5}$$

u_i^j is the numerical approximation of $u(x_i,t_j)$.

The Caputo fractional order derivative is formulated by the structure

$$\frac{\partial^{\alpha} u(x,t)}{\partial x^{\alpha}} = \begin{cases} \frac{1}{\Gamma(1-\alpha)}\int_0^x \frac{u_{\xi}(\xi,t)}{(x-\xi)^{\alpha}}d\xi & \text{if} \quad 0 < \alpha \leq 1, \\ u_x(x,t) & \text{if} \quad \alpha = 1. \end{cases} \tag{6}$$

Initially, as the boundary value problem needs to be discretized to be able to solve (1), it is first necessary to discretize the order space-fractional derivative.

The operator $\left(\frac{\partial u}{\partial \xi}\right)_{i+1}^{s}$ is approximated by the following formula

$$\left(\frac{\partial u}{\partial \xi}\right)_{i+1}^{s} = \frac{u_{i+1}^{s} - u_{i}^{s}}{h} + R\left(h\right). \tag{7}$$

According to (6) and (7), we get

$$\frac{\partial^{\alpha} u\left(x_{i+1}, t_{j}\right)}{\partial x^{\alpha}} = \frac{h^{-\alpha}}{\Gamma(2-\alpha)}\left[u_{i+1}^{j} - u_{i}^{j} + \sum_{n=1}^{i}\left(u_{i-n+1}^{j} - u_{i-n}^{j}\right)B_{i+1}^{j}\left(n\right)\right], \tag{8}$$

where

$$B_{i+1}^{j}\left(n\right) = \left(n+1\right)^{1-\alpha)} - n^{1-\alpha}, \quad j = 0, 1, \ldots, N-1. \tag{9}$$

Then, we use the forward difference approximation of time derivative is follows:

$$\frac{\partial u(x_{i+1}, t_{j})}{\partial t} = \frac{u_{i+1}^{j+1} - u_{i+1}^{j}}{k} + R\left(k\right). \tag{10}$$

Using approximations (8) and (10), and the linear convection-diffusion equations (1)–(3), we obtain

$$-\left(\frac{1+C_{i}}{A_{i}}\right)u_{i+1}^{j+1} + \left(1+\frac{1}{A_{i}}\right)u_{i+1}^{j} = -\frac{C_{i}}{A_{i}}u_{i}^{j+1} + u_{i}^{j} - \sum_{n=1}^{i}\left(u_{i-n+1}^{j} - u_{i-n}^{j}\right)B_{i+1}^{j}\left(n\right) + \frac{k}{A_{i}}p_{i}^{j}, \tag{11}$$

$$i = \overline{1, \mathrm{M}}, j = \overline{1, \mathrm{N}},$$

with the boundary conditions

$$u_{0}^{j} = u_{M}^{j}, \quad j = \overline{0, \mathrm{N}-1}, \tag{12}$$

and the initial condition

$$u\left(x_{i}\right) = f_{0,i}, \quad i = 0, 1, \ldots, \mathrm{M}, \text{ where } \mathrm{A_i} = \frac{\mathrm{d_{i+1}kh^{-\alpha}}}{\Gamma(2-\alpha)} \text{ and } \mathrm{C_i} = \frac{\mathrm{c_{i+1}k}}{\mathrm{h}}. \tag{13}$$

4 Stability of the Approximate Scheme

In this section, we use the method of Fourier analysis to discuss the stability of the approximate scheme (11)–(13). Assume that the solution of the equations (11)–(13) has the form

$$u_{i}^{j} = \zeta_{i} e^{\nu\tau hj}, \quad i = 0, 1, \ldots, \mathrm{M}, \text{ where } \tau = \frac{2\pi \mathrm{m}}{\mathrm{L}} \text{ and } \nu^{2} = -1. \tag{14}$$

After that, we get

$$\zeta_{i+1} = \frac{\zeta_{i}\left(1 - \frac{C_{i}}{A_{i}}e^{\nu\tau h}\right) - \sum_{n=1}^{i}\left(\zeta_{i-n+1} - \zeta_{i-n}\right)B_{i+1}^{j}\left(n\right)}{\left[-\left(\frac{1+C_{i}}{A_{i}}\right)e^{\nu\tau h} + \left(1+\frac{1}{A_{i}}\right)\right]}, \quad i = \overline{0, \mathrm{M}-1}. \tag{15}$$

Theorem 4.1 *The scheme* (11)–(13) *is unconditionally stable for* $0 < \alpha \le 1$ *if*

$$\max_{0 \le i \le \mathrm{M}-1} \left| \frac{3 - \frac{C_i}{A_i} e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_i}\right)} \right| < 1. \tag{16}$$

Proof. We use the proof by recurrence for $i = 1$, in view of (11)–(13),

$$|\zeta_1| = \left| \frac{\xi_0 \left(1 - \frac{C_0}{A_0} e^{\nu\tau h}\right)}{\left[-\left(\frac{1+C_0}{A_0}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_0}\right)\right]} \right|,$$
$$\le T\,|\zeta_0| \le |\zeta_0| \tag{17}$$

where $T = \left| \frac{1 - \frac{C_0}{A_0} e^{\nu\tau h}}{-\left(\frac{1+C_0}{A_0}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_0}\right)} \right| < 1.$

We assume that the statement is true:

$$|\zeta_i| \le |\zeta_0|, \qquad i = \overline{1, \mathrm{M}} \tag{18}$$

and we prove that the statement is true:

$$|\zeta_{i+1}| \le |\zeta_0|, \qquad i = \overline{0, \mathrm{M}-1}. \tag{19}$$

Then, we obtain

$$|\zeta_{i+1}| = \left| \frac{\zeta_i \left(1 - \frac{C_i}{A_i} e^{\nu\tau h}\right) - \sum_{n=1}^{i} \left(\zeta_{i-n+1} - \zeta_{i-n}\right) B_{i+1}^{j}(n)}{\left[-\left(\frac{1+C_i}{A_i}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_i}\right)\right]} \right|,$$
$$\le \left| \frac{\zeta_i \left(1 - \frac{C_i}{A_i} e^{\nu\tau h}\right) + \left|\sum_{s=0}^{i-1} \zeta_{s+1} - \zeta_s\right|}{\left[-\left(\frac{1+C_i}{A_i}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_i}\right)\right]} \right|,$$
$$\le \left| \frac{3 - \frac{C_i}{A_i} e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_i}\right)} \right| |\zeta_0|,$$
$$\le \max_{0 \le i \le \mathrm{M}-1} \left| \frac{3 - \frac{C_i}{A_i} e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_i}\right)} \right| |\zeta_0|,$$
$$\le \left(\max_{0 \le i \le \mathrm{M}-1} \left| \frac{3 - \frac{C_i}{A_i} e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right) e^{\nu\tau h} + \left(1 + \frac{1}{A_i}\right)} \right| \right) |\zeta_0|,$$
$$\le |\zeta_0|. \tag{20}$$

Finally, the approximate scheme (11)–(13) is unconditionally stable.

5 Convergence of the Approximate Scheme

We start by selecting the following Fourier analysis to discuss the convergence of numerical schemes (11). Now, assume that

$$R_i^j = E_i e^{\mu\eta h j} \quad \text{and} \quad E_i^j = u(x_i, t_j) - u_i^j, \quad i = \overline{0, \mathrm{M}}, \;\; j = \overline{0, \mathrm{N}}, \tag{21}$$

where $\eta = \frac{2\pi m}{L}$ and $\mu^2 = -1$.

After that, we get

$$R_{i+1} = \frac{R_i\left(1-\frac{C_i}{A_i}e^{\nu\tau h}\right) - \sum_{n=1}^{i}\left(R_{i-n+1}-R_{i-n}\right)B_{i+1}^{j}\left(n\right)+\zeta_i}{\left[-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right]}, i=\overline{0,\mathrm{M}-1}. \tag{22}$$

Theorem 5.1 *The scheme* (11)–(13) *is convergent for* $0<\alpha<1$ *if*

$$\max_{0\le i\le \mathrm{M}-1}\left\{\frac{1}{\left|-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right|}, \left|\frac{3-\frac{C_i}{A_i}e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)}\right|\right\} < 1. \tag{23}$$

Proof. We use the proof by recurrence for $i=1$, in view of (11)–(13),

$$\begin{aligned} |R_1| &= \left|\frac{R_0\left(1-\frac{C_0}{A_0}e^{\nu\tau h}\right)+\zeta_0}{\left[-\left(\frac{1+C_0}{A_0}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_0}\right)\right]}\right| \\ &\le T\left(|R_0|+|\zeta_0|\right) \le |R_0|+|\zeta_0|, \end{aligned} \tag{24}$$

where $T^* = \max\left\{\frac{1}{\left|-\left(\frac{1+C_0}{A_0}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_0}\right)\right|}, \left|\frac{1-\frac{C_0}{A_0}e^{\nu\tau h}}{-\left(\frac{1+C_0}{A_0}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_0}\right)}\right|\right\} < 1.$

We assume that the statement is true:

$$|R_i|+|\zeta_i| \le |R_0|+|\zeta_0|, \quad \forall i=\overline{1,\mathrm{M}}, \tag{25}$$

and we prove that the statement is true:

$$|R_{i+1}|+|\zeta_{i+1}| \le |R_0|+|\zeta_0|, \quad \forall i=\overline{0,\mathrm{M}-1}. \tag{26}$$

By the convergence of the series on the right-hand side,

$$\exists T^*>0 \quad |R_0|+|\zeta_0| \le T^*\left(k+h\right), \quad i=\overline{0,\mathrm{M}-1}. \tag{27}$$

Then

$$\begin{aligned} |R_{i+1}| &= \left|\frac{R_i\left(1-\frac{C_i}{A_i}e^{\nu\tau h}\right) - \sum_{n=1}^{i}\left(R_{i-n+1}-R_{i-n}\right)B_{i+1}^{j}\left(n\right)+\zeta_i}{\left[-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right]}\right|, \\ &\le \left|\frac{R_i\left(1-\frac{C_i}{A_i}e^{\nu\tau h}\right) + \left|\sum_{s=0}^{i-1}R_{s+1}-R_s\right|+\zeta_i}{\left[-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right]}\right|, \\ &\le \left|\frac{3-\frac{C_i}{A_i}e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)}\right||R_i| + \frac{1}{\left|-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right|}|\zeta_i|, \\ &\le \left|\frac{3-\frac{C_i}{A_i}e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)}\right||R_0| + \frac{1}{\left|-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right|}|\zeta_0|, \\ &\le C\max_{0\le i\le \mathrm{M}-1}\left\{\frac{1}{\left|-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)\right|}, \left|\frac{3-\frac{C_i}{A_i}e^{\nu\tau h}}{-\left(\frac{1+C_i}{A_i}\right)e^{\nu\tau h}+\left(1+\frac{1}{A_i}\right)}\right|\right\}, \\ &\le C' \le \left(k+h\right), \end{aligned} \tag{28}$$

where the constant C' is given by $C' = |R_0| + |\zeta_0|$. Finally, the scheme (11)–(13) is convergent.

6 Numerical Simulation

In this section, we provide illustrative simulations that demonstrate the theoretical aspects related to stability and convergence of the fractional convection-diffusion.

Example 1. Consider the space-fractional diffusion type of problem :

$$\frac{\partial u(x,t)}{\partial t} = \Gamma(1.2)\, x^{\alpha} \frac{\partial^{\alpha} u(x,t)}{\partial x^{\alpha}} + \left(6x^3 - 3x^2\right) e^{-t}, \quad (x,t) \in \Omega \times]0, \mathrm{T}[, \tag{29}$$

with the boundary conditions $u(0, \mathrm{t}) = u(1, \mathrm{t}) = 0,\ x \in \partial\Omega$, and the initial condition $u(x, 0) = x^2 - x^3,\ x \in \Omega$. The exact solution $u(x,t) = \left(x^2 - x^3\right) e^{-t},\ (x,t) \in \Omega \times]0, \mathrm{T}[$. The problem (29) is unconditionally stable and convergent if

$$\|d\|_{\infty} \leq \frac{h^{\alpha} \Gamma(2-\alpha)}{k}. \tag{30}$$

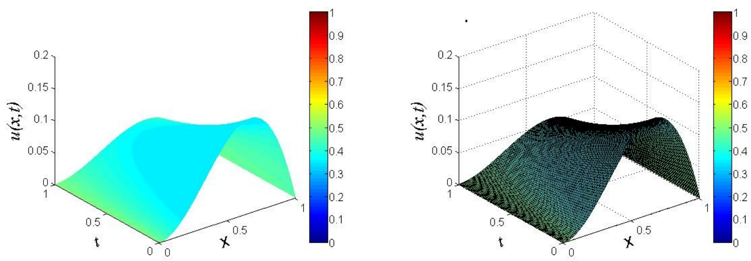


Figure 1: The right figure represents the numerical solution of $u(x,t)$ for $\alpha = 0.93, \mathrm{N} = 100$, while the left figure represents the exact solution.

Example 2. In the second example, we consider the space-fractional diffusion type of problem :

$$\frac{\partial u(x,t)}{\partial t} = x^{\frac{1}{5}} \frac{\partial u(x,t)}{\partial x} + x^{\frac{1}{100}} \frac{\partial^{\alpha} u(x,t)}{\partial x^{\alpha}} + e^{-2t}\left(2\left(x - x^{\alpha}\right) - \Gamma(\alpha) + \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} x^{\alpha-1} - 1\right), \tag{31}$$

with $(x,t) \in \Omega \times]0, \mathrm{T}[$ and the boundary conditions $u(0,t) = u(1,t) = 0$, where $x \in \partial\Omega$ and the initial condition $u(x,0) = x^{\alpha} - x,\ x \in \Omega$. The exact solution $u(x,t) = e^{-2t}\left(x^{\alpha} - x\right),\ (x,t) \in \Omega \times]0, \mathrm{T}[$. In this example, we present different numerical experiments to support the theoretical and numerical analyses of the previous sections. The problem (31) is unconditionally stable and convergent.

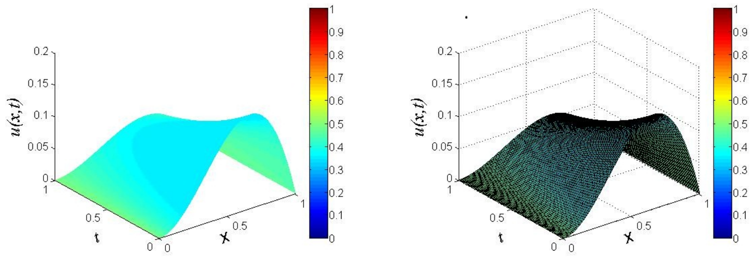


Figure 2: The right figure represents the numerical solution of $u\,(x,t)$ for $\alpha = 0.8, \mathrm{N} = 100$, while the left figure represents the exact solution.

7 Conclusion

In this paper, the one-dimensional space fractional convection-diffusion problem with initial and boundary conditions in a bounded domain is studied by using a finite difference method. The fractional derivative is approximated by the finite difference approximations for space derivatives and Caputo's concept for time-fractional derivatives. Two numerical examples with the known exact solutions are considered to validate theoretical results and demonstrate the accuracy of the method proposed in this paper.

References

[1] Q. M. Al-Mdallal, M. A. Hajji and T. Abdeljawad. On the iterative methods for solving fractional initial value problems: new perspective. *J. Fract. Calculus Nonlinear Syst.* **2** (1) (2021) 76–81.

[2] C. Tadjeran and M. M. Meerschaert. A second-order accurate numerical method for the two dimensional fractional diffusion equation. *J. Comput. Phys.* **2**20 (2007) 813–823.

[3] I. Podlubny. I. *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications.* Elsevier: New York, NY, USA, Vol. 198, 1998.

[4] K. Wang and H. Wang. A fast characteristic finite difference method for fractional advection–diffusion equations. *Adv. Water. Resour* **3**4 (2011) 810–816.

[5] C. Li, and F. Zeng. Finite difference methods for fractional differential equations, *Int. J. Bifurcat. Chaos*, **2**2 (2012) 1230014.

[6] R. J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations.* SIAM: Philadelphia, PA, USA, Vol. 98, 2007.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 582–593

Inducing Chaos through Timescales in a Three-Species Food Chain Model

Khadidja Daas * and Nasreddine Hamri

Laboratory of Mathematics and Their Interactions,
Departement of Mathematics, Boussouf Abdelhafid University Center,
Mila, Algeria

Received: April 1, 2024; Revised: November 2, 2024

Abstract: Over time, considerable attention has been devoted to understanding the complex dynamics of simplified ecosystems containing three trophic levels, revealing that complex behaviors can arise from a simple hierarchy between prey, predator, and top predator. In this study, the extension of a three-species food chain model with a Crowley-Martin-type functional response was examined by introducing discrete timescales to study its impact on system dynamics at various trophic levels. Changes in species abundance were analyzed across three different timescales: slow, fast, and intermediate. The presence of a homoclinic orbit in the subsystem (prey-predator) suggests the existence of period-doubling cascades that eventually lead to chaos in the entire system (prey-predator-top-predator). This study underscores the importance of ecological modeling and trophic interactions in understanding the diverse and intricate dynamics of ecosystems, thus highlighting the significance of research in this domain.

Keywords: *three-level trophic systems; time scales; chaos.*

Mathematics Subject Classification (2010): 37D45, 34C15, 37C75, 65L07, 70Kxx.

* Corresponding author: mailto:khadidja.daas@centre-univ-mila.dz

1 Introduction

A shift in focus towards understanding three-species systems resulted from the recognition that two-species systems are insufficient [7]. In seminal works, it has been demonstrated that models with three or more species can capture complex oscillatory dynamics within certain parameter ranges [4], [11], [8]. However, earlier studies of three-trophic level systems did not account for intraspecific rivalry at higher trophic levels, i.e., in top predators or predators. Thus, it was necessary to consider the potential impact of this rivalry on the system's dynamic properties. In a recent study, Peet et al. [1] added square terms to the equations for top predators and predators, thereby expanding Hastings-Powell's model. They demonstrated the importance of intraspecific rivalry in the evolution of chaotic trajectories by showing the coexistence of a chaotic attractor and a period-one cycle. They also showed how enhancing intraspecific competition among top predators can stabilize the system and pull it out of a chaotic state. Nevertheless, a critical component was absent from earlier research: the examination of various timescales at various trophic levels. Therefore, it is crucial that the modeling technique takes these various timelines into consideration. The system becomes singularly perturbed by adding several timescales, and geometric singular perturbation theory can be used to evaluate the system mathematically. This approach was first used by Rinaldi and Muratori [9] to examine slow-fast cycles in a three-species system on two timescales. They carried on more research and showed that species with slow, intermediate, and fast variables might cohabit oscillatory. In Hastings-Powell's model with several timescales, this was accomplished by employing singular perturbation techniques. Including several timescales can reveal far more complex dynamics like relaxation oscillations and canard cycles. These dynamics provide important new information for researching the occurrence of intricate chaotic oscillations. In our analysis, we used the method described in [6] to include three distinct timescales: slow, intermediate, and fast, and we divided the system into slow, intermediate, and fast subsystems. We investigated the three-species food chain model [10]. The results of each subsystem were then concatenated in order to look into the possibility of solitary slow-intermediate-fast cycles. Because of the intricacy of the model, numerical simulations were used to show that a homoclinic orbit exists inside a subsystem of the entire system. In the whole system, period-doubling cascades to chaos were seen simultaneously. The dynamics of the system changed either gradually or abruptly, depending on the degree of intraspecific rivalry among top predators. As a result, our research provides a possible framework for identifying and evaluating crucial changes that may cause an ecosystem to drastically change. This paper delves into the intricacies of a tri-trophic food web model, dissecting its dynamics and behaviors. Starting with the concept of a dimensionless model, the exploration seamlessly transitions to a meticulous reformulation of the three-species model using two dimensionless positive timescale parameters. In the subsequent section, the focus extends to a linear investigation and dynamic features. Equilibrium points are explored within the context of our study, contributing to a comprehensive understanding. Simultaneously, a detailed stability analysis is conducted, unraveling the nuanced intricacies. Moving through the paper, attention shifts to the dynamics of the subsystems, providing a granular perspective on their interplay and significantly contributing to the overall comprehension of the tri-trophic food web model. The exploration takes an intriguing turn in the final section, where chaos becomes the focal point. Introducing Lyapunov exponents as a lens to understand chaos, the paper concludes with an examination of the gradual entry into chaos. Through this meticulously organized framework, the aim is to provide a holistic understanding of the tri-trophic food web model and its dynamic nuances.

2 A Tri-Trophic Food Web Model

2.1 Dimensionless model

The system under investigation in this work represents a mathematical model of a three-level food chain, which was transformed into a dimensionless model in [10]. The second order Holling pattern and the Crowley-Martin type functional response are combined in this food chain to form a hybrid type of organism.

$$\begin{cases} \dfrac{dX_1}{dT} = a_1 X_1 \left(1 - \dfrac{X_1}{K}\right) - \dfrac{cX_2X_3}{X_1 + D}, \\ \dfrac{dX_2}{dT} = -a_2X_2 + \dfrac{c_1X_1X_3}{X_1 + D_1} - \dfrac{c_2X_2X_3}{1 + dX_2 + bX_3 + bdX_2X_3}, \\ \dfrac{dX_3}{dT} = -mX_3 + \dfrac{c_3X_2X_3}{1 + dX_2 + bX_3 + bdX_2X_3}, \end{cases} \tag{1}$$

where the population densities of the prey, predator, and top predator, respectively, as a function of time T are represented by the variables $X_1(T)$, $X_2(T)$, and $X_3(T)$.

As shown in Table 1, the model (1) is distinguished by the existence of 12 control parameters that regulate the behavior of the system.

Parameter	Description
a_1	The prey's intrinsic growth rate in the absence of predators
K	The prey's carrying capacity
D	Prey environmental protection
D_1	Predator environmental protection
c	The maximum rate of prey reduction per capita
c_1	Similar to c, the maximum rate of prey reduction per capita
c_2, c_3	Characteristics of the Crowley-Martin type functional response
b	Parameter assessing predator interference
a_2	Intermediate predator death rate X_2
m	Top predator death rate X_3

Table 1: Tri-trophic level food chain model parameters.

We simplified this model, ignoring dimensional considerations, to make the mathematical analysis easier. Table 2 contains dimensionless representations of the variables and parameters, where the first line represents the dimensionless variables, and the second line represents the corresponding environmental values.

t	x_1	x_2	x_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
a_1T	$\frac{X_1}{K}$	$\frac{cX_2}{a_1K}$	$\frac{cc_2X_3}{a_1^2dK}$	$\frac{D}{K}$	$\frac{a_2}{a_1}$	$\frac{c_1}{a_1}$	$\frac{D_1}{K}$	$\frac{a_1b}{c_2}$	$\frac{a_1^2bdK}{cc_2}$	$\frac{c}{a_1dK}$	$\frac{c}{a_1}$	$\frac{c_3}{a_1d}$

Table 2: Variables and parameters without dimensions.

As a result, we derived a dimensionless system characterized by nine parameters as detailed below:

$$\begin{cases} \dfrac{dx_1}{dt} = x_1\left[(1-x_1) - \dfrac{x_2}{x_1+c_4}\right] = x_1 g_1(x_1,x_2), \\ \dfrac{dx_2}{dt} = x_2\left[-c_5 + \dfrac{c_6 x_1}{x_1+c_7} - \dfrac{x_3}{x_2+(c_8+c_9x_2)\,x_3+c_{10}}\right] = x_2 g_2(x_1,x_2,x_3), \\ \dfrac{dx_3}{dt} = x_3\left[-c_{11} + \dfrac{c_{12}x_2}{x_2+(c_8+c_9x_2)\,x_3+c_{10}}\right] = x_3 g_3(x_2,x_3). \end{cases} \tag{2}$$

2.2 Reformulation of the three-species model

We employ two dimensionless positive timescale parameters, β_1 and β_2, to rescale the three-species model (2), with $0 < \beta_1, \beta_2 \ll 1$. The rescaling is done so that the growth rate of the predator is $O(\beta_1)$, and the growth rate of the top predator is $O(\beta_2)$. We reformulate the model (2) as follows after making this adjustment:

$$\begin{cases} \dfrac{dx_1}{dt} = x_1\left[(1-x_1) - \dfrac{x_2}{x_1+c_4}\right] = x_1 g_1(x_1,x_2), \\ \dfrac{dx_2}{dt} = \beta_1 x_2\left[-c_5 + \dfrac{c_6 x_1}{x_1+c_7} - \dfrac{x_3}{x_2+(c_8+c_9x_2)\,x_3+c_{10}}\right] = \beta_1 x_2 g_2(x_1,x_2,x_3), \\ \dfrac{dx_3}{dt} = \beta_2 x_3\left[-c_{11} + \dfrac{c_{12}x_2}{x_2+(c_8+c_9x_2)\,x_3+c_{10}}\right] = \beta_2 x_3 g_3(x_2,x_3). \end{cases} \tag{3}$$

The system's fast, intermediate, and slow variables, denoted by the values x_1, x_2, and x_3, respectively, reflect the dimensionless densities of prey, predators, and top predators. By putting the transformation $\tau_1 = \beta_1 t$ into (3), we may define the system and obtain the following results:

$$\beta_1 \frac{dx_1}{d\tau_1} = x_1 g_1(x_1,x_2), \frac{dx_2}{d\tau_1} = x_2 g_2(x_1,x_2,x_3), \beta_1 \frac{dx_3}{d\tau_1} = \beta_2 x_3 g_3(x_2,x_3), \tag{4}$$

and after additional changes $\tau_2 = \beta_2 t$, we obtain

$$\beta_2 \frac{dx_1}{d\tau_2} = x_1 g_1(x_1,x_2), \beta_2 \frac{dx_2}{d\tau_2} = \beta_1 x_2 g_2(x_1,x_2,x_3), \frac{dx_3}{d\tau_2} = x_3 g_3(x_2,x_3), \tag{5}$$

t, τ_1, and τ_2, the dimensionless time variables, represent the fast, intermediate, and slow time scales, respectively. We divide the system into subsystems and use geometric singular perturbation theory to study the dynamics of each one. For the systems (3), (4), and (5), a typical solution trajectory consists of segments corresponding to slow, intermediate, and fast processes. The total solution for the system is then obtained by concatenating the solutions of each subsystem (3). We initially investigate the impact of several time scales on the local dynamics of the system (3) before breaking it down into its component subsystems.

3 Linear Investigation and Dynamic Features

3.1 Equilibrium points

The number of non-negative equilibrium points in the system described by equation (3) can only be four. Table 3 summarizes the outcomes of the equilibrium point values and associated conditions.

Equilibrium point	Values and conditions
P_0	Trivial equilibrium point: $(0, 0, 0)$
P_1	One-species equilibrium point: $(1, 0, 0)$
P_2	Two-species equilibrium point: $(\tilde{x_1}, \tilde{x_2}, 0)$, where $\tilde{x_1}$ and $\tilde{x_2}$ are: $\tilde{x_1} = \frac{c_5 c_7}{c_6 - c_5}$ $\tilde{x_2} = (1 - \tilde{x_1})(\tilde{x_1} + c_4)$ Existence condition: $0 < \frac{c_5 c_7}{c_6 - c_5} < 1$
P_3	Equilibrium point of coexistence of all three species (x_1^*, x_2^*, x_3^*) where: $x_2^* = (1 - x_1^*)(x_1^* + c_4)$ $x_3^* = \frac{(c_{12} - c_{11})x_2^* - c_{10}c_{11}}{c_{11}(c_8 + c_9 x_2^*)}$ Implicit equation for x_1^*: $-c_5 + \frac{c_6 x_1^*}{x_1^* + c_7} - \frac{x_3^*}{x_2^* + \left(c_8 + c_9 x_2^*\right) x_3^* + c_{10}} = 0$ Existence conditions: $0 < x_1^* < 1$, $0 < \frac{c_{10}c_{11}}{c_{12} - c_{11}} < x_2^*$

Table 3: Equilibrium points and conditions.

3.2 Stability analysis

Every equilibrium point's Jacobian matrix was computed, and stability was examined using each matrix's characteristic polynomial. Table 4 summarizes the computation's findings.

Equilibrium Point	Jacobian Matrix	Eigenvalues
P_0	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -\beta_1 c_5 & 0 \\ 0 & 0 & -\beta_2 c_{11} \end{pmatrix}$	$e_1 = 1 > 0$, $e_2 = -\beta_1 c_5 < 0$, $e_3 = -\beta_2 c_{11} < 0$
P_1	$\begin{pmatrix} -1 & \frac{-1}{1+c_5} & 0 \\ 0 & \beta_1(\frac{c_6}{1+c_7} - c_5) & 0 \\ 0 & 0 & -\beta_2 c_{11} \end{pmatrix}$	$e_1 = -1$ $e_2 = \beta_1(\frac{c_6}{1 + c_7} - c_5)$ $e_3 = -\beta_2 c_{11}$
P_2	$\begin{pmatrix} a_{11} & -a_{12} & 0 \\ \beta_1 a_{21} & \beta_1 a_{22} & -\beta_1 a_{23} \\ 0 & 0 & \beta_2 a_{33} \end{pmatrix}$	$e_0 = \beta_2 a_{33}$ $e_{1,2}(\beta_1) = \frac{a_{11} + \beta_1 a_{22}}{2} \pm \frac{1}{2}\sqrt{(a_{11} - \beta_1 a_{22})^2 - 4a_{12}a_{21}}$

Table 4: Table of equilibrium points, Jacobian matrices, and eigenvalues.

Stability results and remarks:

- P_0 is always a saddle point.
- For P_1:

- If $c_5 < \frac{c_6}{1+c_7}$, then $\lambda_2 < 0$, and P_1 is a saddle point.
- If $c_5 > \frac{c_6}{1+c_7}$, then $\lambda_2 > 0$, and P_1 is stable.

- For P_2:
  - The coefficients of the matrix J_{P_2} are

$$a_{11} = 1 - 2\tilde{x_1} - \frac{c_4(1-\tilde{x_1})}{\tilde{x_1}+c_4}, a_{12} = \frac{\tilde{x_1}}{\tilde{x_1}+c_4}, a_{21} = \frac{(1-\tilde{x_1})(\tilde{x_1}+c_4)c_6c_7}{(\tilde{x_1}+c_7)^2},$$

$$a_{22} = -c_5 + \frac{c_6\tilde{x_1}}{\tilde{x_1}+c_7}, a_{23} = \frac{(1-\tilde{x_1})^2(\tilde{x_1}+c_4)^2 + c_{10}(1-\tilde{x_1})(\tilde{x_1}+c_4)}{[(1-\tilde{x_1})(\tilde{x_1}+c_4)+c_{10}]^2},$$

$$a_{33} = -c_{11} + \frac{c_{12}(1-\tilde{x_1})^2(\tilde{x_1}+c_4)^2 + c_{10}c_{12}(1-\tilde{x_1})(\tilde{x_1}+c_4)}{[(1-\tilde{x_1})(\tilde{x_1}+c_4)+c_{10}]^2}$$

  - We select c_5 as the bifurcation parameter in order to find the instability threshold for P_2. The Hopf bifurcation causes the equilibrium point P_2 to lose stability at $c_5 = \hat{c}_5$, where the real parts of the eigenvalues $e_{1,2} = 0$ are located. We select the parameters discussed in [10] and arrange them in Table 5, taking into account $\beta_1 = 1$, $\beta_2 = 1$.

