Ferdowsi University of Mashhad

Iranian Journal of Numerical Analysis and Optimization

Volume 16, Number 4

December 2026

Serial Number: 39

Ferdowsi University of Mashhad, Iran

In the Name of God

Iranian Journal of Numerical Analysis and Optimization (IJNAO)

This journal is authorized under the registration No. 174/853 dated 1386/2/26 (2007/05/16), by the Ministry of Culture and Islamic Guidance.

Volume 16, Number 4, December 2026

ISSN-Print: 2423-6977, **ISSN-Online:** 2423-6969

Publisher: Faculty of Mathematical Sciences, Ferdowsi University of Mashhad

Published by: Ferdowsi University of Mashhad Press

Printing Method: Electronic

Address: Iranian Journal of Numerical Analysis and Optimization

Faculty of Mathematical Sciences, Ferdowsi University of Mashhad

P.O. Box 1159, Mashhad 91775, Iran.

Tel. : +98-51-38806222 , **Fax:** +98-51-38807358

E-mail: ijnao@um.ac.ir

Website: http://ijnao.um.ac.ir

This journal is indexed by:

- SCOPUS
- ZbMATH Open
- ISC
- DOAJ
- Civilica
- Magiran
- Mendeley
- Academia.edu
- Linkedin

- The Journal granted the International degree by the Iranian Ministry of Science, Research, and Technology.

Iranian Journal of Numerical Analysis and Optimization

Volume 16, Number 4, December 2026

Ferdowsi University of Mashhad - Iran

©2026 All rights reserved. Iranian Journal of Numerical Analysis and Optimization

Iranian Journal of Numerical Analysis and Optimization

Director

M. H. Farahi

Editor-in-Chief

Ali R. Soheili

Managing Editor

M. Gachpazan

EDITORIAL BOARD

Abbasbandi, Saeid*

(Numerical Analysis)

Imam Khomeini International University, Iran.

e-mail: abbasbandy@ikiu.ac.ir

Abdi, Ali*

(Numerical Analysis)

University of Tabriz, Iran.

e-mail: a_abdi@tabrizu.ac.ir

Area, Iván*

(Numerical Analysis)

Universidade de Vigo, Spain.

e-mail: area@uvigo.es

Babaie Kafaki, Saman*

(Optimization)

Semnan University, Iran.

e-mail: sbk@semnan.ac.ir

Babolian, Esmail*

(Numerical Analysis)

Kharazmi University, Iran.

e-mail: babolian@khu.ac.ir

Cardone, Angelamaria*

(Numerical Analysis)

Università degli Studi di Salerno, Italy.

e-mail: ancardone@unisa.it

Dehghan, Mehdi*

(Numerical Analysis)

Amirkabir University of Technology, Iran.

e-mail: mdehghan@aut.ac.ir

Effati, Sohrab*

(Optimal Control & Optimization)

Ferdowsi University of Mashhad, Iran.

e-mail: s-effati@um.ac.ir

Emrouznejad, Ali*

(Operations Research)

Aston University, UK.

e-mail: a.emrouznejad@aston.ac.uk

Farahi, Mohammad Hadi*

(Optimal Control & Optimization)

Ferdowsi University of Mashhad, Iran.

e-mail: farahi@um.ac.ir

Gachpazan, Mortaza**
(Numerical Analysis)
Ferdowsi University of Mashhad, Iran.
e-mail: gachpazan@um.ac.ir

Ghanbari, Reza*
(Operations Research)
Ferdowsi University of Mashhad, Iran.
e-mail: rghanbari@um.ac.ir

Hadizadeh Yazdi, Mahmoud*
(Numerical Analysis)
Khaje-Nassir-Toosi University of
Technology, Iran.
e-mail: hadizadeh@kntu.ac.ir

Hojjati, Gholamreza*
(Numerical Analysis)
University of Tabriz, Iran.
e-mail: ghojjati@tabrizu.ac.ir

Hong, Jialin*
(Scientific Computing)
Chinese Academy of Sciences (CAS), China.
e-mail: hjl@lsec.cc.ac.cn

Karimi, Hamid Reza*
(Control)
Politecnico di Milano, Italy.
e-mail: hamidreza.karimi@polimi.it

Khojasteh Salkuyeh, Davod*
(Numerical Analysis)
University of Guilan, Iran.
e-mail: khojasteh@guilan.ac.ir

Lohmander, Peter*
(Optimization)
Swedish University of Agricultural Sciences, Sweden.
e-mail: Peter@Lohmander.com

Lopez-Ruiz, Ricardo*
(Complexity, nonlinear models)
University of Zaragoza, Spain.
e-mail: rilopez@unizar.es

Mahdavi-Amiri, Nezam*
(Optimization)
Sharif University of Technology, Iran.
e-mail: nezamm@sina.sharif.edu

Mirzaei, Davoud*
(Numerical Analysis)
University of Uppsala, Sweden.
e-mail: davoud.mirzaei@it.uu.se

Omrani, Khaled*
(Numerical Analysis)
University of Tunis El Manar, Tunisia.
khaled.omrani@issatso.rnu.tn

Soheili, Ali Reza*
(Numerical Analysis)
Ferdowsi University of Mashhad, Iran.
e-mail: soheili@um.ac.ir

Soleimani Damaneh, Majid*
(Operations Research and Optimization, Finance, and
Machine Learning)
University of Tehran, Iran.
e-mail: m.soleimani.d@ut.ac.ir

Toutounian, Faezeh*

(Numerical Analysis)

Ferdowsi University of Mashhad, Iran.

e-mail: toutouni@um.ac.ir

Türkyılmazoğlu, Mustafa*

(Applied Mathematics)

Hacettepe University, Turkey.

e-mail: turkyilm@hacettepe.edu.tr

Vahidian Kamyad, Ali*

(Optimal Control & Optimization)

Ferdowsi University of Mashhad, Iran.

e-mail: vahidian@um.ac.ir

Xu, Zeshui*

(Decision Making)

Sichuan University, China.

e-mail: xuzeshui@263.net

Vasagh, Zohreh

(English Text Editor)

Ferdowsi University of Mashhad, Iran.

This journal is published under the auspices of Ferdowsi University of Mashhad

* Full Professor

** Associate Professor

We would like to acknowledge the help of Miss Narjes khatoon Zohorian in the preparation of this issue.

Letter from the Editor–in–Chief

I would like to welcome you to the Iranian Journal of Numerical Analysis and Optimization (IJNAO). This journal has been published two issues per year and supported by the Faculty of Mathematical Sciences at the Ferdowsi University of Mashhad. The faculty of Mathematical Sciences with the centers of excellence and the research centers is well-known in mathematical communities in Iran.

The main aim of the journal is to facilitate discussions and collaborations between specialists in applied mathematics, especially in the fields of numerical analysis and optimization, in the region and worldwide. Our vision is that scholars from different applied mathematical research disciplines pool their insight, knowledge, and efforts by communicating via this international journal. In order to assure the high quality of the journal, each article is reviewed by subject-qualified referees. Our expectations for IJNAO are as high as any well-known applied mathematical journal in the world. We trust that by publishing quality research and creative work, the possibility of more collaborations between researchers would be provided. We invite all applied mathematicians especially in the fields of numerical analysis and optimization to join us by submitting their original work to the Iranian Journal of Numerical Analysis and Optimization.

We would like to inform all readers that the Iranian Journal of Numerical Analysis and Optimization (IJNAO), has changed its publishing frequency from "Semiannual" to a "Quarterly" journal since January 2023. The four journal issues per year will be published in the months of March, June, September, and December. One of our goals is to continue to improve the speed of both the review and publication processes, while try continuing to publish the best available international research in numerical analysis and optimization, with the high scientific and publication standards that the journal is known for.

Ali R. Soheili

Editor-in-Chief

Contents

A semi-analytical study on Hepatitis A virus transmission in Indonesia 1189
A. Maheswari, V. Ananthaswamy and J. Anantha Jothi

Second-kind Chebyshev spectral Petrov–Galerkin scheme for solving the time-fractional Euler–Bernoulli beam equation 1221
Y.H. Youssri and M.M. Muttardi

Analysis and optimal control of a proliferating-quiescent Glioma model with combined therapy 1239
T. Jaber and O. Balatif

Traveling-wave patterns and bifurcation dynamics in a spatial predator-prey model 1271
S. Hussain and M. Sen

Practical early stopping for adaptive barrier SGD: Balancing stochastic speed with validation accuracy 1295
A., Fillali, S., Addoune and R., Zeghdane

A novel collocation-based numerical method for higher-order linear ODE systems with pharmacokinetic applications 1319
F. Kaci

Conformable power series method for solving Fisher–Kolmogorov–Petrovskii–Piskunov equation 1341
D.L. Suthar, F. Desalew, M. Aychluh and G. Tirite

Impact of lifestyle interventions and adequate medical infrastructure on diabetes prevention: Dynamics analysis and optimal control 1366
M. Kalita, Z. Mondal, B. Bhuyan, B.J. Nath, H.K. Sarmah

Mathematical modeling and control strategies for meningitis transmission . . . 1401
M. Baroudi, B. Bettioui, M. Belam and A. Labzai

Design of a multi-stage hybrid optimizer combining Bayesian and evolutionary algorithms for modeling and optimizing seizure-related information in EEG signals 1432
M. Khansari and H. Afshari Rad

Coupled fixed points for a class of noncyclic contractions in uniformly convex Banach spaces with error estimates 1464
A. Safari-Hafshejani and Kh. Amiri

A multi-layer SEIRS-V network epidemic model for IoT malware propagation with heterogeneous device types and patch dynamics 1484
A. Tajari Siahmarzkooh

A collocation framework for numerical reconstruction of the source function in a two-term fractional diffusion problem 1518
S.S. Jalalian, A. Babaei and M. Gachpazan

Fractional order Chelyshkov wavelet for solving variable coefficients Caputo time-fractional convection-diffusion equation 1550
S. Sabir and A. Ahmad

Queuing analysis of a VoIP call center under retrial policy with re-service, server maintenance and service disruptions 1591
Preena R. and Sumitha D.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1189–1220
https://doi.org/10.22067/ijnao.2026.97008.1804
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A semi-analytical study on Hepatitis A virus transmission in Indonesia

A. Maheswari, V. Ananthaswamy*, and J. Anantha Jothi

Abstract

A mathematical layout that depicts the epidemiology of Hepatitis A in Indonesia is analyzed in the present research. The framework's specific four compartments —Susceptible, individuals, treatment, and recovered are resolved approximate analytically through the Laplace Adomian decomposition technique and Taylors series method. For each of the four compartments, the approximate analytical answers are stated. The outcomes of multiple kinds of model parameters are demonstrated graphically. The numerical simulation (MATLAB) and our approximate analytical outcomes are contrasted and provided in the table. The approximate analytical outcomes and the numerical simulation fit each other precisely.

AMS subject classifications (2020): 34A05, 34A12, 34E05, 34E10.

*Corresponding author

Received 15 December 2025; revised 4 February 2026; accepted 7 April 2026

Alagarsamy Maheswari
Department of Mathematics, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India. e-mail: ama@vcet.ac.in

Vembu Ananthaswamy
Research Centre and PG Department of Mathematics, The Madura College (Affiliated to Madurai Kamaraj University), Madurai, Tamil Nadu, India. e-mail: ananthu9777@gmail.com

Jeya Kumar Anantha Jothi
Research Scholar, Research Centre and PG Department of Mathematics, The Madura College (Affiliated to Madurai Kamaraj University), Madurai, Tamil Nadu, India. e-mail:ananthajothij98@gmail.com

How to cite this article
Maheswari, A., Ananthaswamy, V. and Anantha Jothi, J., A semi-analytical study on Hepatitis A virus transmission in Indonesia. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1189-1220. https://doi.org/10.22067/ijnao.2026.97008.1804

Keywords: Hepatitis A virus (HAV), $SITR$ model, Nonlinear initial value problem, Laplace Adomian decomposition method (LADM), Taylors series method (TSM), Numerical simulation.

1 Introduction

Acute, transient liver infections are caused by the Hepatitis A virus (HAV) and are typically contracted by the fecal-oral route from tainted beverages or food. The signs and symptoms can include fatigue, nausea, and jaundice, but many people do not experience any of them. Once antibodies are formed, the body's defense system offers lifetime protection, and the infection is typically not persistent. When any individual consumes food or water tainted by an infected person's excrement, it might spread. Hygiene and sanitation deficiencies (e.g., filthy and unwashed hands) and an inadequate supply of potable water are linked to the risk of contracting HAV. For the avoidance of HAV, a vaccine has been developed that is effective as well as secure. Vaccination and proper hygienic are part of safeguarding.

For the sake of the computational estimation of nonlinear and linear stochastic and predictable epidemics incorporating treatment cure and partial immunity, Arif, Abodayeh, and Nawaz [4] introduced and examined a novel finite difference technique. The purpose of the Bae et al. [6] inquiry was to examine HAV persistence on six distinct food-contact surfaces: wood, glass, stainless steel, ceramic, rubber, and plastic. Additionally, it offers helpful information for preventing food-borne illness and controlling HAV on various surfaces that handle food. According to Fiore's research [12], epidemics of Hepatitis A (HA) that are transmitted by food or water are not common in the United States. Nevertheless, people who handle food with HA are often recognized, and assessing their immunoprophylactic needs and putting control measures in place as a significant strain on public health services. In order to determine the epidemiological features of HA cases that take place at High School X as well as the risk variables, Harisma et al. [17] conducted an observational study using a cross-sectional methodology. Ilahi and Fadilaturrohmah [20] presented a mathematical framework that provided a different perspective on how the Hepatitis C virus (HCV) spreads while in medical care in a population with a chronic infection. Sensitivity analysis was required to establish parameters for treating HCV and illness prevention. The risk factors for developing fulminant hepatitis and fatality in Mexico were examined by MacKinney-Novelo et al. [24]. The susceptible-infected-treatment-recovered (SITR) scheme for the dissemination of HA by vaccination was covered in the Sidik, Sari, and Ismail's study [32]. Additionally, the stability with respect to the model for the transmission of HA disease around the sites was examined, along with the interpretation of the model created

through numerical simulations and the analysis of the impact of treatment and vaccination on the human population sick with HA.

People with untreated chronic liver disease have a higher chance of dying. The virus can be transmitted from person to person by the oral route of faeces, by medication delivery in the vein [23, 34, 35], or by consuming tainted food or beverages. Transmission can also occur through routine contact with blood or its components that contain HAV, but not with urine or saliva. According to Staes et al. [33], babies and asymptomatic, nonjaundiced HAV-infected persons are the main sources of HAV dissemination. According to research by Bianco et al. [8], travel is a second common cause of infection in areas with high infection rates, where dissemination can occur both directly and indirectly. According to Koff [22], jaundice, light-colored faces, or dark-colored urine may be signs of HAV at the onset of clinical symptoms or within a day or two of them.

Many mathematical models that describe the processes of HAV dispersion have been established, produced, and analyzed in an attempt to avoid the disease. Guimaraens and Codec [15] created a susceptible-infectious-recovery (SIR) mathematical structure to examine how the epidemiologic dynamics of HA are affected by different infection treatments. Van Effelterre, Marano, and Jacobsen [36] created a mathematical structure called susceptible, latent, infectious, and recovered immune (S-L-I-R) to highlight the importance of considering herd protection caused by early childhood HA immunization. The Fiore, Wasley, and Bell [13] model supported the Immunization Practices Advisory Committee's recommendations for the universal HA vaccination of one-year-olds. Ajelli et al. [1] developed the basic mathematical foundation for the seasonal variations of HAV in regions of Italy with respect to a moderate epidemic status in another study. Similarly, Ajelli et al. [2] developed an individual-based model that was parameterized employing the sociodemographic data and epidemiological and featured an active network of links that comprised millions of humans. The findings from the study indicate that the inadequate vaccination proportion in Italy is adequate to prevent the spread of HA. The Van Effelterre et al. [37] investigation proposed a method that provides insight into the new HAV epidemiology and may be implemented to evaluate the impact of vaccination and other public health initiatives in various settings.

Researchers were also recently investigated the dynamics and outcomes related to a vaccination option that contains HA contamination [5, 9]. For instance, the findings of Brouwer et al. [9] developed a mathematical representation of HAV transmission to assess local spread trends and the effect of vaccination upon the severity and durability of the HAV dissemination in Michigan. They also calculated the proportion of instances that the immunization program prevented in Southeast Michigan and the remaining areas of the state. For the HAV pandemic, Van

Effelterre et al. [37] created and addressed a linear age-organized SEIAR (susceptible-exposed-symptomatic-asymptomatic-recovered) subdivision model.

The research of Febriyanto and Dewi [11] served as the motivation for this work. However, their research, which incorporated the 14th order Runge-Kutta (R-K 14th) approach in MATLAB which lacks the approximate solutions in explicit and did not account for the effect of different model parameter values on HAV control. In order to fill the gap, we used an approximate analytical method and provided the solutions in explicit form. Also, we compared our approximate analytical solutions with the numerical simulation. Therefore, the primary objective of this work was to find the approximate analytical outcomes of the dynamic deterministic epidemiological framework that considers and graphically depicts the influence of several parameters. While other academics now handle these kinds of issues with implicit answers, our semi-analytical results offer explicit expressions. The semi-analytical findings and numerical simulation are subsequently contrasted and graphically shown. We employ numerical simulation, which shows a very good fit to the precision of our method.

The selection of the Laplace Adomian decomposition method (LADM) and Taylors series method (TSM) in this study is based on the need for accuracy of the results with the simulation of the SITR model, which is nonlinear. This method can solve a set of differential equations (DEs) with high precision, especially for short spans of time intervals, making it suitable for predicting the dynamics of infectious diseases over a certain period. Also, these methods provide a much smaller truncation error with the numerical simulation results.

2 Mathematical overview of the problem

The $SITR$ model, which was created by splitting the human population into four categories based on reference sources about mathematical modeling of infectious illnesses is used to explain how vaccination spreads HA disease. The population contains Susceptible (S), a population of healthy people who are at risk for experiencing HA; Infected (I), a group of people who have HA at the moment; Treatment (T), a community of people receiving medical care; and Recovery (R), a group of people that are immune to HA who have recovered. The following presumptions guided the development of the $SITR$ framework to simulate the transmission of HA illness. Populations never change.

a) Because the persons born belong to susceptible groups and each person who passes away has a proportionate rate with other groups, both the birth and death rates are equal. People with an infection or HA will go through the treatment procedures.

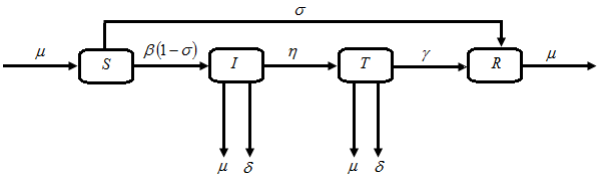


Figure 1: Schematic of the Hepatitis model ($SITR$) a vaccination-related spread

b) The infected and treatment class experienced a death yesterday due to HA infection, but the Susceptible and Recovery squad experienced a natural death.

c) Immunization of infants or kids.

d) HA infection is a possibility for the susceptible individual group.

e) When vulnerable people come into contact with infected people or the virus comes into contact with susceptible people, HA can spread.

f) After getting treatment, people will recover from HA.

g) Only one illness, HA spreads in this situation.

h) Those who recover may be able to spread the infection again.

The mathematical framework for the propagation of HA illness is derived from these hypotheses and is displayed in Figure 1.

Figure 1 provides the DE system for each compartment, which is as follows [11]:

$$\frac{dS}{d\tau} = \mu - (1-\alpha)(\beta)S\frac{1}{N} - \alpha\, S - \mu\, S, \tag{1}$$

$$\frac{dI}{d\tau} = (1-\alpha)(\beta)S\frac{1}{N} - \eta\, I - \delta\, I - \mu\, I, \tag{2}$$

$$\frac{dT}{d\tau} = \eta\, I - \gamma\, T - \delta\, T - \mu\, T, \tag{3}$$

$$\frac{dR}{d\tau} = \gamma\, T + \sigma\, S - \mu\, R. \tag{4}$$

To facilitate model analysis, the equation can be made simpler.

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad t = \frac{T}{N}, \quad r = \frac{R}{N}. \tag{5}$$

So that the governing equations may be found as follows:

$$\frac{ds}{d\tau} = \mu - (1-\sigma)(\beta)s\, i - \sigma\, s - \mu\, s, \tag{6}$$

$$\frac{di}{d\tau} = (1-\alpha)(\beta)s\, i - \eta\, i - \delta\, i - \mu\, i, \tag{7}$$

$$\frac{dt}{d\tau} = \eta\, i - \gamma\, t - \delta\, t - \mu\, t, \tag{8}$$

$$\frac{dr}{d\tau} = \gamma\, t + \sigma\, s - \mu\, r. \tag{9}$$

The corresponding initial circumstances are

$$\left.\begin{aligned} &At \qquad \tau = 0, \quad s(0) = A = 0.8, \quad i(0) = B = 0.2, \\ &t(0) = D = 0.0, \quad r(0) = F = 0.0. \end{aligned}\right\} \tag{10}$$

We identified a subsequent set of values as the initial one for each compartment using the same data source that Febriyanto and Dewi [11] used: $s(0) = 0.8$, $i(0) = 0.2$, $t(0) = 0.0$, and $r(0) = 0.0$, additionally the following parameters were ascertained: $\mu = 0.02$, $\delta = 0.0025$, $\sigma = 0.008$, $\beta = 0.4286$ and $\eta = 0.3$.

3 Approximate analytical expression for the nonlinear initial value problems

DEs, both nonlinear and linear, have a significant influence on many scientific and technical fields and can be used to represent a wide range of behaviors. A number of first-order nonlinear ordinary differential equations (ODEs) lack adequate semi-analytical results. The variational iterative technique [28], Sumudu transform methodology [10], Sumudu Adomian decomposition approach [21], homotopy analysis technique [25, 26, 31], homotopy perturbation technique [3], new homotopy perturbation approach [29], differential transform methodology [16] and Laplace Adomian decomposition method [7, 25] are some of the semi-analytical techniques that can be used to precisely resolve first-order ODEs in nonlinear.

3.1 Approximate analytical expressions for (6) through (9) by employing Laplace Adomian decomposition technique

The LADM, which blends with the Laplace transform (LT) as well as the Adomian decomposition method (ADM), is a powerful mixed technique to deal with differential equations (DEs), especially ones with nonlinearity. Its use results in better precision and effectiveness as compared to traditional approaches. LADM provides several significant advantages by solving nonlinear

DEs without the need for linearization, discrimination, perturbation, or transformation. This attribute makes problem-solving easier and allows for more precise results. LADM delivers approximate analytical outcomes in the form of a fast convergent infinite power series with easily computed terms. Because of this feature, it can be used in a wide range of scientific and engineering disciplines [7, 25, 30]. Its efficacy is shown in several fields, such as fluid dynamics, heat transfer, population dynamics, and more, where LADM has lately been effectively applied to resolve difficult problems and yield useful data [30].

In this sub-section, we demonstrate the application of the LADM to nonlinear ordinary differential systems to attain the semi-analytical expressions. To initiate the process, we commence by applying the LT which is denoted by applying the Laplace transform L for both sides of (6)–(9), we attained

$$L(s) = s(0) + L(\mu) - (1-\sigma)\beta\, L(s\, i) - (\sigma+\mu)L(s), \tag{11}$$

$$L(i) = i(0) + (1-\sigma)\beta\, L(s\, i) - (\eta+\delta+\mu)L(i), \tag{12}$$

$$L(t) = t(0) + \eta\, L(i) - (\gamma+\delta+\mu)L(t), \tag{13}$$

$$L(r) = r(0) + \gamma\, L(t) + \sigma\, L(s) - \mu\, L(r). \tag{14}$$

Assuming that the answers to s, i, t, and r may be represented as infinite series, we proceed.

Additionally, we use Adomian polynomials, whose properties are stated as follows, to split up the nonlinear terms included $s\, i = V$ in the model:

$$s = \sum_{n=0}^{\infty} s_n,\ \ i = \sum_{n=0}^{\infty} i_n,\ \ t = \sum_{n=0}^{\infty} t_n,\ \ r = \sum_{n=0}^{\infty} r_n,\ \ V = \sum_{n=0}^{\infty} V_n(s,\, i). \tag{15}$$

The Adomian polynomials represented by V_n, are described as follows:

$$V_n = \frac{1}{n!}\frac{d^n}{d\lambda^n}\left[\sum_{i=0}^{\infty}\lambda^i s_i \sum_{i=0}^{\infty}\lambda^i i_i\right], \tag{16}$$

$$V_0 = s_0 i_0, \tag{17}$$

$$V_1 = s_1 i_0 + s_0 i_1. \tag{18}$$

Equations (15) through (18) can be substituted into (11) through (14), yielding

$$L\left(\sum_{n=0}^{\infty} s_n\right) = s(0) + L(\mu) - (1-\sigma)\beta\, L\left(\sum_{n=0}^{\infty} V_n(s,\, i)\right) - (\sigma+\mu)L\left(\sum_{n=0}^{\infty} s_n\right), \tag{19}$$

$$L\left(\sum_{n=0}^{\infty} i_n\right) = i(0) + (1-\sigma)\beta\, L\left(\sum_{n=0}^{\infty} V_n(s,i)\right) - (\eta+\delta+\mu)L\left(\sum_{n=0}^{\infty}\right), \tag{20}$$

$$L\left(\sum_{n=0}^{\infty} t_n\right) = t(0) + \eta L\left(\sum_{n=0}^{\infty} i_n\right) - (\gamma+\delta+\mu)\, L\left(\sum_{n=0}^{\infty} t_n\right), \tag{21}$$

$$L\left(\sum_{n=0}^{\infty} r_n\right) = r(0) + \gamma\left(\sum_{n=0}^{\infty} t_n\right) + \sigma L\left(\sum_{n=0}^{\infty} s_n\right) - \mu L\left(\sum_{n=0}^{\infty} r_n\right). \tag{22}$$

The specific properties of the LT derivative are offered by

$$wL(\overline{U}) - \overline{U}(0) = L(\overline{G}(\overline{U})), \tag{23}$$

$$L(\overline{U}) = \frac{1}{w}\overline{U}(0) + \frac{1}{w}L(\overline{G}(\overline{U})). \tag{24}$$

Utilizing the characteristics provided in (24) in (19) through (22), we attain

$$L\left(\sum_{n=0}^{\infty} s_n\right) = \frac{1}{w}s(0) + \frac{1}{w}L(\mu) - \frac{1}{w}(1-\sigma)\beta L\left(\sum_{n=0}^{\infty} V_n(s,i)\right) - \frac{1}{w}(\sigma+\mu)L\left(\sum_{n=0}^{\infty} s_n\right), \tag{25}$$

$$L\left(\sum_{n=0}^{\infty} i_n\right) = \frac{1}{w}i(0) + \frac{1}{w}(1-\sigma)\beta L\left(\sum_{n=0}^{\infty} V_n(s,i)\right) - \frac{1}{w}(\eta+\delta+\mu)L\left(\sum_{n=0}^{\infty} i_n\right), \tag{26}$$

$$L\left(\sum_{n=0}^{\infty} t_n\right) = \frac{1}{w}t(0) + \frac{1}{w}\eta L\left(\sum_{n=0}^{\infty} i_n\right) - \frac{1}{w}(\gamma+\delta+\mu)L\left(\sum_{n=0}^{\infty} t_n\right), \tag{27}$$

$$L\left(\sum_{n=0}^{\infty} r_n\right) = \frac{1}{w}r(0) + \frac{1}{w}\gamma L\left(\sum_{n=0}^{\infty} t_n\right) + \frac{1}{w}\sigma L\left(\sum_{n=0}^{\infty} s_n\right) - \frac{1}{w}\mu L\left(\sum_{n=0}^{\infty} r_n\right). \tag{28}$$

The LADM's zeroth iteration for (25) through (28) is outlined below:

$$L(s_0) = \frac{1}{w}s(0), \tag{29}$$

$$L(i_0) = \frac{1}{w}i(0), \tag{30}$$

$$L(t_0) = \frac{1}{w}t(0), \tag{31}$$

$$L(r_0) = \frac{1}{w} r(0). \tag{32}$$

Now that we have applied the inverse LT to (29) through (32) and replaced (10), we acquire

$$s_0 = A, \tag{33}$$
$$i_0 = B, \tag{34}$$
$$t_0 = D, \tag{35}$$
$$r_0 = F. \tag{36}$$

The LADM's first iteration for (25) through (28) is obtained through contrasting the equation's two sides:

$$L(s_1) = \frac{1}{w} L(\mu) - \frac{1}{w}(1-\sigma)\beta L(V_0(s,i)) - \frac{1}{w}(\sigma+\mu)L(s_0), \tag{37}$$
$$L(i_1) = \frac{1}{w}(1-\sigma)\beta L(V_0(s,i)) - \frac{1}{w}(\eta+\delta+\mu)L(i_0), \tag{38}$$
$$L(t_1) = \frac{1}{w}\eta L(i_0) - \frac{1}{w}(\gamma+\delta+\mu)L(t_0), \tag{39}$$
$$L(r_1) = \frac{1}{w}\gamma L(t_0) + \frac{1}{w}\sigma L(s_0) - \frac{1}{w}\mu L(r_0). \tag{40}$$

The inverse LT are now being applied for (37)–(40) and putting (33)–(36), we get

$$s_1 = [\mu - (1-\sigma)\beta AB - (\sigma+\mu)A]\,\tau, \tag{41}$$
$$i_1 = [(1+\sigma)\beta AB - (\eta+\delta+\mu)B]\,\tau, \tag{42}$$
$$t_1 = [\eta B - (\gamma+\delta+\mu)D]\,\tau, \tag{43}$$
$$r_1 = [\gamma D + \sigma A - \mu F]\,\tau. \tag{44}$$

The second iteration for (25)–(28) of the LADM is attained by compare both sides of the equation, we get

$$L(s_2) = -\frac{1}{w}(1-\sigma)\beta L(V_1(s,i)) - \frac{1}{w}(\sigma+\mu)L(s_1), \tag{45}$$
$$L(i_2) = \frac{1}{w}(1-\sigma)\beta L(V_1(s,i)) - \frac{1}{w}(\eta+\delta+\mu)L(i_1), \tag{46}$$
$$L(t_2) = \frac{1}{w}\eta L(i_1) - \frac{1}{w}(\gamma+\delta+\mu)L(t_1), \tag{47}$$
$$L(r_2) = \frac{1}{w}\gamma L(t_1) + \frac{1}{w}\sigma L(s_1) - \frac{1}{w}\mu L(r_1). \tag{48}$$

Now, taking inverse LT for (45)–(48) and replaced (41)–(44), we get

$$
\begin{aligned}
s_2 =& \left[-(1-\sigma)\beta\left(A((1+\sigma)\beta AB-(\eta+\delta+\mu)B)\right)\right]\frac{\tau^2}{2!}\\
&-(1-\sigma)\beta(B(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A))\frac{\tau^2}{2!}\\
&-(\sigma+\mu)(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A)\frac{\tau^2}{2}, \qquad (49)\\
i_2 =& \left[(1-\sigma)\beta\left(A((1+\sigma)\beta AB-(\eta+\delta+\mu)B)\right)\right]\\
&+(1-\sigma)\beta(B(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A))\frac{\tau^2}{2!}\\
&+(-(\eta+\delta+\mu)(1+\sigma)\beta AB-(\eta+\delta+\mu)B)\frac{\tau^2}{2!}, \qquad (50)\\
t_2 =& \left[\eta((1+\sigma)\beta AB-(\eta+\delta+\mu)B)\right]\frac{\tau^2}{2!}\\
&-\left[(\gamma+\delta+\mu)(\eta B-(\gamma+\delta+\mu)D)\right]\frac{\tau^2}{2!}, \qquad (51)\\
r_2 =& \left[\gamma(\eta B-(\gamma+\delta+\mu)D)-\sigma(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A)\right]\frac{\tau^2}{2!}\\
&-\left[\mu(\gamma D+\sigma A-\mu F)\right]\frac{\tau^2}{2!}. \qquad (52)
\end{aligned}
$$

A similar methodology can then be used to calculate the remaining terms. The responses to the aforementioned systems can be roughly represented as an infinite series using these computed values in the manner described below:

$$s = s_0 + s_1 + s_2 + \cdots, \qquad (53)$$

$$i = i_0 + i_1 + i_2 + \cdots, \qquad (54)$$

$$t = t_0 + t_1 + t_2 + \cdots, \qquad (55)$$

$$r = r_0 + r_1 + r_2 + \cdots. \qquad (56)$$

Thus, by replacing (33)–(36), (41)–(44), and (49)–(52) into (53)–(56), we obtain the approximate analytical expressions of (6)–(9). The findings are as follows:

$$
\begin{aligned}
s =& A+\left[\mu-(1-\sigma)\beta AB-(\sigma+\mu)A\right]\tau\\
&+\left[-(1-\sigma)\beta\left(A((1+\sigma)\beta AB-(\eta+\delta+\mu)B)\right)\right]\frac{\tau^2}{2!}\\
&+\left[-(1-\beta)(B(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A))\right]\frac{\tau^2}{2!}\\
&+\left[-(\sigma+\mu)(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A)\right]\frac{\tau^2}{2!}, \qquad (57)\\
i =& B+\left[(1+\sigma)\beta AB-(\eta+\delta+\mu)B\right]\tau
\end{aligned}
$$

$$+\left[(1-\sigma)\beta\left(A((1+\sigma)\beta AB-(\eta+\delta+\mu)B)\right)\right]\frac{\tau^2}{2!}$$
$$+\left[(1-\sigma)\beta\left(B(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A)\right)\right]\frac{\tau^2}{2!}$$
$$+\left[-(\eta+\delta+\mu)(1+\sigma)\beta AB-(\eta+\delta+\mu)B\right]\frac{\tau^2}{2!}, \tag{58}$$

$$t = D+\left[\eta B-(\gamma+\delta+\mu)D\right]\tau$$
$$+\left[\eta((1+\sigma)\beta AB-(\eta+\delta+\mu)B)\right]\frac{\tau^2}{2!}$$
$$+\left[-(\gamma+\delta+\mu)(\eta B-(\gamma+\delta+\mu)D)\right]\frac{\tau^2}{2!}, \tag{59}$$

$$r = F+\left[\gamma D+\sigma A-\mu F\right]\tau$$
$$+\left[\gamma(\eta B-(\gamma+\delta+\mu)D)-\sigma(\mu-(1-\sigma)\beta AB-(\sigma+\mu)A)\right]\frac{\tau^2}{2!}$$
$$-\left[\mu(\gamma D+\sigma A-\mu F)\right]\frac{\tau^2}{2!}. \tag{60}$$

3.2 Approximate analytical expressions for (6) through (9) by employing Taylor's series approach

One of the easiest and most efficient ways to handle nonlinear equations is the TSM [14, 18, 19, 27]. Furthermore, because it doesn't require an extensive knowledge in mathematical analyzing, TSM is available to a wider range of researchers. Even though there may be certain challenges when applying TSM, such as when dealing with strong nonlinear DEs, these challenges can typically be overcome by employing additional derivatives, obtaining polynomials with higher degree, or adopting the Padé approximant. TSM was employed to provide a structured and quantitative framework for evaluating the study parameters. Its selection is justified by its ability to multiple criteria into a coherent analytical process, ensuring consistency, objectivity, and reproducibility of the results. TSM has been selected deliberately to address the need for a systematic and objective evaluation framework. Compared to alternative approaches, TSM offers improved consistency in multi-criteria assessment and minimizes subjective bias, thereby strengthening the methodological rigor of the study. Its application ensures reliable evaluation and strengthens the robustness of the methodological approach. In this work, there is a complication occurred in solving the nonlinear term. To overcome this situation TSM has been used. In TSM we substitute the initial conditions in the nonlinear term to reduce the complications within three iterations the solution are converges faster. Nonlinear ordinary and fractional DEs, including

Lane-Emden [18], third-order boundary value issues, fractional Bratu-type equations [19], and nonlinear oscillator issues [27], have been extensively solved with TSM in the past few decades.

In this sub-section, we demonstrate the application of the TSM to nonlinear ordinary differential systems to attain the semi-analytical expressions. To initiate the process, for (6) through (9), the extension of the Taylor series is expressed as follows:

$$s(\tau) = \left[\sum_{a=0}^{2} \frac{d^a s}{d\tau^a}\right]_{\tau=0} \frac{\tau^a}{a!} = s(0) + \frac{s'(0)}{1!}\tau + \frac{s''(0)}{2!}\tau^2, \tag{61}$$

$$i(\tau) = \left[\sum_{a=0}^{2} \frac{d^a i}{d\tau^a}\right]_{\tau=0} \frac{\tau^a}{a!} = i(0) + \frac{i'(0)}{1!}\tau + \frac{i''(0)}{2!}\tau^2, \tag{62}$$

$$t(\tau) = \left[\sum_{a=0}^{2} \frac{d^a t}{d\tau^a}\right]_{\tau=0} \frac{\tau^a}{a!} = t(0) + \frac{t'(0)}{1!}\tau + \frac{t''(0)}{2!}\tau^2, \tag{63}$$

$$r(\tau) = \left[\sum_{a=0}^{2} \frac{d^a r}{d\tau^a}\right]_{\tau=0} \frac{\tau^a}{a!} = r(0) + \frac{r'(0)}{1!}\tau + \frac{r''(0)}{2!}\tau^2. \tag{64}$$

By implementing the initial circumstance given in (10), the zeroth step for (61)–(64) for $a = 0$ is as follows:

$$s(0) = A, \tag{65}$$

$$i(0) = B, \tag{66}$$

$$t(0) = D, \tag{67}$$

$$r(0) = F. \tag{68}$$

The initial iteration of (61)–(64) for $a = 1$ through (65)–(68) is as follows:

$$s'(0) = \mu - (1-\sigma)\beta s(0)\, i(0) - (\sigma + \mu)s(0), \tag{69}$$

$$i'(0) = (1-\sigma)\beta\, s(0)\, i(0) - (\eta + \delta + \mu)i(0), \tag{70}$$

$$t'(0) = \eta\, i(0) - (\gamma + \delta + \mu)t(0), \tag{71}$$

$$r'(0) = \gamma\, t(0) + \sigma\, s(0) - \mu\, r(0). \tag{72}$$

The second iterations of (61)–(64) for $a = 2$ applying (65)–(68) and (69)–(72) is as outlined below:

$$s''(0) = -(1-\sigma)\beta\,[s(0)i'(0) + s'(0)i(0)] - (\sigma + \mu)s'(0), \tag{73}$$

$$i''(0) = (1-\sigma)\beta\,[s(0)i'(0) + s'(0)i(0)] - (\eta + \delta + \mu)i'(0), \tag{74}$$

$$t''(0) = \eta i'(0) - (\gamma + \delta + \mu)t'(0), \tag{75}$$

$$r''(0) = \gamma t'(0) + \sigma\, s'(0) - \mu\, r'(0). \tag{76}$$

Adding (65)–(76), we attained the approximate analytical expressions for (6)–(9), which yields:

$$\begin{aligned} s(\tau) =& A + [\mu - (1-\sigma)\beta AB - (\sigma+\mu)A]\,\frac{\tau}{1!} \\ &+ [-(1-\sigma)\beta\,(A((1-\sigma)\beta AB - (\eta+\delta+\mu)B))]\,\frac{\tau^2}{2!} \\ &+ [-(1-\sigma)\beta\,(\mu - (1-\sigma)\beta AB - (\sigma+\mu)A)\,B]\,\frac{\tau^2}{2!} \\ &+ [-(\sigma+\mu)(\mu - (1-\sigma)\beta AB - (\sigma+\mu)A)]\,\frac{\tau^2}{2!}, \end{aligned} \tag{77}$$

$$\begin{aligned} i(\tau) =& B + [(1-\sigma)\beta AB - (\eta+\delta+\mu)B]\,\frac{\tau}{1!} \\ &+ [(1-\sigma)\beta\,(A((1-\sigma)\beta AB - (\eta+\delta+\mu)B))]\,\frac{\tau^2}{2!} \\ &+ [(1-\sigma)\beta\,((\mu - (1-\sigma)\beta AB - (\sigma+\mu)A))\,B]\,\frac{\tau^2}{2!} \\ &+ [-(\eta+\delta+\mu)((1-\sigma)\beta AB - (\eta+\delta+\mu)B)]\,\frac{\tau^2}{2!}, \end{aligned} \tag{78}$$

$$\begin{aligned} t(\tau) =& D + [\eta\beta - (\gamma+\delta+\mu)D]\,\frac{\tau}{1!} + [\eta((1-\sigma)\beta AB - (\eta+\delta+\mu)B)]\,\frac{\tau^2}{2!} \\ &+ [-(\gamma+\delta+\mu)(\eta B - (\gamma+\delta+\mu)D)]\,\frac{\tau^2}{2!}, \end{aligned} \tag{79}$$

$$\begin{aligned} r(\tau) =& F + [\gamma D + \sigma A - \mu F]\,\frac{\tau}{1!} + [\gamma(\eta B - (\gamma+\delta+\mu)B)]\,\frac{\tau^2}{2!} \\ &+ [\sigma(\mu - (1-\sigma)\beta AB - (\sigma+\mu)A) - \mu(\gamma D + \sigma A - \mu F)]\,\frac{\tau^2}{2!}. \end{aligned} \tag{80}$$

4 Numerical simulation

Additionally, the function maingraph in the program run in MATLAB has been employed to numerically resolve the above nonlinear DEs (6) to (9). The numerical simulation has been contrasted with the semi-analytical results. The MATLAB program listed in Appendix A. Figures 2–18 show the numerical simulation carried out using MATLAB, together with the approximate analytical expressions that were obtained by utilizing LADM and TSM. A fairly good consensus exists.

5 Results and discussions

By assigning each parameter to the appropriate value, Figure 2 illustrates the impact of general parameters, such as susceptible class, infected class, treatment, and recovered class. In this figures, we have compared our semi-analytical solutions LADM and TSM with the numerical solution (R-K 14th approach) [11].

For susceptible class $s(\tau)$: Figures 3–4 and 11–12 displays the susceptible class $s(\tau)$ versus time τ by employing (57) and (77). Figures 3 and 11 interline that when the value of the vaccination effectiveness level σ, the transmission rate of an infected person to the suspect β raises, then the susceptible class get drops. From these we noted that the higher vaccine effectiveness σ increases immune protection and prevents individuals from remaining at risk, while a higher transmission rate β accelerates infection among nonimmune individuals together these mechanisms reduce the pool of susceptible individuals in the population. In figs. 4 and 12 interlines that the amount of the proportion of infected individuals who get medical care η raises so does the susceptible class. Therefore, an increase in the proportion of infected individuals receiving medical care η leads to earlier diagnosis, effective treatment, and isolation, which reduces disease transmission and allows fewer susceptible individuals to become infected therefore the susceptible class increases as more people remain uninfected.

For infected class $i(\tau)$: Figures 5–6 and 13–14 demonstrate the infected class $i(\tau)$ against time τ by utilizing (58) and (78). From these figs. 6 and 13, it depicts that the value for the both births or death rates that are not the outcomes of the HA sickness (naturally) μ, and the proportion of infected persons who got therapy η get raises then the infected class get falls. In general, an increase in the natural birth-death rate μ together with a higher proportion of infected individuals receiving therapy η reduces the number of infected individuals by removing infected persons through natural turnover and accelerating recovery through treatment, thereby causing the infected class to decline. Figures 5 and 14 show that as the amounts for the transmission rate of an infected individual to the vulnerable β increases, so does the infected class. Hence, when the transmission rate β increases, infected individuals spread the disease more efficiently, causing more people to become infected, so the infected class increases.

For treatment class $t(\tau)$: Figures 7–8 and 15–16 depict that the treatment class $t(\tau)$ with time τ via (59) and (79). Figures 8 and 16 reveal that the amount for the death rate due to HA infection δ, and the rate of recovery of a person after the medical care for HA infection γ get raised then the treatment class get drops. Form this we noted that the death rate from HA infection δ and the recovery rate after treatment γ increase, patients either die or recover faster, so fewer individuals remain in the treatment class, causing it to decrease. Figures 9–10 and 7 and 15, it exhibit that as the rate of transfer of an infected human to the vulnerable β and

proportion of infected individuals who got the medical care η increases so does the treatment class. Hence, when the infection spreads faster β and more infected individuals receive medical care η, more people enter treatment, so the treatment class increases.

For recovered class $r(\tau)$: Figures 9–10 and 17–18 highlight that the recovered class $r(\tau)$ versus time τ by using (60) and (80). Figures 9–10 and 17–18 it interline that the amount of vaccination effectiveness level σ, the proportion of infected individuals who receive treatment η, rate of recovery of an individual after undergoing treatment for HA infection γ raises then recovered class also get raises. From these we concluded that, when vaccine effectiveness σ improves, more infected individuals receive treatment η, and the recovery rate after treatment γ increases, more people gain immunity or recover from the infection, so the recovered class increases.

For Tables: The contrast for the numerical simulation between the LADM and TSM with the fixed quantities for the variables of vulnerable, diseased, medical care, and recovered is displayed in Tables 1 through 4. This table indicates that the average absolute error percentage was 0.1. The tables reveal that, on a comparatively small scale, our findings match well with the numerical simulation. Compared to our semi-analytical techniques, TSM was somewhat time-consuming, and evaluating values required more time than LADM. It is clear from analyzing the figures and tables that the LADM regularly offers a better approximation than the TSM with the simulations within the specified time interval. Nevertheless, even with a short time span, LADM successfully captures this behavior. This led us to conclude that the numerical simulation yields an approximation with greater precision analytical expression when employing LADM. With the aid of LADM, the solution's convergence is more precise and occurs more quickly. Additionally, second-order ordinary DEs can be resolved via this strategy.

6 Conclusion

The HA pandemic model was served as the basis for a mathematical investigation. By employing the LADM and TSM, the approximate analytical computations for each of the model's four compartments were found. The consequences of the many parameters in the model were illustrated graphically. By comparing the results with a numerical simulation (MATLAB) for each parameter, an ideal fit has been found. The results improve the model's comprehension for learners and highlight how specific parameters affect the populations. Essentially, LADM proved a more useful and reliable strategy for many practical problems, even if both approaches (LADM and TSM) can generate semi-analytical solutions. It was a more effective and dependable method for a variety of nonlinear DEs since it avoids the significant computational effort

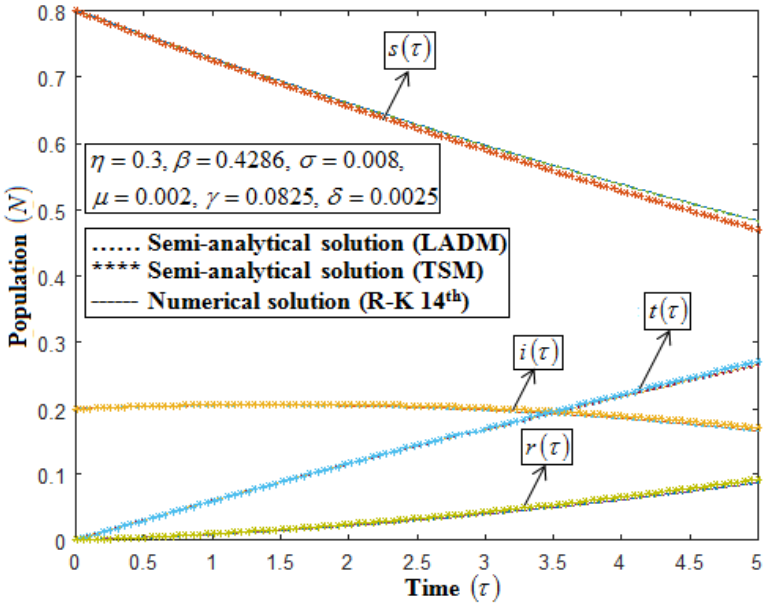


Figure 2: Comparison of approximate analytical expression with the numerical simulation for the considered epidemic model for all parameters.

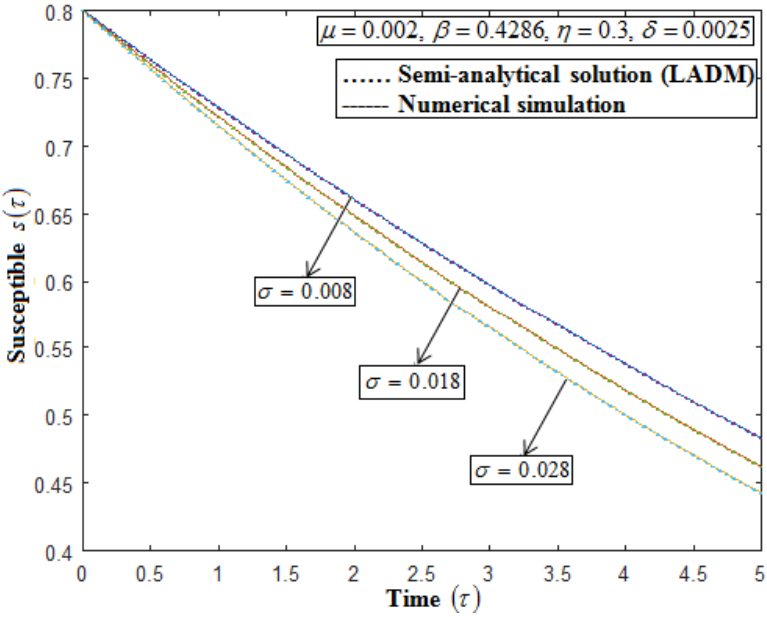


Figure 3: Curves for amount of vaccination effectiveness level σ.