Parameter	c_4	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
Value	0.25	0.8	0.25	0.01	0.1	0.28	0.06	0.25

Table 5: Parameter values.

We find that the real part of the eigenvalues $e_{1,2}$ equals zero at the value $\hat{c}_5 = 0,48$, from this, we reach the following results:

* If $c_5 < 0,48$, the equilibrium point P_2 is unstable.
* If $c_5 > 0,48$, the equilibrium point P_2 is stable.

To emphasize these results further, we provide an example, we select two values for c_5, and based on them, we calculate the equilibrium point P_2, and we analyze the stability in each case.

We present the results in Table 6.

c_5	Equilibrium point P_2	Jacobian matrix	Eigenvalues and stability
0.25	$(0.1136, 0.3223, 0)$	$\begin{pmatrix} -0.1136 & -0.3124 & 0 \\ -0.8254 & -0.9374 & -1.6603 \\ 0 & 0 & -0.3112 \end{pmatrix}$	0.1283, − 1.1794, − 0.3112 P_2 is unstable
0.5	(0.4167,0.3889,0)	$\begin{pmatrix} -0.4167 & -0.6250 & 0 \\ 0.300 & -1.2500 & -1.4950 \\ 0 & 0 & -0.3046 \end{pmatrix}$	$-0.8334+0.1181i$, $-0.8334-0.1181i$, − 0.3046 P_2 is stable

Table 6: Stability study results for the dynamical system for two values of c_5.

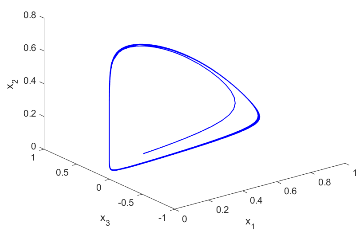
- The interior equilibrium point $P_3 = (x_1^*, x_2^*, x_3^*)$:
We address the instability of the inner equilibrium point P_3 using a numerical example due to the intricacy of the equation involving x_1^*. We select the same set of parameter values with $c_5 = 0.25$, $\beta_1 = 1$, and $\beta_2 = 1$ as shown in Table 5. We use Liu's criteria to investigate the instability of P_3 via the Hopf bifurcation. Assume that the bifurcation parameter is c_{11}. We obtain that the matrix J_{P_3} has the characteristic equation as a function of c_{11}: $\lambda^3 + k_1(c_{11})\lambda^2 + k_2(c_{11})\lambda + k_3(c_{11}) = 0$, where $k_1(c_{11}) = 0.1412c_{11} - 0.0084187$; $k_2(c_{11}) = 7.8822 * 10^{-6}c_{11} - 0.025663$; $k_3(c_{11}) = 0.00023312 - 0.0038824c_{11}$. We consider $\Lambda(c_{11}) = k_1(c_{11})k_2(c_{11}) - k_3(c_{11})$, according to Liu's criteria [12], [3], P_3 becomes unstable through the Hopf bifurcation if there exists a critical value $\hat{c}_{11}$ such that $k_1(\hat{c}_{11}) > 0$, $k_3(\hat{c}_{11}) > 0$, $\Lambda(\hat{c}_{11}) = 0$, and $\left.\frac{d\Lambda}{dc_{11}}\right|_{c_{11}=\hat{c}_{11}} \neq 0$. After selecting the parameter values as previously indicated, we get $\hat{c}_{11} = 0.065964$. For $0 < c_{11} < 0.065964$, the coexistence equilibrium P_3 is stable; for $c_{11} > 0.065964$, it is unstable. We choose $\beta_1 = \beta_2 = 1$ and $c_{11} = 0.06$ as a specific case. Then $P_3 = (0.9240, 0.0893, 0.1412)$ is the only feasible coexistence equilibrium point. The Jacobian matrix evaluated at P_3, which is denoted as J_{P_3}, is given by

$$J_{P_3} = \begin{pmatrix} -0.9241 & -0.7871 & 0 \\ 0.3912 & -1.1659 & -2.6884 \\ 0 & 0.2551 & -0.2556 \end{pmatrix},$$

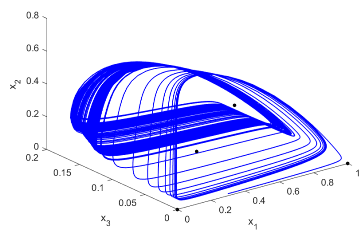
$\lambda^3 + k_1\lambda^2 + k_2\lambda + k_3 = 0$ is the characteristic equation of the matrix J_{P_3}, where $k_1 = 2.3455$, $k_2 = 2.6053$, and $k_3 = 0.9878$. Since $k_1 > 0$, $k_3 > 0$, and $k_1k_2 - k_3 = 5.1229 > 0$, we may conclude that P_3 is stable based on the Routh-Hurwitz criteria.

Choosing $c_{11} = 0.07$ and $\beta_1 = 0.7, \beta_2 = 0.49$ as another example yields unstable values $P_2 = (0.1136, 0.3223, 0)$ and $P_3 = (0.9046, 0.1101, 0.1482)$.

Figures 1, 2, and 3 will be used to illustrate this final example.



(a)



(b)

Figure 1: Two trajectories converging to different periodic attractors, (a) Trajectory for $\beta_1 = 0.7$, $\beta_2 = 0.49$, initial point =$(0.2, 0.1, 0)$, (b) Trajectory for $\beta_1 = 0.7$, $\beta_2 = 0.49$, initial point =$(0.3, 0.1, 0.001)$.

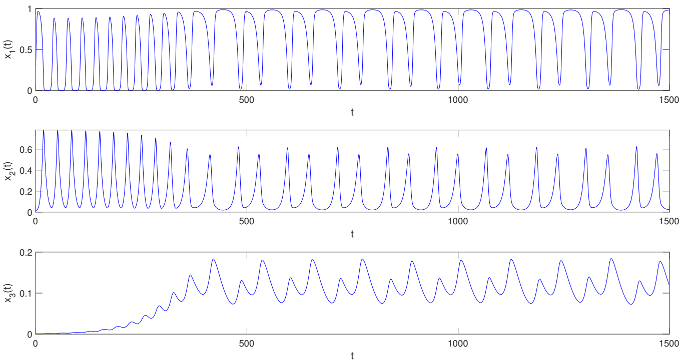
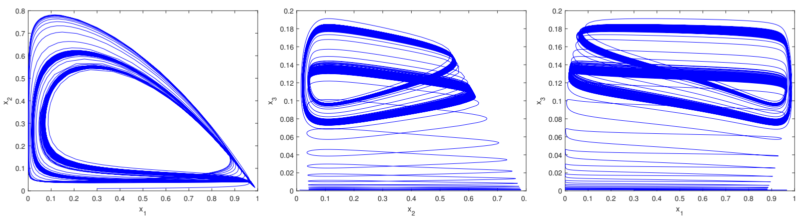


Figure 2: The frequencies of x_1, x_2, and x_3 as a function of time.



(a) (b) (c)

Figure 3: (a) Projection of the trajectory onto the x_1x_2 plane, (b) Projection of the trajectory onto the x_2x_3 plane, (c) Projection of the trajectory onto the x_1x_3 plane.

In the system, two stable limit cycles coexist: one around the equilibrium point P_2 in the x_1x_2 plane, and the other around P_3 in three-dimensional space, illustrating bistability. The trajectories are sensitive to initial conditions, suggesting the possibility of chaos.

4 Behavior of Subsystems

First, we present the model in the singular limit, that is, in the case when either $\beta_1 \to 0$ or $\beta_2 \to 0$, or both may occur. For $0 < \beta_1 < \beta_2 \ll 1$, the trajectory of the entire system (3) is a perturbed solution of subsystems. After time is divided into fast, intermediate, and slow timescales, we list the subsystems of the system (2) in Table 7. Concatenated slow-intermediate-fast flow, or the solutions of the aforementioned subsystems, make up the unique trajectory. A schematic example of a unique slow-intermediate-fast cycle is shown in Figure 4a, which consists of one slow flow segment (the thin black line), three segments of intermediate flows (the medium blue line), and two segments of fast flows (the thick red line). The critical manifold of the fast subsystem is the set of all equilibrium points as follows:

$$M^0 = \{(x_1, x_2, x_3) : x_1 = 0, x_2, x_3 > 0\}, \quad M^1 = \{(x_1, x_2, x_3) : g_1(x_1, x_2) = 0, x_3 > 0\},$$

Condition	Subsystem
$\beta_1 \to 0$ (Fast)	$\begin{cases} \frac{dx_1}{dt} = x_1 \left[(1-x_1) - \frac{x_2}{x_1+c_4} \right] = x_1 g_1(x_1,x_2), \\ \frac{dx_2}{dt} = 0, \\ \frac{dx_3}{dt} = 0. \end{cases}$
$\beta_2 \to 0, \beta_1 > 0$ (Intermediate)	$\begin{cases} \frac{dx_1}{dt} = 0, \\ \frac{dx_2}{dt} = \beta_1 x_2 \left[-c_5 + \frac{c_6 x_1}{x_1+c_7} - \frac{x_3}{x_2+(c_8+c_9x_2)\,x_3+c_{10}} \right] = \beta_1 x_2 g_2(x_1,x_2,x_3), \\ \frac{dx_3}{dt} = 0. \end{cases}$
$\beta_1, \beta_2 \to 0$ (Slow)	$\begin{cases} \frac{dx_1}{dt} = 0, \\ \frac{dx_2}{dt} = 0, \\ \frac{dx_3}{dt} = x_3 \left[-c_{11} + \frac{c_{12} x_2}{x_2+(c_8+c_9x_2)\,x_3+c_{10}} \right] = x_3 g_3(x_2,x_3). \end{cases}$

Table 7: Description of subsystems in the model.

it expressly looks like this $M^1 = \{(x_1, x_2, x_3) : x_2 := \varphi(x_2) = (1-x_2)(x_2+c_4), x_2 > 0, x_3 > 0\}$. There is a fold in this surface, and we can find the fold curve by $\mathcal{C} = \{(x_1, x_2, x_3) : \varphi'(x_1) = 0, x_2 = \varphi(x_1), x_3 \geq 0\}$, implying $x_{1,max} = \frac{1-c_4}{2}$.

With the exception of the fold curve $\mathcal{C}$, where it loses its hyperbolicity, M^1 is hyperbolic everywhere. The non-trivial critical manifold $M^1 = 0$ is divided into typically hyperbolic attracting and repelling sub-manifolds by the fold curve $\mathcal{C}$ as follows: $M_a^1 = \{(x_1, \varphi(x_1), x_3) : x_1 > x_{1,max}, x_3 \geq 0\}$ and $M_r^1 = \{(x_1, \varphi(x_1), x_3) : x_1 < x_{1,max}, x_3 \geq 0\}$, respectively. The trivial critical manifold M^0 (x_2x_3-plane) and the manifold M^1 intersect at a curve that is described by $\mathcal{T}_1 = \{(0, c_4, x_3) : x_3 \geq 0\}$.

The transcritical bifurcation curve, $\mathcal{T}_1$, splits the plane into hyperbolic sub-manifolds that are typically repellent and attractive, respectively. Next, the manifold M_0's attracting and repelling sub-manifolds are $M_a^0 = \{(0, x_2, x_3) : x_2 > \varphi(0), x_3 > 0\}$, and $M_r^0 = \{(0, x_2, x_3) : x_2 < \varphi(0), x_3 > 0\}$, respectively.

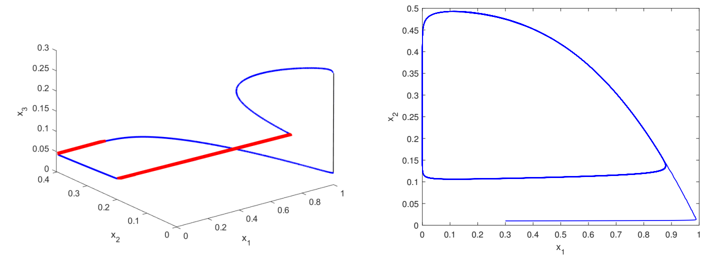
For β_1, β_2, where $0 < \beta_1, \beta_2 \ll 1$, Fenichel's theorem [5] guarantees the existence of locally invariant perturbed sub-manifolds, the essential manifolds M^0 and M^1 have the sub-manidolds $M^0_{\beta_1,\beta_2}$ and $M^1_{\beta_1,\beta_2}$, respectively, with the exception of the non-hyperbolic curves $\mathcal{C}$ and $\mathcal{T}_1$. Additionally, the corresponding attracting M^1_{a,β_1,β_2} and repelling M^1_{r,β_1,β_2} sub-manifolds are perturbed by the attracting (M_a^1) and repelling (M_r^1) sub-manifolds of the critical manifold. Consequently, the perturbed sub-manifolds $M^0_{\beta_1,\beta_2}$ and $M^1_{\beta_1,\beta_2}$ dictate the dynamics of the entire system (3) for $\beta_1, \beta_2 \neq 0$ locally.

The intermediate subsystem is defined on the manifolds M^0 and M^1 with $x_3 =$ constant, therefore we analyze the intermediate system in the plane $x_3 = c'$, parallel to the x_1x_2-plane. Substituting $x_3 = c'$ in the expression of g_3, we obtain an explicit expression of the non-trivial nullcline of the intermediate subsystem as follows:

$$x_1 = \frac{c_5c_7(1+c_9c')x_2 + (c_5c_7c_8 + c_7)c' + c_5c_7c_{10}}{(c_6 + c_9(c_6-c_5)c' - c_5)x_2 + (c_6c_8 - c_5c_8 - 1)c' + c_{10}(c_6-c_5)}. \tag{6}$$

The number of equilibrium points is related to the value c' and other background parameters.

For example, in Figure 4b, we represents that for $c' = 0.1$, the system has a unique unstable equilibrium surrounded by a stable limit cycle.



(a) (b)

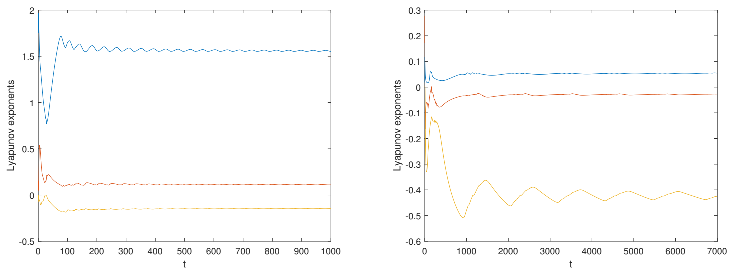
Figure 4: The dynamics of the subsystems: (a) A schematic representation of the slow-intermediate-fast cycle for $\beta_1 = 0.005$ and $\beta_2 = 0.0035$, (b) The dynamics of the intermediate systems for $c = 0.1$, $\beta_1 = 0.1$ and $\beta_2 = 0$.

Understanding the dynamics generated by the intermediate subsystem of the complete system is particularly important for certain specific instances. The primary goal of this work is to examine many chaotic dynamics that the system (3) exhibits and to find out how the system's chaotic regimes may be impacted by the various timescales.

5 Dynamical Analysis of Chaos

5.1 Lyapunov exponents

We represented the Lyapunov exponents by assuming two values for β_1 and β_2 and obtained the following results.



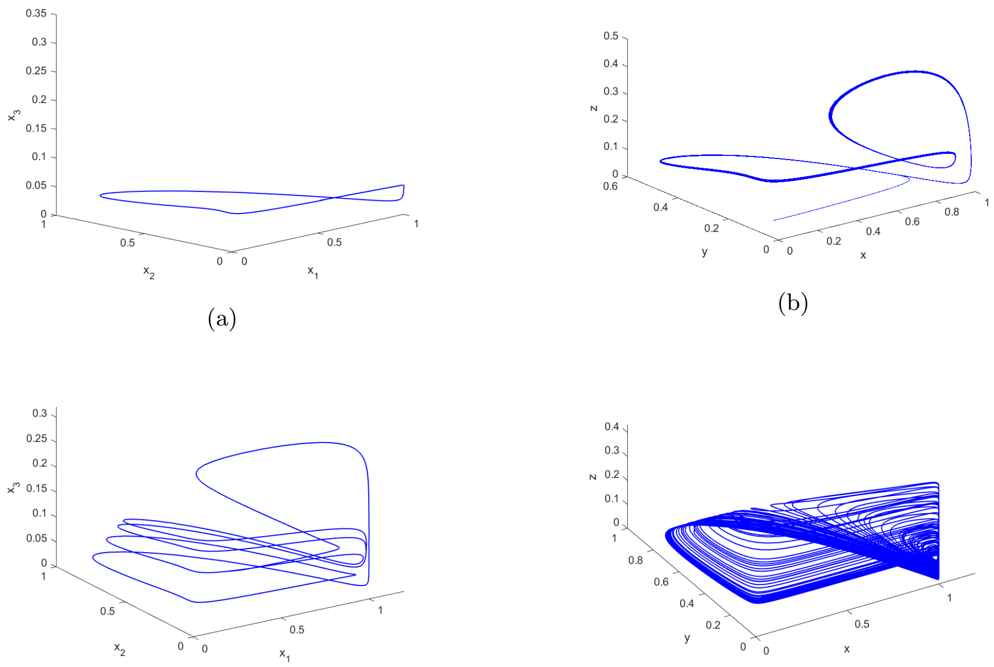
(a) (b)

Figure 5: (a) Lyapunov exponents for $\beta_1 = 1, \beta_2 = 1$, (b) Lyapunov exponents for $\beta_1 = 0.1, \beta_2 = 0.05$.

Comparison: The presence of this type of Lyapunov exponents in Figure 5a indicates greater complexity in the system's behavior, where there is local convergence in one direction with dispersion or chaos in others. This suggests a higher sensitivity of system (2) to initial conditions compared to the system after temporal segmentation (3), resulting in a more complex and chaotic behavior. The role of this study using temporal segmentation becomes evident, as it has made the system's behavior clearer.

5.2 Entering chaos gradually

We demonstrate how the timescales parameters affect the chaotic dynamics. Our simulations begin at $\beta_1 = \beta_2 = 1$ and are progressively decreased to determine the period-doubling cascade that leads to chaos. In Figure 6a, we note that the species live along a periodic orbit in the absence of multiple timescales. On the other hand, in Figure 6b, the 1-periodic orbit experiences a period-doubling bifurcation as β_1 and β_2 decrease, we achieve a 2-periodic orbit. In Figure 6c, we obtain a 4-periodic orbit. Thus, with consecutive period-doubling bifurcations, the system becomes chaotic from periodic (see Figure 6d).



(a) (b) (c) (d)

Figure 6: The periodic-doubling bifurcation with varying β_1 and β_2, (a) Period 1 for $\beta_1 = 1$ and $\beta_2 = 1$, (b) Period 2 for $\beta_1 = 0.25$ and $\beta_2 = 0.125$, (c) Period 4 for $\beta_1 = 0.2$ and $\beta_2 = 0.1$, (d) Chaos for $\beta_1 = 0.1$ and $\beta_2 = 0.05$.

6 Conclusion

The ecological interpretations of the results presented in this paper underscore the importance of understanding the environmental impacts of changes in key parameters of environmental models. As the reproduction rate of the first predator approaches zero, it reflects the environmental response of the food chain and predators. Reducing the reproduction rate leads to a decrease in the population of the first predator, indirectly affecting the higher predator, which relies on the first predator as a food source. It is worth noting that the environmental effects of these changes are not limited to the individual level but also extend to the ecosystem level as a whole. Due to the complex interactions between living organisms and environmental factors, the ecosystem can transition into a state of chaos, where behavior becomes unpredictable and dynamics are unstable. Therefore, this research sheds light on the importance of analyzing the environmental impacts of changes in the key parameters of environmental models and their role in determining the stability and evolution of ecosystems in the long term.

In summary, this research makes a significant contribution to understanding the environmental impacts of changes in ecological systems and identifying factors that influence their stability. This helps in developing strategies for conserving biodiversity and ensuring environmental sustainability.

References

[1] A. Peet, P. Deutsch and E. Peacock-López. Complex dynamics in a three-level trophic system with intraspecies interaction. *Journal of Theoretical Biology* **232** (4) (2005) 491–503.

[2] B. Deng and G. Hines. Food chain chaos due to Shilnikov's orbit. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **12** (3) (2002) 533–538.

[3] D. Sen, S. Ghorai and M. Banerjee. Complex dynamics of a three species prey-predator model with intraguild predation. *Ecological Complexity* **34** (2018) 9–22.

[4] M. Benkara and N. Hamri. Stability and Hopf bifurcation of generalized differential- algebraic biological economic system with the hybrid functional response and predator Harvesting. *Nonlinear Dynamics and Systems Theory* **23**(4) (2023) 398–409.

[5] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *Journal of differential equations* **31** (1) (1979) 53–98.

[6] P. Chowdhury, M. Banerjee and S. Petrovskii. Coexistence of chaotic and non-chaotic attractors in a three-species slow–fast system. *Chaos, Solitons & Fractals* **167** (2023) 113015.

[7] P. Price, C. Bouton, P. Gross, B. McPheron, J. Thompson and A. Weis. Interactions among three trophic levels: influence of plants on interactions between insect herbivores and natural enemies. *Annual review of Ecology and Systematics* **11** (1) (1980) 41–65.

[8] R. Ouahabi and N. Hamri. *Systèmes dynamiques et chaos.* PhD thesis, Universitè Frères Mentouri-Constantine 1, 2018.

[9] S. Rinaldi and S. Muratori. Slow-fast limit cycles in predator-prey models. *Ecological Modelling* **61** (3-4) (1992) 287–308.

[10] U. R. Kumar and N. R Kamel. Dynamics of a three species food chain model with Crowley–Martin type functional response. *Chaos, solitons & fractals* **42** (3) (2009) 1337–1346.

[11] W. Laouira and N. Hamri. New Design of Stability Study for Linear and Nonlinear Feedback Control of Chaotic Systems. *Nonlinear Dynamics and Systems Theory* **22** (4) (2022) 414–423.

[12] W. Liu. Criterion of Hopf bifurcations without using eigenvalues. *Journal of Mathematical Analysis and Applications* **182** (1) (1994) 250–256.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 594–602

Analysis of Customer Satisfaction Survey on E-Commerce Using Simple Additive Weighting Method

P. Katias [1], T. Herlambang [2], M. Tafrikan [3], N. E. Suharno [4*], K. Oktafianto [5], A. A. Firdaus [6] and H. Arof [7]

[1] *Department of Management, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[2] *Department of Information System, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[3] *Department of Mathematics, Walisongo State Islamic University, Indonesia.*
[4*] *Department of Business, Faculty of Vocational Studies, Airlangga University, Indonesia.*
[5] *Department of Mathematics, University of PGRI Ronggolawe, Indonesia.*
[6] *Department of Engineering, Faculty of Vocational Studies, Airlangga University, Indonesia.*
[7] *Department of Electrical Engineering, University of Malaya, Malaysia.*

Received: February 27, 2024; Revised: November 28, 2024

Abstract: E-commerce is used as a transaction medium for buying and selling in digital form, providing many conveniences. The various types of e-commerce that exist make consumers confused about choosing good quality e-commerce. Therefore, this study aims to recommend determining the best e-commerce. One of the models used in this study is the SAW (Simple Additive Weighting) method because this method can provide an accurate assessment based on the criteria values and preference weights that have been determined by the authors. The SAW method can also choose the best alternative from several existing alternatives. Consideration of the use of this method is based not only on decisions made alone but also on considerations from several previous studies. The results obtained from this study using the SAW (Simple Additive Weighting) method with the highest score for customer satisfaction is Tokopedia with a value of 0,992.

Keywords: *e-commerce; simple additive weighting method.*

Mathematics Subject Classification (2010): 90B50, 68U35.

* Corresponding author: mailto:novianto-edi@vokasi.unair.ac.id

1 Introduction

Indonesia is currently experiencing very rapid technological developments compared to those of several years ago. This is proven by many new innovations emerged in developing technology in Indonesia. The very rapid development of technology also affects daily activities [1]. For example, in the fields of business, health and socio-culture. Online media provide information very quickly because the need for information continues to increase. Therefore, many of us use the internet to access information from information providers [2].

One of the influences of increasingly advanced technology is in transactions in the online shopping or e-commerce business sector. E-Commerce is all activities related to transactions or trades carried out using electronic devices and internet networks and is better known as online commerce or online buying and selling [3], [4]. This activity is one of the activities never separated from daily life because the online buying and selling activities create wider opportunities for traders and buyers, starting from production requests, goods demand up to reachability not only between sub-districts but also between cities, provinces and even between countries [5].

The e-commerce system makes it easier for someone to make online transactions, but behind all the convenience gained, there are also negative things arising from e-commerce, for example, many people have bought products, but when the product reaches the buyer's hands, it does not actually match what is stated in the product information, starting from color, size, to the estimated date of delivery. So, commonly, people are now still confused about which e-commerce company is the best to minimize the worry that comes with online transactions. For that reason, a Decision Support System (DSS) is needed [6], [7]. DSS is a computer-based system that makes it easy to produce an objective decision from several alternatives and interconnected criteria [8].

It is necessary to carry out a selection using a decision support system to help speed up the selection process by algorithmic logic or appropriate methods so that the results obtained have a high level of accuracy. In this research, the selection of the best e-commerce was conducted by applying the SAW method. Based on previous studies, the SAW method has often proven useful to other researchers in completing their investigations. Using the SAW method can provide accurate assessments based on the criteria values and preference weights determined by the researchers. The SAW method can also select the best alternative from several existing alternatives because of the ranking process after determining the weights for each attribute [9], [10], [11].

In the research conducted in [12], a fuzzy logic approach was applied in determining computer specifications for a complete computer package, according to the needs of each buyer, in terms of both brand and fuzzy logic such as processor speed, hard disk capacity, memory capacity, monitor size, power supply size, and VGA size. The results of testing the system, with 10 sample users, showed an accuracy of 68%.

2 Research Method

This research was conducted in Semarang. The method used was Simple Addtive Weighting (SAW).

2.1 The simple additive weighting (SAW)

The SAW (Simple Additive Weighting) method is often called the weighted sum method. The basic concept of the SAW method is to find a weighted sum of performance ratings for each alternative on all attributes. The SAW method requires the process of normalizing the decision matrix (x) to a scale that can be compared with all existing alternative ratings [13].

$$r_{ij} \begin{cases} \frac{x_{ij}}{Max\ x_{ij}}\ if\ j : atribute\ of\ benefit, \\ \frac{Min\ x_{ij}}{x_{ij}}\ if\ j : atribute\ of\ cost, \end{cases} \tag{1}$$

where r_{ij} is the normalized performance rating, Max is the maximum value of each row and column, Min is the minimum of each row and column, x_{ij} are the rows and columns of a matrix.

Here, r_{ij} is the normalized performance rating of alternative A_i on attribute $C_i; i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The preference value for each alternative (V_i) is given as

$$V_i = \sum_{j=1}^{n} w_j r_{ij}. \tag{2}$$

A larger V_i value indicates that alternative A_i is more selected.

2.2 The SAW method procedure

1. Determine the criteria to be used as a reference in decision making, namely C_i.
2. Determine the suitability rating of each alternative for each criterion.
3. Create a decision matrix based on criteria (C_i), then normalize the matrix based on equations adjusted to the type of attribute.
4. The final result is obtained from the ranking process, namely the sum of the multiplication of the normalized matrix R with the weight vector, so that the largest value is selected as the best alternative (A_i) as a solution.

3 Results and Discussion

3.1 Determining alternative

The process of determining alternatives is carried out by giving questionnaires directly to random e-commerce customers in the city of Semarang. And the results obtained are as shown in the following tables.

Table 1 shows the alternative names or e-commerce used in selecting online shopping applications.

3.2 Determining criteria

The criteria used in selecting e-commerce are shown in Table 2.

Alternatives	Codes
Blibli	A_1
Bukalapak	A_2
Lazada	A_3
Shopee	A_4
Tokopedia	A_5

Table 1: Alternatives.

Criteria C_i	Description
C_1	Appearance
C_2	Choice of product/fiture
C_3	Access speed
C_4	Service
C_5	Promo
C_6	Delivery

Table 2: Criteria used to select e-commerce.

Value	Rating Scale
1	Very unsatisfied
2	unsatisfied
3	Fairly satisfied
4	Satisfied
5	Very satisfied

Table 3: Rating scale.

3.3 Rating scale

The researchers provide values/rating scale for all existing alternatives. The rating scale is shown in Table 3.

Next, each criterion with its given weight is shown in Table 4.

Criteria C_i	Description	Weight
C_1	Appearance	10%
C_2	Selected product/fitures	20%
C_3	Access speed	15%
C_4	Service	15%
C_5	Promo	25%
C_6	Delivery	15%

Table 4: Weight criteria.

3.4 Case example

Case example :
The authors will determine which e-commerce is most popular among the public using several criteria, that is, appearance, choice of products/features, speed of access, service, promos and delivery.