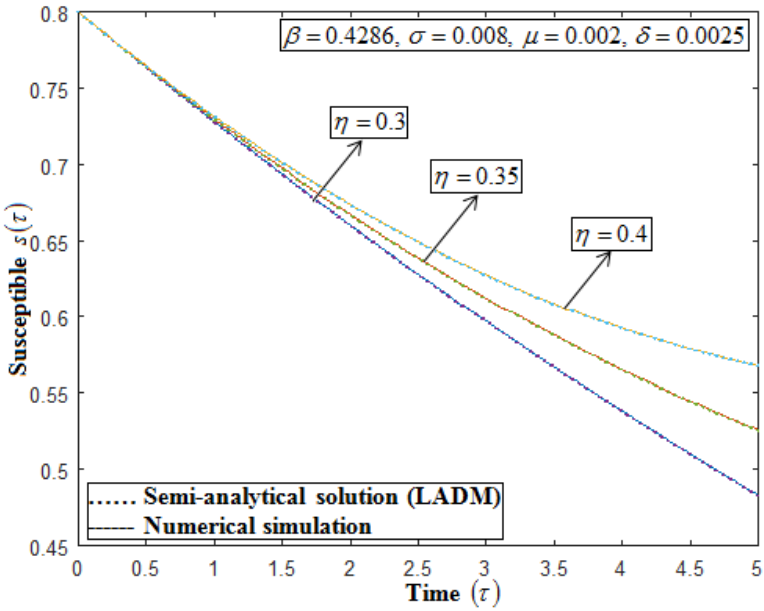


Figure 4: Consequences for proportion of infected person who receive treatment η.

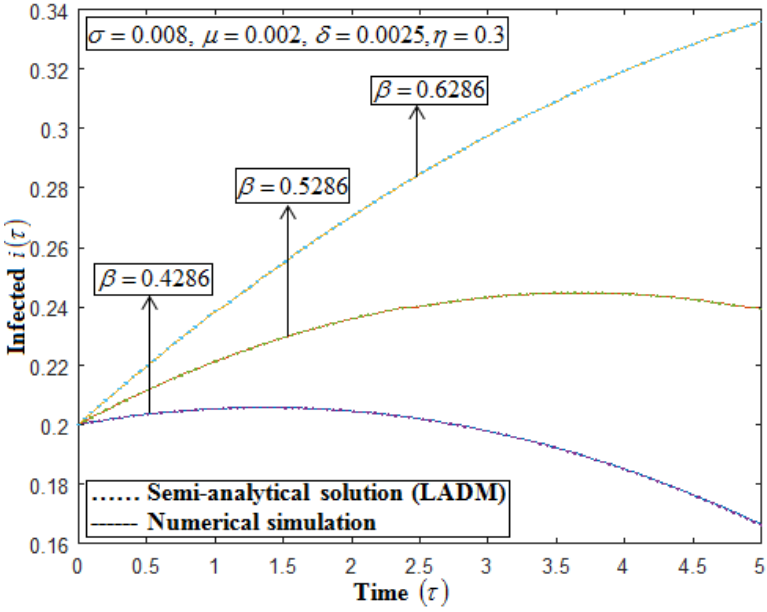


Figure 5: Influence for transmission rate of an infected individual to the suspect β.

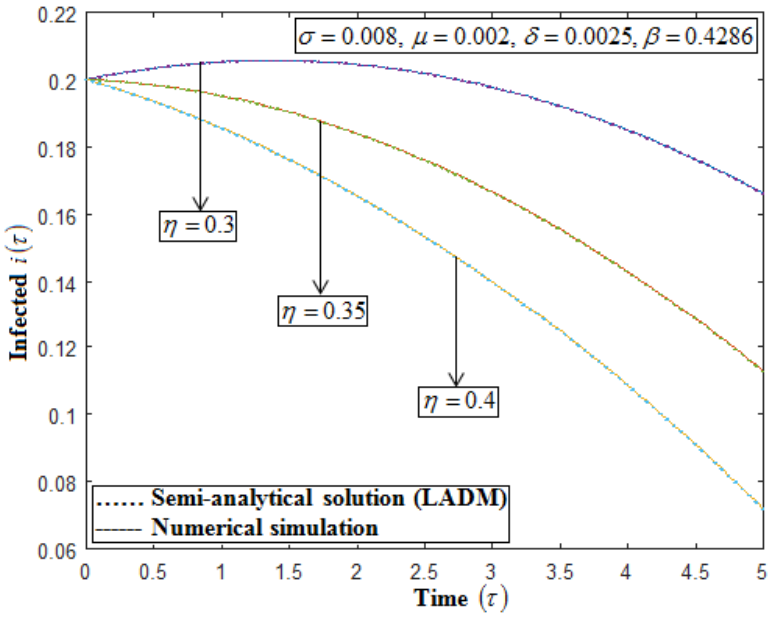


Figure 6: Plots for proportion of infected individuals who receive treatment η.

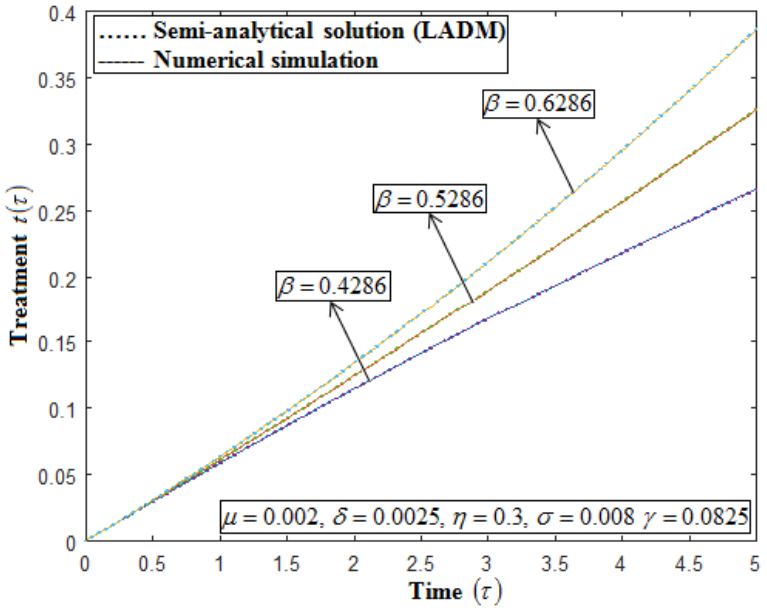


Figure 7: The impacts for transmission rate of an infected individual to the suspect β.

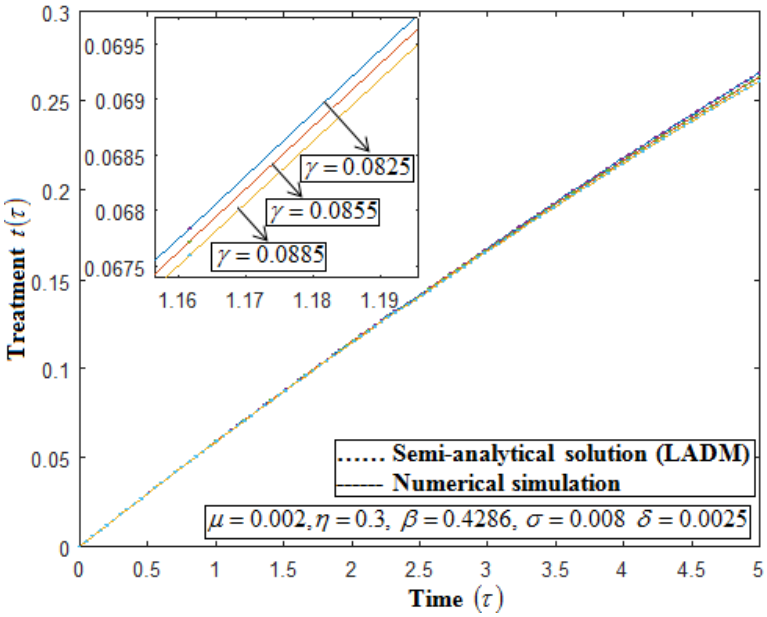


Figure 8: Influence for recovery rate of an individual after undergoing treatment for HA infection γ.

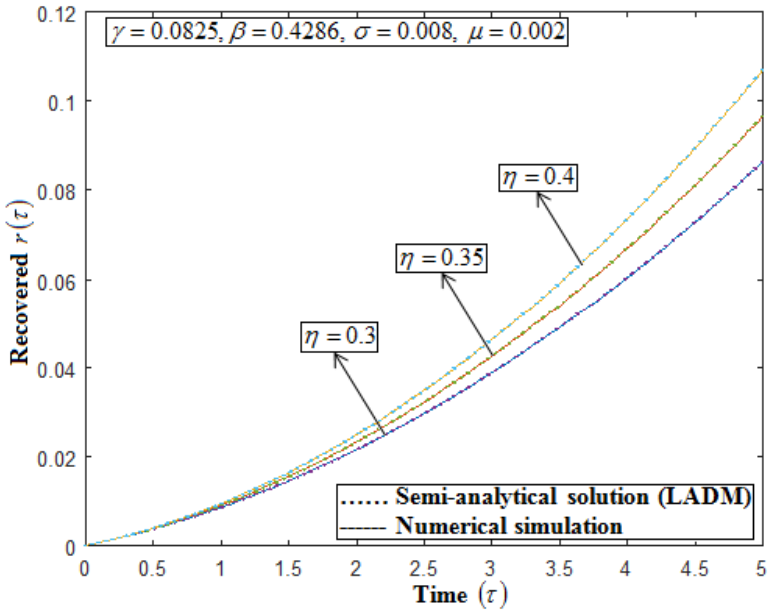


Figure 9: Consequences for proportion of infected individuals who receive treatment η.

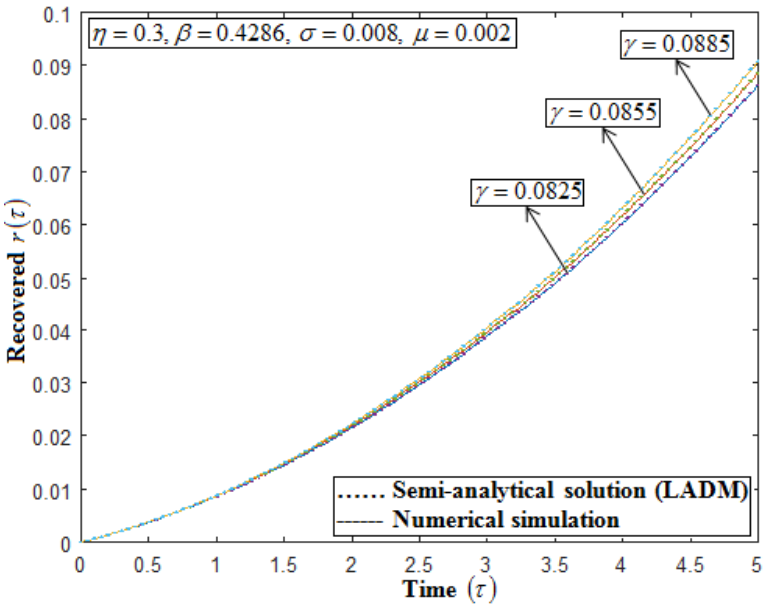


Figure 10: Impacts for recovery rate of an individual after undergoing treatment for HA infection γ.

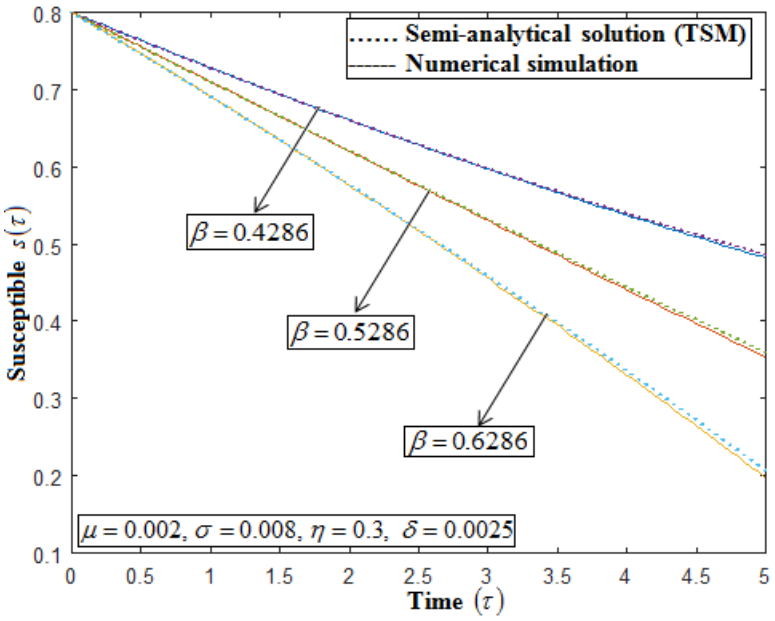


Figure 11: The impacts for transmission rate of an infected individual to the suspect β.

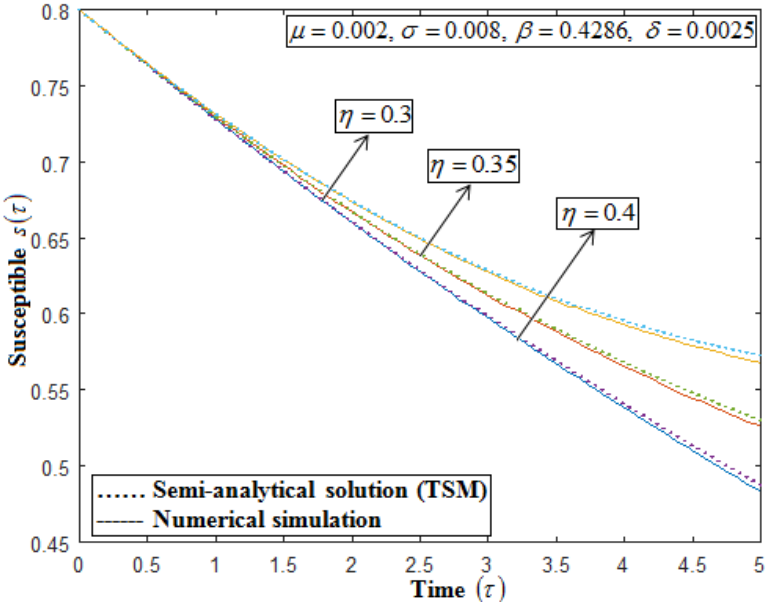


Figure 12: Consequences for proportion of infected individuals who receive treatment η.

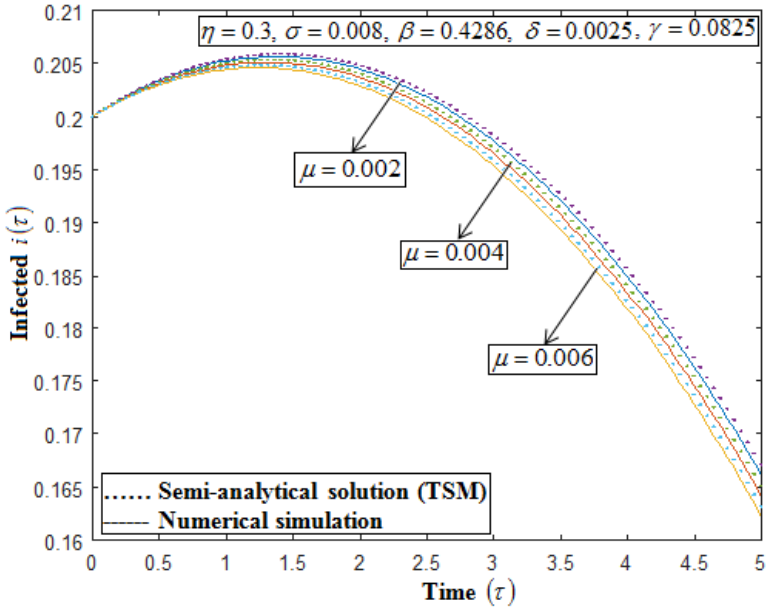


Figure 13: Variance for rate of births or deaths that are not the outcomes of HA infection (naturally) μ.

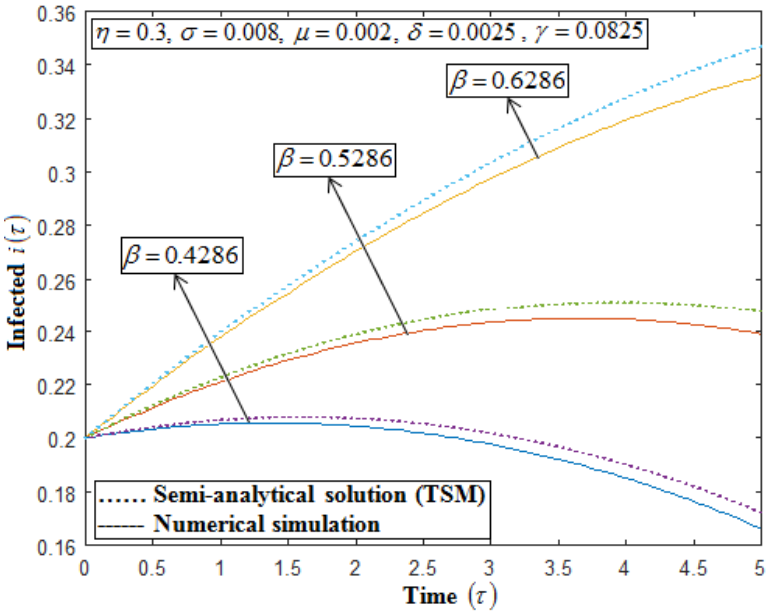


Figure 14: Plots for transmission rate of an infected individual to the suspect β.

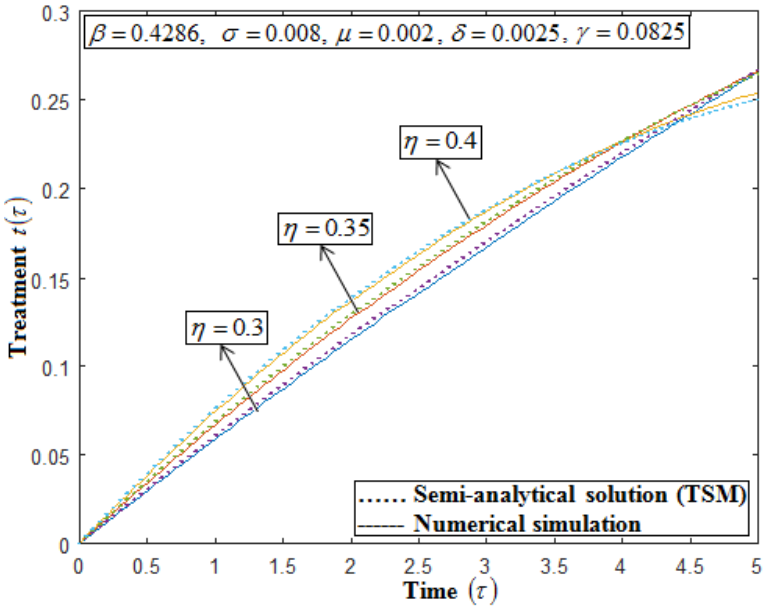


Figure 15: The effect of proportion for infected individuals who receive treatment η.

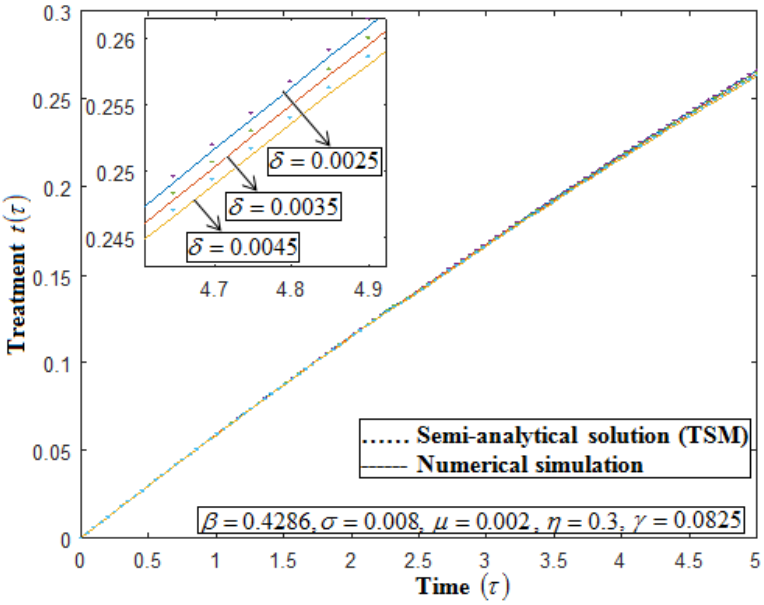


Figure 16: Influence of death rate due to HA infection δ.

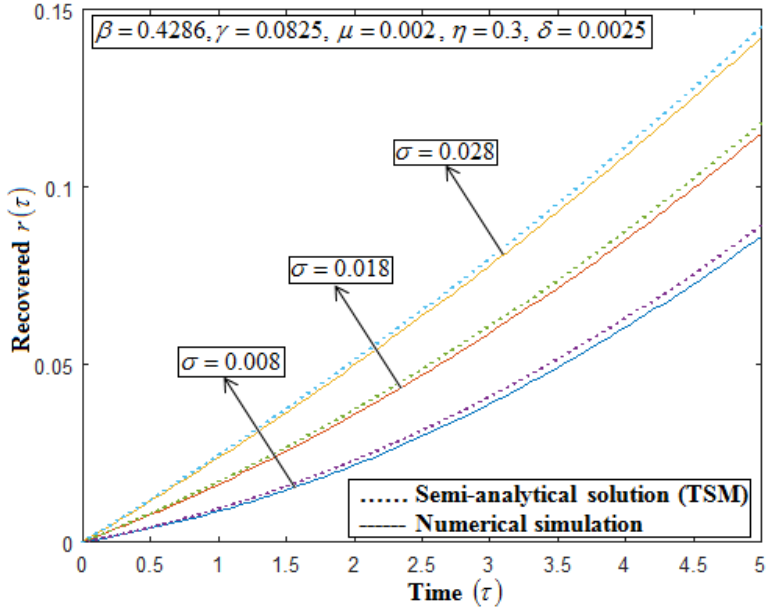


Figure 17: Variance for amount of vaccination effectiveness level σ.

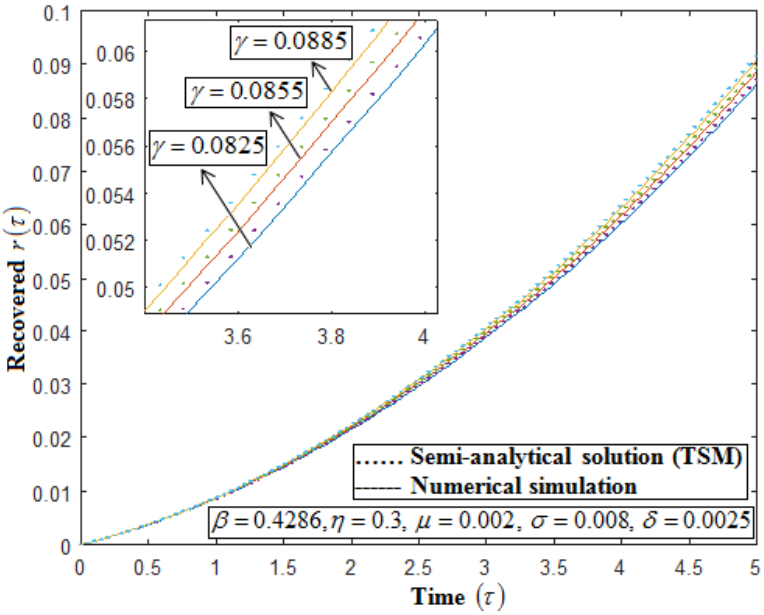


Figure 18: Impacts for recovery rate of an individual after undergoing treatment for HA infection γ.

Table 1: Contrasting numerical simulation for the susceptible class $s(\tau)$ with approximate analytical formulations for LADM and TSM with a given amount of $\mu = 0.02,\ \sigma = 0.008,\ \beta = 0.4286,\ \eta = 0.3,\ \delta = 0.0025,\ \gamma = 0.0825$.

τ	**Numerical simulation**	**LADM** (57)	**Error % of LADM**	**TSM** (77)	**Error % of TSM**
0	0.8000	0.8000	0.0000	0.8000	0.0000
1	0.7280	0.7281	0.0137	0.7283	0.0412
2	0.6602	0.6604	0.0303	0.6608	0.0909
3	0.5970	0.5970	0.0000	0.5973	0.0503
4	0.5376	0.5378	0.0372	0.5380	0.0744
5	0.4826	0.4828	0.0414	0.4830	0.0829
Absolute Average Error %			**0.0204**		**0.0566**

Table 2: The comparison of numerical simulation for the infected class $i(\tau)$ with approximate analytical formulations for LADM and TSM with an assigned value of $\mu = 0.02,\ \sigma = 0.008,\ \beta = 0.4286,\ \eta = 0.3,\ \delta = 0.0025,\ \gamma = 0.0825$.

τ	**Numerical simulation**	**LADM** (58)	**Error % of LADM**	**TSM** (78)	**Error % of TSM**
0	0.2000	0.2000	0.0000	0.2000	0.0000
1	0.2050	0.2052	0.0976	0.2055	0.2439
2	0.2042	0.2044	0.0979	0.2045	0.1469
3	0.1975	0.1977	0.1013	0.1978	0.1519
4	0.1846	0.1849	0.1625	0.1850	0.2167
5	0.1660	0.1661	0.0602	0.1663	0.1807
Absolute Average Error %			**0.0866**		**0.1567**

Table 3: Comparing numerical simulation for the treatment class $t(\tau)$ with approximate analytical formulas for LADM and TSM for a certain value of $\mu = 0.02,\ \sigma = 0.008,\ \beta = 0.4286,\ \eta = 0.3,\ \delta = 0.0025,\ \gamma = 0.0825$.