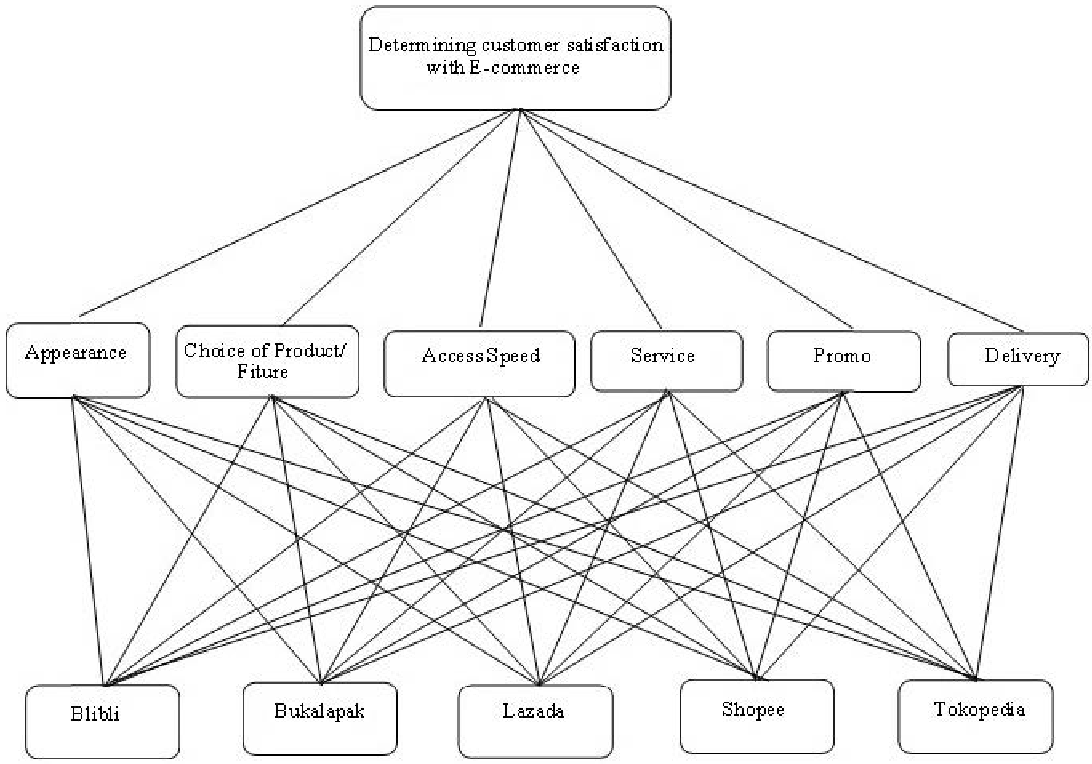


Figure 1: Hierarchy of determining the most preferred e-commerce.

3.5 Application of SAW mehod

The following are the research data used, previously summarized using Microsoft Excel software.

1. Determining the Suitability Rating.
   The next step in determining the suitability rating is shown in Table 5.

2. Determining the Decision Matrix.
   The next step is to form a decision matrix (x) using the suitability rating table for

Alternatives	Average Value					
	C_1	C_2	C_3	C_4	C_5	C_6
A_1	4.1	4	3.9	4.2	3.9	3.6
A_2	4.273	4.182	4.091	4	4.364	4.091
A_3	4.143	3.929	4.071	3.857	3.714	3.571
A_4	4.033	4.067	3.767	4	4.033	4.033
A_5	4.5	4.429	4.5	4.5	4.286	4

Table 5: Suitability rating.

each alternative for each criterion as follows:

$$X = \begin{pmatrix} 4.1 & 4 & 3.9 & 4.2 & 3.9 & 3.6 \\ 4.273 & 4.182 & 4.091 & 4 & 4.364 & 4.091 \\ 4.143 & 3.929 & 4.071 & 3.857 & 3.714 & 3.571 \\ 4.033 & 4.067 & 3.767 & 4 & 4.033 & 4.033 \\ 4.5 & 4.429 & 4.5 & 4.5 & 4.286 & 4 \end{pmatrix}.$$

Next, calculate the normalized value of each alternative using the method in equation (1). It should be noted that researchers here use the benefit attribute because in this research, the criteria determined refer to benefits, not costs.

a. Appearance Criterion (C_1)

$$\begin{aligned} r_{11} &= \frac{4.1}{\max\{4.1; 4.273; 4.143; 4.033; 4.5\}} = \frac{4.1}{4.5} = 0,911, \\ r_{21} &= \frac{4.273}{\max\{4.1; 4.273; 4.143; 4.033; 4.5\}} = \frac{4.273}{4.5} = 0.949. \end{aligned}$$

b. Fiture Criterion (C_2)

$$\begin{aligned} r_{12} &= \frac{4}{\max\{4; 4.182; 3.929; 4.067; 4.429\}} = \frac{4}{4.429} = 0,903, \\ r_{22} &= \frac{4.273}{\max\{4; 4.182; 3.929; 4.067; 4.429\}} = \frac{4.182}{4.429} = 0.944. \end{aligned}$$

c. Access speed Criterion (C_3)

$$\begin{aligned} r_{13} &= \frac{3.9}{\max\{3.9; 4.091; 4.071; 3.767; 4.5\}} = \frac{3.9}{4.5} = 0,867, \\ r_{23} &= \frac{4.091}{\max\{3.9; 4.091; 4.071; 3.767; 4.5\}} = \frac{4.091}{4.5} = 0.909. \end{aligned}$$

d. Service Criterion (C_4)

$$\begin{aligned} r_{14} &= \frac{4.2}{\max\{4.2; 4.3; 3.857; 4; 4.5\}} = \frac{4.2}{4.5} = 0,933, \\ r_{24} &= \frac{4}{\max\{4.2; 4.3; 3.857; 4; 4.5\}} = \frac{4}{4.5} = 0.889. \end{aligned}$$

e. Promo Criterion (C_5)

$$r_{15} = \frac{3.9}{\max\{3.9; 4.364; 3.714; 4.033; 4.286\}} = \frac{3.9}{4.364} = 0,894,$$
$$r_{25} = \frac{4.364}{\max\{3.9; 4.364; 3.714; 4.033; 4.286\}} = \frac{4.364}{4.364} = 1.$$

f. Delivery Criterion (C_6)

$$r_{16} = \frac{3.6}{\max\{3.6; 4.091; 3.571; 4.033; 4\}} = \frac{3.6}{4.091} = 0,879,$$
$$r_{26} = \frac{4.091}{\max\{3.6; 4.091; 3.571; 4.033; 4\}} = \frac{4.091}{4.091} = 1.$$

Then the normalization results are transformed into a normalization matrix, the normalization matrix for this research is as follows:

$$R = \begin{pmatrix} 0.911 & 0.903 & 0.867 & 0.933 & 0.894 & 0.879 \\ 0.949 & 0.944 & 0.909 & 0.889 & 1 & 1 \\ 0.921 & 0.887 & 0.905 & 0.857 & 0.851 & 0.873 \\ 0.896 & 0.918 & 0.837 & 0.889 & 0.924 & 0.986 \\ 1 & 1 & 1 & 1 & 0.982 & 0.978 \end{pmatrix}.$$

3. Ranking.
The final step is to calculate the final preference value (V_i) obtained from the sum of the multiplication of normalized matrix row elements (R) with preference weights (W). The weights used are as follows:

$$W = \{0.10; 0.20; 0.15; 0.15; 0.25; 0.15\}.$$

The formula used is the formula in equation (2),

$$\begin{aligned} V_1 &= (0.10)(0.911) + (0.20)(0.903) + (0.15)(0.867) + (0.15)(0.933) + \\ &\quad (0.25)(0.894) + (0.15)(0.879) = 0.89705\ (\textit{blibli}), \\ V_2 &= (0.10)(0.949) + (0.20)(0.944) + (0.15)(0.909) + (0.15)(0.889) + \\ &\quad (0.25)(1) + (0.15)(1) = 0.9533\ (\textit{bukalapak}), \\ V_3 &= (0.10)(0.921) + (0.20)(0.887) + (0.15)(0.905) + (0.15)(0.857) + \\ &\quad (0.25)(0.851) + (0.15)(0.873) = 0.8775\ (\textit{Lazada}), \\ V_4 &= (0.10)(0.896) + (0.20)(0.918) + (0.15)(0.837) + (0.15)(0.889) + \\ &\quad (0.25)(0.924) + (0.15)(0.986) = 0.911\ (\textit{shopee}), \\ V_5 &= (0.10)(1) + (0.20)(1) + (0.15)(1) + (0.15)(1) + \\ &\quad (0.25)(0.982) + (0.15)(0.978) = 0.9922\ (\textit{tokopedia}). \end{aligned}$$

4. Description of Research Data Analysis Results
Among V_1, V_2, V_3, V_4 and V_5, the highest value is V_5 = Tokopedia with the result of 0,992 from the calculation using the *Simple Additive Weighting* method. It is concluded that Tokopedia is the e-commerce with the highest customer satisfaction based on predetermined criteria. Then the most satisfied criteria or services are C_1 (Appearance), C_3 (Service), and C_4 (Access speed) with a higher average value compared to other criteria or services.

4 Conclusion

4.1 Conclusions

Based on the results of customer satisfaction survey research on e-commerce using the SAW (Simple Additive Weighting) method, several conclusions can be drawn. These conclusions are presented as follows:

1. In terms of the appearance criteria (C_1), Tokopedia has the highest average value with a score of 4.5.
2. In terms of the product/feature choice criteria (C_2), respondents are more satisfied with Tokopedia e-commerce.
3. In terms of the accesses speed criteria (C_3), respondents are more satisfied with Tokopedia e-commerce.
4. In terms of the service criteria (C_4), respondents are more satisfied with Tokopedia, Bukalapak, and Shope e-commerces having the same scores.
5. In terms of the promo criteria (C_5), respondents are more satisfied with Bukalapak e-commerce.
6. In terms of the delivery criteria (C_6), respondents are more satisfied with Bukalapak e-commerce.
7. According to the data obtained by the researchers, the e-commerce with the highest value for customer satisfaction is Tokopedia with a value of 0.992.
8. The e-commerce with the lowest level of customer satisfaction is Lazada, with a value of 0.877.
9. The e-commerce most used by respondents is Shopee with 30 respondents.
10. The e-commerce least used by respondents is Blibli with 10 respondents.
11. The customer satisfaction survey ranking for e-commerce using the SAW (Simple Additive Weighting) method is from top to bottom, respectively, Tokopedia, Bukalapak, Shopee, Blibli, and Lazada.

4.2 Suggestions

Based on the research results, several problems were revealed, so several suggestions were made, these suggestions are as follows:

1. Insufficient number of respondents or less widespread distribution of the g-form.
2. It is suggested that respondents filling out the g-form, receive a prize for the fastest completion or it be drawn randomly after all respondents have completed the g-form.
3. The criteria specified are only a few, they should be added so that respondents can assess e-commerce in more detail.

References

[1] E. A. Tsaniya and R. Sulaiman. Application of Intuitionistic Fuzzy Multisets in Selecting the Best E-commerce. *MATHUnesa: Mathematical Scientific Journal* **9** (3) (2021) 484–493.

[2] D. A. Wulandari and Y. P. Astuti. Selection of the Best E-commerce Using the Intuitionistic Fuzzy Topsis Method. *MATHUnesa : Mathematical Scientific Journal* **8** (2) (2020) 120–129.

[3] A. Wantoro and K. Muludi. Combination of Analytical Hierarchy Process (AHP) and Simple Addive Weight (SAW) Methods to Determine the Best E-commerce Website. *SISTEMASI: Journal of Information Systems* **9** (1) (2020) 131–142.

[4] S. Fauzi and L. F. Lina. The Role of Product Photos, Online Customer Reviews, Online Customer Ratings on Consumer Buying Interest in E-commerce. *Muhammadiyah Journal of Business Management* **2** (1) (2021) 21–26.

[5] Z. Azhar, N. Mulyani, J. Hutahaean and A. Mayhaky. The Best E-commerce Selection Decision Support System Using the MOOSRA Method. *Budidarma Informatics Media* **6** (4) (2022) 2346–2351.

[6] A. Y. Asih, B. Gunawan, N. Hidayati, T. Herlambang, D. Rahmalia and K. Oktafianto. Weights Optimization Using Firefly Algorithm for Dengue Fever Optimal Control Model by Vaccination, Treatment and Abateseae. *Nonlinear Dynamics and Systems Theory* **23** (3) (2023) 237–248.

[7] M. Y. Anshori, I. H. Santoso, T. Herlambang, D. Rahmalia, K. Oktafianto and P. Katias. Forecasting of Occupied Rooms in the Hotel Using Linear Support Vector Machine. *Nonlinear Dynamics and Systems Theory* **23** (2) (2023) 129–140.

[8] J. V. B. Ginting. Application of a Decision Support System in Determining the Best E-commerce Using the SAW Method. *Budidarma Media Informatics Journal* **4** (1) (2020) 225–228.

[9] H. Sucipto. Decision Support System to Majoring High School Student Using SAW Method. *Scientific Journal of Information Systems and Informatics Engineering* **6** (2) (2016) 147–157.

[10] M. Y. Anshori, I. H. Santoso, T. Herlambang, M. Tafrikan, M. Adinugroho, K. Oktafianto and A. A. Firdaus. Analysis of Customer Satisfaction Survey on E-Wallets Using Simple Additive Weighting and TOPSIS Methods. *Nonlinear Dynamics and Systems Theory* **24** (1) (2024) 28–40.

[11] E. Pudjiarti and M. Tabrani. Analysis of Customer Satisfaction Surveys on E-commerce using the Simple Additive Weighting Method. *Scientific Journal of Electronics and Computers* **14** (2) (2021) 286–300.

[12] I. A. Nasution. Decision Support System for Determining Laptop Selection by Applying Fuzzy Tahani. *Pelita Informatika Budi Darma* **6** (1) (2014) 93–96.

[13] S. Kusumadewi, S. Hartati, A. Harjoko and R. Wardoyo. *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM).* Graha Ilmu, Yogyakarta, 2006.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 603–614

A New Generalization of Fuglede's Theorem and Operator Equations

S. Makhlouf [1]* and F. Lombarkia [2]

[1] *Laboratory of Mathematical Techniques (LTM), University of Batna 2, Batna, Algeria.*
[2] *Department of Mathematics, Faculty of Mathematics and Informatics, University of Batna 2, Batna, Algeria.*

Received: February 1, 2024; Revised: December 8, 2024

Abstract: In this paper, the operator equations $AX - XB = C$ and $AXB - X = C$, where A, B, C and X are bounded linear operators on the Hilbert space $\mathcal{H}$, are investigated and criteria of solvability are established. First, in a Hilbertian framework, by extending the famous Fuglede's theorem to a certain class of operators that are not necessarily normal, we show that some classical criteria, as Roth's removal rule for the first equation, remain valid even under assumptions on A and B weaker than usual. Second, in a Banachian framework, we establish our criteria of solvability by using the inner inverses of the operators $\delta_{A,B}$ and $\Delta_{A,B}$ defined on $L(\mathcal{H})$ by $\delta_{A,B}(X) = AX - XB$ and $\Delta_{A,B}(X) = AXB - X$.

Keywords: *Fuglede-Putnam theorem; elementary operators; operator equations; inner inverses.*

Mathematics Subject Classification (2010): 47B47, 47A11, 47A62, 15A09, 93B28.

1 Introduction and Basic Definition

Let $\mathcal{H}$ be an infinite complex Hilbert space and $L(\mathcal{H})$ be the Banach space of all bounded linear operators from $\mathcal{H}$ into $\mathcal{H}$. For $T \in L(\mathcal{H})$, let $\ker(T)$, $\mathcal{R}(T)$, $\sigma(T)$ and $\sigma_p(T)$ stand for the null space, range, spectrum and point spectrum of T, respectively. We recall some definitions of the local spectral theory.

* Corresponding author: mailto:seddik.makhlouf@univ-oeb.dz

Definition 1.1 An operator $T \in L(\mathcal{H})$ is said to have the single-valued extension property at $\lambda_0 \in \mathbb{C}$ (abbreviated SVEP at λ_0) if for every open disc $\mathbb{D}$ centered at λ_0, the only analytic function $f : \mathbb{D} \to X$, which satisfies the equation $(T - \lambda I)f(\lambda) = 0$ for all $\lambda \in \mathbb{D}$, is the function $f \equiv 0$. An operator $T \in L(\mathcal{H})$ is said to have the SVEP if T has the SVEP at every $\lambda \in \mathbb{C}$.

Definition 1.2 An operator $T \in L(\mathcal{H})$ is said to have Bishop's property (β) if for any open subset V of $\mathbb{C}$ and any sequence $\{f_n\}$ of $\mathcal{H}$-valued analytic functions on V, the convergence of $(T - \lambda)f_n(\lambda)$ to zero uniformly on each compact subset of V leads to the convergence of $f_n(\lambda)$ to zero again uniformly on each compact subset of V.

Definition 1.3 An operator $T \in L(\mathcal{H})$ is said to be decomposable if for every open cover $\{U, V\}$ of $\mathbb{C}$, there are T-invariant subspaces $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathcal{H} = \mathcal{X} + \mathcal{Y}$, $\sigma(T|_{\mathcal{X}}) \subset \overline{U}$, and $\sigma(T|_{\mathcal{Y}}) \subseteq \overline{V}$.

The following implications are always satisfied:

$$T \text{ is decomposable} \Rightarrow T \text{ has Bishop's property } (\beta) \Rightarrow T \text{ has the SVEP.}$$

Recall that the ascent $p(T)$ and descent $q(T)$ of T are defined by

$$p(T) = \inf\{n \in \mathbb{N} : \ker T^n = \ker T^{n+1}\},$$

$$q(T) = \inf\{n \in \mathbb{N} : \mathcal{R}(T^n) = \mathcal{R}(T^{n+1})\}$$

with $\inf \emptyset = \infty$. It is well known that if $p(T)$ and $q(T)$ are both finite, then $p(T) = q(T)$. We denote by $\Pi(T) = \{\lambda \in \mathbb{C} : p(T - \lambda I) = q(T - \lambda I) < \infty\}$ the set of poles of the resolvent. In the sequel, we shall denote by $accS$ and $isoS$ the set of accumulation points and the set of isolated points of $S \subset \mathbb{C}$, respectively.

Definition 1.4 We say that $T \in L(\mathcal{H})$ is polaroid if for any isolated point λ in $\sigma(T)$, λ is a pole of the resolvent of T (i.e., $iso\sigma(T) \subseteq \Pi(T)$).

Fuglede's theorem states that if an operator commutes with a normal operator, it also commutes with its adjoint, i.e., if X and A are in $L(\mathcal{H})$ with A normal, then

$$AX = XA \Longrightarrow A^*X = XA^*,$$

this was first proven in 1950 by B. Fuglede [19] and then by C.R.Putnam [23] in a more general version. Thanks to its numerous applications, this theorem has a very effective role in the theory of bounded operators. There are different proofs of this theorem, besides, the first two are due to Fuglede and Putnam, see [20]. Perhaps the most elegant proof is due to Rosenblum [24]. Then, with a wonderful matrix operator trick, S.Berberian [10] showed the equivalence between Fuglede's theorem and that of Putnam. Afterwards, it was called the Fuglede-Putnam theorem. This theorem is therefore stated as follows: if X, A and B are bounded Hilbert space operators such that A and B are normal, then

$$AX = XB \Longrightarrow A^*X = XB^*.$$

This theorem has been extended by relaxing the normality hypotheses on A and B to various classes of non-normal operators. It has also been formulated using the elementary operator $\delta_{A,B}$ as follows: if A and B are normal operators, then $\ker \delta_{A,B} \subset \ker \delta_{A^*,B^*}$, where $\delta_{A,B}$ is the generalized derivation defined on $L(\mathcal{H})$ by $\delta_{A,B}(X) = AX - XB$. The Fuglede-Putnam theorem has a (natural) analogue: if A and B are normal, then $\ker \Delta_{A,B} \subset \ker \Delta_{A^*,B^*}$, where $\Delta_{A,B}$ is the elementary operator defined on $L(\mathcal{H})$ by $\Delta_{A,B}(X) = AXB - X$.

In the following, we will denote by $d_{A,B}$ each of elementary operators $\Delta_{A,B}$ or the generalized derivation $\delta_{A,B}$.

In the second section of this paper, we derive a nice generalization of Fuglede's theorem for decomposable operators $A \in L(\mathcal{H})$ which are polaroid with A and A^* being reduced by each of eigenspaces, using examples of non-normal operators, we justify that the set of such operators strictly contains the normal operators. The third section is devoted to the application of these results to give necessary and sufficient conditions for the existence of solutions to the operator equations $AX - XB = C$ and $AXB - X = C$ in this general framework, which presents a generalization of the results obtained by S. Schweinsberg in [27]. In the last section, independently of the previous ones, using the inner inverse of the elementary operator $d_{A,B}$, we give necessary and sufficient conditions for the existence of solutions to the operator equations $d_{A,B}(X) = C$ and also, the form of these solutions.

2 An Extension of Fuglede's Theorem

In [21], the authors proved the following theorem.

Theorem 2.1 *[21, Theorem 2.2, Theorem 2.3]*
Suppose that $A, B \in L(\mathcal{H})$ satisfies the following conditions:

i) A and B^* are reduced by each of their eigenspaces,

ii) A and B^* are polaroid,

iii) A and B^* have property (β).

Then

$$\ker(d_{A,B} - \lambda I) \subseteq \ker(d_{A^*,B^*} - \overline{\lambda} I)$$

holds for every complex number λ, which means that the Fuglede-Putnam theorem holds.

This theorem is established for many classes of operators, we mention, for example, the operators $A \in L(\mathcal{H})$, which satisfy the equation

$$(A^*)^2A^2 - 2A^*A + I = 0,$$

such A are natural generalizations of isometric operators ($A^*A = I$) and are called 2-isometric operators. It is known that an isometric operator is a 2-isometric operator. 2-isometric operators have been studied by many authors and they have many interesting properties, see [2, 3, 13, 16, 22] for example. In [28, Lemma 2.6], the authors proved that 2-isometric operators have Bishop's property (β) and in [28, Corollary 2.5] they proved that 2-isometric operators are reduced by each of their eigenspaces. In [15, Proposition 2.1], B. P. Duggal proved that 2-isometric operators are polaroid. Then we have the following examples.

Example 2.1 Suppose that A and B^* are 2-isometric operators. Then

$$\ker(d_{A,B} - \lambda I) \subseteq \ker(d_{A^*,B^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}.$$

Example 2.2 Suppose that A and B^* are unilateral weighted shift operators on l_2 defined by $Ae_n = \alpha_n e_{n+1}$ and $B^* e_n = \beta_n e_{n+1}$ for all $n \geq 0$ and such that $\alpha_n^2 \alpha_{n+1}^2 - 2\alpha_n^2 + 1 = 0$ and $\beta_n^2 \beta_{n+1}^2 - 2\beta_n^2 + 1 = 0$ for all $n \geq 0$, where $\{e_n\}_n^\infty$ is a canonical orthogonal basis for l_2 and $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ are bounded sequences of non-negative numbers. Then A and B^* are 2-isometric, it follows that

$$\ker(d_{A,B} - \lambda I) \subseteq \ker(d_{A^*,B^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}.$$

The following theorem forms an interesting generalization of Fuglede's theorem to a set larger than that of the normal operators.

Theorem 2.2 *Suppose that $A \in L(\mathcal{H})$ satisfies the following conditions:*

i) A is decomposable,

ii) A is polaroid,

iii) A and A^* are reduced by each of their eigenspaces ($\ker(A - \lambda I) = \ker(A^* - \overline{\lambda I})$, $\forall \lambda \in \sigma_p(A)$).

Then $\ker(\delta_{A,A} - \lambda I) \subseteq \ker(\delta_{A^*,A^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}.$

Proof. If A is decomposable, it follows that A and A^* have property (β), on the other hand, it is well known that A is polaroid if and only if A^* is polaroid. Then we obtain the result.

We note that normal operators A on a Hilbert space are decomposable, polaroid, A and A^* are reduced by each of their eigenspaces. We note also that the class of operators A which are decomposable, polaroid, A and A^* are reduced by each of their eigenspaces, contains strictly normal operators. Since E. Albrecht in [6, Proposition 5.1] constructed a non normal, subnormal operator S which is decomposable and since subnormal operators (their adjoint too) are hyponormal, it follows that S is polaroid, S and S^* are reduced by each of their eigenspaces. Another interesting class of bounded operators from which the conditions of the previous theorem are satisfied, is the class of compact p-symmetric operators. Now we recall the definition of p-symmetric operators.

Definition 2.1 [11, Definition 1.2] Let $A \in L(\mathcal{H})$, where $\mathcal{H}$ is a separable complex Hilbert space. A is called p-symmetric if $AT = TA$ implies $A^*T = TA^*$ for all trace class operators T.

Proposition 2.1 *Let $A \in L(\mathcal{H})$, where $\mathcal{H}$ is a separable complex Hilbert space. If A is compact and p-symmetric, then A is decomposable polaroid, A and A^* are reduced by each of their eigenspaces. Therefore* $\ker(d_{A,A} - \lambda I) \subseteq \ker(d_{A^*,A^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}.$

Proof. It is well known that compact operators are decomposable, and from [9, Corollary V.10.3], compact operators are polaroid. We have A is compact, it follows that if $\lambda \in \sigma_p(A)$, then $\overline{\lambda} \in \sigma_p(A^*)$, since A is p-symmetric, then from [11], we deduce that A is reduced by each of its eigenspaces. Since A^* is also compact and p-symmetric, we deduce that A^* is reduced by each of its eigenspaces.

Now we give another example which satisfies the conditions of Theorem 2.2. In [7], S.A. Alzraiqi and A.B. Patel introduced the class of n-normal operators, we recall that an operator $A \in L(\mathcal{H})$ is said to be an n-normal operator if $A^n A^* = A^* A^n$.

Example 2.3 Let $A \in L(\mathcal{H})$ such that A is 2-normal, $\sigma(A) \cap (-\sigma(A)) \subset \{0\}$ and $kerA = kerA^*$, then from [14, Theorem 4.4], A is decomposable and from [14, Theorem 2.3], A is polariod and it is reduced by each of its eigenspaces. Since A^* is also 2-normal, then A^* is polariod and it is reduced by each of its eigenspaces and it follows that $\ker(d_{A,A} - \lambda I) \subseteq \ker(d_{A^*,A^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}$.

Example 2.4 Let $A = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}$ be an operator acting in a two-dimensional complex Hilbert space. Then A is 2-normal, $\sigma(A) \cap (-\sigma(A)) \subset \{0\}$ and $kerA = kerA^*$, then A is decomposable, polaroid and A and A^* are reduced by each of their eigenspaces. Then

$$\ker(d_{A,A} - \lambda I) \subseteq \ker(d_{A^*,A^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}.$$

3 Solvability Criteria for the Equation $d_{A,B}(X) = C$ in a Hilbertian Framework

Mathematicians often try to find suitable solutions to problems in a wide range of fields by using various methods, and to study the properties of solutions such as existence, uniqueness, stability, and so on. See, for example, [4] and [5]. The previous results are very useful for solving the equation $d_{A,B}(X) = C$ in a more general setting. Let us first recall that in [26], W. E. Roth proved for finite matrices over a field that $AX - XB = C$ is solvable for X if and only if the matrices $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ are similar. A considerably briefer proof has been given by Flanders and Wimmer [18]. In [25], Rosenblum showed that the result remains true when A and B are bounded self-adjoint operators on a complex separable Hilbert space. In [27], A. Schweinsberg extended the result to include finite rank operators and normal operators on a Hilbert space. In this part, we generalize it to the operators $A, B \in L(\mathcal{H})$ satisfying the conditions given below.

Theorem 3.1 *Suppose that $A, B \in L(\mathcal{H})$ satisfy the following conditions:*

i) A^* have property (β), and B is decomposable.

ii) A^* and B are polaroid,

iii) A^*, B and B^* are reduced by each of their eigenspaces.

Then the operator equation $AX - XB = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ are similar.

Proof. If the equation $AX - XB = C$ has a solution X, then

$$\begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}.$$

Hence $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ are similar.

Suppose that $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ are similar, then there exists an invertible operator $\begin{pmatrix} Q & R \\ S & T \end{pmatrix}$ such that $\begin{pmatrix} Q & R \\ S & T \end{pmatrix}\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}\begin{pmatrix} Q & R \\ S & T \end{pmatrix}$, this implies that

$$QA - AQ = CS, RB - AR = CT$$
$$SA = BS, TB = BT. \tag{3.1}$$

We apply Theorem 2.2 above, we get

$$\ker(d_{B,B} - \lambda I) \subseteq \ker(d_{B^*,B^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C},$$

also from [21, Theorem2.2], we obtain

$$\ker(d_{B,A} - \lambda I) \subseteq \ker(d_{B^*,A^*} - \overline{\lambda} I), \quad \forall \lambda \in \mathbb{C}.$$

Thus, the equality (3.1) gives

$$SA^* = B^*S, \quad TB^* = B^*T, \tag{3.2}$$

and by taking the adjoint in (3.2), we have $AS^* = S^*B, \quad BT^* = T^*B$,
which ensures that B commutes with SS^* and TT^*. We have also

$$C(SS^* + TT^*) = (QS^* + RT^*)B - A(QS^* + RT^*).$$

We apply the result from [27, Lemma 1], we deduce that there exists $X = -(QS^* + RT^*)(SS^* + TT^*)^{-1}$, which is the solution to the operator equation $AX - XB = C$.

Corollary 3.1 *Suppose that $A, B \in L(\mathcal{H})$ such that $\mathcal{H}$ is a separable Hilbert space. If*

i) A^* is 2-isometric,

ii) B is compact and p-symetric,

then the operator equation $AX - XB = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ are similar.

If we set $B = A$ in Theorem 3.1, we get the following corollary.

Corollary 3.2 *Let $A \in L(\mathcal{H})$ satisfy*

i) A is decomposable and polaroid,

ii) A and A^* are reduced by each of their eigenspaces.

Then the operator equation $AX - XA = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & A \end{pmatrix}$ are similar.

Corollary 3.3 *Let $A \in L(\mathcal{H})$ such that $\mathcal{H}$ is a separable Hilbert space. If A is compact and p-symmetric, then the operator equation $AX - XA = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & A \end{pmatrix}$ are similar.*

As a consequence of Corollary 3.2, we obtain a well known theorem of A. Schweinsberg [27, Theorem 1]:

Corollary 3.4 *[27, Theorem 1] Let $A \in L(\mathcal{H})$ be a normal operator. Then the operator equation $AX - XA = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & A \end{pmatrix}$ are similar.*

Example 3.1 Let A^* be a 2-isometric operator and B be a 2-normal operator on a Hilbert space $\mathcal{H}$, $\sigma(A) \cap (-\sigma(A)) \subset \{0\}$ and $ker A = ker A^*$, then the equation $AX - XB = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ are similar.

Example 3.2 Let $A \in L(\mathcal{H})$ such that A is 2-normal, $\sigma(A) \cap (-\sigma(A)) \subset \{0\}$ and $ker A = ker A^*$. Then the operator equation $AX - XA = C$ has a solution if and only if $\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$ and $\begin{pmatrix} A & C \\ 0 & A \end{pmatrix}$ are similar.

We also get similar results for the equation $AXB - X = C$.

Theorem 3.2 *Let $A, B \in L(\mathcal{H})$ such that*

i) A is decomposable and polaroid.

ii) A and A^* are reduced by each of their eigenspaces.

iii) B has property (β), is polaroid and reduced by each of its eigenspaces.

Then the equation $AXB - X = C$ has a solution in $L(\mathcal{H})$ if and only if there exist two invertible operators U and V such that $U\begin{pmatrix} A & C \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}V$ and $U\begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}V$.