τ	**Numerical simulation**	**LADM** (59)	**Error % of LADM**	**TSM** (79)	**Error % of TSM**
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5860	0.5862	0.0341	0.5865	0.0853
2	0.1145	0.1145	0.0000	0.1147	0.1747
3	0.1675	0.1676	0.0597	0.1679	0.2388
4	0.2180	0.2180	0.0000	0.2182	0.0917
5	0.2655	0.2656	0.0377	0.2658	0.1130
Absolute Average Error %			**0.0219**		**0.1173**

Table 4: On contrasting the approximate analytical expressions for LADM and TSM along with the numerical simulation for recovered class $r(\tau)$ with particular amounts of $\mu = 0.02,\ \sigma = 0.008,\ \beta = 0.4286,\ \eta = 0.3,\ \delta = 0.0025,\ \gamma = 0.0825$.

τ	**Numerical simulation**	**LADM** (60)	**Error % of LADM**	**TSM** (80)	**Error % of TSM**
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.8570	0.8572	0.0233	0.8575	0.0583
2	0.2147	0.2149	0.0932	0.2150	0.1397
3	0.3875	0.3875	0.0000	0.3877	0.0516
4	0.6035	0.6036	0.0166	0.6038	0.0497
5	0.8630	0.8631	0.0116	0.8633	0.0348
Absolute Average Error %			**0.0241**		**0.0557**

involved in locating many higher-order derivatives in TSM. As a result, LADM proved better suited for first-order nonlinear DEs with faster convergence and greater precision. Based on the findings, the following inference may be made:

- An increase in the recovery rate from HAV indicates an increase in the recovered rate as well. As a result, the disease's recovery rates aid in preventing its spread.

- As the rate of vaccination rises, so do those receiving treatment for HAV. This led us to conclude that vaccination is necessary for the affected individuals to recover.

- When vaccination rates rise, the number of people who recover from HAV infection rises, and the number of people who are susceptible that is, who are all suspected of having HAV get drops.

This led us to the conclusion that the right vaccination should be given in order to stop the spread of HAV. In order to prevent the spread of the infections, individuals should also be made aware of the importance of getting the right vaccinations and receiving treatment for HAV.

7 Appendices

Appendix A: MATLAB programming for (6) through (9)

$functiongraphmain3$
$options = odeset('RelTol', 1e-6,' Stats',' on');$
$Xo = [0.8, 0.2, 0.0, 0.0];$
$tspan = [0, 5]; ss$
tic
$[t, x] = ode45(@TestFunction, tspan, Xo, options);$
toc
$figure$
$holdon$
$plot(t, X(:, 1),' -')$
$plot(t, X(:, 2),' -')$
$plot(t, X(:, 3),' -')$
$plot(t, X(:, 4),' -')$
$legend('S',' E',' T',' R')$
$ylabel('Population')$
$xlabel('Time')$
$return$
$function[dx_dt] = TestFunction(, x)$
$mu = 0.002, delta = 0.0025, sigma = 0.008, beta = 0.4286, gamma = 0.0825, eta = 0.3;$
$dx_dt(1) = mu - (1 - sigma) * beta * x(1) * x(2) - sigma * x(1) - mu * x(1);$
$dx_dt(2) = (1 - sigma) * beta * x(1) * x(2) - eta * x(2) - delta * x(2) - mu * x(2);$
$dx_dt(3) = eta * x(2) - gamma * x(3) - delta * x(3) - mu * x(3);$
$dx_dt(4) = gamma * x(3) + sigma * x(1) - mu * x(4);$
$dx_dt = dx_dt';$
$return$

Appendix B: Nomenclature

Symbol	Meaning
β	The transmission rate of an infected individual to the suspect
t	Treatment class
γ	The recovery rate of an individual after undergoing treatment for hepatitis A infection
σ	Vaccination effectiveness level
r	Recovered class
η	Proportion of infected individuals who receive treatment
μ	Rate of births or deaths that are not the result of HA infection (naturally)
s	Susceptible class
δ	Death rate due to hepatitis A infection
i	Individuals class
τ	Time
N	The number of the human population

Conflict of interests

The authors declare that there is no conflict of interests.

Acknowledgement

The authors are very grateful to the reviewers for carefully reading the paper and for their comments and suggestions which have improved the paper. Also, the authors are thankful to Sri. S. Natanagopal, Secretary, The Madura College Board, Dr. J. Suresh, The Principal, and Dr. C. Thangapandi, Head of the Department, The Madura College (Autonomous), Madurai, Tamil Nadu, India for their constant support to our research work.

References

[1] Ajelli, M., Fumanelli. L., Manfredi, P., and Merler, S. *Spatiotemporal dynamics of viral hepatitis A in Italy*, Theor. Popul. Biol., 79 (1-2) (2011), 1–11.

[2] Ajelli, M., Iannelli, M., Manfredi, P., and Degli Atti, M.L.C. *Basic mathematical models for the temporal dynamics of HAV in medium-endemicity Italian areas*, Vaccine, 26 (13) (2008), 1697–1707.

[3] Ananthaswamy, V., Meenakshi. S., Sivasankari, S., and Seenith Sivasundaram, *A mathematical study on the transmission dynamics of Hepatitis A virus*, Math. Eng. Sci. Aerosp. (MESA), 15 (2) (2024), 613–631.

[4] Arif, M. S., Abodayeh, K., and Nawaz, Y., *Precision in disease dynamics: Finite difference solutions for stochastic epidemics with treatment cure and partial immunity*, Partial Differ. Equ. Appl. Math., 9 (1) (2024), 1–9.

[5] Ayouni, K., Naffeti, B., and W. Ben Aribi et al., *Hepatitis A virus infection in Central-West Tunisia: an age structured model of transmission and vaccination impact*, BMC Infect. Dis., 20 (1) (2020), 1–13.

[6] Bae, S. C., Park, S. Y., Kim, A. N., Oh, M. H., and Ha, S. D., *Survival of hepatitis A virus on various food-contact surfaces during 28 days of storage at room temperature*, Food Res. Int., 64 (2014), 849–854.

[7] Bazuaye, F. E. O., *A Laplace Decomposition Analysis Of Corona Virus Disease 2019 (Covid 19) Pandemic Model*, Int. J. Math. Sci. Optim.: Theor. Appl., 6 (2) (2020), 847–861.

[8] Bianco, E., Mariano, A., Mele, A., Spada, E., and Tosti, M., *Epidemiology of acute viral hepatitis: twenty years of surveillance through SEIEVA in Italy and a review of the literature*, Rapp. ISTISAN, 12 (2006).

[9] Brouwer, A. F., Zelner, J. L., and Eisenberg, M. C., et al., *The impact of vaccination efforts on the spatiotemporal patterns of the hepatitis A outbreak in Michigan, 2016–18*, Epidemiol., 31 (5) (2020), 628–635.

[10] Farman, M., Azeem, M., and Ahmad, M.O., *Analysis of COVID-19 epidemic model with sumudu transform*, AIMS Public Health, 9 (2) (2022), 316–330.

[11] Febriyanto, E.Y., and Dewi, M. F., *Numerical analysis of Hepatitis A transmission using the susceptible-infected-treatment-recovery (SITR) model with vaccination using 14th-order Runge-Kutta method Runge Kutta method*, Rangkiang Math. J., 4 (1) (2025), 6–15.

[12] Fiore, A.E., *Hepatitis A is transmitted by Food*, Clin. Infect. Dis., 38 (5) (2004),705–715.

[13] Fiore, A.E., Wasley, A., and Bell, B.P., *Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP)*, MMWR Recomm. Rep., 55 (7) (2006), 1–23.

[14] Groza, G., and Razzaghi, M., *A Taylor series method for the solution of the linear initial-boundary-value problems for partial differential equations*, Computers Math. Appl., 66 (2013), 1329–1343.

[15] Guimaraens, M.A.D., and Codec, C.T., *Experiments with mathematical models to simulate hepatitis A population dynamics under different levels of endemicity*, Cadernos de Sa´ude P´ublica, 21 (5) (2005), 1531–1539.

[16] Harir, A., Melliani, S., El Harfi, H., and Chadli, L.S., *Variational iteration method and differential transformation method for solving the SEIR epidemic model*, Int. J. Differ. Equ., 2020 (1) (2020), 6314–6320.

[17] Harisma, F.B., Syahrul, F., Mubawadi, T., and Mirasa, Y. A., *Analisis Kejadian Luar Biasa Hepatitis A di SMA X Kabupaten Lamongan Tahun 2018*, J. Berk. Epidemiol., 6 (2) (2018), 112–121.

[18] He, J.H., and Ji, F.Y., *Taylor series solution for Lane-Emden equation*, J. Math. Chem., 57 (2019), 1932–1934.

[19] He, C.H., Shen, Y., Ji, F.Y., and He, J.-H., *Taylor series solution for fractal Bratu-type equation arising in electrospinning process*, Fract., 28 (2020), 2050011–2052193.

[20] Ilahi, F., and Fadilaturrohmah, N., *Model SEIR Untuk Penyebaran Penyakit Hepatitis C Dengan Pengobatan Pada Populasi Terinfeksi Kronis*, Jurnal Riset Dan Aplikasi Matematika (JRAM), 5 (1) (2021), 19–28.

[21] Jouhda, Z.M. and AL-Azzawi, S.N.A., *Analysis Approximate with Using Sumudu Adomian Decomposition Method for Solving SEIVR Epidemic Model*, J. Engin., 31 (2) (2025), 177–202.

[22] Koff, R. S., *Clinical manifestations and diagnosis of hepatitis A virus infection,* Vaccine, 10, Supplement 1, (1992), S15–S17.

[23] Krugman, S., and Giles, J.P., *Viral hepatitis: New light on an old disease*, JAMA, 212 (6) (1970), 1019–1029.

[24] Mac Kinney-Novelo, I., Barahona-Garrido, J., Castillo-Albarran, F., Santiago-Hernández, J.J., Méndez-Sánchez, N., Uribe, M. and Chávez-Tapia, N. *Clinical course and management of acute hepatitis A infection in adults*, Ann. Hepatol., 11 (5) (2012), 652–657.

[25] Meenakshi, S., Ananthaswamy, V., Anantha Jothi, J., and Seenith Sivasundaram *Mathematical analysis of COVID-19 disease in Malaysia using SIRS model*, Nonlinear Stud., 32 (2) (2025), 675–693.

[26] Meenakshi, S., Ananthaswamy, V., Anantha Jothi, J., and Santhi, V. K., *A Mathematical Study on SIR Epidemic Model During COVID-19*, Commun. Math. Appl., 15 (3) (2024), 1115–1128.

[27] Miletics, E., and Moln´arka, G., *Taylor series method with numerical derivatives for initial value problems*, J. Comput. Methods Sci. Eng., 4 (2004), 105–114.

[28] Mungkasi, S., *Variational iteration and successive approximation methods for a SIR epidemic model with constant vaccination strategy*, Appl. Math. Model. 90 (2021), 1–10.

[29] Narmatha, S., Ananthaswamy, V., Seenith Sivasundaram and Subha, M., *Semi analytical solution to a SIR model using new approach to homotopy perturbation method*, Math. Eng. Sci. Aerosp. (MESA), 13 (4) (2022), p. 1053.

[30] Sahu, I., and Jena, S.R., *SDIQR mathematical modelling for COVID-19 of Odisha associated with influx of migrants based on Laplace Adomian decomposition technique*, Model. Earth Syst. Environ., 9 (4) (2023), 4031– 4040.

[31] Shruthi, M., Ananthaswamy, V., and Seenith Sivasundaram, *A mathematical study on the typhoid disease model*, Math. Eng. Sci. Aerosp. (MESA), 15 (4) (2024), 1213–1231.

[32] Sidik, A. T. R., Sari, L. N., and Ismail, A., *Analisis Kestabilan Model SITR pada Penyebaran Penyakit Hepatitis A dengan Vaksinasi*, Res. Math. Nat. Sci., 1 (1) (2022), 1–11.

[33] Staes, C.J., Schlenker, T.L., Risk, I., Cannon, K.G., Harris, H., Pavia, A.T., Shapiro, C.N. and Bell, B.P. *Sources of infection among persons with acute hepatitis A and no identified*

risk factors during a sustained community-wide outbreak, Pediatrics, 106 (4) (2000), E54–E60.

[34] Stapleton, J.T., and Lemon, S.M., *Infectious Diseases, in (Chapter) Hepatitis A and hepatitis E*, Lippincott Co, Philadelphia, US, (1994), 790–797.

[35] Trujillo-Ochoa, J.L., Viera-Segura, O., and Fiery, N.A., *Challenges in management of hepatitis A virus epidemiological transition in Mexico*, Ann. Hepatol., 18 (1) (2019), 14–22.

[36] Van Effelterre, T., Marano, C., and Jacobsen, K.H., *Modeling the hepatitis A epidemiological transition in Thailand*, Vaccine, 34 (4) (2016), 555–562.

[37] Van Effelterre, T., Zink, T.K., Hoet, B.J., Hausdorff, W.P., and Rosenthal, P., *A mathematical model of hepatitis A transmission in the United States indicates value of universal childhood immunization*, Clin. Infect. Dis., 43 (2) (2006), 158–164.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1221–1238
https://doi.org/10.22067/ijnao.2026.97767.1839
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Second-kind Chebyshev spectral Petrov–Galerkin scheme for solving the time-fractional Euler–Bernoulli beam equation

Y.H. Youssri*, and M.M. Muttardi

Abstract

This work introduces a novel Petrov–Galerkin spectral technique designed to investigate vibrational behavior in fourth-order time-fractional Euler–Bernoulli beams with pinned-pinned boundary conditions. A tailored basis comprising shifted second-kind Chebyshev polynomials is developed, which automatically fulfills homogeneous boundary constraints, thus lowering computational demands. By selecting test functions from the same Chebyshev family, the Petrov–Galerkin framework converts the fractional partial differential equation into a structured algebraic system that can be efficiently resolved through Gaussian elimination. The proposed scheme delivers exponential convergence for sufficiently smooth solutions, exhibits low numerical dispersion, and remains straightforward to implement even for high-order problems. Rigorous error estimates in both L_∞ and L_2 norms ensure theoretical reliability. Computational tests, including a case lacking a closed-form exact solution, validate the precision and computational efficiency of the approach.

*Corresponding author

Received 9 February 2026; revised 14 May 2026; accepted 24 May 2026

Youssri Hassan Youssri
Department of Mathematics, Faculty of Science, Cairo University, Giza 12613, Egypt. e-mail: youssri@cu.edu.eg

Maha Muftah Muttardi
Mathematics and Actuarial Science Department, School of Science and Engineering, The American University in Cairo, New Cairo 11835, Egypt. e-mail: mahamuttardi@aucegypt.edu

How to cite this article
Youssri, Y.H. and Muttardi, M.M., Second-kind Chebyshev spectral Petrov–Galerkin scheme for solving the time-fractional Euler–Bernoulli beam equation. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1221-1238. https://doi.org/10.22067/ijnao.2026.97767.1839

AMS subject classifications (2020): Primary 65M70, 42C05; Secondary 65G99, 35R11.

Keywords: Time-fractional fourth-order partial differential equations; Chebyshev polynomials; Petrov–Galerkin method; Error analysis.

1 Introduction

Research interest in fractional differential equations has grown considerably, owing to their capacity to represent intricate physical processes exhibiting memory effects and hereditary characteristics across diverse domains such as wave dynamics, turbulent flows, signal analysis, and flow through porous structures [17, 23]. For treatments of both fractional and classical differential problems, refer to [14, 25]. Incorporating fractional derivatives of orders in $(0, 1)$ into the temporal component of the classical diffusion equation yields fractional-order differential models. Substantial theoretical and computational efforts have been devoted to the time-fractional fourth-order sub-diffusion equation. Notably, closed-form analytical solutions for most fractional differential equations remain unavailable [28]. Recent progress in numerical techniques for fractional differential equations is documented in [6, 31, 32, 33, 34]. Recent investigations have highlighted the effectiveness of hybrid spectral techniques combined with cardinal functions in delivering highly accurate numerical approximations for challenging fractional models, including multi-dimensional Rayleigh–Stokes systems [18], time-fractional three-dimensional Sobolev equations [19], and variable-order two-dimensional Wu–Zhang systems [16]. Beam models serve to characterize transversely oscillating beams subjected to external forcing, as discussed in [15]. Among these, the Euler–Bernoulli beam formulation represents one of the most fundamental frameworks capable of addressing a wide range of engineering challenges [10]. This model explicitly accounts for undamped transverse oscillations of a straight, flexible beam without support contributions.

A variety of numerical strategies have been proposed to tackle the Euler–Bernoulli equation, such as finite difference schemes [12], alternating group explicit iterative procedures [7], alternating direction implicit methods [13], the Legendre–Laguerre–Galerkin technique [11], Adomian decomposition approaches [30], and Petrov–Galerkin formulations [36].

Spectral methods are widely regarded as among the most accurate numerical or semi-analytical tools for partial differential equations (PDEs), provided they are properly implemented, due to their high-order convergence properties. Standard spectral and weighted residual techniques have been extensively investigated for solving stochastic PDEs, fractional Lane–Emden equations, linear hyperbolic first-order PDEs, one- and two-dimensional heat equations,

second-order diffusion problems, and other recent spectral approaches for diverse PDE classes [1, 2, 3, 5, 8, 20, 26, 35, 38].

Selecting suitable basis functions is critical for the success of spectral approximations. Among the many options, Chebyshev polynomials of the first kind have proven particularly effective due to their advantageous properties and trigonometric representation. Readers interested in Chebyshev polynomials may consult [21] for foundational material. In recent years, Chebyshev polynomials have been successfully employed within spectral frameworks to address numerous differential problems; see, for instance, [9, 22].

The central objective of this study is to develop an explicit Chebyshev Petrov–Galerkin method for solving a time-fractional fourth-order uniform Euler–Bernoulli beam problem with pinned-pinned boundary conditions. The remainder of this manuscript is organized as follows: Section 2 reviews essential concepts from fractional calculus and key properties of shifted second-kind Chebyshev polynomials (S2KCPs). Section 3 presents the core spectral methodology for addressing the underlying fractional PDE. Section 4 addresses the estimation of truncation errors. Section 5 provides numerical illustrations with comparative assessments. Finally, Section 6 offers concluding observations.

2 Fundamental relations

This section introduces the Caputo fractional derivative and relevant characteristics of S2KCPs.

2.1 Caputo fractional derivative

Definition 1. [23] The Caputo fractional derivative of order $r > 0$ is given by

$$D_t^r \chi(t) = \frac{1}{\Gamma(s-r)} \int_0^t (t-z)^{s-r-1} \chi^{(s)}(z)\, dz, \quad t > 0, \tag{1}$$

where $s = \lceil r \rceil$ (the smallest integer greater than or equal to r), ensuring $s - 1 < r \leq s$. This convention aligns with the standard Caputo notation $n - 1 < \alpha \leq n$ commonly adopted in the literature, thereby avoiding ambiguity with the alternative form $s - 1 \leq r < s$.

The operator D_t^r possesses the following properties:

$$D_t^r b = 0 \quad (b \text{ is a constant}), \tag{2}$$

$$D_t^r\, t^s = \begin{cases} 0, & \text{if } s \in \mathbb{N}_0 \text{ and } s < \lceil r \rceil, \\ \frac{\Gamma(s+1)}{\Gamma(s-r+1)}\, t^{s-r}, & \text{if } s \in \mathbb{N}_0 \text{ and } s \geq \lceil r \rceil, \end{cases} \tag{3}$$

where $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$.

2.2 Essential properties of S2KCPs

The S2KCPs defined on $[0, 1]$ are given by $\{U_m^*(\iota) = U_m(2\iota - 1) : m \in \mathbb{N}_0\}$, with the explicit representation [9]:

$$U_m^*(\iota) = m \sum_{k=0}^{m} \frac{(-1)^{m+k}\, 2^{2k}\, (m+k+1)!}{(m-k)!\, (2k+1)!}\, \iota^k, \quad m \geq 0. \tag{4}$$

These polynomials satisfy an orthogonality relation with respect to the weight $\hat{w}(\iota) = \sqrt{\iota(1-\iota)}$:

$$\int_0^1 \hat{w}(\iota)\, U_m^*(\iota)\, U_n^*(\iota)\, d\iota = g_{m,n}, \tag{5}$$

where

$$g_{m,n} = \begin{cases} \frac{\pi}{8}, & m = n, \\ 0, & m \neq n. \end{cases} \tag{6}$$

Remark 1. The weight function $\hat{w}(\iota) = \sqrt{\iota(1-\iota)}$ arises from the standard Chebyshev weight $\sqrt{1-x^2}$ on $[-1, 1]$ via the affine transformation $x = 2\iota - 1$, since $dx = 2\, d\iota$ and $\sqrt{1-x^2} = 2\sqrt{\iota(1-\iota)}$. The factor of 2 is incorporated into the normalization constant $g_{m,n}$.

The three-term recurrence relation reads as follows:

$$U_n^*(\iota) = 2(2\iota - 1)U_{n-1}^*(\iota) - U_{n-2}^*(\iota), \quad U_0^* = 1,\ U_1^* = 2\iota - 1. \tag{7}$$

The inversion formula [9] is expressed as

$$\iota^j = \frac{4\Gamma(j + \frac{3}{2})}{\sqrt{\pi}} \sum_{p=0}^{j} \frac{j!(1+p)!}{p!(j-p)!(2+j+p)!}\, U_p^*(\iota), \quad j \geq 0. \tag{8}$$

3 Petrov–Galerkin approach for fourth-order time-fractional PDEs

We examine the time-fractional fourth-order PDE (TF4PDE) [24]:

$$D_t^{\alpha}\nu(\iota,t)+\beta\,\frac{\partial^4\nu(\iota,t)}{\partial\iota^4}=f(\iota,t),\quad 0<\alpha\le 1,\ 0<\iota,t\le 1, \tag{9}$$

subject to the initial and boundary conditions:

$$\nu(\iota,0)=p_1(\iota),\quad \nu_t(\iota,0)=p_2(\iota),\quad 0<\iota\le 1, \tag{10}$$

$$\nu(0,t)=\nu(1,t)=\nu_\iota(0,t)=\nu_\iota(1,t)=0,\quad 0<t\le 1. \tag{11}$$

Here, f denotes a prescribed source term and $\beta>0$ is a constant parameter.

3.1 Trial functions

We construct basis functions that inherently satisfy the homogeneous boundary constraints:

$$\varphi_i(\iota)=a_1U_i^*+a_2U_{i+1}^*+a_3U_{i+2}^*+a_4U_{i+3}^*+U_{i+4}^*, \tag{12}$$

$$\psi_j(t)=U_j^*(t). \tag{13}$$

Utilizing the identities $U_k^*(0)=(-1)^k(k+1)$, $U_k^{*'}(0)=\frac{2}{3}(-1)^{k-1}(k)_3$, $U_k^*(1)=k+1$, and $U_k^{*'}(1)=\frac{2}{3}(k)_3$, we enforce $\varphi_i(0)=\varphi_i(1)=\varphi_i'(0)=\varphi_i'(1)=0$, which yields a linear system for the coefficients $a_1,\dots,a_4$:

$$\begin{aligned}a_1(-1)^i(i+1)+a_2(-1)^{i+1}(i+2)+a_3(-1)^{i+2}(i+3)\\ +a_4(-1)^{i+3}(i+4)+(-1)^{i+4}(i+5)=0,\end{aligned} \tag{14}$$

$$a_1(i+1)+a_2(i+2)+a_3(i+3)+a_4(i+4)+(i+5)=0, \tag{15}$$

$$\begin{aligned}a_1(i)_3(-1)^{i-1}+a_2(i+1)_3(-1)^i+a_3(i+2)_3(-1)^{i+1}\\ +a_4(i+3)_3(-1)^{i+2}+(i+4)_3(-1)^{i+3}=0,\end{aligned} \tag{16}$$

$$a_1(i)_3+a_2(i+1)_3+a_3(i+2)_3+a_4(i+3)_3+(i+4)_3=0. \tag{17}$$

Solving this 4×4 linear system (for instance, via Gaussian elimination) produces the compact representation:

$$\varphi_i(\iota) = \frac{(i+4)(i+5)}{(i+1)(i+2)} U_i^* - 2\frac{i+5}{i+2} U_{i+2}^* + U_{i+4}^*. \tag{18}$$

The approximate solution takes the form:

$$\nu^M(\iota, t) = \sum_{i=0}^{M} \sum_{j=0}^{M} c_{ij}\, \varphi_i(\iota) \psi_j(t). \tag{19}$$

Substituting this expansion into (9) and applying the Petrov–Galerkin procedure with test functions $\psi_{r,s}(\iota, t) = U_r^*(\iota) U_s^*(t)$ leads to

$$\left(D_t^\alpha \nu^M + \beta \partial_\iota^4 \nu^M - f,\ U_r^* U_s^* \right)_\omega = 0, \tag{20}$$

where $\omega(\iota, t) = \hat{w}(\iota)\hat{w}(t)$. Introducing the matrices,

$$\begin{aligned}
\hat{\boldsymbol{A}} &= (a_{ir}), \quad a_{ir} = (\varphi_i, U_r^*)_{\hat{w}}, \\
\hat{\boldsymbol{B}} &= (b_{sj}), \quad b_{sj} = (U_s^*, D_t^\alpha U_j^*)_{\hat{w}}, \\
\hat{\boldsymbol{H}} &= (\bar{h}_{ir}), \quad \bar{h}_{ir} = (\partial_\iota^4 \varphi_i, U_r^*)_{\hat{w}}, \\
\hat{\boldsymbol{K}} &= (k_{sj}), \quad k_{sj} = (U_s^*, U_j^*)_{\hat{w}}, \\
\hat{\boldsymbol{F}} &= (f_{rs}), \quad f_{rs} = (f, U_r^* U_s^*)_\omega,
\end{aligned}$$

with dimensions $\hat{\boldsymbol{A}}, \hat{\boldsymbol{H}} \in \mathbb{R}^{(M+1)\times(M+1)}$, $\hat{\boldsymbol{B}}, \hat{\boldsymbol{K}} \in \mathbb{R}^{M\times(M+1)}$, $\hat{\boldsymbol{F}} \in \mathbb{R}^{(M+1)\times M}$, and $\hat{\boldsymbol{C}} = (c_{ij}) \in \mathbb{R}^{(M+1)\times(M+1)}$, the resulting system can be written as follows:

$$\hat{\boldsymbol{A}} \hat{\boldsymbol{C}} \hat{\boldsymbol{B}}^{\mathsf{T}} + \beta\, \hat{\boldsymbol{H}} \hat{\boldsymbol{C}} \hat{\boldsymbol{K}}^{\mathsf{T}} = \hat{\boldsymbol{F}}. \tag{21}$$

The matrix products are dimensionally compatible: both $\hat{\boldsymbol{A}}\hat{\boldsymbol{C}}\hat{\boldsymbol{B}}^{\mathsf{T}}$ and $\hat{\boldsymbol{H}}\hat{\boldsymbol{C}}\hat{\boldsymbol{K}}^{\mathsf{T}}$ yield matrices of size $(M+1) \times M$.

Corollary 1. [4] The qth derivative of U_j^* admits the representation:

$$D^q U_j^*(\iota) = \sum_{\substack{p=0 \\ (q+p+j) \text{ even}}}^{j-q} \lambda_{j,p,q}\, U_p^*(\iota), \tag{22}$$

where $\lambda_{j,p,q} = \dfrac{j\, 2^{2q}\, (q)_{\frac{1}{2}(j-p-q)}}{\left(\frac{1}{2}(j-p-q)\right)!\, \left(\frac{1}{2}(q+p+j)\right)_{1-q}}$.

3.2 Petrov–Galerkin solution

Theorem 1. The entries of the matrices appearing in (21) are given by

$$
\begin{aligned}
k_{sj} &= g_{s,j},\\
a_{ir} &= g_{i+4,r} - \frac{2(i+2)(2i^2+8i+15)}{(i+1)(2i^2+4i+3)}g_{i+2,r} + \frac{(i+3)(2i^2+12i+19)}{(i+1)(2i^2+4i+3)}g_{i,r},\\
\bar{h}_{ir} &= \begin{cases} 128\pi(i+1)(i+2)(i+3)(i+4), & i=r,\\ \frac{256\pi i(i+2)(2i^4+18i^3+43i^2+63i+54)}{2i^2+4i+3}, & i-r=2,\\ \frac{384\pi(i-1)i(2i^5+24i^4+81i^3+218i^2+489i+474)}{(i+1)(2i^2+4i+3)}, & i-r=4,\\ 0, & \text{otherwise,}\end{cases}\\
b_{sj} &= (-1)^{j+1}j^2\pi\,\Gamma\!\left(\tfrac{3}{2}-\alpha\right)\,{}_4F_3\!\left(\begin{matrix}1, 1-j, j+1, \frac{3}{2}-\alpha\\ \frac{3}{2}, -s-\alpha+2, s-\alpha+2\end{matrix}\middle|1\right).
\end{aligned}
\tag{23}
$$

Proof. The formulas for k_{sj} and a_{ir} follow directly from the orthogonality relation (5) together with the compact basis expression (18). To obtain $\bar{h}_{ir}$, apply Corollary 1 with $q=4$ to $\partial_\iota^4\varphi_i$ and exploit orthogonality. For b_{sj}, combine (3) and (4) to write

$$
\begin{aligned}
b_{sj} &= \int_0^1 \left[D_t^\alpha U_j^*(t)\right] U_s^*(t)\hat{w}(t)\,dt\\
&= js\sum_{k=0}^{j}\sum_{n=0}^{s}\frac{(-1)^{j-k}2^{2k}(j+k-1)!\,\Gamma(k+1)\,(-1)^{s-n}2^{2n}(s+n-1)!}{(j-k)!(2k)!\,\Gamma(k-\alpha+1)(s-n)!(2n)!}\\
&\quad \int_0^1 \frac{t^{k+n-\alpha}}{\sqrt{t(1-t)}}\,dt.
\end{aligned}
\tag{24}
$$

Evaluating the resulting beta integral and simplifying the expression leads to the unified ${}_4F_3$ hypergeometric representation. □

Remark 2. The Petrov–Galerkin framework provides three principal benefits: (i) the compact basis (18) enforces boundary conditions exactly, thereby eliminating the need for penalty terms; (ii) the operational matrices exhibit sparse, banded structure, enabling assembly with $O(M^2)$ complexity; (iii) exponential convergence is attained for sufficiently smooth solutions, offering superior accuracy compared to finite-difference or finite-element approaches when high precision is required.

4 Error analysis

We examine the approximation error $\nu - \nu^M$ measured in both L_∞ and L_2 norms.

4.1 L_∞-norm error analysis

Let $\hat{\nu}^M$ denote the interpolant of ν at the Chebyshev nodes. Then

$$\|\nu - \nu^M\|_\infty \leq \|\nu - \hat{\nu}^M\|_\infty. \tag{25}$$

Applying the two-dimensional Taylor remainder formula yields

$$\begin{aligned} \nu - \hat{\nu}^M = & \frac{\partial^{M+1}\nu(\eta,t)}{\partial \iota^{M+1}(M+1)!}\prod_{i=0}^{M}(\iota-\iota_i) + \frac{\partial^{M+1}\nu(\iota,\mu)}{\partial t^{M+1}(M+1)!}\prod_{j=0}^{M}(t-t_j) \\ & + \frac{\partial^{2M+2}\nu(\hat{\eta},\hat{\mu})}{\partial \iota^{M+1}\partial t^{M+1}((M+1)!)^2}\prod_{i=0}^{M}(\iota-\iota_i)\prod_{j=0}^{M}(t-t_j), \end{aligned} \tag{26}$$

where the third term carries a *positive* sign, consistent with the standard remainder expression. Bounding the derivatives by constants g_1, g_2, g_3 and minimizing the nodal polynomials through Chebyshev node selection leads to

$$\|\nu - \nu^M\|_\infty \leq g_1\frac{(1/2)^{M+1}\Upsilon_M}{\bar{\zeta}_M(M+1)!} + g_2\frac{(1/2)^{M+1}}{\hat{U}_M(M+1)!} + g_3\frac{(1/4)^{M+1}\Upsilon_M}{\bar{\zeta}_M\hat{U}_M((M+1)!)^2}. \tag{27}$$

4.2 L_2-norm error analysis

Theorem 2. Assume that $\nu^M \in \Lambda^M$ represents the approximate solution and that $\partial^{i+j}\nu/\partial\iota^i\partial t^j \in C(\Omega)$ for $i, j \leq M+1$. Define

$$\mathcal{L}_M = \sup_{(\iota,t)\in\Omega}\left|\frac{\partial^{2(M+1)}\nu}{\partial\iota^{M+1}\partial t^{M+1}}\right|. \tag{28}$$

Then

$$\|\nu - \nu^M\|_2 \leq \mathfrak{n}\,\frac{\mathcal{L}_M}{M^{3/4}(\Gamma(M+2))^2}, \tag{29}$$

where $\mathfrak{n} > 0$ denotes a generic constant independent of M (explicitly, $\mathfrak{n} = \sqrt{\pi}\,\sup_{z\geq 1}\mathbf{o}_z^{a,b}$ as given in [39]).

Proof. Exploiting the best approximation property together with a Taylor expansion gives

$$\|\nu-\nu^M\|_2^2 \leq \int_0^1\int_0^1 \frac{\mathcal{L}_M^2\,\iota^{2(M+1)}t^{2(M+1)}}{(\Gamma(M+2))^4}\hat{w}(\iota)\hat{w}(t)\,d\iota\,dt. \tag{30}$$

Evaluating the beta integrals and applying the gamma ratio bound from [39] yield the stated result with the explicit constant $\mathfrak{n}$. □

Theorem 3. Under the assumptions of Theorem 2, define

$$\mathcal{U}_{M,m} = \sup_{\Omega}\left|\frac{\partial^{2M-m+2}\nu}{\partial\iota^{M-m+1}\partial t^{M+1}}\right|, \quad m=1,\ldots,4. \tag{31}$$

Then

$$\left\|\frac{\partial^m(\nu-\nu^M)}{\partial\iota^m}\right\|_2 \leq \mathfrak{n}\,\frac{\mathcal{U}_{M,m}}{(2(M-m)+3)^{1/2}M^{1/4}\Gamma(M+2)\Gamma(M-m+2)}. \tag{32}$$

Theorem 4. If $D_t^\alpha\nu \in C(\Omega)$ and

$$\mathcal{P}_{M,\alpha} = \sup_{\Omega}\left|\frac{\partial^{2M-\alpha+2}\nu}{\partial\iota^{M+1}\partial t^{M-\alpha+1}}\right|, \tag{33}$$

then

$$\|D_t^\alpha(\nu-\nu^M)\|_2 \leq \mathfrak{n}\,\frac{\mathcal{P}_{M,\alpha}}{(2M+3)^{1/2}(M-\alpha)^{1/4}\Gamma(M+2)\Gamma(M-\alpha+2)}. \tag{34}$$

Theorem 5. The residual $\mathbf{R} = D_t^\alpha\nu^M + \beta\partial_\iota^4\nu^M - f$ satisfies the bound:

$$\begin{aligned}\|\mathbf{R}\|_2 \leq \mathfrak{n}\Bigg[&\frac{\mathcal{P}_{M,\alpha}}{(2M+3)^{1/2}(M-\alpha)^{1/4}\Gamma(M+2)\Gamma(M-\alpha+2)}\\ &+\beta\,\frac{\mathcal{U}_{M,4}}{(2M-5)^{1/2}M^{1/4}\Gamma(M+2)\Gamma(M-2)}\Bigg].\end{aligned} \tag{35}$$

Consequently, $\|\mathbf{R}\|_2 \to 0$ as $M\to\infty$, with the constant β explicitly retained in the fourth-derivative contribution.

5 Illustrative examples

This section demonstrates the practical applicability of the proposed method through three numerical test cases. All computations were carried out using Mathematica 12 on a workstation equipped with an Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz and 16.0 GB of RAM.

Example 1. [24, 27, 29] Consider the TF4PDE given by

$$D_t^{\alpha}\nu(\iota,t)+\beta\,\frac{\partial^4\nu(\iota,t)}{\partial\iota^4}=f(\iota,t),\quad 0<\alpha\leq 1, \tag{36}$$

subject to the initial and boundary conditions

$$\begin{aligned}&\nu(\iota,0)=\sin(\pi\,\iota),\quad 0<\iota\leq 1,\\&\nu(0,t)=\nu(1,t)=\nu_{\iota}(0,t)=\nu_{\iota}(1,t)=0,\quad 0<t\leq 1,\end{aligned} \tag{37}$$

where the source term is $f(\iota,t)=\left(\frac{1}{\Gamma(2-\alpha)}\,t^{1-\alpha}+\beta\,\pi^4\,(t+1)\right)\sin(\pi\,\iota)$ and the exact solution is $\nu(\iota,t)=(t+1)\,(1-\iota)\,\sin(\pi\,\iota)$.

Table 1 presents a comparison of the maximum absolute error (MAE) between the proposed scheme and existing methods from several references [24, 27, 29] at $\alpha=0.5$. Table 2 examines how the MAE evolves for the proposed method as M (the number of basis terms) varies, with α fixed at 0.5.

Table 3 explores the variation of MAE across different time instances (t) within the interval $0<t<1$, using the proposed method with $\alpha=0.1$ and $M=12$. This table underscores the capability of the method to deliver highly accurate results even with a relatively small number of basis functions. Figures 1, 2, and 3 display the exact and computed solutions for $M=12$ at various values of α. Figure 4 illustrates the absolute error distribution for Example 1 with $M=12$.

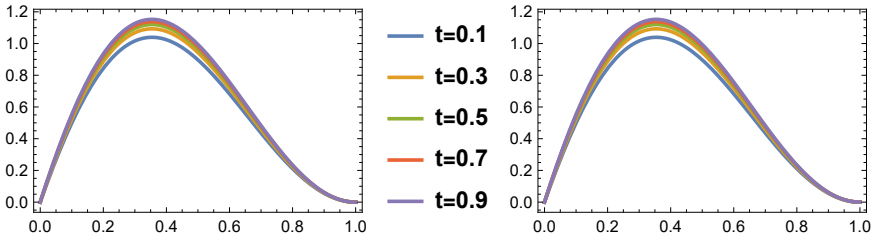


Figure 1: The exact (left) and approximate (right) solutions of Example 1 for $\alpha=0.1$

Table 1: Comparison of the MAE for Example 1.

At $\Lambda t=0.00001$, $m=40$, $n=500$			At $M=12$
Method in [24]	Method in [29]	Method in [27]	Our method
4.69×10^{-8}	1.43×10^{-4}	2.32×10^{-4}	2.22×10^{-11}

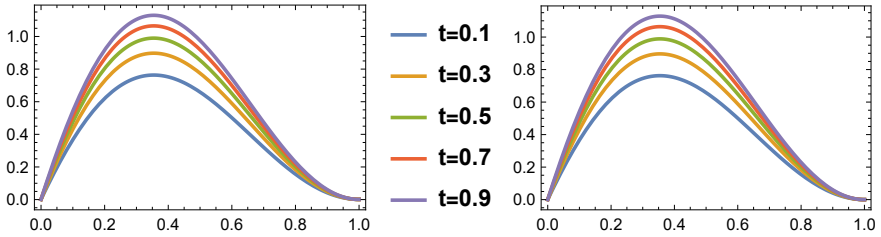


Figure 2: The exact (left) and approximate (right) solutions of Example 1 for $\alpha = 0.5$

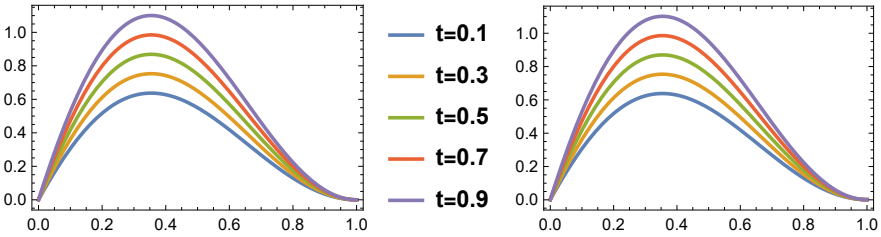


Figure 3: The exact (left) and approximate (right) solutions of Example 1 for $\alpha = 1$

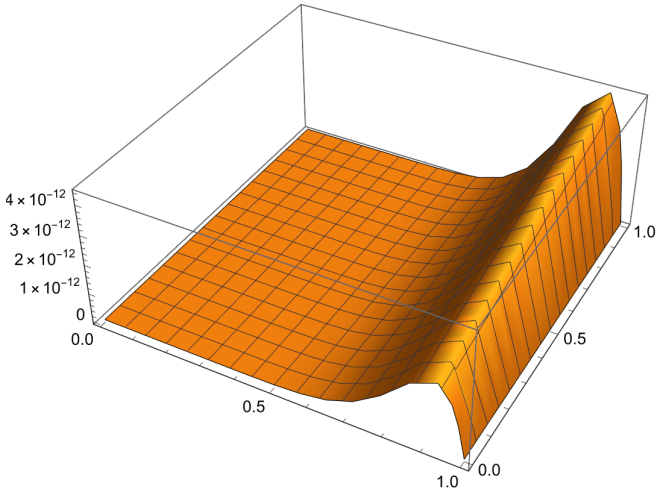


Figure 4: Absolute error example 1 when $M = 12$

Table 2: MAE for Example 1.

M	2	4	6	8	10	12
Error	2.51×10^{-2}	4.62×10^{-4}	7.82×10^{-6}	5.35×10^{-8}	7.06×10^{-10}	2.22×10^{-11}

Example 2. [24] Consider the TF4PDE of the form

Table 3: MAE for Example 1.

(x,t)	$(x,0.1)$	$(x,0.3)$	$(x,0.5)$	$(x,0.7)$	$(x,0.9)$
Error	2.43×10^{-13}	4.52×10^{-12}	4.92×10^{-12}	2.22×10^{-12}	2.22×10^{-12}

$$D_t^{\alpha}\nu(\iota,t)+\beta\,\frac{\partial^4\nu(\iota,t)}{\partial\iota^4}=f(\iota,t),\quad 0<\alpha\leq 1, \tag{38}$$

subject to the initial and boundary conditions

$$\begin{aligned}&\nu(\iota,0)=0,\quad 0<\iota\leq 1,\\ &\nu(0,t)=\nu(1,t)=\nu_{\iota}(0,t)=\nu_{\iota}(1,t)=0,\quad 0<t\leq 1,\end{aligned} \tag{39}$$

where $f(\iota,t)=\left(\frac{2}{\Gamma(3-\alpha)}\,t^{2-\alpha}+16\,\beta\,\pi^4\,t^2\right)\sin(2\,\pi\,\iota)$ and the exact solution is $\nu(\iota,t)=t^2\,\sin(2\,\pi\,\iota)$. Table 4 compares the MAE obtained by the proposed method with that from [24] for several values of α. Figures 5, 6, and 7 show the exact and computed solutions for $M=12$ at different α values. Figure 8 displays the absolute error for Example 2 with $M=12$.

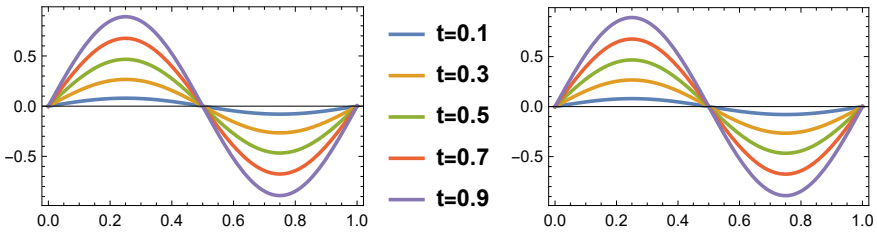


Figure 5: The exact (left) and approximate (right) solutions of Example 2 for $\alpha=0.1$

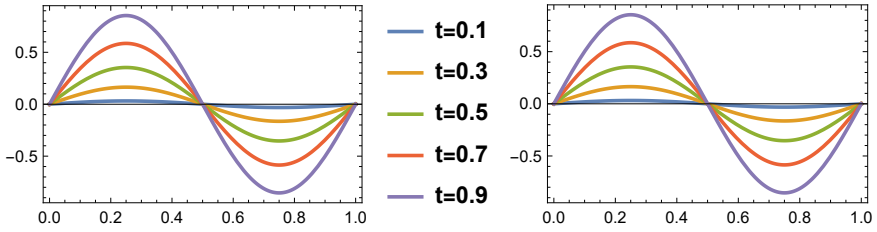


Figure 6: The exact (left) and approximate (right) solutions of Example 2 for $\alpha=0.5$

Example 3. Consider the TF4PDE given by

$$D_t^{\alpha}\nu(\iota,t)+\beta\,\frac{\partial^4\nu(\iota,t)}{\partial\iota^4}=f(\iota,t),\quad 0<\alpha\leq 1, \tag{40}$$

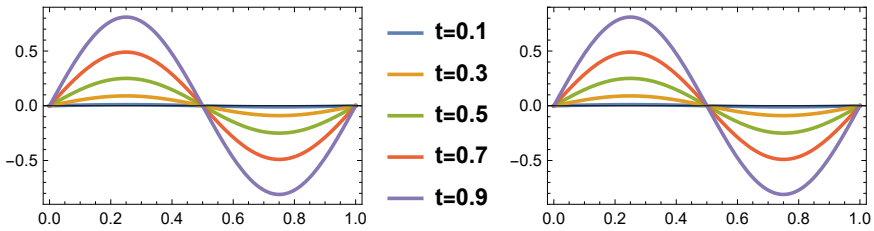


Figure 7: The exact (left) and approximate (right) solutions of Example 2 for $\alpha = 1$

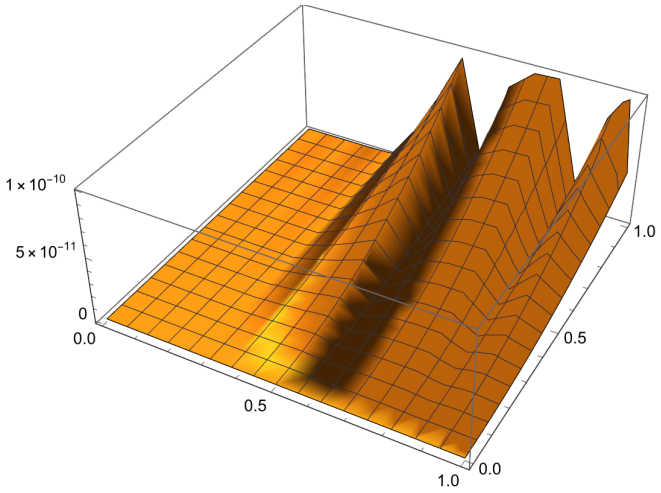


Figure 8: Absolute error example 2 when $M = 12$

Table 4: Comparison of the MAE for Example 2.

α	Method in [24] at $m = 200$, $n = 160$	Our method at $M = 12$
0.4	7.3899×10^{-8}	2.36×10^{-10}
0.5	1.4865×10^{-7}	5.29×10^{-10}
0.7	5.7410×10^{-7}	3.67×10^{-10}

subject to the initial and boundary conditions

$$\begin{aligned} &\nu(\iota, 0) = \sin(\pi\,\iota), \quad 0 < \iota \leq 1, \\ &\nu(0, t) = \nu(1, t) = \nu_\iota(0, t) = \nu_\iota(1, t) = 0, \quad 0 < t \leq 1, \end{aligned} \tag{41}$$

where the source term $f(\iota, t)$ is selected so that the exact solution is $\nu(\iota, t) = e^{\alpha\, t} \sin(\pi\,\iota)$. Since a closed-form expression for f is not readily available, we evaluate it numerically using the definition of the Caputo derivative. Table 5 reports the MAE for various values of M with $\alpha = 0.9$, demonstrating the method's ability to achieve high accuracy even with modest M.

Table 6 presents the MAE at different time levels t for $\alpha = 0.5$ and $M = 12$ over $0 < \iota < 1$, confirming that the computed solution closely matches the exact one.

Table 5: MAE for Example 3.

M	2	4	6	8	10	12
Error	6.36×10^{-2}	1.00×10^{-3}	5.92×10^{-6}	5.00×10^{-8}	3.70×10^{-10}	2.08×10^{-12}

Table 6: MAE for Example 3.

(x,t)	$(x,0.1)$	$(x,0.3)$	$(x,0.5)$	$(x,0.7)$	$(x,0.9)$
Error	1.32×10^{-12}	1.46×10^{-12}	1.62×10^{-12}	1.79×10^{-12}	1.16×10^{-12}

6 Concluding remarks

This study has addressed the vibrational analysis of a fourth-order time-fractional Euler–Bernoulli beam with pinned-pinned boundary conditions. A novel Petrov–Galerkin spectral framework was developed for this purpose. The approach employs a compact basis formed from Chebyshev polynomials of the second kind to represent the beam displacement. The governing fractional PDE was then transformed into a system of algebraic equations via the Petrov–Galerkin procedure. These equations, involving unknown expansion coefficients characterizing the beam response, were solved efficiently using Gaussian elimination, which is well-suited to the resulting structured system. The method was validated through three numerical examples, including one lacking a known closed-form solution. The computational results confirm that the proposed scheme delivers both high accuracy and computational efficiency.

Potential directions for future work:

- Generalization to variable-order fractional derivatives and nonlinear beam formulations incorporating geometric or material nonlinearities.
- Development of adaptive refinement strategies for solutions exhibiting low regularity or sharp boundary layers.
- Implementation of parallel algorithms and GPU acceleration to handle large-scale three-dimensional beam networks and enable real-time simulations.

- Extension to inverse problems, such as the identification of fractional orders or material parameters from measured vibration data.
- Integration with machine learning methodologies for structural health monitoring and damage detection in fractional viscoelastic structures.

Conflict of interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

References

[1] Abbasov, Z.D. and Youssri, Y.H. *Fibonacci collocation spectral method and error analysis for heat conduction waves in an n-dimensional domain*, Open J. Math. Sci., 10 (2026), 341–364.

[2] Abd-Elhameed, W.M. and Alkenedri, A.M. *Spectral solutions of linear and nonlinear BVPs using certain Jacobi polynomials generalizing third- and fourth-kinds of Chebyshev polynomials*, CMES-Comp. Model. Eng. Sci., 126(3) (2021), 955–989.