Proof. If X is a solution of $AXB - X = C$, then $AXB = C + X$.
Let $U = \begin{pmatrix} I & X \\ O & I \end{pmatrix}$ and $V = \begin{pmatrix} I & XB \\ O & I \end{pmatrix}$, it is clear that U and V are invertible, in addition, we have

$$U\begin{pmatrix} A & C \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}V \text{ and } U\begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}V.$$

Conversely, assume that there exist two invertible operators
$U = \begin{pmatrix} Q & R \\ S & T \end{pmatrix}$ and $V = \begin{pmatrix} Q_1 & R_1 \\ S_1 & T_1 \end{pmatrix}$ such that
$U\begin{pmatrix} A & C \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}V$ and $U\begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}V$,
so

$$\begin{cases} QA = AQ_1, & (1) \\ SA = S_1, & (2) \\ QC + R = AR_1, & \\ SC + T = T_1. & (3) \end{cases} \quad \text{and} \quad \begin{cases} Q = Q_1, & (4) \\ RB = R_1, & \\ S = BS_1, & (5) \\ TB = BT_1. & \end{cases}$$

From (1) and (4), we have $AQ = QA$, then, according to Theorem 2.2 and by taking the adjoint, we have $AQ^* = Q^*A$, consequently, we get $AQ^*Q = Q^*QA$. From (2) and (5), we have $BSA = S$; knowing that if A is decomposable, then A^* has property (β), moreover, if A is polaroid, then A^* is too, which allows us to obtain from [21, Theorem 2.3] that $B^*SA^* = S$, and by taking the adjoint, we get $AS^*B = S^*$, which implies that $S^*SA = (AS^*B)SA = (AS^*B)S_1 = AS^*S$. Therefore, A commutes with the sum Q^*Q+S^*S and so with the inverse $(Q^*Q + S^*S)^{-1}$, which exists according to [27, Lemma 1].

In addition, from (3), we have $S^*SC = S^*T_1 - S^*T = A(S^*T)B - S^*T$. Therefore

$$\begin{aligned}(Q^*Q + S^*S)\,C &= Q^*(AR_1 - R) + A(S^*T)B - S^*T,\\ &= Q^*ARB - Q^*R + A(S^*T)B - S^*T,\\ &= A(Q^*R + S^*T)B - (Q^*R + S^*T),\end{aligned}$$

and so

$$\begin{aligned}C &= (Q^*Q + S^*S)^{-1}\, A(Q^*R + S^*T)B - (Q^*Q + S^*S)^{-1}\,(Q^*R + S^*T),\\ &= A\,(Q^*Q + S^*S)^{-1}\,(Q^*R + S^*T)B - (Q^*Q + S^*S)^{-1}\,(Q^*R + S^*T),\end{aligned}$$

which means that $X = (Q^*Q + S^*S)^{-1}\,(Q^*R + S^*T)$ is a solution of the equation $\Delta_{A,B}(X) = C$, and the proof is complete.

Corollary 3.5 *Suppose that $A, B \in L(\mathcal{H})$ such that $\mathcal{H}$ is a separable Hilbert space. If*

i) A is compact and p-symmetric,

ii) B is 2-isometric,

then the operator equation $AXB - X = C$ has a solution if and only if there exist two invertible operators U and V such that $U\begin{pmatrix} A & C\\ 0 & I\end{pmatrix} = \begin{pmatrix} A & 0\\ 0 & I\end{pmatrix}V$ *and* $U\begin{pmatrix} I & 0\\ 0 & B\end{pmatrix} = \begin{pmatrix} I & 0\\ 0 & B\end{pmatrix}V$.

If we set $B = A$ in Theorem 3.2, we get the following corollary.

Corollary 3.6 *Let $A \in L(\mathcal{H})$ satisfy*

i) A is decomposable and polaroid,

ii) A and A^* are reduced by each of their eigenspaces.

Then the operator equation $AXA^* - X = C$ has a solution if and only if there exist two invertible operators U and V such that $U\begin{pmatrix} A & C\\ 0 & I\end{pmatrix} = \begin{pmatrix} A & 0\\ 0 & I\end{pmatrix}V$ *and* $U\begin{pmatrix} I & 0\\ 0 & A^*\end{pmatrix} = \begin{pmatrix} I & 0\\ 0 & A^*\end{pmatrix}V$.

Corollary 3.7 *Suppose that $A \in L(\mathcal{H})$ such that $\mathcal{H}$ is a separable Hilbert space. If A is compact and p-symmetric, then the operator equation $AXA^* - X = C$ has a solution if and only if there exist two invertible operators U and V such that* $U\begin{pmatrix} A & C\\ 0 & I\end{pmatrix} = \begin{pmatrix} A & 0\\ 0 & I\end{pmatrix}V$ *and* $U\begin{pmatrix} I & 0\\ 0 & A^*\end{pmatrix} = \begin{pmatrix} I & 0\\ 0 & A^*\end{pmatrix}V$.

Example 3.3 Let $A \in L(\mathcal{H})$ such that A is 2-normal, $\sigma(A) \cap (-\sigma(A)) \subset \{0\}$ and $kerA = kerA^*$. Then the operator equation $AXA^* - X = C$ has a solution if and only if there exist two invertible operators U and V such that $U \begin{pmatrix} A & C \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} V$ and $U \begin{pmatrix} I & 0 \\ 0 & A^* \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & A^* \end{pmatrix} V$.

4 Solvability Criteria for the Equation $d_{A,B}(X) = C$ in a Banachian Framework

Let E be a Banach space, and $d_{A,B} \in L(L(E))$, in this section, we give necessary and sufficient conditions for regularity of the elementary operator $d_{A,B}$, then we deduce necessary and sufficient conditions for the existence of solutions to the operator equations $d_{A,B}(X) = C$, using the inner inverses of the elementary operator $d_{A,B}$. First, we recall the following definitions.

Definition 4.1 Let $A \in L(E)$. An operator $B \in L(E)$ is said to be an inner inverse of A if it satisfies the equation

$$ABA = A.$$

We denote the inner inverse by A^-. An operator with an inner inverse will be called regular.

Remark 4.1 We note that

1. $A \in L(E)$ has an inner inverse if and only if $\ker(A)$ and $\mathcal{R}(A)$ are closed and complemented subspaces of E.

2. If A has an inverse A^{-1} in $L(E)$, then A^{-1} is the only inner inverse of A.

Theorem 4.1 *Suppose that $A, B \in L(E)$ are polaroid, $p(\delta_{A,B}) \leq 1$ and $\delta^*_{A,B}$ has the SVEP at 0, then the following conditions are pairwise equivalent:*

1. *$\delta_{A,B}$ has a closed range,*

2. *$L(E) = \ker(\delta_{A,B}) \oplus \mathcal{R}(\delta_{A,B})$,*

3. *$0 \in iso\sigma(\delta_{A,B})$,*

4. *$\delta_{A,B}$ is regular.*

Proof. The equivalences $1 \Leftrightarrow 2 \Leftrightarrow 3$ have been proven by the authors in [17, Theorem 3.2]. The condition (4) is equivalent to (1). Indeed, if $\delta^-_{A,B}$ is an inner inverse of $\delta_{A,B}$, then $\delta_{A,B}\delta^-_{A,B}\delta_{A,B}\delta^-_{A,B} = \delta_{A,B}\delta^-_{A,B}$, i.e., $\delta_{A,B}\delta^-_{A,B}$ is a projection on the closed subspace $\mathcal{R}(\delta_{A,B}\delta^-_{A,B})$. Moreover, $\mathcal{R}(\delta_{A,B}) = \mathcal{R}(\delta_{A,B}\delta^-_{A,B}\delta_{A,B}) \subseteq \mathcal{R}(\delta_{A,B}\delta^-_{A,B}) \subseteq \mathcal{R}(\delta_{A,B})$, so $\mathcal{R}(\delta_{A,B}\delta^-_{A,B}) = \mathcal{R}(\delta_{A,B})$, and it is therefore closed. Conversely, if $\mathcal{R}(\delta_{A,B})$ is closed, then $P_{\mathcal{R}(\delta_{A,B})}$ is a bounded linear operator and, by the Douglas theorem, the equation $\delta_{A,B}X = P_{\mathcal{R}(\delta_{A,B})}$ admits a solution; that is, there exists B in $L(E)$ such that $\delta_{A,B}B = P_{\mathcal{R}(\delta_{A,B})}$. Then $\delta_{A,B}B\delta_{A,B} = \delta_{A,B}$ and therefore $\delta_{A,B}$ has an inner inverse.

Corollary 4.1 *Suppose that $A, B \in L(E)$ are polaroid, $p(\delta_{A,B}) \leq 1$ and $\delta^*_{A,B}$(the dual of $\delta_{A,B}$) has the SVEP at 0 and $C \in L(E)$. If $0 \in iso\sigma(\delta_{A,B})$, then the operator equation $\delta_{A,B}(X) = C$ has a solution if and only if*

$$\delta_{A,B}\delta^-_{A,B}C = C.$$

In this case, the general solution is

$$X = \delta^-_{A,B}C + (I_{L(E)} - \delta^-_{A,B}\delta_{A,B})U,$$

where $U \in L(E)$ is an arbitrary operator.

Proof. We apply Theorem 4.1, we deduce that $\delta_{A,B}$ is regular, and from [12], we get the result.

Corollary 4.2 *Suppose that $A, B \in L(\mathcal{H})$ are normal operators and $C \in L(\mathcal{H})$. If $0 \in iso\sigma(\delta_{A,B})$, then the operator equation $\delta_{A,B}(X) = C$ has a solution if and only if*

$$\delta_{A,B}\delta^-_{A,B}C = C.$$

In this case, the general solution is

$$X = \delta^-_{A,B}C + (I_{L(\mathcal{H})} - \delta^-_{A,B}\delta_{A,B})U,$$

where $U \in L(\mathcal{H})$ is an arbitrary operator.

Proof. If A and B are normal operators, it follows that A and B are polaroid, $p(\delta_{A,B} - \lambda) \leq 1$ holds for every complex number λ, and $\delta^*_{A,B}$ has the SVEP at 0. Hence the result follows from Theorem 4.1 and Corollary 4.1.

Theorem 4.2 *Suppose that $A, B \in L(E)$ are contractions, then the following conditions are pairwise equivalent:*

1. *$\Delta_{A,B}$ has a closed range,*
2. *$L(E) = \ker(\Delta_{A,B}) \oplus \mathcal{R}(\Delta_{A,B})$,*
3. *$0 \in iso\sigma(\Delta_{A,B})$,*
4. *$\Delta_{A,B}$ is regular.*

Proof. The equivalences $1 \Leftrightarrow 2 \Leftrightarrow 3$ have been proven by the authors in [17, Theorem 3.2], and in the same way as in Theorem 4.1, we show that (4) is equivalent to (1).

Corollary 4.3 *Suppose that $A, B \in L(E)$ are contractions and $C \in L(E)$. If $0 \in iso\sigma(\Delta_{A,B})$, then the operator equation $\Delta_{A,B}(X) = C$ has a solution if and only if $\Delta_{A,B}\Delta^-_{A,B}C = C$. In this case, the general solution is $X = \Delta^-_{A,B}C + (I_{L(E)} - \Delta^-_{A,B}\Delta_{A,B})U$, where $U \in L(E)$ is an arbitrary operator.*

5 Conclusion

Many researchers have focused on studying equations of the form $AX - XB = C$ and $AXB - X = C$ due to their significance in solving various problems in many fields such as physics, biology, economics, etc. They have achieved considerable results in this regard.

This work is part of the same context where we presented, in the first part, important results, represented by the provision of the necessary and sufficient conditions for these equations to have solutions in the Hilbertian framework, through an important extension of the theorem of Fugleg, while giving the general form of expression of these solutions.

These results represent a natural and important extension of many previously known results to much broader classes of operators than usual. Examples of applications have been included, as well as some corollaries of these results.

In the second part, within the Banachian framework and using generalized inverse operators, we provided the necessary and sufficient conditions for these equations to have solutions, as well as the expression in general form of these solutions, and also their important implications.

Acknowledgment

The authors thank the referees for their careful reading and their precious comments. Their help is much appreciated.

References

[1] J. Agler. A disconjugacy theorem for Toeplitz operators. *Amer. J. Math.* **112** (1990) 1–14.

[2] J. Agler and M. Stankus. M-isometric transformations of Hilbert space. I. *Integral Equ. Oper. Theory* **21** (1995) 383–429.

[3] J. Agler and M. Stankus. M-isometric transformations of Hilbert space. II .*Integral Equ. Oper. Theory* **23** (1995) 1–48.

[4] S. Ait Temghart, H. El Hammar, C. Allalou and K. Hilal. Existence of Weak Solutions for a Class of (p(b(u)),q(b(u)))-Laplacian Problems. *Nonlinear Dynamics and Systems Theory* **23** (1) (2023) 107–118.

[5] Y. Akdim, S. Lalaoui Rhali and Y. Oumouacha. Solvability of Nonlinear Elliptic Problems with Degenerate Coercivity in Weighted Sobolev Space. *Nonlinear Dynamics and Systems Theory* **23** (5) (2023) 461–474.

[6] E. Albrecht. Generalized spectral operators. *North-Holland Mathematics Studies. North-Holland.* **27** (1977) 259–277.

[7] S. A. Alzuraiqi and A. B. Patel. On n-normal operators. *General Math. Notes* **1** (2010) 61–73.

[8] M.L. Arias, G. Corach and M.C. Gonzalez. Generalized inverses and Douglas equations. *Proc. Amer. Math. Soc.* **136** (2008) 3177–3183.

[9] E. T. Angus and D. C. Lay. *Introduction to functional Analysis.* New York, Wiley 1980.

[10] S. K. Berberian. Extensions of a theorem of Fuglede and Putnam. *Proc. Amer. Math. Soc.* **71** (1978) 113-114.

[11] S. Bouali and J. Charles. Extension de la notion d'operateur D-symetrique. I. *Acta Scientiarum Mathematicarum* **58** (1) (1994) 517-526.

[12] S. R. Caradus. *Generalised Inverse and operator Theory*, Queen's Paper in Pure and Appl Math, Quenn's Univ. Kingston, 1978.

[13] M. Chō, S. Ota, K. Tanahashi and A. Uchiyama. Spectral properties of m-isometric operators. *Funct. Anal. Approx. Comput.* **4** (2012) 33–39.

[14] M. Chō, J. E. Lee, K. Tanahashi and A. Uchiyama. Remarks on n-normal operators. *Filomat* **32** (15) (2018) 5441–5451.

[15] B. P. Duggal. Tensor product of n-isometries. *Linear algebra and its applications* **437** (1) (2012) 307–318.

[16] B. P. Duggal. An elementary operator with log-hyponormal, p-hyponormal entries. *Linear algebra and its applications* **437** (2008) 1109–1116.

[17] B. P. Duggal, S. D. Djordjevic and C. S. Kubrusly. Elementary operators, finite ascent, range clodure and compactness. *Linear algebra and its applications* **449** (2014) 334–340.

[18] H. Flanders and H.K. Wimmer. On the matrix equations $AX - XB = G$ *and* $AX - YB = C^*$, *SIAM J. Appl. Math.* **32** (4) (1977) 707–710.

[19] B. Fuglede. A commutativity theorem for normal operators. *PNAS* **36** (1) (1950) 35–40.

[20] P. R. Halmos. Commutativity and spectral properties of normal operators. *Acta Sci. Math. Szeged. Leopoldo Fejér Frederico Riesz LXX annos natis dedicatus, Pars B* **12** (1950) 153–156.

[21] F. Lombarkia and M. Amouch. Asymetric Fuglede Putnam's Theorem for operators reduced by their eigenspaces. *Filomat* **20** (31) (2017) 6409–6417.

[22] S. M. Patel. 2-isometry operators. *Glasnik Mat.* **37** (2002) 143–147.

[23] C. R. Putnam. On normal operators in Hilbert space. *Amer. J. Math.* **73** (1951) 357–362.

[24] M. Rosenblum. On a theorem of Fuglede and Putnam. *J. Lond. Math. Soc.* **33** (1958) 376–377.

[25] M. Rosenblum. On the opertor equation $AX - XB = Q$ *with self adjoint A,B, Proc. Amer. Math. Soc.* **20** (1969) 115–120.

[26] W. E. Roth. The eqution $AX - YB = C$ and $AX - XB = C$ in matrices. *Proc. Amer. Math. Soc.* **3** (1952) 392–396.

[27] A. Schweinsberg. The operator eqution $AX - XB = C$ with normal A and B. *Pacific Journal of Mathematics* **102** (2) (1982) 447–453.

[28] J. Shen and J. Guoxing. On an elementary operator with 2-isometric operator entries. *Filomat* **14** (2) (2018) 5083–5088.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 615–624

Boundedness in Nonlinear Oscillatory Systems over a Given Time Interval

A. A. Martynyuk, T. O. Lukyanova * and Yu. A. Martynyuk-Chernienko

S. P. Tymoshenko Institute of Mechanics, National Academy of Sciences of Ukraine, Ukraine

Received: December 20, 2023; Revised: December 12, 2024

Abstract: This paper considers three types of perturbed motion equations with a stable linear (nonlinear) approximation. New sufficient conditions are established for the boundedness of motion on a finite interval with respect to a given Lyapunov function. The conditions are obtained on the basis of the direct Lyapunov method and the method of integral inequalities.

Keywords: *equations of perturbed motion; stable approximation; boundedness with respect to given function.*

Mathematics Subject Classification (2010): 34D40, 34D20, 70K40.

1 Introduction

Non-autonomous systems of equations, applicable in nonlinear mechanics [1], are studied by various methods (see [2–7] and the bibliography therein). The Lyapunov function method [8], combined with the method of integral inequalities (see [1, 9]), allows establishing new conditions for the boundedness of motion over a specified time interval. This paper is structured as follows.

Section 2 discusses a system of two scalar equations with nonlinear stable approximation. Definitions of motion boundedness with respect to a positive definite function are provided.

In Section 3, an estimation of the Lyapunov function is established.

Section 4 presents conditions for the boundedness of motion with respect to a positive definite function.

Section 5 addresses the problem of boundedness of solutions to equation systems with autonomous stable approximation.

In Section 6, conditions for boundedness are established in the case of stability of non-autonomous linear approximation.

Section 7 provides conditions for the boundedness of solutions over a specified time interval for perturbed motion equations in the normal Cauchy form.

The concluding section offers comments on the obtained results.

* Corresponding author: mailto:lukianovatetyana@gmail.com

2 Formulation of the Problem

Consider a system of perturbed motion equations in the form

$$\frac{dx}{dt} = f(t,x,y) + \sum_{i=1}^{m} \mu^i X_i(t,x,y) + \sum_{i=0}^{\infty} \mu^i \varphi_i(t),$$

$$\frac{dy}{dt} = g(t,x,y) + \sum_{i=1}^{m} \mu^i Y_i(t,x,y) + \sum_{i=0}^{\infty} \mu^i \psi_i(t), \tag{1}$$

where $t \in R_\tau$, $x, y \in D \subset \mathbb{R}$, $0 < \mu^i < \mu_0$ is a small parameter, $f : R_\tau \times D \times D \to R$, $g : R_\tau \times D \times D \to R$. The coefficients of the polynomials X_i and Y_i, and the functions φ_i and ψ_i are bounded functions of time $t \in R_\tau$, where τ is a finite number or the symbol $+\infty$.

Together with the systems of equations ((1) and others), we will consider a positive definite continuously differentiable Lyapunov function $V(t,x,y)$ and its total derivative along the solutions of system (1) and other systems of equations investigated in this paper.

Taking into account certain results from [10,11], we provide the following definitions.

Definition 2.1 A solution $(x(t), y(t))^T$ of system (1) is called bounded for given $t_0 \geq 0$ and $\beta > 0$ with respect to the positive definite function $V(t,x,y)$ on the interval R_τ if from the condition $V(t_0, x_0, y_0) = \beta^* \leq \beta$, it follows that $V(t, x(t), y(t)) \leq \beta$ for all $t \in R_\tau$.

Definition 2.2 A solution $(x(t), y(t))^T$ of system (1) is called bounded on a given interval for a given $t_0 \in R_\tau$ if there exists a positive number $\tau > 0$ and a positive definite function $V(t,x,y)$ such that with respect to it, the solution $(x(t), y(t))^T$ of system (1) is bounded on the finite interval R_τ.

Let us obtain conditions for the boundedness of solutions of system (1) in the sense of Definitions 2.1 and 2.2.

3 Estimation of the Lyapunov Function on Solutions of System (1)

For the Lyapunov function $2V_1(x,y) = x^2 + y^2$, let us compute the total derivative with respect to time:

$$\frac{d}{dt}V_1(x,y) = x\Big(f(t,x,y) + \sum_{i=1}^{m} \mu^i X_i(t,x,y) + \sum_{i=0}^{\infty} \mu^i \varphi_i(t)\Big) +$$

$$+ y\Big(g(t,x,y) + \sum_{i=1}^{m} \mu^i Y_i(t,x,y) + \sum_{i=0}^{\infty} \mu^i \Psi_i(t)\Big). \tag{2}$$

Suppose there exist a non-negative function $a_1(t,\mu)$ and a continuous function $a_2(t,\mu)$, as well as values $\mu_1 \in (0,\mu_0]$, $\mu_2 \in (0,\mu_0]$ such that

$H_1.$: $xf(t,x,y) + yg(t,x,y) \leq 0$ for all $t \in \mathbb{R}_\tau, (x,y) \in D \times D$;

$H_2.$: $x \sum\limits_{i=1}^{m} \mu^i X_i(t,x,y) + y \sum\limits_{i=1}^{m} \mu^i Y_i(t,x,y) \leq a_1(t,\mu)(x^2 + y^2)$ for $\mu < \mu_1$;

$H_3.\ :\ x\sum\limits_{i=0}^{\infty}\mu^i\varphi_i(t)+y\sum\limits_{i=0}^{\infty}\mu^i\Psi_i(t)\le a_2(t,\mu)$, for $(x,y)\in D\times D$ for $\mu<\mu_2$.

Let us show that the following statement holds.

Lemma 3.1 *If conditions* H_1–H_3 *are satisfied for system* (1) *and the function* $V_1(x,y)$*, then*

$$V_1(x(t),y(t))\le V_1(x_0,y_0)\exp\Big(\int\limits_0^t a_1(s,\mu)ds\Big)+\int\limits_{t_0}^t\exp\Big[\int\limits_\tau^t a_1(s,\mu)ds\Big]a_2(\tau,\mu)d\tau \tag{3}$$

for all $t\in R_\tau$ *and* $0<\mu<\min(\mu_1,\mu_2)$*, where* $x_0=x(t_0),\ y_0=y(t_0)$.

Proof. From conditions H_1–H_3 and relation (2), it follows that

$$\frac{d}{dt}V_1(x(t),y(t))\le a_1(t,\mu)V_1(x(t),y(t))+a_2(t,\mu)$$

for all $t\in R_\tau$ and $0<\mu<\min(\mu_1,\mu_2)$.

Let us compute the derivative with respect to time of the product of two functions:

$$\frac{d}{dt}\Big\{V_1(x(t),y(t))\exp\Big[-\int\limits_{t_0}^t a_1(s,\mu)ds\Big]\Big\}=$$

$$=\Big\{\frac{d}{dt}V_1(x(t),y(t))-a_1(t,\mu)V_1(x(t),y(t))\Big\}\exp\Big[-\int\limits_{t_0}^t a_1(s,\mu)ds\Big]. \tag{4}$$

From equation (4), upon integration from t_0 to t, we obtain

$$V_1(x(t),y(t))\exp\Big[-\int\limits_{t_0}^t a_1(s,\mu)ds\Big]-V_1(x_0,y_0)=$$

$$=\int\limits_{t_0}^t\Big[\frac{d}{dt}V_1(x(\tau),y(\tau))-a_1(\tau,\mu)V(x(\tau),y(\tau))\Big]\times$$

$$\times\Big[-\int\limits_{t_0}^t a_1(s,\mu)ds\Big]d\tau\le\int\limits_{t_0}^t a_2(\tau,\mu)\Big[\int\limits_\tau^{t_0} a_1(s,\mu)ds\Big]d\tau. \tag{5}$$

From inequality (5), we deduce the estimate (3). □

4 Conditions for the Boundedness of Solutions to System (1)

The estimate (3) allows us to establish the following conditions for the boundedness of solutions to system (1).

Theorem 4.1 *To ensure that the solution* $(x(t), y(t))^T$ *of system* (1) *is bounded on a given interval with respect to the function* $V_1(x, y)$*, it suffices that conditions* H_1*–*H_3 *hold, and if* $V_1(x_0, y_0) = \beta^* < \beta$*, the following estimate holds:*

$$\exp\Big(\int\limits_0^t a_1(s,\mu)ds\Big) + \frac{1}{\beta^*}\int\limits_0^t a_2(s,\mu)\exp\Big(\int\limits_s^t a_1(\tau,\mu)d\tau\Big)ds \le \frac{\beta}{\beta^*} \tag{6}$$

for all $t \in R_\tau$ *and* $0 < \mu < \min(\mu_1, \mu_2)$.

Proof. From estimate (3) under condition (6), we obtain that $V_1(x(t), y(t)) \le \beta$ for all $t \in R_\tau$. This, according to Definition 2.1, proves the statement of Theorem 4.1. □

5 System of Equations with Autonomous Stable Approximation

We consider a system of perturbed motion equations in the form (see [12])

$$\frac{dy_s}{dt} = -\lambda_s z_s + \sum_{i=1}^{\infty}\mu^i Y_{s_i}(t, x, z) + \sum_{i=0}^{\infty}\mu^i \varphi_{s_i}(t),$$

$$\frac{dz_s}{dt} = \lambda_s y_s + \sum_{i=1}^{\infty}\mu^i Z_{s_i}(t, x, z) + \sum_{i=0}^{\infty}\mu^i \psi_{s_i}(t), \ s = 1, 2, \dots, n. \tag{7}$$

In system (7), the coefficients of the polynomials Y_{s_i} and Z_{s_i}, as well as the functions $\varphi_{s_i}(t)$, $\psi_{s_i}(t)$, are bounded functions of time $t \in R_\tau$, where τ is a finite number.

It is assumed that there are no external or internal resonances in system (7).

For system (7), we choose the Lyapunov function as $2V_2(y, z) = \sum\limits_{s=1}^{n}(y_s^2 + z_s^2)$ and compute its total derivative with respect to $t \in R_\tau$. Specifically,

$$\frac{d}{dt}V_2(y, z) = \sum_{s=1}^{n}\Big\{y_s\Big(-\lambda_s z_s + \sum_{i=1}^{\infty}\mu^i Y_{s_i}(t, y, z) +$$

$$+\sum_{i=0}^{\infty}\mu^i \varphi_{s_i}(t)\Big) + z_s\Big(\lambda_s y_s + \sum_{i=1}^{\infty}\mu^i Z_{s_i}(t, y, z) + \sum_{i=0}^{\infty}\mu^i \psi_{s_i}(t)\Big)\Big\}. \tag{8}$$

Let there exist a non-negative function $\bar a_1(t, \mu)$ and a continuous function $\bar a_2(t, \mu)$, as well as values μ_1, $\mu_2 \in (0, 1]$ such that

$H_4.$: $\sum\limits_{s=1}^{n}\Big(y_s \sum\limits_{i=1}^{\infty}\mu^i Y_{s_i}(t, y, z) + z_s \sum\limits_{i=0}^{\infty}\mu^i Z_{s_i}(t, y, z)\Big) \le \bar a_1(t, \mu)\sum\limits_{s=1}^{n}(y_s^2 + z_s^2)$
for $\mu < \mu_1$ and $t \in R_\tau$;

$H_5.$: $\sum\limits_{s=1}^{n}\Big(y_s \sum\limits_{i=1}^{\infty}\mu^i \varphi_{s_i}(t) + z_s \sum\limits_{i=0}^{\infty}\mu^i \psi_{s_i}(t)\Big) \le \bar a_2(t, \mu)$
for $\mu < \mu_2$, $t \in R_\tau$ and $|y_s| < k < +\infty$, $|z_s| < k < \infty$, $s = 1, 2, \dots, n$.

Theorem 5.1 *To ensure that the solution* $(y(t), z(t))^T$ *of system* (7) *is bounded on a given interval with respect to the function* $V_2(y, z)$*, it is sufficient that conditions* H_4*,* H_5 *hold, and if* $V_2(y_0, z_0) = \beta^* < \beta$*, the following estimate holds:*

$$\exp\Big(\int\limits_0^t \bar a_1(s,\mu)ds\Big) + \frac{1}{\beta^*}\int\limits_0^t \bar a_2(s,\mu)\exp\Big(\int\limits_s^t \bar a_1(\tau,\mu)d\tau\Big)ds \le \frac{\beta}{\beta^*} \ \ \textit{for all} \ \ t \in R_\tau, \tag{9}$$

and for $\mu \in (0, \mu^*)$, *where* $\mu^* = \min(\mu_1, \mu_2)$.

Proof. Under conditions H_4, H_5, it follows from equation (8) that

$$\frac{d}{dt} V_2(y(t), z(t)) \leq \bar{a}_1(t, \mu) V_2(y(t), z(t)) + \bar{a}_2(t, \mu)$$

for all $t \in R_\tau$ and $0 < \mu < \mu^*$. Hence, we find that

$$V_2(y(t), z(t)) \leq V_2(y_0, z_0) + \int_0^t (\bar{a}_1(s, \mu) V_2(y(s), z(s)) + \bar{a}_2(s, \mu)) ds. \tag{10}$$

Applying Lemma 3.1 to inequality (10), we obtain for the function $V_2(y, z)$, an estimate similar to estimate (3). This estimate, together with condition (9), leads to the statement of Theorem 5.1. □

6 Boundedness of Solutions of a Quasilinear System with Stable Nonautonomous Approximation

Let us consider a quasilinear nonautonomous system of equations

$$\frac{dx_i}{dt} = \sum_{s=1}^{n} p_{si}(t) x_s + X_i(t, x_1, \ldots, x_n, \mu) + \psi_i(t), \ i = 1, 2, \ldots, n, \tag{11}$$

where the functions $X_i(t, x_1, \ldots, x_n, \mu)$ have expansions in powers of the parameter μ, $\psi_i(t)$ are bounded functions on any specified time interval $t \in R_\tau$. Let us assume that for system (11), a positive definite function $V_3(t, x)$ differentiable with respect to t has been constructed.

The total derivative of the function $V_3(t, x)$ due to system (11) can be represented as

$$\frac{dV_3}{dt}(t, x) = \frac{\partial V_3}{\partial t}(t, x) + \sum_{i=1}^{n} \sum_{s=1}^{n} \frac{\partial V_3}{\partial x_s}(t, x) p_{si}(t) x_s +$$
$$+ \sum_{s=1}^{n} \frac{\partial V_3}{\partial x_s}(t, x) X_s(t, x, \mu) + \sum_{s=1}^{n} \frac{\partial V_3}{\partial x_s}(t, x) \psi_s(t). \tag{12}$$

Let us assume that for system (11), there exist a positive function $\tilde{a}_1(t, \mu)$ and a bounded function $\tilde{a}_2(t)$ such that the following conditions hold:

- H_6. : $\frac{\partial V_3}{\partial t}(t, x) + \sum_{i=1}^{n} \sum_{s=1}^{n} \frac{\partial V_3}{\partial x_s}(t, x) p_{si}(t) x_s \leq 0$ for all $(t, x) \in R_\tau \times D$, where $D \subset \mathbb{R}^n$ is an open set;
- H_7. : $\sum_{s=1}^{n} \frac{\partial V_3}{\partial x_s}(t, x) X_s(t, x, \mu) \leq \tilde{a}_1(t, \mu) V_3(t, x)$ for $0 < \mu < \mu_1$ and $(t, x) \in R_\tau \times D$;
- H_8. : $\sum_{s=1}^{n} \frac{\partial V_3}{\partial x_s}(t, x) \psi_s(t) \leq \tilde{a}_2(t)$ for all $t \in R_\tau$ and $|x_s| < h$, where $h = const > 0$.

The condition H_6, together with the positive definiteness of the function $V_3(t, x)$, ensures the stability of the zero solution of the linear approximation system within the system of equations (11). Taking into account conditions H_6–H_8, we obtain an estimation from equation (12):

$$\frac{dV_3}{dt}(t,x) \le \tilde{a}_1(t,\mu)V_3(t,x) + \tilde{a}_2(t)$$

for all $t \in R_\tau$ and $0 < \mu < \mu_1$. Hence, we find the estimate of the change of the function $V_3(t, x(t))$ as

$$V_3(t,x(t)) \le V(t_0,x_0) + \int_{t_0}^{t} (\tilde{a}_1(s,\mu)V_3(s,x(s)) + \tilde{a}_2(s))ds \tag{13}$$

for all $t \in R_\tau$ and $0 < \mu < \mu_0$.

Applying Lemma 3.1 to the inequality (13), we can easily obtain the estimate of the function $V_3(t, x(t))$ in the form of (3). The following statement holds.

Theorem 6.1 *To ensure that the solution $x(t)$ of system (11) with a stable linear approximation is bounded on a given interval with respect to the function $V_3(t, x)$, it is sufficient that conditions H_6–H_8 hold, and for a given $\beta > 0$, the inequalities $V_3(t_0, x_0) = \beta^* < \beta$ are satisfied, as well as the inequality*

$$exp\Big(\int_{t_0}^{t} \tilde{a}_1(s,\mu)ds\Big) + \frac{1}{\beta^*}\int_{t_0}^{t} \tilde{a}_2(s)\exp\Big(\int_{s}^{t} \tilde{a}_1(\tau,\mu)d\tau\Big)ds < \frac{\beta}{\beta^*}$$

at all $t \in R_\tau$ and $0 < \mu < \mu_1$.

The proof of Theorem 6.1 is similar to the proof of Theorem 5.1.

7 Conditions for the Boundedness of Solutions of a System in Normal Form

Let us consider the differential equations of perturbed motion

$$\frac{dx_s}{dt} = X_s(t, x_1, \dots, x_n), \quad s = 1, 2, \dots, n, \tag{14}$$

$$x_s(t_0) = x_{s_0}, \tag{15}$$

where $X_s(t, 0, \dots, 0) \ne 0$ for all $t \in R_\tau$. We associate with the system (14) a differentiable function $V(t, x_1, \dots, x_n) > 0$, for which we write a Lyapunov relation

$$V(t,x(t)) = V(t_0,x_0) + \int_{t_0}^{t} \dot{V}(s,x(s))ds, \tag{16}$$

where $\dot{V}(t, x(t))$ is the total derivative of the function $V(t, x)$ due to the system of equations (14) and $x(t) = (x_1(t), \dots, x_n(t))^T$.

Let $V(t, x(t)) = v(t)$, and suppose that the following condition is satisfied:

$$H_9. : \; v(t_0) + \int_{t_0}^{t} \dot{V}(s,x(s))ds \le w(t) + \int_{t_0}^{t} p(s)v(s)ds, \tag{17}$$

where $w(t)$ and $p(t)$ are non-negative bounded functions on the given interval R_τ. The following statement holds.

Lemma 7.1 *If the perturbed motion equations* (14) *admit a differentiable function* $V(t,x)$*, and condition* H_9 *is satisfied, then the function* $V(t,x(t)) = v(t)$ *satisfies the inequality*

$$v(t) \le w(t) + \int_{t_0}^{t} \exp\Big(\int_{\tau}^{t} p(s)ds\Big) p(\tau)w(\tau)d\tau \tag{18}$$

for all $t \in R_\tau$.

Proof. From equation (16) under condition (17), we obtain the inequality

$$v(t) \le w(t) + \int_{t_0}^{t} p(s)v(s)ds$$

for all $t \in R_\tau$. Let us denote $z(t) = \int_{t_0}^{t} p(s)v(s)ds$ and note that $z(t_0) = 0$. Obviously,

$$\frac{dz}{dt} = p(t)v(t) \le p(t)[w(t) + z(t)] = p(t)w(t) + p(t)z(t).$$

From here, it follows that

$$z(t) \le \int_{t_0}^{t} \exp\Big[\int_{\tau}^{t} p(s)ds\Big] p(\tau)w(\tau)d\tau. \tag{19}$$

Since $v(t) \le w(t) + z(t)$, taking (19) into account yields the statement of Lemma 7.1. □

Theorem 7.1 *For the solution* $x(t)$ *of the normal system of equations* (14) *to be bounded on a given interval with respect to the function* $V(x,x)$*, it is sufficient that Lemma 7.1 holds and for a given* $\beta > 0$*, if* $V(t_0, x_0) = \beta^* < \beta$*, then the inequality applies*

$$w(t) + \int_{t_0}^{t} \exp\Big[\int_{\tau}^{t} p(s)ds\Big] p(\tau)w(\tau)d\tau \le \beta \tag{20}$$

for all $t \in R_\tau$.

Proof. If the conditions of Lemma 7.1 are satisfied, then the estimate for the function $V(t,x(t))$ given by (18) holds. From condition (20) and the fact that $V(t_0,x_0) = \beta^*$, it follows that $V(t,x(t)) \le \beta$ for all $t \in R_\tau$. This proves the statement of Theorem 7.1. □

Corollary 7.1 *If in condition* H_9*, we set* $w(t) = \beta^*$*, then the estimate* (18) *takes the form*

$$v(t) \le \beta^* \exp\Big[\int_{t_0}^{t} p(s)ds\Big] \tag{21}$$

for all $t \in R_\tau$.

Based on the estimate (21), the conditions for the boundedness of solutions of system (14) with respect to the function $V(t,x)$ take the form

$$\exp\Big[\int\limits_{t_0}^{t} p(s)ds\Big] \le \frac{\beta}{\beta^*} \tag{22}$$

for all $t \in R_\tau$. If condition (22) is satisfied, then we have the estimate $V(t,x(t)) \le \beta$ for all $t \in R_\tau$.

8 Example

Consider a non-autonomous oscillatory system of the second order [13]

$$\ddot{x} + p(t)\dot{x} + [a^2 + q(t)]x = f(t,x,y), \quad a = \text{const} \ne 0, \tag{23}$$

where $p(t) \ge 0$ for all $t \in R_\tau$ and $\int\limits_0^\infty q(s)ds < +\infty$. The functions $p(t)$, $q(t)$, $f(t,0,0)$ are continuous on $t \in R_\tau$ and $f(t,0,0) \ne 0$ for all $t \in R_\tau$.

Let us rewrite the equation (23) in the form of a system

$$\begin{cases} dx/dt = y, & x(t_0) = x_0, \\ dy/dt = -p(t)y - [a^2 + q(t)]x + f(t,x,y), & y(t_0) = y_0, \end{cases} \tag{24}$$

and for the total derivative of the function $V(x,y) = a^2x^2 + y^2$ on the solutions of system (24), we obtain the estimate

$$\begin{gathered} \frac{d}{dt}V(x(t),y(t)) = -2p(t)y^2(t) - 2q(t)x(t)y(t) + 2yf(t,x,y) \le \\ \le 2|q(t)||x(t)y(t)| - 2p(t)y^2(t) + 2y(t)f(t,x,y) \le \\ \le \frac{|q(t)|}{|a|}\big(a^2x^2(t) + y^2(t)\big) + \big|2y(t)f(t,x,y) - 2p(t)y^2(t)\big| = \overline{a}_1(t)V(x(t),y(t)) + \overline{a}_2(t), \end{gathered} \tag{25}$$

where $a_1(t) = \frac{|q(t)|}{|a|}$, $a_2(t) = \big|2y(t)f(t,x,y) - 2p(t)y^2(t)\big|$.

From inequality (25), it follows that

$$\frac{d}{dt}V(x(t),y(t)) \le \overline{a}_1(t)V(x(t),y(t)) + \overline{a}_2(t)$$

for all $t \in R_\tau$. Hence, we find the estimate of the change of the function $V(x(t),y(t))$ as

$$V(x(t),y(t)) \le V(x_0,y_0) + \int\limits_{t_0}^{t} (\overline{a}_1(s)V(x(s),y(s)) + \overline{a}_2(s))ds \tag{26}$$

for all $t \in R_\tau$.

Applying Lemma 3.1 to the inequality (26), we can easily obtain the estimate of the function $V(x(t),y(t))$ in the form

$$V(x(t),y(t)) \le V(x_0,y_0)\exp\Big(\int\limits_0^t \overline{a}_1(s)ds\Big) + \int\limits_{t_0}^{t} \exp\Big[\int\limits_{\tau}^{t} \overline{a}_1(s)ds\Big]\overline{a}_2(\tau)d\tau \tag{27}$$

for all $t \in R_\tau$, where $x_0 = x(t_0)$, $y_0 = y(t_0)$.

The following statement holds.

Applying Theorem 6.1 to inequality (27), we find that the solutions of system (24) are bounded in the sense of Definitions 2.1 and 2.2 if, for given estimates $0 < \beta < \beta^*$ and for $V(x_0, y_0) < \beta$, the following inequality holds:

$$\exp\Big(\int_0^t \overline{a}_1(s)ds\Big) + \int_{t_0}^t \exp\Big[\int_\tau^t \overline{a}_1(s)ds\Big]\overline{a}_2(\tau)d\tau < \frac{\beta^*}{\beta}$$

for all $t \in R_\tau$.

Note that if $a_2(t) = 0$ for all $t \in R_\tau$, then the boundedness of solutions of system (24) occurs under the conditions $V(x_0, y_0) < \beta$ and

$$\int_{t_0}^t \overline{a}_1(s)ds < \ln\left(\frac{\beta^*}{\beta}\right)$$

for all $t \in R_\tau$, where $0 < \beta < \beta^*$ are predefined values.

9 Conclusion

For systems of perturbed motion equations with stable nonlinear or linear approximations, conditions for the boundedness of solutions over a given time interval with respect to a positive definite function have been obtained. This new property of motion applies to nonlinear non-autonomous systems and has broad applications in nonlinear mechanics and system theory.

References

[1] A. A. Martynyuk. *Stability Analysis: Nonlinear Mechanics Equations.* Gordon and Breach Publishers, Amsterdam, 1995.

[2] R. Bellman. *Stability Theory of Differential Equations.* Dover Publ. Inc., New York, 1953.

[3] W. Hahn. *Stability of Motion.* Translated from the German manuscript by Arne P. Baartz. Die Grundlehren der mathematischen Wissenschaften. Band 138. Springer-Verlag New York, Inc., New York, 1967.

[4] F. C. Hoppensteadt. *Analysis and Simulation of Chaotic Systems.* Springer-Verlag, Berlin, 1993.

[5] J. H. Hubbard and B. H. West. *Differential Equations.* Springer-Verlag, Berlin, 1991.

[6] A. A. Martynyuk. Analysis of equi-boundedness and stability of essentially nonlinear systems. *Int. Appl. Mech.* **59** (1) (2023) 69–78.

[7] F. Verhulst. *Nonlinear Differential Equations and Dynamical Systems.* Springer-Verlag, Berlin, 1990.

[8] A. M. Lyapunov. The general problem of the stability of motion. *Int. J. Control* **55** (3) (2092) 531–773.

[9] D. R. Anderson and D. J. Ulness. *Results for conformable differential equations.* Working Paper, January, 2016.

[10] T. Yoshizawa. *Stability Theory by Liapunov's Second Method*. The Math. Sci. of Japan, Tokyo, 1966.

[11] V. I. Zubov. *Mathematical Methods for the Study of Automatic Control Systems*. Pergamon Press, London, 1962.

[12] V. G. Veretennikov. *Stability and Oscillations of Nonlinear Systems*. Provided by the SAO/NASA Astrophysics Data System, 1984.

[13] J. La Salle and S. Lefschetz. *Stability by Liapunov's Direct Method with Applications.* Academic Press, New York, 1961.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 625–634

Analysis of the Best Laptop Selection System Using Simple Additive Weighting (SAW) Method and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) Method

D. Novita [1], T. Herlambang [2], M. Tafrikan [3], K. Oktafianto [4], R. A. Sinulingga [5*] and H. Arof [6]

[1] *Department of Management, Muhammadiyah University Surabaya, Indonesia.*
[2] *Department of Information Systems, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[3] *Department of Mathematics, Walisongo State Islamic University, Indonesia.*
[4] *Department of Mathematics, University of PGRI Ronggolawe, Indonesia.*
[5*] *Department of Business, Faculty of Vocational Studies, University of Airlangga, Indonesia.*
[6] *Department of Electrical Engineering, University of Malaya, Malaysia.*

Received: April 22, 2024; Revised: November 22, 2024

Abstract: Every laptop has different specifications, and of course, the differences in specifications will affect the performance of the laptop when in use. The need to choose the right laptop depends on your needs. Therefore, we need an appropriate laptop recommendation system for prospective buyers. Choosing the optimal laptop according to your needs can be solved with a Decision Support System (DSS). The DSS has a mathematical model that can be used as a solution to these problems. There are several methods commonly used in solving problems, including the Simple Additive Weighting Method (SAW), Weighted Product (WP), and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). In this study, the SAW and TOPSIS methods were used, then the results were compared to those of the previous studies by using the WP method with the same data and criteria. The results of this study indicate that differences in laptop recommendations are only found in the second and third order. When using the SAW method, the second and third recommended laptops in a row are A6 (HP 14-G1024 U) and A3 (Acer Aspire E5-551). When using the TOPSIS method, the second and third recommendations for laptops in a row are A3 (Acer Aspire E5-551) and A6 (HP 14-G1024 U). The results of this study indicate that the SAW method gives the same laptop recommendation results as the WP method.

Keywords: *selection; laptop; TOPSIS; SAW; WP.*

Mathematics Subject Classification (2010): 90B50, 68U35.

* Corresponding author: mailto:rizkyamalia@vokasi.unair.ac.id

1 Introduction

Each laptop has different specifications, of course, and the differences in specifications surely affect the performance of the laptop when you use it. Currently, the main needs of the average student are limited to office applications and taking online courses, merely requiring middle to lower class laptops. However, those who work as graphic designers or gamers require devices with high specifications to meet their needs. A frequent problem occurring is buying a laptop whose specifications do not meet your needs. Lack of understanding by the user of laptop specifications makes the purchase not optimal. This can be minimized by contacting the store directly, but is limited to the store staff's knowledge or available inventory. There are several features that serve as benchmarks for choosing a laptop, that is, the Central Processing Unit (CPU), Graphics Processing Unit (GPU), Random Access Memory (RAM), storage, display, and price. Some of these features result in laptop buying recommendations.

Therefore, a system that recommends the right laptop for you is needed so that the purchase of a laptop will meet your needs optimally for home use. Choosing the optimal laptop according to your needs can be effectively done by using a Decision Support System (DSS), a discipline of operations research that can be utilized for decision making support in the form of mathematical models. DSS is an interactive software-based system designed to help decision makers collect, analyze, and process information from raw data, documents, frameworks, and business models to identify problems, solve them, and make decisions. SPK is computer software used in specific situations to analyze and present business data to help users make business decisions.

DSS has a mathematical model used as a solution to the problems. The model is Multi Criteria Decision-Making (MCDM). MCDM is one of the methods developed and used to help decision makers choose out of several decision options to take by several criteria to be considered to make the right and optimal decision [6]. Fuzzy MCDM is a decision support method whose purpose is to determine predicted alternatives out of several alternatives based on certain criteria used in the Fuzzy Multi Criteria decision method [7].

In terms of usefulness, MCDM is grouped into two models. They are Multi Objective Decision Making (MODM) used to solve problems in continuous space and Multi Attribute Decision Making (MADM) used to solve problems in discrete space. And the method used in this study is MADM.

There are several methods commonly employed in solving MADM problems, that is, the Simple Additive Weighting Method (SAW), Weighted Product (WP), and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). These three methods are used in helping decision making for laptop selection.

The previous research conducted by [11] contributed results able to help make laptop selection decisions employing the WP method. And in this study, the researchers used the SAW and TOPSIS methods by using the same data and criteria as those the previous research used [11]. The researchers compare the results obtained by both methods to those obtained by the WP method.

2 Research Method

2.1 Research method

The Simple Additive Weighting (SAW) method and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) can assist for laptop selection decision making. The basic concept of the SAW method is to find out the weighted sum of the performance ratings for each alternative on all attributes. The SAW method requires a process of normalizing the decision matrix (X) to a scale that can be compared to all existing alternative ratings.

$$r_{ij} = \begin{cases} \frac{x_{ij}}{Max\ x_{ij}}\ if\ j : atribute\ of\ benefit, \\ \frac{Min\ x_{ij}}{x_{ij}}\ if\ j : atribute\ of\ cost, \end{cases} \tag{1}$$

where r_{ij} is the normalized performance rating of alternative A_i on attribute $C_i; i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The preference value for each alternative (V_i) is given as

$$V_i = \sum_{j=1}^{n} w_j r_{ij}, \tag{2}$$

where the greater value of V_i indicates that alternative A_i is preferred or more frequently chosen.

The TOPSIS concept is based on the concept that the best selected alternative has not only the shortest distance from the positive ideal solution but also the longest distance from the negative ideal solution. This concept is frequently used to solve decision making problems in several MADM models because the concept is simple and easy to understand, computationally efficient and has the ability to measure the relative performance of decision alternatives in a simple mathematical form.

TOPSIS requires the performance rating of each alternative A_i on each normalized criterion C_j, that is,

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}. \tag{3}$$

The positive ideal solution A^+ and the negative ideal solution A^- can be determined based on the normalized weight rating (y_{ij}) as follows:

$$y_{ij} = w_i r_{ij}, \tag{4}$$

$i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

$$A^+ = \left(y_1^+, y_2^+, \ldots, y_n^+\right), \tag{5}$$

$$A^- = \left(y_1^-, y_2^-, \ldots, y_n^-\right) \tag{6}$$

with

$$y_j^+ = \begin{cases} max\ y_{ij}; if\ j : atribute\ of\ benefit, \\ min\ y_{ij}; if\ j : atribute\ of\ cost, \end{cases}$$

$$y_j^- = \begin{cases} max\ y_{ij}; if\ j : atribute\ of\ benefit, \\ min\ y_{ij}; if\ j : atribute\ of\ cost. \end{cases}$$

The distance between the alternative A_i and the positive ideal solution is formulated as follows:

$$D_i^+ = \sqrt{\sum_{j=1}^{n} \left(y_i^+ - y_{ij}\right)^2}; i = 1, 2, \ldots, m. \tag{7}$$

The distance between the alternative A_i and the negative ideal solution is formulated as follows:

$$D_i^- = \sqrt{\sum_{j=1}^{n} \left(y_{ij} - y_i^-\right)^2}; i = 1, 2, \ldots, m. \tag{8}$$

The preference value of each alternative (V_i) is given as

$$V_i = \frac{D_i^-}{D_i^- + D_i^+}; i = 1, 2, \ldots, m. \tag{9}$$

The higher value of V_i indicates that A_i is the preferred value.

2.2 Research material

The data and weighting used in this study are the same as those in the previous research [11]. The data in question can be seen in Table 1.

No	Alternative	Criteria				
		C_1	C_2	C_3	C_4	C_5
1	Axioo Neon TNW C825	Intel Celeron N2940	2 GB	500 GB	Intel HD Family	4.100.000
2	Axioo Neon TNN C825	Intel Celeron Quad Core N2920	2 GB	500 GB	Intel HD Family	4.000.000
3	Acer Aspire E5-551	AMD A10-7300	4 GB	1 TB	AMD Raden R7 M265	6.699.000
4	Lenovo Ideapad 100	Intel Core i3-5005U	2 GB	500 GB	NVIDIA GeForce 920A DDR3L 2 GB	5.399.000
5	Toshiba S40 A	Intel Core i3-3227u	2 GB	500 GB	NVIDIA GoForce GT 740 M	6.200.000
6	HP 14- U G1024 U	AMD A4-500	2 GB	500 GB	AMD Radeon HD 833	3.830.000

Table 1: Criteria.

From Table 1, coding is made as shown in Table 2.

In solving the selection of the best laptop by the SAW and TOPSIS methods, criteria and weights are required to perform calculations so that the best alternative will be obtained. The following are the criteria for decision making, based on the parameters in determining the best laptop at SMK Mandiri Bekasi as in Table 3.

In these criteria, a level of importance of the criteria is determined based on the predetermined weight value. The rating of each alternative on each criterion can be seen in Table 4.

Based on the criteria from the rating of each alternative (L_i) on each criterion (K_i) already determined, the weight of each criterion (K_i) is then determined.

No	Codes	Alternatives
1	A_1	Axioo Neon TNW C825
2	A_2	Axioo Neon TNN C825
3	A_3	Acer Aspire E5-551
4	A_4	Lenovo Ideapad 100
5	A_5	Toshiba S40 A
6	A_6	HP 14-G1024 U

Table 2: Alternative Codes.

Criteria	Description
K_1	Prosesor
K_2	RAM
K_3	Harddisk
K_4	VGA
K_5	Harga

Table 3: Atribute Codes.

Value	Alternative
1	Very low
2	Low
3	Fair
4	High
5	Very High

Table 4: Alternative Rating.

a) Processor Weight Value (K_1).

The weight value (W) of each processor criterion has been determined by the SMK Mandiri Bekasi school.

Processor	Very low	1
	Low	2
	Fair	3
	High	4
	Very High	5

Table 5: Processor Criteria.

b) RAM Weight Criteria (K_2).

The weight value (W) of each RAM criterion has been determined by the SMK Mandiri Bekasi school.

RAM Capacity	1 GB	1
	RAM Weight Criteria	2
	RAM Weight Criteria	3
	8 GB	4
	16 GB	5

Table 6: RAM Criteria.

c) Harddisk weight criteria (K_3).

The weight value (W) of each Harddisk criterion has been determined by the SMK

Harddisk Capacity	250 GB	1
	320 GB	2
	500 GB	3
	750 GB	4
	>750 GB	5

Table 7: Harddisk Criteria.

Mandiri Bekasi school.

d) VGA Weight Criteria (K_4).

Processor	Very low	1
	Low	2
	Fair	3
	High	4
	Very High	5

Table 8: VGA Criteria.

e) Price Weight Criteria (K_5).

Price capacity	$3-4$ M	1
	$4-6$ M	2
	$6-8$ M	3
	$8-15$ M	4
	≥ 15 M	5

Table 9: Price Criteria.

f) Weight Value Criteria.

W_1	Processor	5
W_2	RAM	4
W_3	Harddisk	3
W_4	VGA	5
W_5	Price	3

Table 10: Weight Criteria.

3 Results and Discussion

3.1 Solving by SAW method

To determine the normalization matrix, the elements can first be determined using equation (1) or (2):

$$
\begin{aligned}
r_{11} &= \frac{2}{\max\{\ 2\ \ 4\ \ 1\ \ 4\ \ 4\ \ 4\ \}} = \frac{2}{4} = 0.5, \\
r_{21} &= \frac{4}{\max\{\ 2\ \ 4\ \ 1\ \ 4\ \ 4\ \ 4\ \}} = \frac{4}{4} = 1, \\
r_{31} &= \frac{1}{\max\{\ 2\ \ 4\ \ 1\ \ 4\ \ 4\ \ 4\ \}} = \frac{1}{4} = 0.25, \\
\vdots &= \vdots \\
r_{12} &= \frac{1}{\max\{\ 1\ \ 4\ \ 3\ \ 1\ \ 1\ \ 1\ \}} = \frac{1}{4} = 0.25, \\
r_{22} &= \frac{4}{\max\{\ 1\ \ 4\ \ 3\ \ 1\ \ 1\ \ 1\ \}} = \frac{4}{4} = 1 \\
\vdots &= \vdots
\end{aligned}
$$

and so on. Based on the results obtained, a matrix is formed as displayed in Table 11.

No	Alternative	Criteria				
		K_1	K_2	K_3	K_4	K_5
1	A_1	0.5	0.25	0.75	0.4	0.2
2	A_2	1	1	0.75	0.4	0.5
3	A_3	0.25	0.75	1	1	0.25
4	A_4	1	0.25	0.75	0.4	0.2
5	A_5	1	0.25	0.75	0.4	0.25
6	A_6	1	0.25	0.75	0.4	1

Table 11: Calculation of Matrix Normalization.

Then each element of the normalization matrix and weight criteria are substituted in equation (3).

$$\begin{aligned} V_1 &= 5(0.5)+4(0.25)+3(0.75)+5(0.4)+3(0.2)=8.35, \\ V_2 &= 5(1)+4(1)+3(0.75)+5(0.4)+3(0.5)=14.75, \\ V_3 &= 5(0.25)+4(0.75)+3(1)+5(1)+3(0.25)=13, \\ V_4 &= 5(1)+4(0.25)+3(0.75)+5(0.4)+3(0.2)=10.85, \\ V_5 &= 5(1)+4(0.25)+3(0.75)+5(0.4)+3(0.25)=11, \\ V_6 &= 5(1)+4(0.25)+3(0.75)+5(0.4)+3(1)=13.25. \end{aligned}$$

The V value shows the order of laptop recommendations ranging from the largest to smallest. Based on the Simple Additive Weighting (SAW) method applied, the results and order of selection priorities are as displayed in Table 12. Table 12 shows that the priority order of the first laptop selection is A_2 (Axioo Neon TNN C825), that of the second laptop selection is A_6 (HP 14-G1024 U), and so on.

Alternative	Results	Ranking
A_1	8.35	6
A_2	14.75	1
A_3	13	3
A_4	10.85	5
A_5	11	4
A_6	13.25	2

Table 12: Alternative Priority.

3.2 Solving by TOPSIS Mehod

By using equation (4), the normalized matrix is obtained as in Table 13 below.

No	Alternative	Criteria				
		K_1	K_2	K_3	K_4	K_5
1	A_1	0.2408	0.1857	0.3841	0.2981	0.5361
2	A_2	0.4815	0.7428	0.3841	0.2981	0.2144
3	A_3	0.1204	0.5571	0.5121	0.7454	0.4288
4	A_4	0.4815	0.1857	0.3841	0.2981	0.5361
5	A_5	0.4815	0.1857	0.3841	0.2981	0.4288
6	A_6	0.4815	0.1857	0.3841	0.2981	0.1072

Table 13: Calculation of Matrix Normalization.

Then, from the normalized matrix, the weighted matrix is obtained as in Table 14.

Table 14 is obtained by multiplying the elements of each row in Table 13 by the corresponding weight criteria.

The positive and negative ideal solution matrix is obtained from equation (5) or (6). In the positive ideal solution, the largest value is selected for the profit attribute and the smallest value for the cost attribute. Meanwhile in the negative ideal solution, it applies vice versa. Then by using equations (7) and (8), the results are obtained as in Table 16.

Alternative	Criteria				
	K_1	K_2	K_3	K_4	K_5
A_1	1.204	0.7428	1.1523	1.4905	1.6083
A_2	2.4075	2.9712	1.1523	1.4905	0.6432
A_3	0.602	2.2284	1.5363	3.727	1.2864
A_4	2.4075	0.7428	1.1523	1.4905	1.6083
A_5	2.4075	0.7428	1.9205	1.4905	1.2864
A_6	2.4075	0.7428	1.1523	1.4905	0.3216

Table 14: Calculation of Weighted Matrix Normalization.

Alternative	Criteria				
	K_1	K_2	K_3	K_4	K_5
$A(+)$	2.075	2.9712	1.9205	3.727	0.3216
$A(-)$	0.602	0.7428	1.1523	1.4905	1.6083

Table 15: Calculation of Positive and Negative Ideal Matrix.

Alternative	Ideal Solution Distance	
	$D(+)$	$D(-)$
A_1	3.696193	0.602
A_2	2.386523	3.026056
A_3	2.211341	2.731303
A_4	3.494771	1.8055
A_5	3.301293	1.988361
A_6	3.249281	2.217076

Table 16: Calculation of Alternative Distance Matrix to Positive and Negative Ideal Solutions.

By using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method, the results and order of selection priorities are as in Table 17.

Alternative	Results	Ranking
A_1	0.140059	6
A_2	0.559078	1
A_3	0.5526	2
A_4	0.340643	5
A_5	0.375896	4
A_6	0.405586	3

Table 17: Alternative Priority.

Based on Table 17, it can be seen that the order of priority for choosing the first laptop is A_2 (Axioo Neon TNN C825), and that for choosing the second laptop is A_3 (Acer Aspire E5-551), and so on.

4 Conclusion

The application of SAW and TOPSIS methods provides different priority orders for the second and third laptop recommendations. By using the SAW method, the second and third laptop recommendations are A_6 (HP 14-G1024 U) and A_3 (Acer Aspire E5-551). At the same time, when using the TOPSIS method, the second and third laptop recommendations are A_3 (Acer Aspire E5-551) and A_6 (HP 14-G1024 U). When compared to the results of the previous studies, it can be seen that the SAW method provides the same laptop recommendation sequence results as the WP method, that is, A_6 (HP 14-G1024 U) and A_3 (Acer Aspire E5-551). Meanwhile the TOPSIS method gives different results in the order of recommendations for the second and third laptops, that is, A_6 (HP 14-G1024 U) and A_3 (Acer Aspire E5-551) by the WP method and A_3 (Acer Aspire E5-551) and A_6 (HP 14-G1024 U) by the TOPSIS method. The difference occurs due to differences in calculation methods among SAW, WP, and TOPSIS.