[3] Abd-Elhameed, W.M. and Alkenedri, A.M. *New formulas for the repeated integrals of some Jacobi polynomials: Spectral solutions of even-order boundary value problems*, Int. J. Appl. Comput. Math., 7(4) (2021), 166.

[4] Abd-Elhameed, W.M., Machado, J.A.T. and Youssri, Y.H. *Hypergeometric fractional derivatives formula of shifted Chebyshev polynomials: Tau algorithm for a type of fractional delay differential equations*, Int. J. Nonlinear Sci. Numer. Simul., 23(7-8) (2022), 1253–1268.

[5] Abdelghany, E.M., Abd-Elhameed, W.M., Moatimid, G.M., Youssri, Y.H. and Atta, A.G. *A tau approach for solving time-fractional heat equation based on the shifted sixth-kind Chebyshev polynomials*, Symmetry, 15(3) (2023), 594.

[6] Ahmad, S., Ullah, A., Akgül, A. and De la Sen, M. *A novel homotopy perturbation method with applications to nonlinear fractional-order KdV and Burgers equation with exponential-decay kernel*, J. Function Spaces, 2021(1) (2021), 8770488.

[7] Andrade, C. and McKee, S. *High accuracy a.d.i. methods for fourth-order parabolic equations with variable coefficients*, J. Comput. Appl. Math., 3(1) (1977), 11–14.

[8] Atta, A.G., Abd-Elhameed, W.M., Moatimid, G.M. and Youssri, Y.H. *Shifted fifth-kind Chebyshev Galerkin treatment for linear hyperbolic first-order partial differential equations*, Appl. Numer. Math., 167 (2021), 237–256.

[9] Atta, A.G. and Youssri, Y.H. *Advanced shifted first-kind Chebyshev collocation approach for solving the nonlinear time-fractional partial integro-differential equation with a weakly singular kernel*, Comput. Appl. Math., 41(8) (2022), 381.

[10] Aziz, T., Khan, A. and Rashidinia, J. *Spline methods for the solution of fourth-order parabolic partial differential equations*, Appl. Math. Comput., 167(1) (2005), 153–166.

[11] Bassuony, M.A., Doha, E.H., Abd-Elhameed, W.M. and Youssri, Y.H. *A Legendre-Laguerre-Galerkin method for the uniform Euler-Bernoulli beam equation*, East Asian J. Appl. Math., 8(2) (2018), 280–295.

[12] Evans, D.J. *A stable explicit method for the finite-difference solution of a fourth-order parabolic partial differential equation*, Comput. J., 8(3) (1965), 280–287.

[13] Evans, D.J. and Yousif, W.S. *A note on solving the fourth-order parabolic equation by the age method*, Int. J. Comput. Math., 40(1-2) (1991), 93–97.

[14] Fernandez, A., Restrepo, J.E. and Suragan, D. *A new representation for the solutions of fractional differential equations with variable coefficients*, Mediterr. J. Math., 20(1) (2023), 27.

[15] Han, S.M., Benaroya, H. and Wei, T. *Dynamics of transversely vibrating beams using four engineering theories*, J. Sound Vib., 225(5) (1999), 935–988.

[16] Heydari, M.H. and Hosseininia, M. *An efficient Jacobi spectral method for variable-order time fractional 2D Wu-Zhang system*, Comput. Math. Appl., 140 (2023), 89–106.

[17] Hilfer, R. *Applications of Fractional Calculus in Physics*, World Scientific, 2000.

[18] Hosseininia, M., Heydari, M.H., Baleanu, D. and Bayram, M. *A hybrid method based on the classical/piecewise Chebyshev cardinal functions for multi-dimensional fractional Rayleigh–Stokes equations*, Results Appl. Math., 25 (2025), 100541.

[19] Hosseininia, M., Heydari, M.H. and Razzaghi, M. *A hybrid spectral approach based on 2D cardinal and classical second-kind Chebyshev polynomials for time-fractional 3D Sobolev equation*, Math. Methods Appl. Sci., 46(18) (2023), 18768–18788.

[20] Jie, S. *Efficient spectral-Galerkin II. Direct solvers for second- and fourth-order equations by using Chebyshev polynomials*, SIAM J. Sci. Comput., 16 (1) (1995), 74–87.

[21] Mason, J.C. and Handscomb, D.C. *Chebyshev polynomials*, Chapman and Hall/CRC, 2002.

[22] Moustafa, M., Youssri, Y.H. and Atta, A.G. *Explicit Chebyshev-Galerkin scheme for the time-fractional diffusion equation*, Int. J. Mod. Phys. C, 35 (2024).

[23] Podlubny, I. *Fractional differential equations: An introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*, Elsevier, 1998.

[24] Roul, P. and Goura, V.M.K.P. *A high-order numerical method and its convergence for time-fractional fourth-order partial differential equations*, Appl. Math. Comput., 366 (2020), 124727.

[25] Sajjadi, S.A., Najafi, H.S. and Aminikhah, H. *A numerical study on the non-smooth solutions of the nonlinear weakly singular fractional Volterra integro-differential equations*, Math. Methods Appl. Sci., 46(4) (2023), 4070–4084.

[26] Salama, M.H., Zedan, H.A., Abd-Elhameed, W.M., Youssri, Y.H. *Shifted Lucas polynomial-driven Tau approach for the telegraph equation: theoretical and numerical insights into accuracy and stability*, Comp. Appl. Math., 45 (2026), 397.

[27] Siddiqi, S.S. and Arshed, S. *Numerical solution of time-fractional fourth-order partial differential equations*, Int. J. Comput. Math., 92(7) (2015), 1496–1518.

[28] Sun, Z.Z. and Gao, G.H. *Fractional differential equations*, De Gruyter, 2020.

[29] Tariq, H. and Akram, G. *Quintic spline technique for time fractional fourth-order partial differential equation*, Numer. Methods Partial Differ. Equations, 33(2) (2017), 445–466.

[30] Wazwaz, A.M. *Analytic treatment for variable coefficient fourth-order parabolic partial differential equations*, Appl. Math. Comput., 123(2) (2001), 219–227.

[31] Xu, C., Farman, M. and Shehzad, A. *Analysis and chaotic behavior of a fish farming model with singular and non-singular kernel*, Int. J. Biomathematics, (2023), 2350105.

[32] Xu, C., Liao, M., Farman, M. and Shehzad, A. *Hydrogenolysis of glycerol by heterogeneous catalysis: a fractional order kinetic model with analysis*, MATCH Commun. Math. Comput. Chem., 91(3) (2024), 635–664.

[33] Xu, C., Lin, J., Zhao, Y., Cui, Q., Ou, W., Pang, Y., Liu, Z., Liao, M. and Li, P. *New results on bifurcation for fractional-order octonion-valued neural networks involving delays*, Network: Comput. Neur. Syst., 36 (2024), 545-597.

[34] Xu, C., Zhao, Y., Lin, J., Pang, Y., Liu, Z., Shen, J., Liao, M., Li, P. and Qin, Y. *Bifurcation investigation and control scheme of fractional neural networks owning multiple delays*, Comput. Appl. Math., 43(4) (2024), 1–33.

[35] Youssri, Y.H., Abd-Elhameed, W.M. and Atta, A.G. *Spectral Galerkin treatment of linear one-dimensional telegraph type problem via the generalized Lucas polynomials*, Arab J. Math., 11(3) (2022), 601–615.

[36] Youssri, Y.H., Abd-Elhameed, W.M., Elmasry, A.A. and Atta, A.G. *An efficient Petrov-Galerkin scheme for the Euler-Bernoulli beam equation via second-kind Chebyshev polynomials*, Fractal Fract., 9(2) (2025), 78.

[37] Youssri, Y.H. and Atta, A.G. *Spectral collocation approach via normalized shifted Jacobi polynomials for the nonlinear Lane-Emden equation with fractal-fractional derivative*, Fractal Fract., 7(2) (2023), 133.

[38] Youssri, Y.H. and Muttardi, M.M. *A mingled tau-finite difference method for stochastic first-order partial differential equations*, Int. J. Appl. Comput. Math., 9(2) (2023), 14.

[39] Zhao, X., Wang, L. and Xie, Z. *Sharp error bounds for Jacobi expansions and Gegenbauer–Gauss quadrature of analytic functions*, SIAM J. Numer. Anal., 51(3) (2013), 1443–1469.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1239–1270
https://doi.org/10.22067/ijnao.2026.97425.1826
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Analysis and optimal control of a proliferating-quiescent Glioma model with combined therapy

T. Jaber* and O. Balatif

Abstract

Low-grade gliomas (LGGs) are slow-growing but infiltrative brain tumors characterized by long-term persistence and frequent progression. In this work, we develop and analyze a compartmental mathematical model describing LGG dynamics under combination therapy involving cytotoxic chemotherapy and targeted molecular treatment. The tumor population is structured into proliferating, quiescent, and damaged compartments, and therapeutic interventions are incorporated through bilinear control terms accounting for chemo-therapies. We first establish the well-posedness of the model by proving existence, uniqueness, positivity, and boundedness of solutions. The dynamics of untreated glioma growth are then analyzed, revealing that the tumor-free equilibrium (TFE) is always unstable in the absence of therapy, leading to exponential tumor expansion. Under constant treatment regimens, we derive an explicit controlled reproduction number that governs tumor persistence or eradication. Using linearization and Lyapunov–LaSalle stability techniques, we establish necessary and sufficient conditions for both

*Corresponding author

Received 29 January 2026; revised 19 March 2026; accepted 17 April 2026

Tawfik Jaber
Laboratory of fundamental mathematics and their application, Department of Mathematics, Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco. e-mail: jaber.tawfik@ucd.ac.ma

Omar Balatif
Laboratory of fundamental mathematics and their application, Department of Mathematics, Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco. e-mail: balatif.maths@gmail.com

How to cite this article
Jaber, T. and Balatif, O., Analysis and optimal control of a proliferating-quiescent Glioma model with combined therapy. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1239-1270. https://doi.org/10.22067/ijnao.2026.97425.1826

local and global asymptotic stability of the tumor-free state. Finally, we formulate an optimal control problem to identify treatment strategies that minimize tumor burden while limiting therapeutic toxicity. The existence of optimal controls is rigorously proven using classical results from optimal control theory. Our results provide analytical insight into the stabilizing role of combination therapy and offer a quantitative guidance for designing optimized treatment protocols for LGG.

AMS subject classifications (2020): Primary 45D05; Secondary 42C10, 65G99.

Keywords: Bilinear models; Reproduction number; Glioma dynamics; Optimality conditions; Numerical simulations.

1 Introduction

Low-grade gliomas (LGGs) are a heterogeneous group of slow-growing primary brain tumors that originate from glial cells [15], typically classified as WHO grade II gliomas [27]. These tumors, although initially indolent, have the potential to undergo malignant transformation and progress to high-grade gliomas (e.g., glioblastoma) [41, 39]. LGGs commonly affect young adults, often presenting with seizures or neurological deficits and are associated with long-term morbidity [35]. Their infiltrative nature, resistance to complete surgical resection, and variable response to therapy make LGGs a challenging clinical and biological entity [5, 38]. Understanding their progression dynamics and developing effective, personalized treatment protocols are of crucial importance in neuro-oncology [39].

In the early years, treatment for LGGs mainly involved surgical resection, where doctors aimed to remove as much of the tumor as possible [18, 34]. After surgery, a watchful waiting approach was commonly used, especially in younger or lower-risk patients, because LGGs tend to grow slowly, and early surgical resection resulted in a clinically relevant survival benefit [20]. As more long-term data became available, it became clear that many LGGs eventually progress, and delaying treatment could reduce survival in high-risk patients. This led to the increased use of adjuvant treatments given after surgery, such as radiotherapy and chemotherapy. A major shift happened after the results of the RTOG 9802 clinical trial [37], which showed that combining radiotherapy with PCV chemotherapy (procarbazine, lomustine, and vincristine) significantly improved both progression-free survival and overall survival in patients with high-risk LGGs. In the 2000s and 2010s, temozolomide (TMZ) gained popularity as a chemotherapy option due to its oral administration, better tolerability, and promising clinical results [33]. It became the preferred chemotherapy for many patients, especially in cases where PCV was too toxic or

not tolerated. By the late 2010s and early 2020s, TMZ combined with radiotherapy became a common first-line treatment in clinical practice, especially for symptomatic or progressing tumors [29].

From a mathematical perspective, several models have been proposed to understand and simulate LGG growth and treatment response [17, 36]. Early models were primarily empirical, focusing on fitting exponential or logistic growth curves to tumor volume [4]. Later approaches incorporated compartmental models that distinguish between proliferative and quiescent tumor states [6]. Mathematical modeling has increasingly contributed to the understanding of LGG dynamics and treatment strategies. Several mathematical studies have been devoted to modeling the dynamics of LGGs and evaluating treatment strategies. In particular, the work in [8] investigates the nonlinear behavior of LGG growth through a system of differential equations, provides a detailed analysis of the system's stability in the absence of treatment, and demonstrates, through numerical simulations, how treatment timing and temozolomide dosing influence tumor progression. In [2], a mathematical model is constructed to investigate optimal combinations of chemotherapy and radiotherapy. Patient-specific parameters derived from imaging data inform in silico trials on a large virtual cohort, leading to the proposal of improved treatment schedules with potentially enhanced survival outcomes. The role of immunotherapy is explored in [21], where a coupled ordinary differential equation model captures the interactions between glioma cells, immune effectors, and adoptive cellular immunotherapy. Using Pontryagin's maximum principle, the authors formulate and solve an optimal control problem aimed at minimizing tumor burden and therapy-induced toxicity. Their analysis includes a characterization of the optimality system and demonstrates, through numerical solutions, the effectiveness of dynamically adjusted treatment strategies. Additionally, in [26], the authors examine the stability properties of a three-compartment model representing tumor cell-cycle dynamics under chemotherapy. They identify a critical drug dosage threshold above which global asymptotic stability of the TFE is guaranteed, and below which only local stability is achieved, with the existence of a bifurcation point separating treatment success from persistence.

However, many existing models fail to incorporate the bilinear structure of drug-tumor interaction, which is biologically relevant given that treatment efficacy often depends on both the drug concentration and tumor cell population [19, 3]. Additionally, few models combine the dynamics of multiple therapies (e.g., TMZ and IDH inhibitors) within a control-theoretic framework that respects pharmacokinetic constraints and clinical realities.

The remainder of the paper is organized as follows. In Section 2, we introduce the mathematical model describing the tumor tissue compartments and the therapeutic control mechanisms. Section 3 is devoted to the stability analysis of the uncontrolled system, including the characterization of the TFE. In Section 4, we investigate the local and global stability properties of the

controlled system under constant treatment strategies. Section 5 formulates the optimal control problem and derives the associated adjoint system together with the characterization of the optimal controls. Numerical simulations illustrating the theoretical results and highlighting the impact of optimal treatment strategies on tumor dynamics are presented in Section 6. Finally, Section 7 concludes the paper with a summary of the main findings and perspectives for future research.

2 Mathematical model of glioma dynamics

We present a compartmental mathematical model that describes the temporal evolution of LGG under the influence of combination therapy involving cytotoxic chemotherapy and targeted molecular therapy. The formulation is inspired by the chemotherapy-driven model introduced in [2], and is extended here to explicitly incorporate a second therapeutic modality through an additional bilinear control term, targeting isocitrate dehydrogenase (IDH) mutations, which are present in approximately 80% of LGG cases [16].

The tumor population is partitioned into three biologically distinct and mutually interacting compartments:

- $G_p(t)$: Proliferating tumor cells actively undergoing cell division.
- $G_q(t)$: Quiescent tumor cells in a reversible, nondividing state.
- $G_d(t)$: Damaged quiescent cells undergoing clearance or repair.

The dynamics of the system, illustrated by the flow diagram in Figure 1, are governed by the following system of ordinary differential equations:

$$\begin{aligned}
\frac{dG_p}{dt} &= \underbrace{\kappa G_p}_{\text{Proliferation}} - \underbrace{\tau_1 G_p}_{\text{Transition to quiescence}} + \underbrace{\tau_2 G_q}_{\text{Re-entry from quiescence}} \\
&\quad - \underbrace{\phi_1 u_1 G_p}_{\text{Chemo-induced damage}} - \underbrace{\phi_2 u_1 G_p}_{\text{Chemo-induced death}} - \underbrace{\kappa\omega u_2 G_p}_{\text{Targeted therapy inhibition}}, \\
\frac{dG_q}{dt} &= \underbrace{\tau_1 G_p}_{\text{Transition to quiescence}} - \underbrace{\tau_2 G_q}_{\text{Re-entry from proliferation}}, \\
\frac{dG_d}{dt} &= \underbrace{\phi_1 u_1 G_p}_{\text{Chemo-induced damage}} - \underbrace{\tau_3 G_d}_{\text{Clearance of damaged cells}},
\end{aligned} \tag{1}$$

where $u_1(t)$ and $u_2(t)$ represent time-dependent therapeutic controls bounded by $0 \leq u_i(t) \leq u_{i,\max} < 1$, with $i = 1, 2$.

2.1 Biological interpretation of model parameters

All state variables are assumed to be nonnegative and continuous for $t \geq 0$. The model parameters, all strictly positive, and have the following biological interpretations:

- **Proliferation rate ($\kappa > 0$)**: Represents the net growth rate of proliferating tumor cells, accounting for both birth and natural death processes. This parameter reflects the intrinsic aggressiveness of the glioma.
- **Transition rates ($\tau_1 > 0$, $\tau_2 > 0$)**:
  - τ_1: Rate at which proliferating cells enter quiescence, modeling reversible cell cycle arrest
  - τ_2: Rate at which quiescent cells re-enter proliferation, representing tumor plasticity
- **Therapeutic efficacy parameters**:
  - $\phi_1 > 0$: Rate of chemotherapy-induced DNA damage leading to damaged quiescent state
  - $\phi_2 > 0$: Rate of direct cytotoxic elimination of proliferating cells by chemotherapy
  - $\omega > 0$: Inhibition rate of proliferative capacity by IDH-targeted therapy
- **Clearance rate ($\tau_3 > 0$)**: Represents the rate at which damaged cells are removed from the system, either through apoptosis or immune-mediated clearance.

The total tumor burden is defined by $T(t) = G_p(t) + G_q(t) + G_d(t)$, and its evolution is implicitly regulated by the balance between proliferation, phenotypic transitions, therapeutic pressure, and damaged cell clearance, as described by system (1). Figure 1 provides a schematic representation of these interactions.

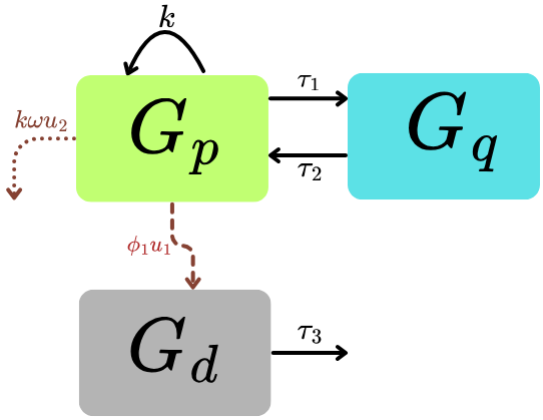


Figure 1: Flow diagram of the three-compartment tumor model. Compartments are colored, and arrows indicate transitions.

2.2 Mathematical well-posedness

From a mathematical perspective, the system (1) is well posed. Under standard assumptions on the continuity and local Lipschitz regularity of the right-hand side, the existence and uniqueness of solutions follow from classical results in the theory of ordinary differential equations [24]. Moreover, for any nonnegative initial condition, the solution remains nonnegative for all $t \geq 0$. This invariance property is essential for biological consistency and follows from the cooperative structure of the system and the absence of negative source terms acting on vanishing states [1, 10].

Theorem 1. For any initial condition $(G_p(0), G_q(0), G_d(0)) \in \mathbb{R}^3_+$ and measurable control functions $u_1, u_2 \in L^\infty([0,T])$ with $0 \leq u_i(t) \leq u_{i,\max}$, system (1) admits a unique absolutely continuous solution on $[0,T]$.

Proof. Let $\mathbf{x}(t) = (G_p(t), G_q(t), G_d(t))^\top$. The system (1) can be written as

$$\dot{\mathbf{x}}(t) = A(\mathbf{u}(t))\mathbf{x}(t),$$

where

$$A(\mathbf{u}) = \begin{bmatrix} \kappa(1-\omega u_2) - \tau_1 - (\phi_1+\phi_2)u_1 & \tau_2 & 0 \\ \tau_1 & -\tau_2 & 0 \\ \phi_1 u_1 & 0 & -\tau_3 \end{bmatrix}.$$

Define $f(t, \mathbf{x}, \mathbf{u}(t)) := A(\mathbf{u}(t))\mathbf{x}$. We verify the Carathéodory conditions for existence of solution to system (1)

(C1) Measurability in t: For each fixed $\mathbf{x}$, the map $t \mapsto f(t, \mathbf{x}, \mathbf{u}(t))$ is measurable because $\mathbf{u}(\cdot)$ is measurable and $A(\cdot)$ is continuous.

(C2) Continuity in $\mathbf{x}$: For each fixed t and $\mathbf{u}(t)$, the map $\mathbf{x} \mapsto f(t, \mathbf{x}, \mathbf{u}(t))$ is linear, hence continuous.

(C3) Lipschitz continuity in $\mathbf{x}$: Let $\|\cdot\|$ denote the Euclidean norm. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ and almost every $t \in [0, T]$, we have

$$\begin{aligned} \|f(t, \mathbf{x}, \mathbf{u}(t)) - f(t, \mathbf{y}, \mathbf{u}(t))\| &= \|A(\mathbf{u}(t))(\mathbf{x} - \mathbf{y})\| \\ &\leq \|A(\mathbf{u}(t))\| \cdot \|\mathbf{x} - \mathbf{y}\|, \end{aligned}$$

where $\|A\|$ denotes the induced matrix norm.
Let $M := \max\{|\kappa|, |\tau_1|, |\tau_2|, |\tau_3|, |\phi_1|, |\phi_2|, |\omega|\}$, and let $U_{\max} := \max\{u_{1,\max}, u_{2,\max}\}$. Using the Frobenius norm $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$, we obtain the uniform bound

$$\|A(\mathbf{u}(t))\|_F \leq \sqrt{9M^2(1 + 4U_{\max}^2)} =: L.$$

Since all norms on $\mathbb{R}^{3\times 3}$ are equivalent, there exists $C > 0$ such that

$$\|A(\mathbf{u}(t))\| \leq C\|A(\mathbf{u}(t))\|_F \leq CL \quad \text{for a.e. } t \in [0, T].$$

Hence, f is Lipschitz continuous in $\mathbf{x}$ with constant CL.

(C4) Linear growth condition: For almost every $t \in [0, T]$, we have

$$\|f(t, \mathbf{x}, \mathbf{u}(t))\| \leq \|A(\mathbf{u}(t))\| \cdot \|\mathbf{x}\| \leq CL\|\mathbf{x}\|.$$

Since conditions (C1)–(C4) are satisfied, by the Carathéodory existence and uniqueness theorem for ordinary differential equations (see, e.g., [9]), there exists a unique absolutely continuous function $\mathbf{x} : [0, T] \to \mathbb{R}^3$ satisfying

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t f(s, \mathbf{x}(s), \mathbf{u}(s))\, ds \quad \text{for all } t \in [0, T].$$

□

Proposition 1. For nonnegative initial conditions, the solutions of (1) remain nonnegative for every $t \geq 0$.

Proof. We proceed by showing that the nonnegative orthant $\mathbb{R}^3_+$ is positively invariant. Consider the boundary faces of $\mathbb{R}^3_+$.

- If $G_p = 0$, then
$$\dot{G}_p = \tau_2 G_q \geq 0,$$
since $G_q \geq 0$ by the initial condition and continuity of solutions.
- If $G_q = 0$, then
$$\dot{G}_q = \tau_1 G_p \geq 0.$$
- If $G_d = 0$, then
$$\dot{G}_d = \phi_1 u_1 G_p \geq 0.$$

Equivalently, the system matrix $A(\mathbf{u}(t))$ is *quasi-positive* (Metzler) for all t, that is, all off-diagonal entries are nonnegative. Thus, by By Nagumo's invariance theorem [32], the set $\mathbb{R}^3_+$ is positively invariant. Consequently, for any nonnegative initial condition, the solution satisfies

$$\mathbf{x}(t) \in \mathbb{R}^3_+ \quad \text{for every } t \in [0, T].$$

□

Proposition 2. For bounded controls $0 \leq u_i(t) \leq u_{i,\max}$, the solutions of (1) remain bounded on any finite time interval $[0, T]$.

Proof. Let $M(t) = G_p(t) + G_q(t) + G_d(t)$. We get

$$\frac{dM}{dt} = \kappa(1 - \omega u_2)G_p - \phi_2 u_1 G_p - \tau_3 G_d \leq \kappa G_p \leq \kappa M.$$

By Gronwall's inequality, $M(t) \leq M(0)e^{\kappa t}$ for $t \in [0, T]$, establishing boundedness. □

3 Stability analysis of untreated glioma dynamics

Understanding the natural progression of glioma in the absence of therapeutic intervention is crucial for establishing baseline dynamics and quantifying the necessity of treatment. In this section, we conduct a stability analysis of the TFE for the uncontrolled system. This analysis serves two primary purposes: (1) to characterize the inherent growth dynamics of untreated glioma, and (2) to provide a reference point against which therapeutic efficacy can be evaluated.