References

[1] A. A. Chamid. Application of the Topsis Method to Determine Home Condition Priorities. *Journal of SIMETRIS.* **7** (2) (2016) 537–544.

[2] T. Hartanto and M. I. Prasetiyowati. Web-Based Laptop Selection Decision Support System with Analytical Hierarchy Process Method (Case Study: SAMCO COMPUTER). *ULTIMATICS* **4** (2) (2012) 7–15.

[3] I. E. Herlambang. Decision Making System for Product Purchase Recommendations Using the Fuzzy MCDM Method (Case Study of PT. Nerangi Sarana Karya). *Jurnal Ilmiah Sistem Informasi* **1** (1) (2022) 51–61.

[4] H. Hertyana, E. Mufida and A. A. Kaafi. Laptop Selection Decision Support System Using the Topsis Method. *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)* **6** (1) (2021) 36–44.

[5] A. Y. Kungkung and R. H. Kiswanto. Comparative Analysis of SAW, WP, and TOPSIS Methods Using Hamming Distance. *National Conference on Information Systems* (2018) 836–841.

[6] S. Kusumadewi, S. Hartati, et al. *Fuzzy Multi-Atttribute Decision Making (FUZZY MADM).* Graha Ilmu, Yogyakarta, 2006.

[7] R. Lissa, A. R. Tanaamah and A. D. Wowor. Combination of the Muslim Index Forecasting Algorithm and Fuzzy-MCDM Development in Predicting the Suitability of Food Crops in Salatiga. *National Seminar on Indonesian Information* (2015) 355–362.

[8] A. P. Lubis. Selection of Superior Types of Palm Oil Seeds Using the Fuzzy MCDM Method. *Royal National Seminar (SENAR)* (2018) 115–120.

[9] G. P. Sanyoto, R. I. Handayani and E. Widanengsih. Decision Support System for Selection of Laptops for Operational Needs Using the AHP Method (Case Study: Directorate of Course Development and Training, Ministry of Education and Culture). *Journal of Computing and Information System* **13** (2) (2017) 167–174.

[10] H. Supriyono. Selection of Residential Houses Using the Weighted Product Method. *Journal of Computer Science and Information Technology* **1** (1) (2015) 23–28.

[11] S. Susliansyah, R. R. Aria and S. Susilowati. The Best Laptop Selection System Using the Weighted Product (WP) Method. *Techno Nusa Mandiri: Journal of Computing and Information Technology* **16** (1) (2019) 15–20.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 635–647

Analysis and Optimal Control of a Mathematical Model of Malaria

L. Ouattara [1]*, D. Ouedraogo [1], O. Diop [2] and A. Guiro [1]

[1] *Nazi Boni University, Bobo-Dioulasso, Burkina Faso.*
[2] *Virtual University of Senegal, Dakar.*

Received: April 18, 2024; Revised: December 1, 2024

Abstract: In this work, we propose a mathematical model of malaria which takes into account the vector class represented by S_v, I_v and humans class represented by S_h, E_h, I_h and R_h. The basic reproduction number R_0 of the model is determined. We introduce two controls in our initial model. Therefore, the model with control will be presented and studied. The objective of the model with optimal control is to observe the effect of preventive measures, represented here by control u_1, and curative measures, represented by control u_2, on the evolution of malaria disease. The controls u_1 and u_2 will be characterized. Then we use the Python software for the numerical simulation of the model.

Keywords: *malaria; reproduction number; vector; simulation; optimal control.*

Mathematics Subject Classification (2010): 93B05, 93A30, 49J15, 49N90.

1 Introduction

Malaria is an acute febrile illness caused by Plasmodium parasites, which are spread to people through the bites of infected female Anopheles mosquitoes. It is preventable and curable. Malaria is a life-threatening disease primarily found in tropical countries. It was first discovered in India in the 15^{th} century. However, without prompt diagnosis and effective treatment, a case of uncomplicated malaria can progress to a severe form of the disease, which is often fatal without treatment. Malaria is not contagious and cannot spread from one person to another; the disease is transmitted through the bites of female Anopheles mosquitoes. The world's population is at risk of exposure to malaria [9]. In 2021, an estimated 247 million people contracted malaria in 85 countries. That same year, the disease claimed approximately 619000 lives [9]. The first symptoms of malaria

* Corresponding author: mailto:lassmaths2@gmail.com

usually appear within 10 to 15 days after the infectious bite of mosquitoes. Fever, headache and chills are usually signs of malaria, although these symptoms can hardly be attributed to malaria. WHO recommends rapid diagnostic testing for anyone suspected of having malaria. If Plasmodium falciparum malaria is not treated within 24 hours, the infection can progress to severe illness and death [9]. Malaria can be diagnosed using tests that determine the presence of the parasites causing the disease. There are two main types of tests, firstly, microscopic examination of blood smears and rapid diagnostic test called (TDR). Artemisinin based combination therapies (ACT) are the most effective antimalarial medicines available today. The resistance of the parasite to the different treatments leads to an endemic situation. Given all these threats, a mathematical model of malaria has been proposed to eradicate the disease. The most used means are the preventive ones. A number of recent studies of malaria show the significant direct effect of climatic factors such as temperature and rainfall on the transmission dynamics of vectors [11, 13–16, 18, 19]. B. Traore et al. [20] studied a mathematical model of malaria taking into account mosquito larvae and transmission of malaria in a periodic environment with a constant recruitment of vector and human population. Abba B. Gumel et al. [11] studied a malaria model taking into account seasonality and temperature variation in the mode of malaria transmission. Our model takes into account the preventive measures (distribution of mosquito nets, preventive medicine for children under 5 years old, spraying of areas etc.) represented by the control (u_1) and taken during a year. We use a control of cured persons (u_2) in order to allow the government to support population.

The structure of the paper is as follows. We present the mathematical model in Section 2. In Section 3, we present and study the mathematical model with control. We conclude in Section 4.

2 The Formulation of Mathematical Model

In this model, the human population is divided into four classes: the susceptible S_h, the exposed E_h, the infected I_h and cured R_h. The vector population (mosquitoes) is subdivided into two classes: susceptible vectors S_v and infected I_v. $\mu_h N_h$ is the dynamic recruitment of the human population. $\gamma_1 S_h$, $\gamma_1 E_h$, $\gamma_1 I_h$ and $\gamma_1 R_h$ are the number of susceptible, exposed, infected and cured individuals, respectively, that die naturally. $\dfrac{\beta I_v S_h}{N_h}$ is the proportions of susceptible humans that can encounter female Anopheles with a β rate. $\dfrac{\alpha_{E_h} E_h S_v}{N_h}$ and $\dfrac{\alpha_{E_h} I_h S_v}{N_h}$ are the respective proportions of Anopheles that bite exposed (E_h) and infected humans (I_h) that can infect them. γE_h is the total exposed population that manifests malaria disease at time t (exposed individuals who pass into the I_h class). θI_h is the set of sick humans who recover from malaria (humans who enter the R_h class). Λ_v is the recruitment of mosquitoes. μS_v and μI_v are the mosquitoes that die naturally, respectively, in classes S_v and I_v. The individual cured ($\alpha_1 R$) of malaria disease recontacts malaria disease in recruitment.

Remark 2.1 In our model, mosquitoes do not recover from malaria. Each person cured of malaria (R_h) is brought back into the susceptible population.

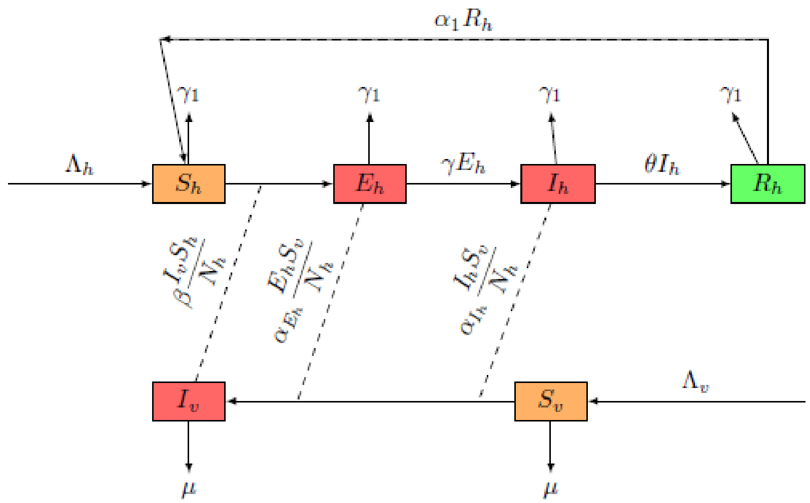


Figure 1: Transfer diagram, the black dashed arrows indicate the direction of the infection, the solid arrows represent the transition from one class to another.

The mathematical model without control is given

$$\begin{cases} \dot{S}_h = \Lambda_h + \alpha_1 R_h - \beta \dfrac{I_v S_h}{N_h} - \gamma_1 S_h, \\ \dot{E}_h = \beta \dfrac{I_v S_h}{N_h} - \gamma E_h - \gamma_1 E_h, \\ \dot{I}_h = \gamma E_h - \theta I_h - \gamma_1 I_h, \\ \dot{R}_h = \theta I_h - (\gamma_1 + \alpha_1) R_h, \\ \dot{S}_v = \Lambda_v - \alpha_{E_h} \dfrac{E_h S_v}{N_h} - \alpha_{I_h} \dfrac{I_h S_v}{N_h} - \mu S_v, \\ \dot{I}_v = \alpha_{E_h} \dfrac{E_h S_v}{N_h} + \alpha_{I_h} \dfrac{I_h S_v}{N_h} - \mu I_v \end{cases} \tag{1}$$

with the initial conditions

$$S_h(0) > 0, \quad S_v(0) > 0, \quad E_h(0) > 0, \quad I_h(0) > 0,$$

$$I_v(0) > 0, \quad R_h(0) > 0.$$

Total human population and the number of vectors are described by the following equations:

$$\dot{N}_h = \Lambda_h - \gamma_1 N_h(t) \tag{2}$$

and

$$\dot{N}_v = \Lambda_v - \mu N_v(t). \tag{3}$$

Symbols	Description	values	Sources
Λ_h	Constant recruitment rate for humans	5000	[20]
γ_1	Natural mortality rate of humans	0.00167	[20]
μ	Natural mortality rate of mosquitoes	0.01	estimate
Λ_v	Constant recruitment rate for mosquitoes	150	[20]
θ	Transfer rate of humans from I_h to R_h	0.3	estimate
α_{I_h}	Contact rate of susceptible mosquitoes with humans infected with malaria	0.05	estimate
α_{E_h}	Contact rate of susceptible mosquitoes with humans exposed to malaria	0.07	estimate
β	Contact rate of infected mosquitoes with susceptible humans	0.425	estimate
γ	Transition rate from E_h to I_h.	0.52	estimate

Table 1: The parameters of model (1).

Estimation of the total vector population at time t. Let us consider equations (3) and (2). The total vector population is estimated at time t by

$$N_v = \frac{\Lambda_v}{\mu} + \left(N_v(0) - \frac{\Lambda_v}{\mu}\right)\exp(-\mu t).$$

The total human population is estimated at time t by

$$N_h = \frac{\Lambda_h}{\gamma_1} + \left(N_h(0) - \frac{\Lambda_h}{\gamma_1}\right)\exp(-\gamma_1 t); \quad t \geq 0.$$

3 The Optimal Control Problem

In this section, we introduce two controls u_1 (prevention) and u_2 (treatment) into the model (1). Furthermore, we first prove the existence of two optimal controls u_1^*, u_2^* and then give the characterization of these two controls. So far, there is no preventive vaccine against malaria. In our study, we use the means of prevention other than vaccine.
The basic reproduction number R_0 is given by

$$R_0 = \sqrt{\frac{\beta\Lambda_v\gamma_1}{\Lambda_h\mu^2(\gamma+\gamma_1)}\left(\alpha_{E_h} + \frac{\alpha_{I_h}\gamma}{\theta+\gamma_1}\right)}.$$

3.1 Presentation of the problem

The controls u_1 and u_2 are defined as follows:

- $u_1(t) \in [0,1]$ is the control corresponding to the distribution of mosquito nets, preventive medication for children under five years and other means to prevent malaria. The rate of people sleeping under mosquito nets or protecting themselves against mosquitoes and/or preventing malaria is denoted by $u_1(t) \in [0,1]$ with $t \in [0, t_f]$. The ideal is to get the entire population to sleep under a mosquito net and to warn all children, in this case $u_1 = 1$. In reality, this is not possible, we seek to protect the maximum number of people ($u_1 = u_{1max}$).

- The second control $u_2(t) \in [0,1]$ represents the treatment of patients over the interval $[0;t_f]$. The control u_2 that we consider here can therefore represent the treatment of symptoms or the isolation of patients in hospitals to avoid possible new infection. If all patients are treated, then $u_2 = 1$.
  For all positive t, we unambiguously denote $u_i(t)$ simply by u_i for $i = 1, 2$.

By inserting the controls into the model (1), we get the following controlled equations:

$$\begin{cases} \dot{S}_h = \Lambda_h + \alpha_1 R_h - \beta(1-u_1)\dfrac{I_v S_h}{N_h} - \gamma_1 S_h, \\ \dot{E}_h = \beta(1-u_1)\dfrac{I_v S_h}{N_h} - \gamma E_h - \gamma_1 E_h, \\ \dot{I}_h = \gamma E_h - (\theta + u_2) I_h - \gamma_1 I_h, \\ \dot{R}_h = (\theta + u_2) I_h - \gamma_1 R_h, \\ \dot{S}_v = \Lambda_v - \alpha_{E_h}\dfrac{E_h S_v}{N_h} - \alpha_{I_h}\dfrac{I_h S_v}{N_h} - \mu S_v, \\ \dot{I}_v = \alpha_{E_h}\dfrac{E_h S_v}{N_h} + \alpha_{I_h}\dfrac{I_h S_v}{N_h} - \mu I_v \end{cases} \tag{4}$$

with the initial conditions

$$S_h(0) > 0, \quad S_v(0) > 0, \quad E_h(0) > 0, \quad I_h(0) > 0,$$

$$I_v(0) > 0, \quad R_h(0) > 0.$$

The basic reproduction number (R_0^c) of the model (4) is the number of cases generated by the primary infected individual under controls u_1 and u_2. This demonstrates that controls u_1 and u_2 play a role in combating malaria disease. We observe that if $u_1 = u_2 = 0$ (absence of all malaria control strategies), then we recover the same reproduction number R_0 as in the model (1) without control ($R_0^c = R_0$, if $u_1 = u_2 = 0$),

$$R_0^c = \sqrt{\frac{\beta \Lambda_v \gamma_1 (1-u_1)}{\Lambda_h \mu^2 (\gamma + \gamma_1)} \left(\alpha_{E_h} + \frac{\alpha_{I_h} \gamma}{\theta + u_2 + \gamma_1} \right)}.$$

Remark 3.1 The objective of these controls is to observe the effect of malaria treatments and the effect of preventive measures in the fight against malaria. Furthermore, we aim to propose strategies to minimize the infected population (I_h) while maximizing the recovered population (R_h) and the susceptible population (S_h).

3.2 Study of optimal control problem

In this section, we define the Hamiltonian associated with the control problem (4). Then we characterize the solutions of control problem (4) after proving their existence. Mathematically, for a fixed terminal time t_f , the problem is to minimize the functional objective J on $[0, t_f]$.

$$J(u_1, u_2) = \int_0^{t_f} \left(I_h(t) - S_h(t) - R_h(t) + \frac{A_1}{2} u_1^2(t) + \frac{A_2}{2} u_2^2(t) \right) dt. \tag{5}$$

The first terms represent the gain for the I_h that we wish to reduce. The constants A_1 and A_2 are positive, and correspond to the weights that regularize the control for prevention and treatment, respectively. As given in the literature, the costs are assumed to be quadratic functions. Indeed, costs are rarely linear, and are often presented as non -linear functions of control. Other types of functions exist in the literature [2, 4, 6, 10]. The most natural thing to do is to consider quadratic functions. These also allow us to make the analogy with the energy that is expanded here for all these measurements. Our objective is to limit the transmission of the disease by reducing the number of mosquitoes and infected humans.

We determine the optimal control (u_1^*, u_2^*) such that

$$J(u_1^*, u_2^*) = min\left\{J(u_1, u_2) : (u_1, u_2) \in \Gamma\right\}, \tag{6}$$

where

$$\Gamma = \left\{ \ (u_1, u_2), \begin{cases} u_i(t) \text{ is a continuous function by pieces on } [0,t_f] \\ a_i \leq u_i(t) \leq b_i \end{cases} \right\} \tag{7}$$

is the set of controls and a_i, b_i are constants belonging to [0;1], $i = 1, 2$. The optimal control problem is then solved when we determine $(u_1^*, u_2^*) \in \Gamma$ which minimizes the function (5).

Definition 3.1 (the Hamiltonian of the minimization problem) Pontryagin's maximum principle [12] converted (4), (5) and (6) into the problem of minimizing the Hamiltonian H defined by

$$H = -S_h - R_h + I_h + \frac{A_1}{2}u_1^2 + \frac{A_2}{2}u_2^2 + \sum_{i=1}^{6} \lambda_i f_i, \tag{8}$$

where

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \end{pmatrix} = \begin{pmatrix} \Lambda_h + \alpha_1 R_h - \beta(1-u_1)\dfrac{I_v S_h}{N_h} - \gamma_1 S_h \\ \beta(1-u_1)\dfrac{I_v S_h}{N_h} - \gamma E_h - \gamma_1 E_h \\ \gamma E_h - (\theta + u_2) I_h - \gamma_1 I_h \\ (\theta + u_2) I_h - (\gamma_1 + \alpha_1) R_h \\ \Lambda_v - \alpha_{E_h}\dfrac{E_h S_v}{N_h} - \alpha_{I_h}\dfrac{I_h S_v}{N_h} - \mu S_v \\ \alpha_{E_h}\dfrac{E_h S_v}{N_h} + \alpha_{I_h}\dfrac{I_h S_v}{N_h} - \mu I_v \end{pmatrix}$$

is the righ-hand side of the differential equation (1), the state variable and λ_i, $i = 1, ..., 6$, are the adjoint variables associated with their respective states.

Theorem 3.1 *Consider the optimal control problem* (4) *subject to* (5). *Then there exist an optimal pair of controls* (u_1^*, u_2^*) *and corresponding optimal states* $(S_h, E_h, I_h, R_h, S_v, I_v)$ *that minimize the objective function* $J(u_1, u_2)$ *over the set of admissible controls* Γ.

Proof. The existence of optimal control can be proved by using the results from [7] (see Theorem 2.1) and Fleming's results [3] (Theorem III.4.1), we must check the following conditions:

- the set of controls and solutions present is nonempty,
- the admissible set Γ is convex and closed,
- the vector field of the state system is bounded by a linear function of control,
- the objective function is convex,
- there exist constants $c_1, c_2 > 0$ such that the integrated part of the objective function is bounded by $c_1(|u_1|^2 + |u_2|^2)^{\frac{p}{2}} - c_2$.

(1) We verify these conditions thanks to a result of Lukes et al. [8], which assures the existence of solutions for the state system (1).

(2) The set Γ is convex and bounded by definition.

(3) The right-hand side of the state system (4) is bounded by a linear function in the state and control variables.

(4) The integrated part of the objective functional is

$$f^0(X, u_1, u_2) = I_h - S_h - R_h + \frac{A_1}{2}u_1^2 + \frac{A_2}{2}u_2^2.$$

The Hessian matrix of $f^0(X, u_1, u_2)$ is given by

$$M_{f^0} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

$Spec(M_{f^0}) = \{A_1, A_2\} \subset \mathbb{R}_+^*$.
So, by using [1], f^0 is strictly convex over U.

(5) We have

$$\begin{aligned} f^0(X, u_1, u_2) &= I_h(t) - S_h(t) - R_h(t) + \frac{A_1}{2}u_1^2(t) + \frac{A_2}{2}u_2^2(t), \\ &= N_h - E_h - 2R_h - 2S_h + \frac{A_1}{2}u_1^2(t) + \frac{A_2}{2}u_2^2(t), \\ &\geq -E_h - 2R_h - 2S_h + \frac{A_1}{2}u_1^2(t) + \frac{A_2}{2}u_2^2(t), \\ &\geq \frac{1}{2}\min\{A_1, A_2\}\left(|u_1(t)|^2 + |u_2(t)|^2\right)^{k/2} - (E_h + 2R_h + 2S_h), \\ &\geq c_1\left(|u_1(t)|^2 + |u_2(t)|^2\right)^{\frac{k}{2}} - c_2, \end{aligned}$$

where $c_1 = \frac{1}{2}\min\{A_1, A_2\} > 0$, $c_2 = E_h + 2R_h + 2S_h$ and $k \geq 1$, so the last assertion is verified.

Since the state variables are bounded, we deduce the existence of an optimal control (u_1^*, u_2^*) which minimizes the objective function $J(u_1, u_2)$ in (5). □

We are now interested in the characterization of a control $u^* = (u_1^*, u_2^*)$, the solution of (7). Pose $Z = (S_h, E_h, I_h, R_h, S_v, I_v)$, $U = (u_1, u_2)$ and $L = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$, being the adjoint variables. We define the Lagrangian associated with the problem (this one corresponds to the Hamiltonian increased by the penalties).

$$\begin{aligned}
\mathcal{L}(Z,U,L) = \quad & I_h - R_h - S_h + \frac{A_1}{2}u_1^2 + \frac{A_2}{2}u_2^2 \\
+ \quad & \lambda_1 \left(\Lambda_h + \alpha_1 R_h - \beta(1-u_1)\frac{I_v S_h}{N_h} - \gamma_1 S_h \right) \\
+ \quad & \lambda_2 \left(\beta(1-u_1)\frac{I_v S_h}{N_h} - \gamma E_h - \gamma_1 E_h \right) \\
+ \quad & \lambda_3 \left(\gamma E_h - (\theta + u_2) I_h - \gamma_1 I_h \right) + \lambda_4 \left((\theta + u_2) I_h - (\gamma_1 + \alpha_1) R_h \right) \\
+ \quad & \lambda_5 \left(\Lambda_v - \alpha_{E_h}\frac{E_h S_v}{N_h} - \alpha_{I_h}\frac{I_h S_v}{N_h} - \mu S_v \right) \\
+ \quad & \lambda_6 \left(\alpha_{E_h}\frac{E_h S_v}{N_h} + \alpha_{I_h}\frac{I_h S_v}{N_h} - \mu I_v \right) - w_{11}(u_1 - a_1) - w_{12}(b_1 - u_1) \\
- \quad & w_{21}(u_2 - a_2) - w_{22}(b_1 - u_2),
\end{aligned} \tag{9}$$

where $w_{ij}(t) \geq 0, \quad i,j = 1,2$, are the penalty coefficients verifying

$$\begin{aligned}
w_{11}(u_1 - a_1) = w_{12}(b_1 - u_1) = 0 \quad \text{for optimal control} \quad u_1^* \text{ and} \\
w_{21}(u_2 - a_2) = w_{22}(b_2 - u_2) = 0 \quad \text{for optimal control} \quad u_2^*.
\end{aligned} \tag{10}$$

Theorem 3.2 *Consider an optimal control $u^* = (u_1^*, u_2^*) \in \Gamma$ and corresponding states $X = (S_h,\ E_h,\ I_h,\ R_h,\ S_v,\ I_v)$ of system (4), there exist adjoint functions (λ_i, $i = 1, ..., 6$) satisfying*

$$\begin{aligned}
\dot{\lambda}_1 &= -\left(-1 - \lambda_1 \left((1-u_1)\frac{\beta I_v}{N_h} - \gamma_1 \right) + \beta(1-u_1)\frac{I_v}{N_h}\lambda_2 \right), \\
\dot{\lambda}_2 &= -\left(-(\gamma_1 + \gamma)\lambda_2 + \gamma\lambda_3 - \frac{\alpha_{E_h} S_v}{N_h}\lambda_5 + \frac{\alpha_{E_h} S_v}{N_h}\lambda_6 \right), \\
\dot{\lambda}_3 &= -\left(1 - \lambda_3(\theta + u_2 + \gamma_1) + \lambda_4(\theta + u_2) - \frac{\alpha_{I_h} S_v}{N_h}\lambda_5 + \frac{\alpha_{I_h} S_v}{N_h}\lambda_6 \right), \\
\dot{\lambda}_4 &= -\left(\alpha_1\lambda_1 - (\gamma_1 + \alpha_1)\lambda_4 - 1 \right), \\
\dot{\lambda}_5 &= -\left(-\lambda_5 \left(\mu + \frac{\alpha_{E_h} E_h}{N_h} + \frac{\alpha_{I_h} I_h}{N_h} \right) + \lambda_6 \left(\frac{\alpha_{E_h} E_h}{N_h} + \frac{\alpha_{I_h} I_h}{N_h} \right) \right), \\
\dot{\lambda}_6 &= -\left(-\frac{\beta S_h}{N_h}\lambda_1 + \frac{\beta S_h}{N_h}\lambda_2 - \mu\lambda_6 \right)
\end{aligned} \tag{11}$$

with the transversality conditions given by $\lambda_i(t_f) = 0, \quad i = 1, ..., 6$. Furthermore, the optimal controls are characterized by

$$\begin{aligned}
u_1^* &= \max\left\{ a_1, \min\left\{ b_1, \left(\frac{\lambda_2(t) - \lambda_1(t)}{A_1} \right) \beta\frac{S_h}{N_h} \right\} \right\}, \\
u_2^* &= \max\left\{ a_2, \min\left\{ b_2, \frac{(\lambda_3(t) - \lambda_4(t))}{A_2} I_h \right\} \right\}.
\end{aligned} \tag{12}$$

Proof. The differential equations for the adjoint variables are standard results from Pontryagin's maximum principle [17]. The right-hand sides of the differential equations can be easily computed. Let $w^* = (u_1^*, u_2^*)$ be the corresponding solution $X = (S_h, E_h, I_h, R_h, S_v, I_v)$ that minimizes $J(u_1, u_2)$ over Γ. By Pontryagin's maximum principle [17], there exist adjoint functions

$$p(t) = (\lambda_1(t), \lambda_2(t), \lambda_3(t), \lambda_4(t), \lambda_5(t), \lambda_6(t)), \quad t \in [0, t_f]),$$

verifying the following conditions:

$$\frac{dp(t)}{dt} = -\frac{\partial H}{\partial X}, \tag{13}$$

$$\frac{dX(t)}{dt} = \frac{\partial H}{\partial p}, \tag{14}$$

$$\frac{\partial \mathcal{L}}{\partial u_1} = \frac{\partial \mathcal{L}}{\partial u_2} = 0. \tag{15}$$

The condition (13) yields the system (11) and condition (14) yields the system (4). The optimality condition (15) gives the following system:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial u_1}\Big|_{(u_1=u_1^*)} &= A_1 u_1^* + \lambda_1 \beta \frac{I_v S_h}{N_h} - \lambda_2 \beta \frac{I_v S_h}{N_h} - w_{11} + w_{12} = 0, \\ \frac{\partial \mathcal{L}}{\partial u_2}\Big|_{(u_2=u_2^*)} &= A_2 u_2^* + \lambda_1 \beta \frac{I_v S_h}{N_h} - \lambda_2 \beta \frac{I_v S_h}{N_h} - w_{21} + w_{22} = 0. \end{aligned} \tag{16}$$

By solving (16) and using (10), we obtain the result (12) □

3.3 Numerical simulation

First, note that the optimality system is a problem with two boundary conditions. Indeed, the state system is solved in the direction with the initial conditions $X(0) = (100, 90, 70, 100, 60)$. The adjoint functions are solved in the opposite direction [5], with the transversality conditions $\lambda_i(t_f) = 0$, $i = 1, ..., 6$, where $t_f = 12$ months. The numerical simulations are obtained by using Python. The control curves in Figure

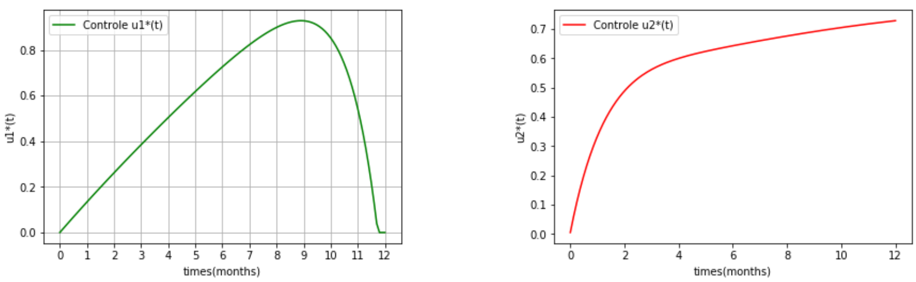


Figure 2: u_1 and u_2 control curves.

2 show the period of application for the controls u_1 and u_2. The curve of u_1 shows that the period of implementation for control u_1 are the first nine months of the year. The measurement control has no effect during the last quarter (October, November and

December) of the year. The best period for implementing preventive measures (u_1) is the month of June. Furthermore, the government's efforts should be focused on the month of June. The best time of the year for distributing mosquito nets, spraying public space and vaccinating the susceptible human population is June. For better prevention results, 50% of the population should be involved. The curve of (u_2) in Figure 2 shows that the control (u_2) of malaria patient care should be applied continuously throughout the year. For better results, over +70% of malaria patients should be supported by the government.

Figure 3: $S_h(t)$

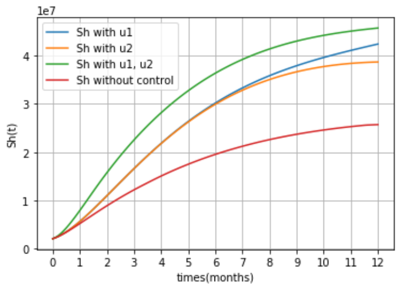


Figure 4: $E_h(t)$

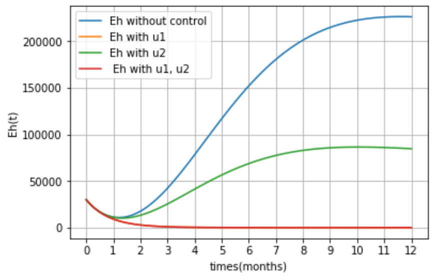


Figure 5: $I_h(t)$

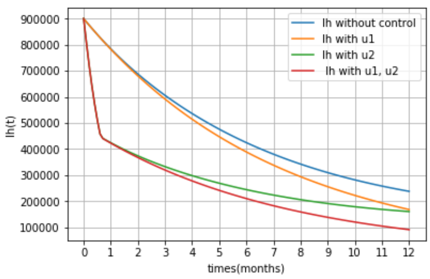


Figure 6: $R_h(t)$

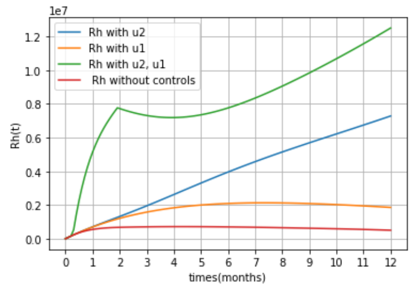


Figure 7: $S_v(t)$

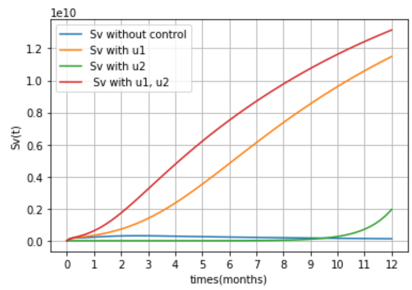
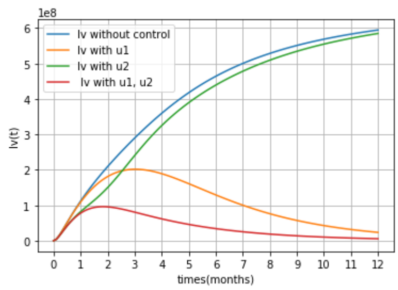


Figure 8: $I_v(t)$,



Population dynamics with different control aspects u_1 and u_2.

Comment

The curves of Figure 3 describe the dynamics of the susceptible human population (S_h) by using the treatment (u_2) and prevention (u_1) controls. This proves that treating only malaria patients (control u_2) has a negligible effect in the fight against malaria. If we apply only treatment (control u_2), we observe a smaller effect on the susceptible population. Preventive measures are the best way to preserve the susceptible population (S_h) from malaria disease. Furthermore, if prevention and treatment are applied simultaneously, the susceptible population remains protected from malaria.

The curves of Figure 4 describe the dynamics of E_h by using u_1 and u_2 controls. This proves that treating malaria patients has a great effect in the exposed population (E_h). We have found that if all the controls are applied, the number of malaria exposed cases (E_h) falls and comes to zero after two months. We also note that if all the controls are applied, no individual is exposed to the disease after two month. Providing treatment is the most effective control strategy (control u_2) to be applied to the exposed population. Moreover, if we simultaneously apply treatment (u_2) and preventive measures (u_1) to the exposed population, then the number of the exposed individuals decreases to zero after two months.

The curves in Figure 5 show the dynamics of I_h by using u_1 and u_2 controls. This proves that treating malaria patients and using preventive measures have a great effect in the fight against malaria. We have found that if all the controls are applied, the number of malaria infected (I_h) cases falls and becomes zero after three months. We also note that if all the controls are applied, no individual is infected with the disease after three months. To observe the effect of controls on the infected human population (I_h), we need to simultaneously apply both treatment (control u_2) and preventive measures (control u_1).

The curves of Figure 6 show the dynamics of R_h by using u_1 and u_2 controls. This proves that treating malaria patients (control u_2) has a great effect in the fight against malaria. We have found that if all the controls are applied, the number of malaria cured persons increases. Treatment (control u_2) has a significant effect on the recovery of malaria patients. Preventive measures (control u_1) have a less considerable effect on recovered individuals (R_h).

The curves of Figure 7 describe the dynamics of S_v by using u_1 and u_2 controls. This proves that treating (u_2) malaria patients and using preventive measures (u_1) have a great effect in the fight against malaria. We have found that if all the controls are applied, the number of susceptible vectors increases. If preventive measures (u_1) are applied, susceptible mosquitoes cannot become infected from the infected human population (I_h, E_h).

The curves of Figure 8 describe the dynamics of (I_v) by using u_1 and u_2 controls. This proves that treating malaria patients (I_h) has a great effect in the evolution of infected vectors (I_v). We have found that if all the controls are applied, the number of infected falls and becomes zero after ten months.

4 Conclusion

In this paper, we have developed a SEIRS malaria mathematical model. We considered model (1), in which we introduced two controls, u_1 and u_2. The control u_1 in our model shows that the best way to prevent malaria disease is to prevent contact between susceptible people (S_h) and infected vectors (I_v) by using impregnated mosquito nets and spraying public spaces. If all these preventive measures are followed, malaria will disappear after a long period of application of measures. The government should support the susceptible population (by distributing mosquito nets, spraying public areas and distributing preventive pharmaceutical products to children under the age of five years). The control u_1 in our model proves that, in order to fight against malaria, it is necessary to develop diverse strategies:

- Subsidize the access of malaria patients to hospitals or take care of all malaria patients or take care of 90% of malaria patients. Otherwise, through the infected people who are not treated for malaria, many Anopheles become infected and then spread the malaria disease.

- Reduce the population's exposure to malaria through awareness raising, free distribution of impregnated mosquito nets, access to preventive care for children under five years of age, free access to anti-malarial treatment, malaria testing, etc. We can also, in the framework of the fight against malaria, take into account the mortality rate of mosquitoes, that is to say, increase the mortality of mosquitoes by killing infected Anopheles around the population (by using insecticides or other means).

References

[1] M. Barro, A. Guiro and D. Ouedraogo. Optimal control of a SIR epidemic model with general incidence function and time delays. *Cubo (Temuco)* **20** (2) (2018) 53–66.

[2] R. V. Culshaw, S. Ruan and R. J. Spiteri. Optimal HIV treatment by maximising immune response. *Journal of mathematical biology* **48** (5) (2004) 545–562.

[3] W. H. Fleming and R. W. Rishel. *Deterministic and stochastic optimal control*, vol. 1. Springer Science & Business Media, 2012.

[4] L. M. Hocking. *Optimal control: an introduction to the theory with applications.* Oxford University Press, 1991.

[5] J. Karrakchou, M. Rachik and S. Gourari. Optimal control and infectiology: application to an hiv/aids model. *Applied mathematics and computation* **177** (2) (2006) 807–818.

[6] S. Lenhart and J. T. Workman. *Optimal control applied to biological models.* CRC press, 2007.

[7] M. Lhous, M. Rachik and A. Larrache. Free optimal time control problem for a seir-epidemic model with immigration of infective. *International Journal of Computer Applications* **159** (02) (2017).

[8] D. L. Lukes. Lie groups underlying fault avoidance in dynamical control systems. In: *Advances in Computing and Control*, pages 174–181. Springer, 2006.

[9] WHO. `https://creativecommons.org/licenses/by-nc-sa/3.0/igo/deed.fr`.

[10] S. Nanda, H. Moore and S. Lenhart. Optimal control of treatment in a mathematical model of chronic myelogenous leukemia. *Mathematical biosciences* **210** (1) (2007) 143–156.

[11] K. Okuneye and A. B. Gumel. Analysis of a temperature-and-rainfall-dependent model for malaria transmission dynamics. *Mathematical biosciences* **287** (2017) 72–92.

[12] N. Ostianu, L. Pontryagin and R. Gamkrelidze. Geometry - 1. *Journal of Mathematical Sciences (New York)* (01) (1998).

[13] H. Ouedraogo and A. Guiro. Analysis of dengue disease transmission model with general incidence functions. *Nonlinear Dynamics and Systems Theory* **23** (1) (2023) 79–94.

[14] K. P. Paaijmans, S. S. Imbahale, M. B. Thomas and W. Takken. Relevant microclimate for determining the development rate of malaria mosquitoes and possible implications of climate change. *Malaria Journal* **9** (1) (2010) 1–8.

[15] P. E. Parham and E. Michael. Modeling the effects of weather and climate change on malaria transmission. *Environmental health perspectives* **118** (5) (2010) 620–626.

[16] P. E. Parham, J. Waldock, G. K. Christophides, D. Hemming, F. Agusto, K. J. Evans, N. Fefferman, H. Gaff, A. Gumel, S. LaDeau, et al. Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370** (1665) (2015) 20130551.

[17] L. S. Pontryagin. *Mathematical theory of optimal processes.* CRC press, 1987.

[18] X. Rodó. Nitrile splitting in 2-amino-1-(arylmethyleneamino)-5-aryl-3, 4-dicyanopyrroles. *Chemistry of Heterocyclic Compounds* (12) (2013) 1627–1629.

[19] J. R. Rohr, A. P. Dobson, P. T. Johnson, A. M. Kilpatrick, S. H. Paull, T. R. Raffel, D. Ruiz-Moreno, and M. B. Thomas. Frontiers in climate change–disease research. *Trends in ecology & evolution* **26** (6) (2011) 270–277.

[20] S. B. Traore Bakary and T. Sado. A mathematical model of malaria transmission in a periodic environment. *Journal of Biological Dynamics*, 2018.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 648–662

Well-Posedness of Boundary Control System of Nonlinear Chemical Reaction

S. Sutrima *, S. Wibowo and R. Setiyowati

Department of Mathematics, University of Sebelas Maret, Ir. Sutami, no.36 A Kentingan, 57126, Surakarta, Indonesia.

Received: May 1, 2024; November 9, 2024

Abstract: We consider mixed boundary control systems inducted by non-isothermal axial dispersion chemical tubular reactors. We characterize the well-posedness, approximate controllability, and transfer function of the mixed boundary control system. There exists an admissible operator control such that the mixed boundary control system is well-posed. By constructing an extended space, the classical solution can be obtained explicitly. Sufficient conditions for approximate controllability of the mixed boundary control system are identified by the eigenvalues and eigenvectors of the Sturm-Loiuville operator using an equivalence in the extended space. A proper transfer function of the associated boundary control system equipped with an output can be constructed. The proper transfer function shows that the associated boundary control system is well-posed.

Keywords: *chemical tubular reactor; boundary control system; well-posed; approximately controllable; Sturm-Liouville operator; transfer function.*

Mathematics Subject Classification (2010): 93B18, 93B60, 93C15.

1 Introduction

A non-isothermal reaction is a reaction in the process taking place at a temperature varying from one point to another. Dynamical analysis of non-isothermal tubular reactors has been studied massively recently, see [1–6]. The dynamics of non-isothermal axial dispersion chemical tubular reactors are described by nonlinear partial differential equations (PDEs) derived from mass and energy balance equations. The nonlinearities are usually located in the kinetic terms due to the Arrhenius law for non-isothermal reactors. In particular, let L be the length of the tubular reactor and if the reaction is

* Corresponding author: mailto:sutrima@mipa.uns.ac.id

characterized by first-order kinetics with respect to the reactant concentration $C = 1$, then the reactor temperature $T(\xi, \tau)$ at the position ξ, $0 \leq \xi \leq L$, and the time τ govern the boundary problem of the nonlinear PDEs:

$$\begin{aligned} \frac{\partial T}{\partial \tau} &= D\frac{\partial T}{\partial \xi^2} - \upsilon\frac{\partial T}{\partial \xi} - \frac{\Delta H}{\rho C_p}k_0 e^{-E/RT} - \frac{4h}{\rho C_p d}(T - T_c(\tau)), \\ D\frac{\partial T}{\partial \xi}(0, \tau) &= \upsilon[T(0, \tau) - T_{\text{in}}(\tau)], \\ \frac{\partial T}{\partial \xi}(L, \tau) &= 0, \end{aligned} \tag{1}$$

where D, υ, ΔH, ρ, C_p, k_0, E, R, h, d, T_c, and T_{in} denote the energy dispersion coefficients, the superficial fluid velocity, the heat of reaction, the density, the specific heat, the kinetic constant, the activation energy, the ideal gas constant, the wall heat transfer coefficient, the reactor diameter, the coolant temperature, and the inlet temperature, respectively.

Well-posedness for the nonlinear problem (1) is crucial. There is a complex problem when dealing with the nonlinearity of infinite dimensional system. Linearization about the steady state is an approximate solution to the problem. The linearization transforms the system into a boundary control system, specifically, a mixed boundary control system with inner and boundary controls u and v, respectively, see (5). Therefore, we will focus on the linearized system, addressing its well-posedness, approximate controllability, and the well-posedness of the related input-output system.

In general, the well-posedness for the control systems is determined by the well definability and boundedness of the mappings of input to state, input to output, initial state to input, and initial state to final state [7]. For the mixed boundary control system of the linearized system of system (1), the well-posedness requires the boundedness of the existence of the admissible control operator (input-state map) B, see Definition 2.1 below. Thus, the well-posedness of the linearized system depends on the well-posedness of the state-space formulation (A, B). In the state-space, sufficient and necessary conditions for (approximate) controllability have been investigated, see [8–11]. However, for sufficiently smooth inputs, we can redefine the state space to be an extended state space, for illustration, see (15). By this construction, the sufficiency for the well-posedness and controllability of the associated problems (systems) with respect to some state space have been investigated, see [8, 12–14]. These facts guide investigations to the well-posedness and controllability for the linearized system of system (1).

Henceforth, we consider the associated state-space (A, B, C) of the linearized system of system (1), where C is the input-output map. In this space, the boundedness of C implies the well-posedness for the control system [15]. On the other hand, the boundedness of C can be identified by a system transfer function. Curtain and Weiss [16] proved that C is bounded if and only if the transfer function is uniformly bounded in a right half-plane. Therefore, to prove the well-posedness, it is enough to show that the transfer function is bounded in some right half-plane. Unfortunately, this approach is for only a few systems. In a class of structural control systems that measure acceleration at a point, the boundedness of C is proved by showing that the transfer function is proper [17]. In the paper, the justification of the transfer function was not computed directly but the properness of the transfer function is shown due to the fact that the infinitesimal generator generates an analytic semigroup. Now, one should justify the transfer function of the state-space (A, B, C) for the linearized system of system (1).

2 A Mixed Boundary Control Problem of Chemical Reactor

To facilitate analysis, the dynamic model (1) will be converted into an equivalent dimensionless distributed parameter system. By putting the new state variables

$$
\begin{aligned}
t &= \frac{\tau \upsilon}{L}, & x &= \frac{\xi}{L}, & P_e &= \frac{\upsilon}{D}, \\
B &= -\frac{\Delta H k_0 L D e^{-E/RT_0}}{\upsilon T_0}, & \gamma &= \frac{E}{RT_0}, & \beta &= \frac{4hD^2 L}{d\upsilon}, \\
z &= \frac{T - T_0}{T_0}, & u &= \frac{T_c - T_0}{T_0}, & v &= \frac{T_{\text{in}} - T_0}{T_0},
\end{aligned}
$$

where T_0 is a reference temperature and P_e is a Peclet number, we get the nonlinear dimensionless model

$$
\begin{aligned}
z_t(x,t) &= \tfrac{1}{P_e} z_{xx}(x,t) - z_x(x,t) + Be^{\gamma z(x,t)/(1+z(x,t))} + \beta[u(t) - z(x,t)], \\
z(0,t) - \tfrac{1}{P_e} z_x(0,t) &= v(t), \\
z_x(1,t) &= 0.
\end{aligned} \tag{2}
$$

Let $z_s(x), u_s, v_s$ be the steady states of the system (2), so these functions satisfy

$$
\begin{aligned}
0 &= \frac{1}{P_e}\frac{d^2 z_s}{dx^2}(x) - \frac{dz_s}{dx}(x) + Be^{\gamma z_s(x)/(1+z_s(x))} + \beta[u_s - z_s(x)], \\
v_s &= z_s(0) - \frac{1}{P_e}\frac{dz_s}{dx}(0), \\
0 &= \frac{dz_s}{dx}(1).
\end{aligned} \tag{3}
$$

By linearizing the nonlinear system (2) about the steady states and using the same symbols again, we have the linearized system

$$
\begin{aligned}
z_t(x,t) &= \tfrac{1}{P_e} z_{xx}(x,t) - z_x(x,t) + J(x)z(x,t) + \beta u(t), \\
z(0,t) - \tfrac{1}{P_e} z_x(0,t) &= v(t), \\
z_x(1,t) &= 0,
\end{aligned} \tag{4}
$$

where $J(x) = e^{\gamma z_s(x)/(1+z_s(x))}/(1 + z_s(x))^2$. We note that this problem is a distributed parameter system controlled both internally and at the boundary, and referred to as the mixed boundary control problem. We will focus on analyzing the mixed boundary control problem (4) with the interior control $u = u(t)$ and the boundary control $v = v(t)$, $t \geq 0$.

We recall the abstract mixed boundary control problem [8, 18]

$$
\begin{aligned}
\dot{z}(t) &= \mathcal{A}z(t) + B_d u(t), \quad z(0) = z_0, \\
\mathcal{B}z(t) &= v(t),
\end{aligned} \tag{5}
$$

where $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset Z \to Z$, $B_d \in \mathcal{L}(U, Z)$, $\mathcal{B} : \mathcal{D}(\mathcal{B}) \subset Z \to V$ such that $\mathcal{D}(\mathcal{A}) \subset \mathcal{D}(\mathcal{B})$, and Z, U, V are separable Hilbert spaces. We simplify the mixed boundary control system (5) by $(\mathcal{A}, \mathcal{B})$.

Definition 2.1 The mixed boundary control system $(\mathcal{A}, \mathcal{B})$ (5) is said to be well-posed if:

(a) The operator $A : \mathcal{D}(A) \to Z$, where $\mathcal{D}(A) = \mathcal{D}(\mathcal{A}) \cap \ker(\mathcal{B})$ and

$$Az = \mathcal{A}z, \quad \text{for all} \quad z \in \mathcal{D}(A), \tag{6}$$

is the infinitesimal generator of a C_0-semigroup $T(t)$ on Z;

(b) There is an admissible control operator $B \in \mathcal{L}(V, Z)$ for $T(t)$ such that for each $v \in V$, $Bv \in \mathcal{D}(\mathcal{A})$, $\mathcal{A}B \in \mathcal{L}(V, Z)$ and

$$\mathcal{B}Bv = v. \tag{7}$$

Condition (b) implies that the operator $\mathcal{B}$ is onto on V. Therefore, $\mathcal{B}$ has at least one bounded right inverse $\mathcal{F} \in \mathcal{L}(V, Z)$. In this case, we can put $B = (\mathcal{A} - A)\mathcal{F}$. Further, we can show that

$$\mathcal{A} = A + B\mathcal{B} \quad \text{and} \quad \mathcal{B}(sI - A)^{-1}B = I \tag{8}$$

for all $s \in \rho(A)$.

We begin to analyze the linearized system (4). We set $Z = L_2(0, 1)$, $U = V = \mathbb{C}$, and define the operator $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset Z \to Z$ as

$$\mathcal{A} := \frac{1}{P_e}\frac{d^2}{dx^2} - \frac{d}{dx} + J(x) \tag{9}$$

with

$$\mathcal{D}(\mathcal{A}) = \left\{h \in L_2(0, 1) : h \text{ and } \tfrac{dh}{dx} \text{ are a.c., } \tfrac{d^2h}{dx^2} \in L_2(0, 1), \tfrac{dh}{dx}(1) = 0\right\},$$

where a.c. denotes "absolutely continuous".

We define an operator $\mathcal{B} : \mathcal{D}(\mathcal{B}) \subset Z \to V$ by

$$\mathcal{B}h := h(0) - \frac{1}{P_e}\frac{dh}{dx}(0) \quad \text{with} \quad \mathcal{D}(\mathcal{B}) = \mathcal{D}(\mathcal{A}) \tag{10}$$

and an operator $A : \mathcal{D}(A) \subset Z \to Z$ by

$$A := \frac{1}{P_e}\frac{d^2}{dx^2} - \frac{d}{dx} + J(x) \tag{11}$$

with $\mathcal{D}(A) = \mathcal{D}(\mathcal{A}) \cap \ker \mathcal{B}$.

Theorem 2.1 *The linearized system* (4) *is a well-posed mixed boundary control problem.*

Proof. Let $Z = L_2(0, 1)$, $U = V = \mathbb{C}$. We consider the operators $\mathcal{A}, \mathcal{B}$, and A defined in (9), (10), and (11) on their domains, respectively. It is clear that $Az = \mathcal{A}z$ for all $z \in \mathcal{D}(A)$. We see that $A = -A_0$, where A_0 is the Sturm-Liouville operator, where $\rho(x) = P_e e^{-P_e x}$ and $p(x) = e^{-P_e x}$, see [8]. Therefore, A is closed, negative, and self-adjoint with respect to the weighted inner product

$$\langle h_1, h_2 \rangle_\rho := \int_0^1 h_1(x)\overline{h_2(x)}\rho(x)\, dx.$$

Additionally, the eigenvalues of A are real, simple, and form a decreasing sequence. The corresponding eigenfunctions are also orthogonal with respect to the weight function w. Therefore, A is the infinitesimal generator of an exponentially stable semigroup $T(t)$ on Z and $T(t) \geq 0$ for all $t \geq 0$. Let (λ_n, ϕ_n), $n \in \mathbb{N} \cup \{0\}$, be the pairs of the eigenvalue and the corresponding eigenfunction of A, we have

$$T(t)z = \sum_{n=0}^{\infty} e^{\lambda_n t} \langle z, \phi_n \rangle_\rho \phi_n \quad \text{for all} \quad z \in Z. \tag{12}$$

Henceforth, if we define an operator $Bv = b(x)v$ for all $v \in V$, where $b(x) = 1 + ce^{P_e x}$ for some constants c, then B satisfies (7). We confirm that the operators $\mathcal{A}$, A, $\mathcal{B}$, and B satisfy Definition 2.1 on Z, U, and V. We conclude that the mixed boundary control problem (4) is well-posed.

To investigate the solution explicitly, we need to reformulate equation (5) to be an abstract Cauchy problem. In this context, we have a relationship of the solution of the mixed boundary control problem (5) and the solution of the related Cauchy problem. For this purpose, several assumptions are required.

Consider the mixed boundary control system $(\mathcal{A}, \mathcal{B})$ (5) of problem (4) for $\dot{v} \in L_1([0,\tau], V)$ and $u \in L_1([0,\tau], U)$, where $\mathcal{A}, \mathcal{B}$ and A are defined in (9), (10) and (11), respectively. The related abstract Cauchy problem of (5) is

$$\begin{aligned} \dot{w}(t) &= Aw(t) - B\dot{v}(t) + \mathcal{A}Bv(t) + B_d u(t), \\ w(0) &= w_0. \end{aligned} \tag{13}$$

The assumptions guarantee the existence and uniqueness of a classical solution to problem (13) for $w_0 \in \mathcal{D}(A)$.

Theorem 2.2 *If $v \in C^2([0,\tau], V)$, $u \in C^1([0,\tau], U)$, and $w_0 \in \mathcal{D}(A)$, where $w_0 = z_0 - Bv(0)$, then problems (5) and (13) have the classical solutions related by*

$$w(t) = z(t) - Bv(t). \tag{14}$$

Moreover, problem (5) has a unique classical solution.

Proof. Let w be the classical solution of problem (13). This gives $w(t) \in \mathcal{D}(A) \subset \mathcal{D}(\mathcal{A}) \cap \mathcal{D}(\mathcal{B})$ and $Bv(t) \in \mathcal{D}(\mathcal{B})$. Since $w(t) \in \ker(\mathcal{B})$, (7) gives

$$\mathcal{B}z(t) = \mathcal{B}[w(t) + Bv(t)] = \mathcal{B}w(t) + \mathcal{B}Bv(t) = v(t).$$

Further, from equations (13) and (14), we have

$$\dot{z}(t) = \dot{w}(t) + B\dot{v}(t) = \mathcal{A}z(t) + B_d u(t).$$

Thus, the function z in (14) is the classical solution of problem (5) when w is the classical solution of problem (13).

The converse is shown similarly and the uniqueness of z follows from the uniqueness of w.

Alternately, we can reformulate problem (5) to be the abstract Cauchy problem (13) without the derivative of the boundary control term. In this context, we define an extended state space $\mathcal{P} := V \oplus Z$ and reformulate problem (13) on $\mathcal{P}$:

$$\begin{aligned} \dot{p}(t) &= \mathfrak{A}p(t) + \mathfrak{B}\mathfrak{u}(t), \\ p(0) &= p_0, \end{aligned} \tag{15}$$

where $\mathfrak{A} = \begin{bmatrix} 0 & 0 \\ \mathcal{A}B & A \end{bmatrix}$, $\mathfrak{B} = \begin{bmatrix} I & 0 \\ -B & B_d \end{bmatrix}$, $p(t) = \begin{bmatrix} v(t) \\ w(t) \end{bmatrix}$, $\mathfrak{u}(t) = \begin{bmatrix} \dot{v}(t) \\ u(t) \end{bmatrix}$, and $p_0 = \begin{bmatrix} v_0 \\ w_0 \end{bmatrix}$. We verify that the operator $\mathfrak{A}$ generates a C_0-semigroup $K(t)$ on $\mathcal{P}$, given by

$$K(t) = \begin{bmatrix} I & 0 \\ S(t) & T(t) \end{bmatrix}, \tag{16}$$

where $S(t)p_1 = \int_0^t T(t-s)\mathcal{A}Bp_1\, ds$, $p_1 \in V$.

Theorem 2.3 *If $v \in C^2([0,\tau], V)$, $u_d \in C^1([0,\tau], U)$, and $w_0 \in \mathcal{D}(A)$, then $p(t) = \begin{bmatrix} v(t) \\ w(t) \end{bmatrix}$ is the unique classical solution of problem* (15)*, where w is a unique classical solution of problem* (13)*. Moreover, if $z_0 = w_0 + Bv(0)$, then the classical solution of problem* (5) *is defined by*

$$\begin{aligned} z(t) &= \begin{bmatrix} B & I \end{bmatrix} p(t) \\ &= Bv(t) - T(t)Bv(0) + T(t)z_0 - \int_0^t T(t-s)B\dot{v}(s)\, ds + \int_0^t T(t-s)\mathcal{A}Bv(s)\, ds \\ &\quad + \int_0^t T(t-s)B_d u(s)\, ds. \end{aligned} \tag{17}$$

Proof. We see that $\mathfrak{A} := \begin{bmatrix} 0 & 0 \\ \mathcal{A}B & A \end{bmatrix}$ on $\mathcal{D}(\mathfrak{A}) = V \oplus \mathcal{D}(A)$ is the infinitesimal generator of a C_0-semigroup on $\mathcal{P}$ and $\begin{bmatrix} I & 0 \\ B & B_d \end{bmatrix} \in \mathcal{L}(V \oplus U, \mathcal{P})$. This gives that system (15) is well-defined. Moreover, the mild solution of problem (15) is given by

$$\begin{aligned} p(t) = &\begin{bmatrix} I & 0 \\ S(t) & T(t) \end{bmatrix} \begin{bmatrix} v_0 \\ w_0 \end{bmatrix} \\ &+ \int_0^t \begin{bmatrix} I & 0 \\ S(t-s) & T(t-s) \end{bmatrix} \begin{bmatrix} I & 0 \\ -B & B_d \end{bmatrix} \begin{bmatrix} \dot{v}(s) \\ u(s) \end{bmatrix} ds, \end{aligned} \tag{18}$$

where $S(t)p_1 = \displaystyle\int_0^t T(t-s)\mathcal{A}Bp_1\, ds$, $p_1 \in V$. The first component of (18) is

$$p_1(t) = v_0 + \int_0^t \dot{v}(s)\, ds = v(0) + \int_0^t \dot{v}(s)\, ds = v(t).$$

Since $p(0) = p_0 = \begin{bmatrix} v_0 \\ w_0 \end{bmatrix} \in \mathcal{D}(\mathfrak{A})$, the hypothesis and uniqueness theorem gurantee that $p(t)$ is the unique classical solution of problem (15) satisfying $\dot{p}_1(t) = \dot{v}(t)$. Also, from the second component of (18), we have

$$\dot{p_2}(t) = \mathcal{A}Bv(t) + Aw(t) - B\dot{v}(t) + B_d u(t).$$

Since $p_2(0) = w_0$, problem (13) has a unique classical solution $p_2(t) = w(t)$.

Next, if $z_0 = w_0 + Bv(0)$, then Theorem 2.2 gives

$$\begin{bmatrix} B & I \end{bmatrix} p(t) = Bv(t) + p_2(t) = Bv(t) + w(t) = z(t).$$

On the other hand, the mild solution of (13) is given by

$$w(t) = T(t)w_0 - \int_0^t T(t-s)B\dot{v}(s)\,ds + \int_0^t T(t-s)\mathcal{A}Bv(s)\,ds + \int_0^t T(t-s)B_d u(s)\,ds.$$

The last two results give (17).

3 Approximate Controllability for Mixed Boundary Control System

We will investigate an approximate controllability for the mixed boundary control system (4). If v is a differentiable control satisfying $\dot{v} \in L_2([0,\tau], V)$, then the mild solution of problem (4) is well-defined. We define the customized reachability subspace:

$$\mathcal{R}_b = \Big\{ z \in Z : \text{there are a } \tau > 0 \text{ and a differentiable control } v, \text{ with } v(0) = 0, \\ v, \dot{v} \in L_2([0,\tau], V) \text{ and } z(\tau) \text{ is the classical solution (17)} \Big\}.$$

The mixed boundary control system is said to be approximately controllable if $\mathcal{R}_b$ is dense in Z. Let $\mathcal{R}^e$ be the reachability subspace of the extended system $(\mathfrak{A}, \mathfrak{B})$ on $\mathcal{P}$, i.e.,

$$\mathcal{R}^e = \Big\{ p \in \mathcal{P} : \text{there exists a } \tau > 0 \text{ and } \mathfrak{u} \in L_2([0,\tau], V \oplus U) \text{ such that} \\ p(\tau) = \int_0^\tau K(\tau - s)\mathfrak{B}\mathfrak{u}(s)ds \Big\}.$$

Theorem 3.1 *If the extended system* $(\mathfrak{A}, \mathfrak{B})$ *is approximately controllable, then the mixed boundary control system* (4) *is also approximately controllable.*

Proof. Refer to the proof of Theorem 2.3, we have $\mathcal{R}_b = \begin{bmatrix} B & I \end{bmatrix} \mathcal{R}^e$. This implies that if $\mathcal{R}^e$ is dense in $\mathcal{P}$, then $\mathcal{R}_b$ is dense in Z.

We recall that A has real eigenvalues $\{\lambda_n : n \geq 1\}$ and a biorthogonal pair $\{(\phi_n, \phi_n) : n \geq 1\}$ due to A is self-adjoint. We consider that the operator B is finite-rank defined by

$$Bv = \sum_i^m b_i v_i, \quad b_i \in Z, \tag{19}$$

where $v = (v_1, v_2, \ldots, v_m) \in V = \mathbb{C}^m$ and $b_i(x) = 1 + c_i e^{P_e x}$ for some constants c_i. In the following, we have some results of system (15) on $\mathcal{P} = V \oplus Z$.