Setting both therapeutic controls to zero ($u_1(t) = u_2(t) \equiv 0$) in the full model (1) yields the following autonomous system describing untreated glioma progression:

$$\begin{cases} \dfrac{dG_p}{dt} = \kappa G_p - \tau_1 G_p + \tau_2 G_q, \\ \dfrac{dG_q}{dt} = \tau_1 G_p - \tau_2 G_q, \\ \dfrac{dG_d}{dt} = -\tau_3 G_d. \end{cases} \tag{2}$$

Here all parameters are positive constants. We begin by identifying all steady-state solutions of (2).

Definition 1. A point $E^* = (G_p^*, G_q^*, G_d^*) \in \mathbb{R}^3_+$ is an equilibrium of system (2) if it satisfies

$$\left.\frac{dG_p}{dt}\right|_{E^*} = \left.\frac{dG_q}{dt}\right|_{E^*} = \left.\frac{dG_d}{dt}\right|_{E^*} = 0. \tag{3}$$

3.1 TFE

The TFE (tumor-free equilibrium), representing complete eradication of all tumor subpopulations, is given by

$$E_0 = (0, 0, 0). \tag{4}$$

This equilibrium exists unconditionally for all parameter values and represents the desired clinical outcome.

3.2 Existence of positive equilibria

To investigate the possibility of nontrivial steady states, we solve the system of algebraic equations obtained from (2):

$$0 = (\kappa - \tau_1)G_p^* + \tau_2 G_q^*, \tag{5}$$

$$0 = \tau_1 G_p^* - \tau_2 G_q^*, \tag{6}$$

$$0 = -\tau_3 G_d^*. \tag{7}$$

From (7), we immediately obtain $G_d^* = 0$, indicating that in the absence of therapy, no damaged cells persist at equilibrium. Equation (6) establishes a proportional relationship between quiescent and proliferating cells:

$$G_q^* = \frac{\tau_1}{\tau_2} G_p^*. \tag{8}$$

Substituting (8) into (5) yields

$$0 = (\kappa - \tau_1) G_p^* + \tau_2 \left(\frac{\tau_1}{\tau_2} G_p^* \right) = \kappa G_p^*. \tag{9}$$

Since $\kappa > 0$ by biological necessity (tumor cells possess positive proliferative capacity), (9) implies $G_p^* = 0$. Consequently, $G_q^* = 0$ from (8). This leads to the following important result.

Proposition 3. For the uncontrolled glioma system (2) with $\kappa > 0$, the TFE $E_0 = (0, 0, 0)$ is the only nonnegative equilibrium.

3.3 Local stability analysis of the TFE

We now investigate the local stability properties of E_0 using linearization techniques. The local stability determines whether infinitesimally small perturbations from the tumor-free state will decay (stable) or grow (unstable).

Theorem 2. The TFE E_0 of the uncontrolled system (2) is locally asymptotically unstable for all $\kappa > 0$. Specifically, E_0 is a saddle point with a one-dimensional unstable manifold and a two-dimensional stable manifold.

Proof. We linearize system (2) around E_0 by computing the Jacobian matrix:

$$J(E_0) = \begin{bmatrix} \frac{\partial f_1}{\partial G_p} & \frac{\partial f_1}{\partial G_q} & \frac{\partial f_1}{\partial G_d} \\ \frac{\partial f_2}{\partial G_p} & \frac{\partial f_2}{\partial G_q} & \frac{\partial f_2}{\partial G_d} \\ \frac{\partial f_3}{\partial G_p} & \frac{\partial f_3}{\partial G_q} & \frac{\partial f_3}{\partial G_d} \end{bmatrix}_{E_0} = \begin{bmatrix} \kappa - \tau_1 & \tau_2 & 0 \\ \tau_1 & -\tau_2 & 0 \\ 0 & 0 & -\tau_3 \end{bmatrix}, \tag{10}$$

where f_1, f_2, f_3 are the right-hand sides of equations in (2).

The eigenvalues of $J(E_0)$, denoted by $\lambda_1, \lambda_2, \lambda_3$, determine the local stability. Due to the block-triangular structure of $J(E_0)$, we can analyze the eigenvalues separately.

Eigenvalue from the damaged compartment:

The lower-right element gives $\lambda_1 = -\tau_3$. Since $\tau_3 > 0$, $\lambda_1 < 0$, indicating local stability in the G_d-direction.

Eigenvalues from the proliferating-quiescent subsystem:

The remaining eigenvalues come from the 2×2 submatrix:

$$J_{pq} = \begin{bmatrix} \kappa - \tau_1 & \tau_2 \\ \tau_1 & -\tau_2 \end{bmatrix}. \tag{11}$$

The characteristic polynomial of J_{pq} is

$$\begin{aligned} \det(J_{pq} - \lambda I_2) &= \det \begin{bmatrix} \kappa - \tau_1 - \lambda & \tau_2 \\ \tau_1 & -\tau_2 - \lambda \end{bmatrix} \\ &= (\kappa - \tau_1 - \lambda)(-\tau_2 - \lambda) - \tau_1\tau_2 \\ &= -\tau_2(\kappa - \tau_1 - \lambda) - \lambda(\kappa - \tau_1 - \lambda) - \tau_1\tau_2 \\ &= \lambda^2 + (\tau_1 + \tau_2 - \kappa)\lambda - \kappa\tau_2 = 0. \end{aligned} \tag{12}$$

Let λ_2 and λ_3 be the roots of (12). By Vieta's formulas,

$$\lambda_2 + \lambda_3 = -(\tau_1 + \tau_2 - \kappa) = \kappa - \tau_1 - \tau_2, \tag{13}$$

$$\lambda_2\lambda_3 = -\kappa\tau_2. \tag{14}$$

Since $\kappa > 0$ and $\tau_2 > 0$, (14) gives $\lambda_2\lambda_3 < 0$, which implies that λ_2 and λ_3 have opposite signs. Without loss of generality, assume $\lambda_2 < 0$ and $\lambda_3 > 0$.

To confirm the existence of a positive eigenvalue, we can evaluate the characteristic polynomial at $\lambda = 0$:

$$P(0) = -\kappa\tau_2 < 0.$$

Since $P(\lambda) \to +\infty$ as $\lambda \to +\infty$, by the Intermediate Value Theorem, there exists $\lambda_3 > 0$ such that $P(\lambda_3) = 0$.

Alternatively, applying Descartes' rule of signs to $P(\lambda) = \lambda^2 + (\tau_1 + \tau_2 - \kappa)\lambda - \kappa\tau_2$, the coefficient sequence is $[1, (\tau_1 + \tau_2 - \kappa), -\kappa\tau_2]$. There is exactly one sign change regardless of the sign of $(\tau_1 + \tau_2 - \kappa)$, guaranteeing exactly one positive real root.

all eigenvalues:

The complete set of eigenvalues is

$$\lambda_1 = -\tau_3 < 0, \tag{15}$$

$$\lambda_2 < 0, \tag{16}$$

$$\lambda_3 > 0. \tag{17}$$

Since there exists at least one eigenvalue with positive real part ($\lambda_3 > 0$), by Lyapunov's indirect method [22], the equilibrium E_0 is unstable. Moreover, with two negative eigenvalues and one positive eigenvalue, E_0 is classified as a saddle point. The stable manifold is two-dimensional (spanned by eigenvectors corresponding to λ_1 and λ_2), while the unstable manifold is one-dimensional (corresponding to λ_3).

This completes the proof of Theorem 2. □

3.4 Explicit growth rate and biological interpretation

The positive eigenvalue λ_3 can be computed explicitly from the quadratic formula applied to (12):

$$\lambda_3 = \frac{-(\tau_1 + \tau_2 - \kappa) + \sqrt{(\tau_1 + \tau_2 - \kappa)^2 + 4\kappa\tau_2}}{2}. \tag{18}$$

This expression represents the *net growth rate* of the glioma population in the absence of therapy.

Proposition 4. The growth rate λ_3 defined in (18) satisfies the following conditions:

1. $\lambda_3 > 0$ for all $\kappa > 0$, $\tau_1 > 0$, $\tau_2 > 0$.
2. λ_3 is increasing in κ: $\frac{\partial \lambda_3}{\partial \kappa} > 0$.
3. λ_3 satisfies $0 < \lambda_3 < \kappa$.

Proof. 1. The positivity follows from the analysis in Theorem 2.

2. Differentiating (18) with respect to κ, we have

$$\frac{\partial \lambda_3}{\partial \kappa} = \frac{1}{2}\left(1 + \frac{\tau_1 + \tau_2 - \kappa + 2\tau_2}{\sqrt{(\tau_1 + \tau_2 - \kappa)^2 + 4\kappa\tau_2}}\right) > 0.$$

3. From the characteristic polynomial $P(\lambda) = \lambda^2 + (\tau_1 + \tau_2 - \kappa)\lambda - \kappa\tau_2$, we have $P(\kappa) = \kappa^2 + (\tau_1 + \tau_2 - \kappa)\kappa - \kappa\tau_2 = \kappa\tau_1 > 0$. Since $P(0) = -\kappa\tau_2 < 0$ and $P(\kappa) > 0$, by the Intermediate Value Theorem, the positive root λ_3 satisfies $0 < \lambda_3 < \kappa$. □

3.5 Global dynamics and tumor progression

While local stability analysis characterizes behavior near equilibrium, we are also interested in the global behavior for biologically relevant initial conditions.

Proposition 5. The G_d equation in (2) is decoupled and can be solved analytically:

$$\frac{dG_d}{dt} = -\tau_3 G_d \quad \Rightarrow \quad G_d(t) = G_d(0)e^{-\tau_3 t}.$$

Thus, $G_d(t) \to 0$ exponentially for any initial condition $G_d(0) \geq 0$.

The dynamics of proliferating and quiescent cells are governed by the planar linear system:

$$\frac{d}{dt}\begin{bmatrix} G_p \\ G_q \end{bmatrix} = \begin{bmatrix} \kappa - \tau_1 & \tau_2 \\ \tau_1 & -\tau_2 \end{bmatrix}\begin{bmatrix} G_p \\ G_q \end{bmatrix}. \tag{19}$$

Theorem 3. For any initial condition $(G_p(0), G_q(0)) \neq (0, 0)$ with nonnegative components, the solution of (19) satisfies

$$\lim_{t\to\infty} \frac{1}{t} \log \|(G_p(t), G_q(t))\| = \lambda_3 > 0,$$

where λ_3 is given by (18). That is, the total tumor burden grows exponentially at rate λ_3.

Proof. System (19) is linear and homogeneous. Let $\mathbf{v}_2$ and $\mathbf{v}_3$ be eigenvectors corresponding to eigenvalues $\lambda_2 < 0$ and $\lambda_3 > 0$, respectively. The general solution is

$$\begin{bmatrix} G_p(t) \\ G_q(t) \end{bmatrix} = c_2 e^{\lambda_2 t}\mathbf{v}_2 + c_3 e^{\lambda_3 t}\mathbf{v}_3,$$

where c_2, c_3 are constants determined by initial conditions.

For generic initial conditions not aligned with the stable manifold (i.e., $c_3 \neq 0$), as $t \to \infty$,

$$\begin{bmatrix} G_p(t) \\ G_q(t) \end{bmatrix} \sim c_3 e^{\lambda_3 t}\mathbf{v}_3,$$

since $e^{\lambda_3 t}$ dominates $e^{\lambda_2 t}$ ($\lambda_3 > 0 > \lambda_2$). Therefore,

$$\|(G_p(t), G_q(t))\| \sim |c_3|\|\mathbf{v}_3\|e^{\lambda_3 t},$$

and

$$\lim_{t\to\infty} \frac{1}{t} \log \|(G_p(t), G_q(t))\| = \lambda_3 > 0.$$

□

Corollary 1. As $t \to \infty$, the ratio of quiescent to proliferating cells approaches

$$\lim_{t\to\infty} \frac{G_q(t)}{G_p(t)} = \frac{\tau_1}{\tau_2 + \lambda_3},$$

which is determined by the eigenvector $\mathbf{v}_3$ corresponding to λ_3.

4 Stability under constant therapy

We investigate the stability of the glioma dynamics under constant therapy inputs, where the control functions are fixed $u_1(t) = \tilde{u}_1 \in [0, u_{1,\max}]$ and $u_2(t) = \tilde{u}_2 \in [0, u_{2,\max}]$ reflecting a continuous and sustained treatment regime. The main objectives of this analysis are as follows: Establish conditions for tumor eradication by assessing the local stability of the TFE, characterize scenarios of treatment failure through the identification of endemic equilibria, and evaluate the stabilizing impact of IDH-targeted therapy on the proliferative tumor population. System (1) can reduced to a nonlinear autonomous system of ordinary differential equations:

$$\begin{aligned} \frac{dG_p}{dt} &= (\kappa(1-\omega\tilde{u}_2) - \tau_1 - (\phi_1+\phi_2)\tilde{u}_1)G_p + \tau_2 G_q, \\ \frac{dG_q}{dt} &= \tau_1 G_p - \tau_2 G_q, \\ \frac{dG_d}{dt} &= \phi_1\tilde{u}_1 G_p - \tau_3 G_d. \end{aligned} \tag{20}$$

We analyze the system's equilibrium points and their local stability. The system admits a TFE at the origin $(G_p^*, G_q^*, G_d^*) = (0, 0, 0)$, which corresponds biologically to the eradication of all tumor subpopulations. In the case where $G_p^* \neq 0$, solving the steady-state conditions $\dot{G}_p = \dot{G}_q = \dot{G}_d = 0$ yields

$$\begin{aligned} G_q^* &= \frac{\tau_1}{\tau_2} G_p^*, \\ G_d^* &= \frac{\phi_1\tilde{u}_1}{\tau_3} G_p^*. \end{aligned}$$

Substitution into the equation for G_p leads to the following critical condition for the existence of a nontrivial equilibrium:

$$\kappa(1-\omega\tilde{u}_2) = (\phi_1+\phi_2)\tilde{u}_1, \tag{21}$$

Remark 1. Equation (21) defines the threshold where therapeutic effects exactly balance tumor proliferation.

4.1 Local stability analysis

Theorem 4. The TFE point $(0,0,0)$ is locally asymptotically stable if and only if the controlled reproduction number [40, 11]

$$\mathcal{R}_0^{\text{controlled}} := \frac{\kappa(1-\omega\tilde{u}_2)}{(\phi_1+\phi_2)\tilde{u}_1},$$

satisfies $\mathcal{R}_0^{\text{controlled}} < 1$.

Proof. The Jacobian matrix J_0 of system (1) evaluated at the origin is therefore given by

$$J_0 = \begin{bmatrix} \kappa(1-\omega\tilde{u}_2) - \tau_1 - (\phi_1+\phi_2)\tilde{u}_1 & \tau_2 & 0 \\ \tau_1 & -\tau_2 & 0 \\ \phi_1\tilde{u}_1 & 0 & -\tau_3 \end{bmatrix}. \tag{22}$$

This Jacobian J_0 has an upper block-triangular structure, and the eigenvalues are thus composed of the following items:

1. $\lambda_1 = -\tau_3$, which is always negative, hence stable.

2. The remaining eigenvalues are obtained from the 2×2 upper-left submatrix:

$$J_2 = \begin{bmatrix} a & \tau_2 \\ \tau_1 & -\tau_2 \end{bmatrix}, \quad \text{where} \quad a = \kappa(1-\omega\tilde{u}_2) - \tau_1 - (\phi_1+\phi_2)\tilde{u}_1. \tag{23}$$

The characteristic polynomial for this submatrix is

$$\mathcal{X}_{J_2}(\lambda) = det(J_2 - \lambda.I_2) = \lambda^2 + (\tau_2 - a)\lambda - (a\tau_2 + \tau_1\tau_2) = 0. \tag{24}$$

Applying the Routh–Hurwitz criteria [12], local stability requires the following conditions:

- The trace condition: $\text{tr}(J_2) = a - \tau_2 < 0$,

- The determinant condition: $\det(J_2) = -a\tau_2 - \tau_1\tau_2 > 0$.

These two inequalities are simultaneously satisfied if and only if $a < -\tau_1$. This yields a critical threshold for the combined effect of the therapies. Defining the effective controlled reproduction rate $\mathcal{R}_0^{\text{controlled}}$ via

$$\mathcal{R}_0^{\text{controlled}} := \frac{\kappa(1-\omega\tilde{u}_2)}{(\phi_1+\phi_2)\tilde{u}_1}, \tag{25}$$

the condition $a < -\tau_1$ becomes equivalent to $\mathcal{R}_0^{\text{controlled}} < 1$. Thus, the TFE is locally asymptotically stable.

□

Remark 2. Although this threshold is derived directly from the Routh–Hurwitz criterion applied to the Jacobian of the linearized system, it exactly determines the local asymptotic stability of the TFE.

4.2 Global stability of the TFE

We now establish that, under a sufficient therapeutic regimen, the tumor-free state is the unique attractor for all biologically admissible initial conditions.

Theorem 5. If $\mathcal{R}_0^{\text{controlled}} < 1$, then the equilibrium $E_0 = (0,0,0)$ is globally asymptotically stable in the nonnegative orthant.

Proof. We work in $\mathbb{R}^3_+$ and introduce the linear Lyapunov function

$$V(G_p, G_q, G_d) \;=\; G_p \;+\; \frac{\tau_2}{\tau_1} G_q \;+\; \eta G_d,$$

where $\eta > 0$ will be chosen later. Clearly, $V > 0$ for any $(G_p, G_q, G_d) \neq (0,0,0)$, and $V = 0$ at E_0. Under constant controls $\tilde{u}_1, \tilde{u}_2$, the model reads

$$\begin{aligned}
\dot{G}_p &= a\, G_p + \tau_2\, G_q, \\
\dot{G}_q &= \tau_1\, G_p - \tau_2\, G_q, \\
\dot{G}_d &= \phi_1\, \tilde{u}_1\, G_p - \tau_3\, G_d,
\end{aligned}$$

where

$$a \;=\; \kappa(1-\omega\tilde{u}_2) - \tau_1 - (\phi_1+\phi_2)\,\tilde{u}_1.$$

Differentiating V along solutions gives

$$\begin{aligned}\dot{V} &= \dot{G}_p + \frac{\tau_2}{\tau_1}\dot{G}_q + \eta\dot{G}_d \\ &= aG_p + \tau_2 G_q + \frac{\tau_2}{\tau_1}(\tau_1 G_p - \tau_2 G_q) + \eta(\phi_1\tilde{u}_1 G_p - \tau_3\, G_d).\end{aligned}$$

We note by the construction that

$$\tau_2 G_p - \frac{\tau_2^2}{\tau_1} G_q = 0.$$

So transition terms cancel exactly, we obtain

$$\dot{V} = (a + \tau_2 + \eta\phi_1\tilde{u}_1)G_p - \eta\tau_3 G_d.$$

By taking $\eta = \frac{-(a+\tau_2)}{2\phi_1\tilde{u}_1} > 0$, small enough so that $a + \tau_2 + \eta\phi_1\tilde{u}_1 < 0$. Since $\kappa(1-\omega\tilde{u}_2) < (\phi_1+\phi_2)\,\tilde{u}_1$, which is exactly $\mathcal{R}_0^{\text{controlled}} < 1$. Under this hypothesis, $\dot{V}$ is strictly negative on $\mathbb{R}^3_+ \setminus \{E_0\}$. By LaSalle's invariance principle [25], all trajectories originating in $\mathbb{R}^3_+$ converge to the largest invariant subset of $\{\dot{V}=0\}$, namely the singleton $\{E_0\}$.

Hence, E_0 is globally asymptotically stable. □

Remark 3. The condition $a < -\tau_1$ (or equivalently, $\mathcal{R}_0^{\text{controlled}} < 1$ i.e. when the total therapeutic effect meets or exceeds the tumor proliferation rate) guarantees both local and global eradication of the tumor cell populations under constant therapy.

5 The optimal control problem

Among the chemotherapeutic agents commonly used in the treatment of LGGs, temozolomide remains a frontline therapy due to its ability to cross the blood-brain barrier and its cytotoxic activity against rapidly dividing tumor cells [31]. TMZ is an alkylating agent that introduces methyl groups into DNA, leading to DNA damage, replication stress, and apoptosis, particularly in proliferating tumor cells. In parallel, recent clinical advances have led to the approval of IDH inhibitors as targeted therapies for LGGs harboring IDH1 or IDH2 mutations [13]. IDH inhibitors act by suppressing the production of the oncometabolite 2-hydroxyglutarate, thereby reversing epigenetic dysregulation and promoting differentiation of tumor cells [30, 7]. Notably, IDH inhibitors exhibit a cytostatic effect that slows proliferation and may enhance the effectiveness of cytotoxic agents like TMZ when used in combination. In our model, we represent the cytotoxic control (TMZ) by the function $u_1(t)$, which governs the dose of chemotherapy administered over

time. A value of $u_1 = 0$ denotes no drug use, while $u_1 = u_{1,\max}$ corresponds to the maximum tolerable dose, constrained within the range $0 \leq u_1(t) \leq u_{1,\max} < 1$. The targeted IDH inhibitor is modeled by the control $u_2(t)$, with $u_2 = 0$ indicating absence of the therapy and $u_2 = u_{2,\max}$ corresponding to full inhibition. The dosage range is similarly bounded as $0 \leq u_2(t) \leq u_{2,\max} < 1$. Both therapies are assumed to act bilinearly on the proliferating tumor compartment, with TMZ reducing the growth and survival of dividing cells and the IDH inhibitor modulating the tumor microenvironment and the metabolic state of resistant or quiescent cells. By integrating these two mechanisms into our bilinear model, we aim to capture the synergistic dynamics of dual-agent treatment and better inform optimized therapy schedules for LGG patients. Unlike prior studies that evaluate system behavior under fixed or time-invariant treatment intensities, our analysis investigates the tumor dynamics under time-dependent optimal control strategies. Specifically, we consider therapeutic inputs $u_1(t)$ and $u_2(t)$ as measurable functions of time, optimized to minimize tumor burden and treatment cost over a fixed treatment horizon $[0, T]$.

The objective function to be minimized is defined as

$$J(u_1, u_2) = \int_0^T \left(A_1 G_p(t) + A_2 G_q(t) + \tfrac{1}{2} B_1 u_1^2(t) + \tfrac{1}{2} B_2 u_2^2(t)\right) dt, \tag{26}$$

where $G_p(t)$ and $G_q(t)$ represent, respectively, the proliferating and quiescent tumor cell populations at time t, and $A_i > 0$ and $B_i > 0$ $(i = 1, 2)$ are fixed weighting coefficients that reflect the relative importance of reducing tumor burden and minimizing the therapeutic dosing effort.

We seek optimal control functions $u_1^*(t)$ and $u_2^*(t)$ that minimize the cost functional J, that is,

$$J(u_1^*, u_2^*) = \min_{(u_1, u_2) \in \mathcal{U}_{ad}} J(u_1, u_2), \tag{27}$$

where the set of admissible controls $\mathcal{U}_{ad}$ is defined as follows:

$$\mathcal{U}_{ad} = \{(u_1(t), u_2(t)) \in L^\infty(0, T)^2; \quad 0 \leq u_i(t) \leq u_{i,\max} < 1, \text{ for } i = 1, 2\}. \tag{28}$$

This optimal control framework allows us to identify therapy protocols that minimize the tumor size and overall tumor burden throughout the treatment period, while penalizing high doses of TMZ and the IDH inhibitor to reduce toxicity and cost.

5.1 Existence of optimal control solutions

We establish the existence of an optimal control pair $(u_1^*, u_2^*) \in \mathcal{U}_{ad}$ for the control problem governed by the bilinear system (1) and the cost functional (26), using a standard result from optimal control theory (see [14, Corollary 4.1]).

Theorem 6. Consider the optimal control problem defined by system (1) with the objective functional (26). There exists an optimal control pair $(u_1^*, u_2^*) \in \mathcal{U}_{ad}$ such that

$$J(u_1^*, u_2^*) = \min_{(u_1,u_2)\in\mathcal{U}_{ad}} J(u_1, u_2).$$

Proof. To guarantee existence of optimal controls, we verify the following standard conditions:

1. **Nonemptiness of admissible controls and corresponding state trajectories:** Let the state vector be $x = (G_p, G_q, G_d)$ with the right-hand side of the system (1) given by $\dot{x} = F(t, x, u_1, u_2)$. Since the system is bilinear in u_1 and u_2, and all coefficients and parameters are continuous and bounded, and the controls are bounded measurable functions, the right-hand side is Lipschitz continuous in x. Therefore, by the Picard–Lindelof theorem, for any admissible control $(u_1, u_2) \in \mathcal{U}_{ad}$, there exists a unique solution $x(t) \in C^1([0, T]; \mathbb{R}^3)$ satisfying the initial conditions. Hence, the set of admissible control-state pairs is nonempty.

2. **Convexity and closedness of the control set:** From the definition of the controls set (28), it is clear that $\mathcal{U}_{ad}$ is a product of closed intervals in L^∞, hence convex and closed.