Lemma 3.1 *The operator* $\mathfrak{A}$ *in* (15) *has the biorthogonal pair* $\{(\tilde{\phi}_n, \tilde{\psi}_n) : n \geq 1\}$, *where*

$$\tilde{\phi}_n = \begin{bmatrix} 0 \\ \phi_n \end{bmatrix} \quad \textit{and} \quad \tilde{\psi}_n = \begin{bmatrix} \frac{1}{\lambda_n}(\mathcal{A}B)^*\phi_n \\ \phi_n \end{bmatrix},$$

whenever $\lambda_n \neq 0$.

Proof. From the hypothesis, we see that $A\phi_n = \lambda_n \phi_n$, where λ_n is real for all $n \in \mathbb{N}$. For $\mu_n \neq 0$, let $\tilde{\phi}_n = \begin{bmatrix} \tilde{\phi}_n^1 & \tilde{\phi}_n^2 \end{bmatrix}^{tr}$ and $\tilde{\psi}_n = \begin{bmatrix} \tilde{\psi}_n^1 & \tilde{\psi}_n^2 \end{bmatrix}^{tr}$, where Q^{tr} denotes the transpose of Q. Taking into account $\mathfrak{A}\tilde{\phi}_n = \mu_n \tilde{\phi}_n$ gives $\tilde{\phi}_n^1 = 0$, $\tilde{\phi}_n^2 = \phi_n$ and $\mu_n = \lambda_n$.

Similarly, $\mathfrak{A}^*\tilde{\psi}_n = \overline{\mu_n}\tilde{\psi}_n = \lambda_n\tilde{\psi}_n$ gives $(\mathcal{A}B)^*\tilde{\psi}_n^2 = \lambda_n\tilde{\psi}_n^1$ and $A^*\tilde{\psi}_n^2 = \lambda_n\tilde{\psi}_n^2$. This forces that $\tilde{\psi}_n^2 = \phi_n$ and $\tilde{\psi}_n^1 = \frac{1}{\lambda_n}(\mathcal{A}B)^*\phi_n$.

This lemma stresses that the operators A and $\mathfrak{A}$ have common nonzero eigenvalues. However, if 0 is not an eigenvalue of A ($0 \in \rho(A)$), where $\rho(A)$ is the resolvent set of A, we have the following.

Lemma 3.2 *If $0 \in \rho(A)$, then $\lambda_0 = 0$ is the eigenvalue with multiplicity m of $\mathfrak{A}$ and the corresponding biorthogonal pair*

$$\tilde{\phi}_0^i = \begin{bmatrix} e_i \\ -A^{-1}(\mathcal{A}B)e_i \end{bmatrix}, \quad \tilde{\psi}_0^i = \begin{bmatrix} e_i \\ 0 \end{bmatrix},$$

where $\{e_i : i = 1, 2, \dots, m\}$ is the usual orthonormal basis of $V = \mathbb{C}^m$.

Proof. The fact that $0 \in \rho(A)$ implies that A is invertible. Let $\tilde{\phi}_0^i = \begin{bmatrix} \tilde{\phi}_0^{i1} & \tilde{\phi}_0^{i2} \end{bmatrix}^{tr}$ and $\tilde{\psi}_0^i = \begin{bmatrix} \tilde{\psi}_0^{i1} & \tilde{\psi}_0^{i2} \end{bmatrix}^{tr}$, $i = 1, 2, \dots, m$, be the corresponding biorthogonal pair of $\mathfrak{A}$ associated with $\lambda_0 = 0$. The equation $\mathfrak{A}\tilde{\phi}_0^i = 0$ gives $\tilde{\phi}_0^{i2} = -A^{-1}(\mathcal{A}B)\tilde{\phi}_0^{i1}$. Chooshing $\tilde{\phi}_0^{i1} = e_i$ gives the assertion. Finally, $\mathfrak{A}^*\tilde{\psi}_0^i = 0$ gives $\tilde{\psi}_0^{i2} = 0$ and $\tilde{\psi}_0^{i1} = e_i$ for $i = 1, 2, \dots, m$.

Next, we assume the interior control B_d in (5) is given by

$$B_d u = \sum_{i=1}^{m} d_i u_i, \quad d_i \in Z, \tag{20}$$

where $u = (u_1, u_2, \dots, u_m) \in V = \mathbb{C}^m$. The following theorem gives the sufficiency for the approximate controllability of system (5).

Theorem 3.2 *Let $0 \in \rho(A)$. The mixed boundary control system* (5) *with the interior control* (20) *is approximately controllable if for each $n \in \mathbb{N}$,*

$$\text{rank}\,(\langle \mathcal{A}b_1 - \lambda_n b_1 + \lambda_n d_1, \phi_n\rangle_\rho, \cdots, \langle \mathcal{A}b_m - \lambda_n b_m + \lambda_n d_m, \phi_n\rangle_\rho) = 1. \tag{21}$$

Proof. According to Theorem 3.1, we prove the approximate controllability of system (15) in $\mathcal{P}$. Following the proof of Theorem 4.2.3 of [8], we need to prove that $\overline{\mathcal{R}^e} = \mathcal{P}$. However, $\overline{\mathcal{R}^e} = \mathcal{P}$ is equivalent to for each $n \geq 1$, there is a $\mathfrak{u}(t) = \begin{bmatrix} \dot{v}(t) \\ u(t) \end{bmatrix}$, where $v(t) = te_0$, $u(t) = e_0$, $e_0 = (1, 1, \dots 1) \in \mathbb{C}^m$, implying $\langle \mathfrak{B}\mathfrak{u}, \tilde{\psi}_n\rangle \neq 0$. From the definitions of $\mathfrak{B}$, $\mathfrak{u}$ in (15) and $\tilde{\psi}_n$ in Lemma 3.1, we have

$$\begin{aligned} \langle \mathfrak{B}\mathfrak{u}, \tilde{\psi}_n\rangle &= \left\langle \dot{v}, \tfrac{1}{\lambda_n}(\mathcal{A}B)^*\phi_n \right\rangle_\rho + \langle -B\dot{v} + B_d u, \phi_n\rangle_\rho \\ &= \langle \dot{v}, (\mathcal{A}B)^*\phi_n\rangle_\rho + \langle -\lambda_n B\dot{v} + \lambda_n B_d u, \phi_n\rangle_\rho \\ &= \langle \mathcal{A}B\dot{v}, \phi_n\rangle_\rho + \langle -\lambda_n B\dot{v} + \lambda_n B_d u, \phi_n\rangle_\rho \\ &= \langle \mathcal{A}B\dot{v} - \lambda_n B\dot{v} + \lambda_n B_d u, \phi_n\rangle_\rho\,. \end{aligned}$$

Equations (19) and (20) give $\langle \mathcal{A}b_i - \lambda_n b_i + \lambda_n d_i, \phi_n\rangle_\rho \neq 0$, $i = 1, 2, \dots, m$, for each $n \in \mathbb{N}$, and (21) follows.

Lemma 3.3 *Assume that A has the eigenvalue $\lambda_1 = 0$ with the eigenvector ϕ_1.*

(a) If $\langle \mathcal{A}b_i, \phi_1\rangle_\rho = 0$ *for* $i = 1, 2, \ldots, r$*, then* $\mathfrak{A}$ *has the corresponding eigenvectors* $\tilde{\phi}_0^i = \begin{bmatrix} e_i \\ y_i \end{bmatrix}$ *and biorthogonal pairs* $\tilde{\psi}_0^i = \begin{bmatrix} e_i \\ 0 \end{bmatrix}$*, where* $y_i = -\sum_{n=2}^{\infty} \frac{1}{\lambda_n}\langle \mathcal{A}b_i, \phi_n\rangle_\rho \phi_n$*, for* $i = 1, 2, \ldots, r$*.*

(b) If $\langle \mathcal{A}b_i, \phi_1\rangle_\rho \neq 0$ *for* $i = r + 1, \ldots, m$*, then* $\mathfrak{A}$ *has the generalized eigenvectors* $\tilde{\phi}_0^i = \begin{bmatrix} \langle \mathcal{A}b_i, \phi_1\rangle_\rho^{-1} e_i \\ x_i \end{bmatrix}$ *of order 2 satisfying* $\mathfrak{A}\tilde{\phi}_0^i = \begin{bmatrix} 0 \\ \phi_1 \end{bmatrix}$ *with the biorthogonal pair* $\tilde{\psi}_0^i = \begin{bmatrix} \langle \mathcal{A}b_i, \phi_1\rangle_\rho e_i \\ 0 \end{bmatrix}$*, where* $x_i = -\sum_{n=2}^{\infty} \frac{\langle \mathcal{A}b_i, \phi_n\rangle_\rho}{\lambda_n \langle \mathcal{A}b_i, \phi_1\rangle_\rho}\phi_n$ *for* $i = r + 1, \ldots, m$*.*

Proof. (a) Lemma 3.2 implies that $\lambda_1 = 0$ is the eigenvalue of $\mathfrak{A}$ with multiplicity r. Let the corresponding eigenvector of $\mathfrak{A}$ have the form $\tilde{\phi}_0^i = \begin{bmatrix} e_i \\ y_i \end{bmatrix}$. It gives $\mathcal{A}b_i + Ay_i = 0$. Multiplying this equation by $\phi_n\rho$ and simplifying, we have $\langle y_i, \phi_n\rangle_\rho = -\frac{1}{\lambda_n}\langle \mathcal{A}b_i, \phi_n\rangle_\rho$. The form of y_i follows. Then the direct calculation of $\mathfrak{A}^*\tilde{\psi}_0^i = 0$ gives $\tilde{\psi}_0^i = \begin{bmatrix} e_i \\ 0 \end{bmatrix}$.

(b) For i fixed, $r < i \leq m$, we can verify that

$$\mathfrak{A}^2\tilde{\phi}_0^i = \begin{bmatrix} 0 \\ \frac{A(\mathcal{A}b_i)}{\langle \mathcal{A}b_i, \phi_1\rangle_\rho} + A^2 x_i \end{bmatrix}.$$

The facts that $\mathcal{A}b_i \in \mathcal{D}(A)$ and A is self-adjoint with respect to the inner product $\langle \cdot, \cdot\rangle_\rho$ give the second component of $\mathfrak{A}^2\tilde{\phi}_0^i$ is 0. This confirms that $\tilde{\phi}_0^i$ is the generalized eigenvector of $\mathfrak{A}$ of order 2. Further, $\frac{A(\mathcal{A}b_i)}{\langle \mathcal{A}b_i, \phi_1\rangle_\rho} + A^2 x_i = 0$ implies that $\langle x_i, \phi_n\rangle_\rho = -\frac{\langle A(\mathcal{A}b_i), \phi_n\rangle_\rho}{\lambda_n \langle \mathcal{A}b_i, \phi_1\rangle_\rho}$. The required form of x_i follows. The biorthogonal $\tilde{\psi}_0^i$ is found easily. Finally, since $\mathcal{A}b_i$ can be expanded in ϕ_n, we have

$$\mathfrak{A}\tilde{\phi}_0^i = \begin{bmatrix} 0 \\ \frac{\mathcal{A}b_i}{\langle \mathcal{A}b_i, \phi_1\rangle_\rho} + Ax_i \end{bmatrix} = \begin{bmatrix} 0 \\ \phi_1 \end{bmatrix}.$$

Remark 3.1 For $r < i \leq m$, the set $\{\tilde{\phi}_n, \tilde{\phi}_0^i : n \in \mathbb{N}, i = 1, 2, \ldots, m\}$ generates a Riesz basis of $\mathcal{P} = \mathbb{C}^m \oplus Z$ and the operator $\mathfrak{A}$ has a spectral decomposition

$$\mathfrak{A}p = \sum_{i=r+1}^{m} \langle p, \tilde{\psi}_0^i\rangle \begin{bmatrix} 0 \\ \phi_1 \end{bmatrix} + \sum_{n=2}^{\infty} \lambda_n \langle p, \tilde{\psi}_n\rangle \begin{bmatrix} 0 \\ \phi_n \end{bmatrix}, \quad p \in \mathcal{P}.$$

Theorem 3.3 *Let* $\lambda_1 = 0$ *and* $\mathfrak{A} = \tilde{A} + A_h$*, where* $\tilde{A}$ *is a Riesz operator on* $\tilde{Y} = \text{span}_{n\geq 2}\{\tilde{\phi}_n\}$ *and the operator* A_h *is finite-rank on* $Y_h = \text{span}_{i=1,\ldots,m}\{\tilde{\phi}_1, \tilde{\phi}_0^i\}$*. If* (21) *holds for all* $n \geq 2$ *and* $\langle \mathcal{A}b_i, \phi_1\rangle \neq 0$ *for all* $i \geq r + 1$*, then the mixed boundary control system* (5) *with the interior control* (20) *is approximately controllable.*

Proof. From Exercise 4.17 of [8], the extended system $(\mathfrak{A}, \mathfrak{B})$ is approximately controllable if and only if the systems $(\tilde{A}, \tilde{B})$ and (A_h, B_h) are approximately controllable. We verify that (A_h, B_h) is exactly controllable. By Theorem 3.2, $(\tilde{A}, \tilde{B})$ is approximately controllable when (21) holds for all $n \geq 2$ and $\langle \mathcal{A}b_i, \phi_1\rangle \neq 0$ for all $i \geq r+1$. The assertion follows by Theorem 3.1.

4 Transfer Function for Boundary Control Systems

The mixed boundary control system allows to be converted to the boundary control system. Therefore, without loss of generality, we consider the boundary control system (4) with $u(t) = 0$. Moreover, we assume that the output $y(x,t)$ is a substrate concentration measured at x, $0 \leq x \leq 1$. The system equations are

$$\begin{aligned} z_t(x,t) &= \tfrac{1}{P_e} z_{xx}(x,t) - z_x(x,t) + J(x)z(x,t), \\ z(0,t) - \tfrac{1}{P_e} z_x(0,t) &= v(t), \\ z_x(1,t) &= 0, \\ y(x,t) &= Kz(x,t), \end{aligned} \tag{22}$$

where $K \in \mathcal{L}(Z,Y)$ and Y is the output space. System (22) is rewritten as the abstract boundary control system

$$\begin{aligned} \dot{z}(t) =& \mathcal{A}z(t), \quad z(0) = z_0, \\ \mathcal{B}z(t) =& v(t), \\ y(t) =& Kz(t). \end{aligned} \tag{23}$$

The triple $(\mathcal{A}, \mathcal{B}, K)$ denotes the boundary control system (23) with the output operator K.

In this paper, we focus on the boundedness of the input-output map from $v \in L_2([0,\tau], V)$ to $y \in L_2([0,\tau], Y)$.

Definition 4.1 Let $\hat{y}(s)$ and $\hat{v}(s)$ be the Laplace transform of the output and input of system (23), respectively. A system transfer function is an operator $G(s)$ such that

$$\hat{y}(s) = G(s)\hat{v}(s)$$

for all s, $\operatorname{Re} s > \sigma$ for some real σ.

The definition implies that the input-output map is well-defined and the output is Laplace transformable. Further, the system transfer function can be used to determine the boundedness of the input-output map.

Theorem 4.1 ([16]) *Let $(\mathcal{A}, \mathcal{B}, K)$ be any boundary control system. The input-output map of the system is bounded if and only if there exists a real number σ such that the transfer function $G(s)$ associated with $(\mathcal{A}, \mathcal{B}, K)$ satisfies*

$$\sup_{\operatorname{Re} s > \sigma} \|G(s)\|_{\mathcal{L}(V,Y)} < \infty.$$

The function $G(s)$ is said to be proper if the above inequality holds.

The boundary control system (23) can be written in the state-space form (A, B, C), see [19]. Here, the operators A and B satisfy Definition 2.1. The operator $C \in \mathcal{L}(W, Y)$ is defined by $C = K|_W$, where $W = \ker(\mathcal{B})$. The following refers to Theorem 2.6 of [19].

Theorem 4.2 *The input-output map of boundary control system (23) is well-defined for all inputs $v \in \mathcal{H}^2([0,\tau], V)$, $v(0) = 0$. The output can be written as*

$$y(t) = g(t) * v(t),$$

where $g(t)$ is a distribution with the Laplace transform $G(s)$. For each $s \in \rho(A)$, the operator $G(s) \in \mathcal{L}(V, Y)$ is the system transfer function given by

$$G(s) = K(sI - A)^{-1}B.$$

Proof. We consider the state-space formulation (A, B, C) of the boundary control system $(\mathcal{A}, \mathcal{B}, K)$ (23) constructed by the procedure above:

$$\begin{aligned} \dot{z}(t) &= Az(t) + Bv(t), \quad z(0) = z_0, \\ y(t) &= Cz(t), \end{aligned} \tag{24}$$

where A is the infinitesimal generator of a C_0-semigroup $T(t)$ on Z. Let z be the solution of (24). For any $\mu \in \rho(A)$, we can rewrite

$$\begin{aligned} z(t) &= (\mu I - A)^{-1}(\mu I - A)z(t) \\ &= (\mu I - A)^{-1}(\mu z(t) - \dot{z}(t)) + (\mu I - A)^{-1}Bv(t). \end{aligned} \tag{25}$$

For all initial conditions $z(0) = 0$ and smooth controls $v \in \mathcal{H}^2([0, \tau], V)$ with $v(0) = 0$, the first term in (25) is in $W \subset Z$ for each time t because $Bv \in \mathcal{D}(\mathcal{A})$, see Definition 2.1. Since A is the infinitesimal generator on Z with the domain $\mathcal{D}(A)$, $(\mu I - A)^{-1}B \in \mathcal{L}(V, \mathcal{D}(A))$. Further, for any $\mu \in \rho(A)$, Range$(\mu I - A)^{-1}B \subset Z$ and so $(\mu I - A)^{-1}B \in \mathcal{L}(V, Z)$. By applying the operator K to the solution z, we obtain the output y:

$$y(t) = K(\mu I - A)^{-1}(\mu z(t) - \dot{z}(t)) + K(\mu I - A)^{-1}Bv(t). \tag{26}$$

Since $W \subset Z$, $K(\mu I - A)^{-1} \in \mathcal{L}(\mathcal{D}(A), Y)$ and $K(\mu I - A)^{-1}B \in \mathcal{L}(V, Y)$. Since both v and z are Laplace transformable and due to the fact that z is the solution of (24), the Laplace transform of both sides of (26) gives

$$\hat{y}(s) = K(\mu I - A)^{-1}(\mu - s)(sI - A)^{-1}B\hat{v}(s) + K(\mu I - A)^{-1}B\hat{v}(s).$$

This gives the system transfer function

$$G(s) = K(\mu I - A)^{-1}(\mu - s)(sI - A)^{-1}B + K(\mu I - A)^{-1}B.$$

By replacing $\mu = s$, we obtain

$$G(s) = K(sI - A)^{-1}B \tag{27}$$

for any $s \in \rho(A)$.

On the other hand, the input-output map of system (23) is

$$y(t) = K\int_0^t T(t - r)u(r)\,dr.$$

From (27), the distribution of $G(s)$ is $g(t) = KT(t)B$. Therefore, the output can be written as

$$y(t) = \int_0^t g(t - r)u(r)\,dr = g(t) * u(t).$$

Remark 4.1 The resolvent operator $R(s) := (sI - A)^{-1}$, $s \in \rho(A)$, is given explicitly by Theorem 7.1 in [20]. The reference initiated the study of the maximal and minimal Sturm–Liouville operators, all self-adjoint restrictions of the maximal operator $T_{\max}$. Moreover, the spectrum properties of A have been also comprehensively characterized.

The form of the transfer function can be based entirely on the boundary control description (23) not on the construction of a state-space realization. The transfer function is defined in terms of an elliptic problem associated with the boundary control system.

Definition 4.2 The abstract elliptic problem $(\mathcal{A},\mathcal{B})_e$ corresponding to the boundary control system $(\mathcal{A},\mathcal{B})$ in (23) is

$$\begin{aligned} \mathcal{A}z =& sz, \quad s\in\mathbb{C},\\ \mathcal{B}z =& v. \end{aligned} \tag{28}$$

The solution $z\in Z$ is denoted by $z(s)$.

Let α be the growth bound of the semigroup associated with $(\mathcal{A},\mathcal{B})$. The elliptic problem (28) has a unique solution $z(s)$ for all v and $\operatorname{Re} s>\alpha$. The system transfer function may be described through the solution to the abstract elliptic problem (28).

Theorem 4.3 *If $(\mathcal{A},\mathcal{B},K)$ is the boundary control system (23), then there exists an $\alpha\in\mathbb{C}$ such that the system transfer function $G(s)$ is given by*

$$G(s)v = Kz(s) \quad \textit{for all} \quad s\in\mathbb{C},\ \textit{with}\ \operatorname{Re}\, s>\alpha, \tag{29}$$

where $z(s)$ is the solution to the abstract elliptic problem (28) with input v.

Proof. Let α be the growth bound of the C_0-semigroup $T(t)$ generated by A. From Theorem 4.2, for all $s\in\mathbb{C}$ with $\operatorname{Re}\, s>\alpha$, $s\in\rho(A)$, we have the transfer function $G(s)=K(sI-A)^{-1}B$. Here, we note that $\mathcal{A}=A+B\mathcal{B}$ and $\mathcal{B}(sI-A)^{-1}B=I$, see (8). Therefore, we obtain $\mathcal{A}(sI-A)^{-1}B=s(sI-A)^{-1}B$. This implies that $z(s)=(sI-A)^{-1}Bv$ is the solution of abstract elliptic problem (28). The assertion follows.

Alternatively, since $\mathcal{B}$ is onto, for any given $v\in V$, we can choose $z\in Z$ such that $\mathcal{B}z=v$. We define $G\in\mathcal{L}(V,Y)$ by

$$G(s)\mathcal{B}z := Kz - C(sI-A)^{-1}(sz-\mathcal{A}z). \tag{30}$$

The definitions of A and C guarantee that G is well-defined. If z solves the associated elliptic problem, then for any $v\in V$ and $s\in\mathbb{C}$ with $\operatorname{Re}\, s>\alpha$, (30) gives

$$G(s)v = Kz(s).$$

Example 4.1 If the desired steady-state temperature profile of system (22) is essentially uniform and the output $y(t)$ is measured at x_1, $0\le x_1\le 1$, this shows that the control system is well-posed.

From the assumptions, we may linearize about a uniform temperature $z_s(x)=$ constant, so $J(x)$ becomes a constant. Let $J(x)=c$. We use the notations relating with (9), (10), and (11) for $J(x)=c$, $Z=L_2(0,1)$ and $V=\mathbb{C}$. We have the associated C_0-semigroup $T(t)$ is given in the form (12), where λ_n is the solution of the transcendental equation

$$\begin{aligned} \tan\beta_n &= \frac{4P_e\beta_n}{4\beta_n^2-P_e^2},\\ \phi_n(x) &= B_n e^{\frac{P_e}{2}x}\left[\tfrac{2\beta_n}{P_e}\cos\beta_n x+\sin\beta_n x\right],\\ \beta_n^2 &= \frac{4P_e(c-\lambda_n)-P_e^2}{4}, \end{aligned}$$

where B_n are some constants. We note that the eigenvalues λ_n, $n \in \mathbb{N} \cup \{0\}$, form the decreasing sequence of negative real numbers, see Lemma 5.1 of [1].

The elliptic problem corresponding to system (22) when the output $y(t)$ is measured at x_1, $0 \le x_1 \le 1$, is

$$\begin{aligned} \frac{1}{P_e}\frac{d^2 z}{dx^2} - \frac{dz}{dx} + cz = sz, \quad s \in \mathbb{C}, \\ z'(1) = 0, \\ z(0) - \frac{1}{P_e} z'(0) = v, \end{aligned} \tag{31}$$

with the output equation

$$y = Kz(x_1).$$

The solution of the elliptic problem (31) is

$$z(x,s) = e^{\frac{P_e}{2}x}\left[A(v,s)\cos\beta x + B(v,s)\sin\beta x\right],$$

where

$$\begin{aligned} A(v,s) &= \frac{v[4P_e\beta\cos\beta + 2(P_e^2 - 4\beta^2 + 4\beta)\sin\beta]}{4P_e\beta\cos\beta + (P_e^2 - 4\beta^2)\sin\beta}, \\ B(v,s) &= \frac{2P_e v[2\sin\beta - P_e\cos\beta]}{4P_e\beta\cos\beta + (P_e^2 - 4\beta^2)\sin\beta}, \\ \beta^2 &= \frac{4P_e(c-s) - P_e^2}{4}. \end{aligned}$$

In this problem, we have $\|T(t)\| \le Me^{\lambda_0 t}$, where $\lambda_0 = \sup_{n\ge 0}\lambda_n$. Therefore, the growth bound $\alpha = \lambda_0$. For all $s \in \mathbb{C}$ with $\operatorname{Re} s > \lambda_0$, Theorem 4.3 gives the system transfer function

$$G(s) = Ke^{\frac{P_e}{2}x_1}\left[A(s)\cos\beta x_1 + B(s)\sin\beta x_1\right],$$

where

$$\begin{aligned} A(s) &= \frac{4P_e\beta\cos\beta + 2(P_e^2 - 4\beta^2 + 4\beta)\sin\beta}{4P_e\beta\cos\beta + (P_e^2 - 4\beta^2)\sin\beta}, \\ B(s) &= \frac{2P_e[2\sin\beta - P_e\cos\beta]}{4P_e\beta\cos\beta + (P_e^2 - 4\beta^2)\sin\beta}. \end{aligned}$$

We see that the transfer function is proper, and in virtue of Theorem 4.1, the input-output map is bounded. This implies that the control system is well-posed.

Remark 4.2 In the case $J(x)$ is not a constant, using the fact that the set $\{\phi_n\}$ is a basis for the state space Z, we can assume that $J(x)z(x,t)$ can be expressed by

$$J(x)z(x,t) = \sum_{n=0}^{\infty} f_n(t)\phi_n(x).$$

Therefore, by solving for f_n, we can use the procedure above to find the transfer function.

5 Conclusions

We concern with the mixed boundary control systems inducted by non-isothermal axial dispersion chemical tubular reactors. The well-posedness, approximate controllability, and transfer function of the mixed boundary control system can be determined. The well-posedness is identified by the operator control $B = 1 + ce^{P_e x}$. Moreover, the classical solution can be obtained using the extended space. The sufficiency for approximate controllability is justified by the eigenvalues and eigenvectors of the Sturm-Loiuville operator using the equivalence in the extended space. The proper transfer function of the associated boundary control system equipped with an output can be constructed. The transfer function ensures the well-posedness for the associated boundary control system.

Acknowledgments

The authors would like to thank the referees for their valuable comments improving this paper and Sebelas Maret University for supporting through Research Group Grant under Grant No.194.2/UN27.22/PT.01.03/2024.

References

[1] J. J. Winkin, D. Dochain and P. Ligarius. Dynamical analysis of distributed parameter tubular reactors. *Automatic* **36** (2000) 349–361.

[2] M. Laabissi, M. E. Achhab, J. J. Winkin and D. Dochain. Trajectory analysis of nonisothermal tubular reactor nonlinear models. *Syst. Control. Lett.* **42** (2001) 169–184.

[3] M. Laabissi, J. J. Winkin, D. Dochain and M. E. Achhab. Dynamical analysis of a tubular biochemical reactor infinite-dimensional nonlinear model. In: *Proc. European Control Conf.* Seville, Spain, December 12–15, 2005.

[4] B. Aylaj, M. E. Achhab and M. Laabissi. State trajectories analysis for a class of tubular reactor nonlinear nonautonomous models. *Abstr. App. Anal.* **2008** (2008) Article ID 127394 1–13.

[5] A. Hastir, F. Lamoline, J. J. Winkin and D. Dochain. Analysis of the existence of equilibrium profiles in nonisothermal axial dispersion tubular reactors. *IEEE Trans. Automat. Contr.* **65** (4) (2020) 1525–1536.

[6] S. Khatibi, G. O. Cassol and S. Dubljevic. Model predictive control of a non-isothermal axial dispersion tubular reactor with recycle. *Comput. Chem. Eng.* **145** (2021) 107159.

[7] D. Salamon. Infinite-dimensional linear systems with unbounded control and observation: A functional analytic approach. *Trans. Amer. Math. Soc.* **300** (1987) 383–431.

[8] R. F. Curtain and H. J. Zwart. *Introduction to Infinite-dimensional Linear Systems Theory.* Springer, New York, 1995.

[9] A. Bensoussan, G. Da Prato, M. C. Delfour and S. K. Mitter. *Representation and control of infinite dimensional systems.* Birkhaäuser, Boston, 2007.

[10] X. Zhao and G. Weiss. Controllability and observability of a well-posed system coupled with a finite-dimensional system. *IEEE Trans. Automat. Contr.* **56** (1) (2011) 1–12.

[11] S. Sutrima, C. R. Indrati and L. Aryati. Controllability and observability of non-autonomous Riesz-spectral systems. *Abstr. Appl. Anal.* **4210135** (2018) 1–10.

[12] J. Deutscher. Finite-dimensional dual state feedback control of linear boundary control systems. *Internat. J. Control*, **86** (2013), 41-53.

[13] I. Karafyllis and M. Krstic. Small-gain-based boundary feedback design for global exponential stabilization of one-dimensional semilinear parabolic PDEs. *SIAM J. Control Optim.* **57** (2019) 2016–2036.

[14] I. Karafyllis and M. Krstic. Global stabilization of a class of nonlinear reaction-diffusion partial differential equations by boundary feedback. *SIAM J. Control Optim.* **57** (2019) 3723–3748.

[15] D. Salamon. Realization theory in Hilbert space. *Math. Systems Theory* **21** (1989) 147–164.

[16] R. F. Curtain and G. Weiss. Well-posedness of triples of operators (in the sense of linear systems theory). *Internat. Ser. Numer. Math.* **91** (1989) 41–59.

[17] H. T. Banks and K. A. Morris. Input-output stability for accelerometer control systems. *Control Theory Adv. Tech.* **10** (1994) 1–17.

[18] S. Sutrima, M. Mardiyana, Respatiwulan, W. Sulandari and M. Yunianto. Approximate Controllability of Non-autonomous Mixed Boundary Control Systems. AIP Conference Proceedings 2326, 020037 (2021), https://doi.org/10.1063/5.0039274.

[19] A. Cheng and K. Morris. Well-posedness of boundary control systems. *SIAM J. Control Optim.* **42** (4) (2003) 1244–1265.

[20] J. Eckhardt, F. Gesztesy, R. Nichols and G. Teschl, Weyl–Titchmarsh theory for Sturm–Liouville operators with distributional potentials. *Opuscula Math.* **33** (2013) 467–563.

Nonlinear Dynamics and Systems Theory, **24** (6) (2024) 663–666

Contents of Volume 24, 2024

Volume 24 **Number 5** **2024**

Volume 24 **Number 6** **2024**