3. **Bilinearity of the system in control variables:**
The state system (1) is bilinear in u_1 and u_2 with coefficients depending continuously on the state and time. Since the state trajectories are bounded (due to biologically relevant constraints and bounded control functions), the function $F(t, x, u_1, u_2)$ is continuous and satisfies the required regularity conditions.

4. **Convexity and lower boundedness of the integrand in the cost functional:**
The cost functional is of the form:

$$J(u_1, u_2) = \int_0^T \left(\sum_{i=1}^{2} A_i x_i(t) + \frac{1}{2} B_1 u_1(t)^2 + \frac{1}{2} B_2 u_2(t)^2 \right) dt.$$

This integrand is convex in the controls u_1, u_2, because it is a positive definite quadratic form in u_1 and u_2. Moreover, since $L(t, x, u_1, u_2) \geq (B1/2)u_1^2 + (B_2/2)u_2^2 \geq c_2(|u_1^2| + |u_2^2|)$

with $c_2 = \min(B_1/2, B_2/2)$, $\xi = 2$, and $L \geq 0$, together with the uniform boundedness of trajectories on $[0, T]$(Proposition 2), the standard coercivity and compactness conditions of optimal control theory are satisfied (e.g., [14]).

All hypotheses of the Fleming–Rishel existence theorem are verified. Therefore, there exists an optimal control pair $(u_1^*, u_2^*) \in \mathcal{U}_{ad}$ minimizing the objective functional $J(u_1, u_2)$. □

5.2 Characterization of an optimal controls

To derive the necessary conditions for optimality in the LGG therapy model, we employ Pontryagin's maximum principle [23]. This principle introduces a set of adjoint variables that link the state dynamics with the cost functional through a Hamiltonian function. The optimization problem is thus reformulated as a two-point boundary value problem: The state variables evolve forward in time with known initial conditions, while the adjoint variables evolve backward from known terminal conditions. The optimality system consists of the state equations, adjoint equations, and pointwise minimization of the Hamiltonian with respect to the control variables, subject to their admissible bounds. This structured approach provides a powerful framework to characterize the optimal therapeutic strategy.

The Hamiltonian associated with the problem is defined as

$$\mathcal{H}(t) = A_1 G_p(t) + A_2 G_q(t) + \frac{1}{2} B_1 u_1^2(t) + \frac{1}{2} B_2 u_2^2(t) + \lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3, \tag{29}$$

where the functions f_i for $i = 1, 2, 3$ represent the right-hand sides of the system of differential equations (1), and λ_i are the adjoint functions. Therefore

$$\begin{aligned}
f_1(G_p, G_q, G_d, u_1, u_2) &= \kappa(1 - \omega u_2) G_p - \tau_1 G_p + \tau_2 G_q - \phi_1 u_1 G_p - \phi_2 u_1 G_p, \\
f_2(G_p, G_q, G_d, u_1, u_2) &= \tau_1 G_p - \tau_2 G_q, \\
f_3(G_p, G_q, G_d, u_1, u_2) &= \phi_1 u_1 G_p - \tau_3 G_d.
\end{aligned}$$

Theorem 7. Let $(u_1^*, u_2^*) \in \mathcal{U}_{ad}$ be an optimal control pair minimizing the cost functional $J(u_1, u_2)$, subject to the system dynamics (1) and the admissible control set $\mathcal{U}_{ad}$. Then, there exist adjoint functions $\lambda_1(t), \lambda_2(t), \lambda_3(t)$ satisfying the adjoint system

$$\begin{cases} \dot{\lambda}_1 = -A_1 - \lambda_1(\kappa(1-\omega u_2) - \tau_1 - (\phi_1+\phi_2)u_1) - \lambda_2\tau_1 - \lambda_3\phi_1 u_1, \\ \dot{\lambda}_2 = -A_2 - \lambda_1\tau_2 + \lambda_2\tau_2, \\ \dot{\lambda}_3 = \lambda_3\tau_3, \end{cases} \tag{30}$$

with terminal conditions

$$\lambda_1(T) = \lambda_2(T) = 0.$$

Furthermore, the optimal controls $u_1^*(t), u_2^*(t)$ are characterized by

$$\begin{aligned} u_1^*(t) &= \min\left(1, \max\left(0, \frac{((\phi_1+\phi_2)\lambda_1(t) - \phi_1\lambda_3(t))G_p(t)}{B_1}\right)\right), \\ u_2^*(t) &= \min\left(1, \max\left(0, \frac{\kappa\omega\lambda_1(t)G_p(t)}{B_2}\right)\right). \end{aligned} \tag{31}$$

Proof. For $t \in [0,T]$, the adjoint equations and the transversality conditions can be derived using the the Pontryagin's maximum principle [23], such that

$$\begin{aligned} \dot{\lambda}_1 &= -\frac{\partial \mathcal{H}}{\partial G_p} = -A_1 - \lambda_1(\kappa(1-\omega u_2) - \tau_1 - (\phi_1+\phi_2)u_1) - \lambda_2\tau_1 - \lambda_3\phi_1 u_1, \\ \dot{\lambda}_2 &= -\frac{\partial \mathcal{H}}{\partial G_q} = -A_2 - \lambda_1\tau_2 + \lambda_2\tau_2, \\ \dot{\lambda}_3 &= -\frac{\partial \mathcal{H}}{\partial G_d} = \lambda_3\tau_3, \end{aligned} \tag{32}$$

The Hamiltonian minimization condition requires that, at the optimal solution, derive $\mathcal{H}$ with respect to the controls and setting the derivatives to zero gives the following system:

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial u_1} &= B_1 u_1 - \lambda_1(\phi_1+\phi_2)G_p + \lambda_3\phi_1 G_p = 0, \\ \frac{\partial \mathcal{H}}{\partial u_2} &= B_2 u_2 - \lambda_1\kappa\omega G_p = 0. \end{aligned} \tag{33}$$

It follows

$$\begin{aligned} u_1 &= \frac{((\phi_1+\phi_2)\lambda_1 - \phi_1\lambda_3)G_p}{B_1}, \\ u_2 &= \frac{\kappa\omega\lambda_1 G_p}{B_2}. \end{aligned}$$

Applying the control constraints, we project each expression onto the admissible interval:

$$\begin{aligned} u_1^*(t) &= \min\left(1, \max\left(0, \frac{((\phi_1+\phi_2)\lambda_1 - \phi_1\lambda_3)G_p}{B_1}\right)\right), \\ u_2^*(t) &= \min\left(1, \max\left(0, \frac{\kappa\omega\lambda_1 G_p}{B_2}\right)\right). \end{aligned}$$

□

6 Numerical simulations

To assess the performance of the proposed optimal control strategy, we carried out numerical simulations using the forward-backward sweep method (FBSM), a well-established iterative algorithm for solving optimal control problems governed by systems of ordinary differential equations [28]. The method alternates between forward integration of the state equations and backward integration of the adjoint equations, followed by an update of the control variables based on the optimality conditions derived from Pontryagin's maximum principle. This iterative process is repeated until convergence, defined in terms of the normed difference between successive control iterates falling below a prescribed tolerance.

The state system was solved forward in time using a fourth-order Runge–Kutta scheme, while the adjoint system was integrated backward in time under the terminal conditions $\lambda_i(T) = 0$. At each iteration, the optimal controls were updated via projection onto the admissible control set, ensuring biologically feasible treatment intensities. Convergence was typically achieved within a small number of iterations, indicating numerical stability and robustness of the proposed algorithm.

Model parameters were calibrated using clinical data reported in [36], which provide longitudinal measurements of LGG size in patients treated with temozolomide. Tumor burden was quantified through the mean tumor diameter (in millimeters), consistent with clinical practice. When necessary, tumor volumes were reconstructed assuming spherical geometry. The model was formulated in normalized (dimensionless) form by scaling tumor cell densities with respect to a reference carrying capacity K, representing the maximal viable tumor cell density in brain tissue. Consequently, the state variables $0 \leq P, Q, D \leq 1$. Initial conditions were inferred from early diagnostic imaging and chosen to reflect biologically realistic compartment proportions observed in LGG. Specifically, the initial proliferative fraction was set to $P(0) = 0.01$, and the quiescent fraction to $Q(0) = 0.09$, yielding an initial total tumor burden corresponding to 10% of the carrying capacity. No damaged cells were assumed at treatment initiation, so $D(0) = 0$.

This choice reflects the known predominance of quiescent cells in LGGs, together with a smaller actively proliferating subpopulation prior to therapy.

The simulation horizon was fixed at 25 months in order to capture both short-term treatment effects and long-term tumor dynamics. The parameter values reported in Table 1 are used for illustrative purposes; the qualitative behavior of the system depends primarily on the structural

interactions among proliferation, quiescence, and treatment effects rather than on the precise numerical values.

Table 1: The parameters κ, τ_1, τ_2, τ_3, ϕ_1, ϕ_2 are taken from the clinical tumor growth inhibition model developed by Ribba et al. in their LGG TMZ-only work [36]. These parameters were originally estimated by fitting longitudinal MRI tumor size data under temozolomide therapy. All rates are expressed in month^{-1}.

Parameter	Description	Value
κ	Natural proliferation rate of tumor cells	0.114
τ_1	Transition rate from proliferative to quiescent state	0.0226
τ_2	Transition rate from quiescent to proliferative state	0.0045
τ_3	Clearance rate of damaged tumor cells	0.0214
ϕ_1	TMZ-induced damage rate on proliferative cells	0.842
ϕ_2	Cytotoxic effect of TMZ on proliferative cells	0.32
ω	Effect rate of IDH inhibitor on proliferative cells	0.3

The IDH-inhibitor efficacy $\omega = 0.2$ month^{-1} is chosen as a conservative estimate of cytostatic slowdown based on emerging IDH-inhibitor data (e.g., reduced proliferation without strong cytotoxicity). Sensitivity to ω is explored in simulations. Here, we investigate two distinct control strategies over a 30-month treatment horizon:

- **Strategy 1:** IDH-inhibitor monotherapy (cytostatic control $u_2(t)$ only, $u_1(t) \equiv 0$),
- **Strategy 2:** Combined therapy with temozolomide (cytotoxic control $u_1(t)$) and IDH-inhibitor (cytostatic control $u_2(t)$).

The objective is to compare the qualitative and quantitative tumor responses under purely proliferation-slowing treatment versus dual cytotoxic–cytostatic intervention.

Uncontrolled dynamics

In the absence of treatment, the proliferating tumor cell population (G_p) grows exponentially (Figure 2), reflecting the intrinsic instability of the untreated system. This behavior reproduces the characteristic slow but persistent progression of LGG. Simultaneously, the quiescent compartment (G_q) gradually accumulates (Figure 3) , forming a latent reservoir capable of sustaining future tumor regrowth. As expected, the damaged compartment (G_d) remains identically zero without therapy (Figure 4).

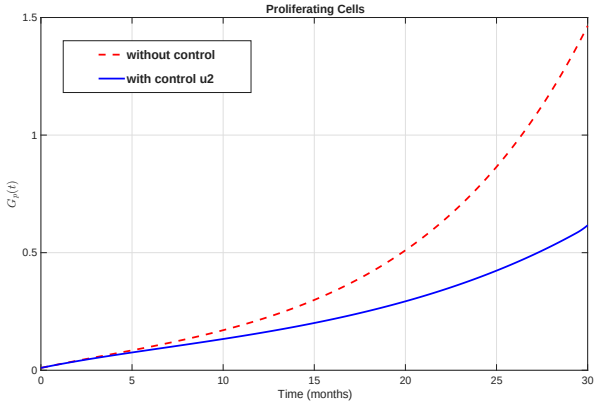


Figure 2: Dynamics of proliferating compartment (G_p) with control u_2 only.

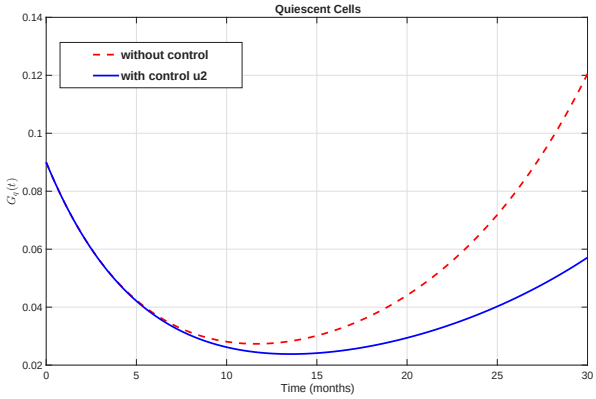


Figure 3: Dynamics of quiescent compartment (G_q) with control u_2 only.

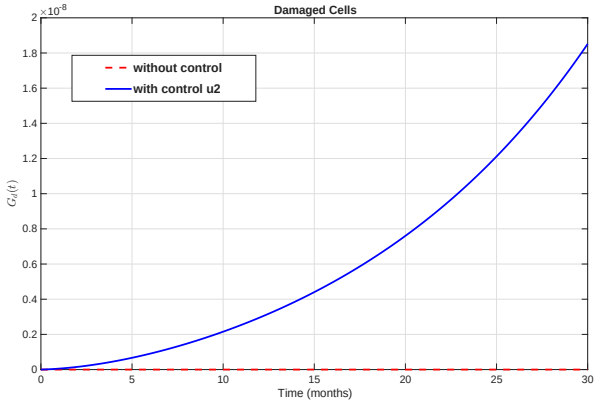


Figure 4: Dynamics of damaged compartment (G_d) with control u_2 only.

Strategy 1: IDH-inhibitor monotherapy (cytostatic control)

In this part of the simulations, we consider the optimal cytostatic control $u_2(t)$ alone, with $u_1(t) \equiv 0$. As shown in Figure 2, the proliferating cell population G_p exhibits a marked reduction compared to the uncontrolled case. The IDH-inhibitor acts by decreasing the effective proliferation rate, thereby slowing tumor expansion rather than inducing direct cell death. Consequently, G_p does not undergo complete eradication but instead converges toward a lower stabilized level, reflecting tumor growth containment rather than elimination. Quantitatively, the proliferating population decreases from a peak value exceeding 1.5 units to approximately 0.6 units after treatment, demonstrating substantial but partial suppression of tumor activity.

As shown in Figure 3, the quiescent compartment G_q decreases indirectly under IDH-inhibitor monotherapy, dropping from approximately 0.12 to less than 0.06. This reduction results from the cytostatic action of $u_2(t)$, which lowers the effective proliferation rate and consequently limits the flux of newly generated cells transitioning into the quiescent state. Importantly, because the IDH inhibitor is modeled as a purely cytostatic agent, the damaged compartment G_d remains negligible during therapy.

Strategy 2: Combined TMZ and IDH Therapy

When both controls are active, a markedly stronger therapeutic effect is observed. The proliferating compartment G_p rapidly declines toward near-zero levels (Figure 5), demonstrating the synergistic effect of cytotoxic killing $u_1(t)$ and proliferation slowdown $u_2(t)$. This combined action not only halts tumor growth but actively reduces tumor burden.

The quiescent population G_q also progressively decreases to near zero, as shown in (Figure 6). Although its decline is slower than that of G_p, the sustained reduction reflects both diminished inflow from proliferating cells and indirect effects of cytotoxic depletion. The gradual elimination of this dormant reservoir is critical to minimizing recurrence risk.

In contrast to monotherapy, the damaged compartment G_d (Figure 7) emerges exclusively under TMZ administration. Its monotonic increase during treatment serves as a marker of irreversible tumor cell inactivation. The accumulation of G_d, together with its clearance dynamics, illustrates the long-term reduction in the tumor's effective reproductive capacity achieved through cytotoxic intervention.

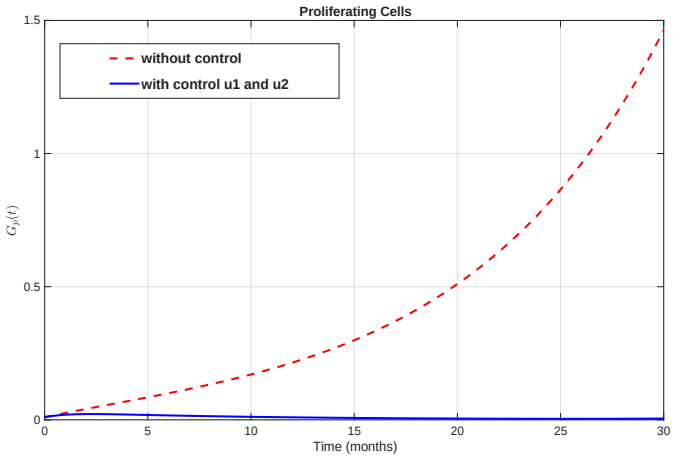


Figure 5: Dynamics of proliferating compartment (G_p) with controls u_1 and u_2.

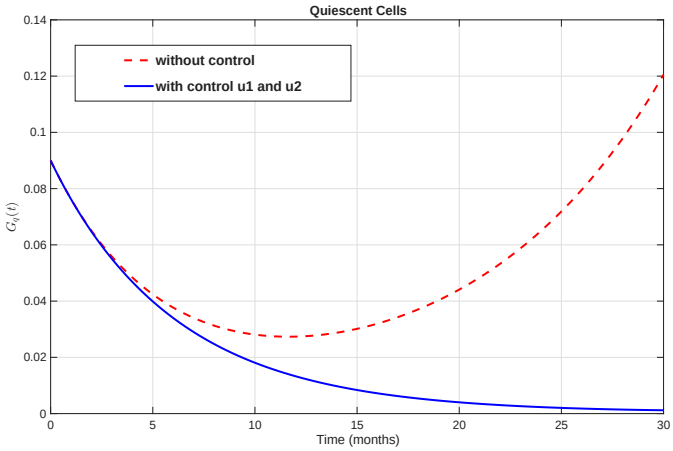


Figure 6: Dynamics of quiescent compartment (G_q) with controls u_1 and u_2.

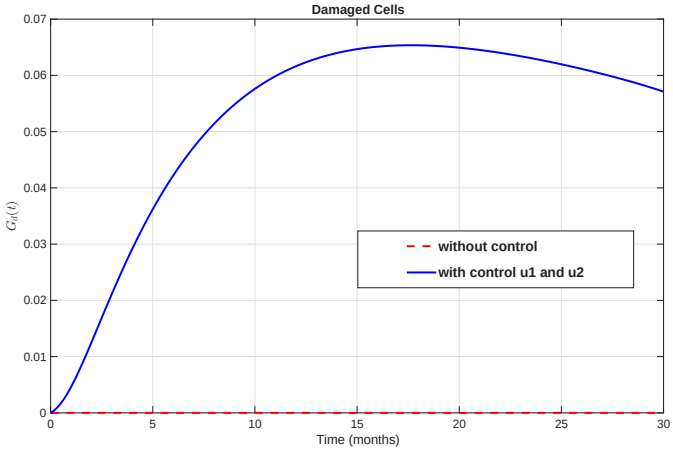


Figure 7: Dynamics of damaged compartment (G_d) with controls u_1 and u_2.

Remark 4. The simulations highlight a fundamental difference between the two strategies. IDH-inhibitor monotherapy achieves tumor growth stabilization through cytostatic slowing of proliferation but does not induce substantial irreversible tumor reduction. In contrast, the combined TMZ–IDH therapy simultaneously suppresses active proliferation, depletes the quiescent reservoir, and generates irreversible cellular damage, leading to sustained tumor control.

From a clinical perspective, these results suggest that cytostatic maintenance alone may contain tumor progression, whereas combination therapy is required to achieve significant tumor burden reduction and long-term disease stabilization.

7 Conclusion

In this study, we developed a bilinear compartmental model to describe LGG dynamics under combined therapy with temozolomide (TMZ) chemotherapy and IDH-targeted treatment. The model includes important biological processes such as the reversible transition between proliferating and quiescent tumor cells, therapy-induced damage, and removal of damaged cells.

Mathematically, we showed that the system is well-posed: Solutions exist, are unique, remain positive, and are bounded for any admissible control. The analysis of the system without treatment showed that the TFE is unstable, meaning the tumor will grow if left untreated. This highlights the need for therapy to control tumor growth.

For constant treatments, we defined a threshold parameter, called the controlled reproduction number, which measures the balance between tumor growth and therapy. When this number is below one, the TFE becomes stable, showing that constant therapy at sufficient intensity can stop tumor growth. This demonstrates the benefit of combining TMZ and IDH-targeted therapy.

We also solved an optimal control problem to find the best time-dependent treatment schedules that reduce tumor size while keeping drug use reasonable. The numerical results show that the optimal controls start with a strong initial dose, then remain nearly constant for a long period, and finally decrease gradually. This pattern matches the theoretical stability results: Once the tumor is under control, maintaining it with a steady treatment is enough. The long constant phase suggests that a maintenance-like strategy can be effective while limiting side effects.

Overall, this study shows that combining mathematical modeling, stability analysis, and optimal control can help design effective treatment strategies for LGG. The results could be useful for developing personalized treatment plans that balance tumor control with patient well-being.

Funding

This research received no external funding.

Conflicts of Interest

All authors jointly worked on the results and they read and approved the final manuscript, and also declare that they have no competing interests.

References

[1] Andrica, D., and Rassias, T. *Differential and integral inequalities*, Cham: Springer, 2019.

[2] Ayala-Hernández, L.E., Gallegos, A., Schucht, P., Murek, M., Pérez-Romasanta, L., Belmonte-Beitia, J. and Pérez-García, V.M. *Optimal combinations of chemotherapy and radiotherapy in low-grade gliomas: a mathematical approach*, J. Pers. Med., 11(10) (2021), 1036.

[3] Balatif, O., and Rachik, M., and Rachik, Z. *Optimal control for multi-input bilinear systems with an application in cancer chemotherapy*, Int. J. Sci. Innov. Math. Res. (IJSIMR), 3 (2015), 22–31.

[4] Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J.M., Hlatky, L. and Hahnfeldt, P. *Classical mathematical models for description and prediction of experimental tumor growth*, PLoS Comput. Biol., 10(8) (2014), e1003800.

[5] Bodnar, M., Vela-Pérez, M. and Tryniecka-Maciek, A. *Analyzing the impact of proliferation and treatment parameters on low-grade glioma growth using mathematical models*, Physica D, 472 (2025), 134491.

[6] Bull, J.A., Mech, F., Quaiser, T., Waters, S.L. and Byrne, H.M. *Mathematical modelling reveals cellular dynamics within tumour spheroids*, PLoS Comput. Biol., 16(8) (2020), e1007961.

[7] Cai, Z., Yang, H., Yu, Z., Su, J., Zhang, J., Ye, Z., Hu, K., Huang, T. and Zhou, H. *Efficacy and safety of IDH inhibitors in IDH-mutated cancers: a systematic review and meta-analysis of 4 randomized controlled trials*, World J. Surg. Oncol., 22(1) (2024), 295.

[8] Carmona-Moreno, F.J., Gallegos, A., Barba-Franco, J.J., Urenda-Cázares, E., Jiménez-Guerrero, E.V. and Macías-Díaz, J.E. *Dynamics analysis of nonlinear differential equation systems applied to low-grade gliomas and their treatment*, Foundations, 5(1) (2025), 6.

[9] Coddington, E.A., and Levinson, N. *Theory of ordinary differential equations*, New York: McGraw-Hill, 1955.

[10] Davarian, N., Heydari, A., Hashemi Mehne, S.H. and Heydari, A.A. *The investigation of stability and optimal control of the multigroup Covid-19 epidemic model in Iran*, J. Math. Exten., 19(1) (2025), 1–32.

[11] Diekmann, O., Heesterbeek, J.A.P., and Metz, J.A.J. *On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol., 28(4) (1990), 365–382.

[12] Edelstein-Keshet, L. *Mathematical models in biology*, Philadelphia: SIAM, 2005.

[13] Eisenstein, M. *FDA approves first IDH-targeted glioma drug*, Nature Portfolio News, 2024.

[14] Fleming, W.H., and Rishel, R.W. *Deterministic and stochastic optimal control*, New York: Springer, 2012.

[15] Forst, D.A., Nahed, B.V., Loeffler, J.S. and Batchelor, T.T., *Low-grade gliomas*, Oncologist, 19(4) (2014), 403–413.

[16] Han, S., Liu, Y., Cai, S.J., Qian, M., Ding, J., Larion, M., Gilbert, M.R. and Yang, C. *IDH mutation in glioma: molecular mechanisms and potential therapeutic targets*, Br. J. Cancer, 122(11) (2020), 1580–1589.

[17] Handoko, H., Wahyudi, S.T., Setyawan, A.A. and Kartono, A. *A dynamical model of combination therapy applied to glioma*, J. Biol. Phys., 48(4) (2022), 439–459.

[18] Hollon, T., Hervey-Jumper, S.L., Sagher, O. and Orringer, D.A. *Advances in the surgical management of low-grade glioma*, Semin. Radiat. Oncol., 25(3) (2015), 181–188.

[19] Jaber, T., and Balatif, O. *A mathematical model of drug-resistant cancer with cell cycle structure: stability, optimal control, and cost-effectiveness analysis*, Int. J. Dyn. Control, 13(10) (2025), 354.

[20] Jakola, A.S., Skjulsvik, A.J., Myrmel, K.S., Sjåvik, K., Unsgård, G., Torp, S.H., Aaberg, K., Berg, T., Dai, H.Y., Johnsen, K. and Kloster, R. *Surgical resection versus watchful waiting in low-grade gliomas*, Ann. Oncol., 28(8) (2017), 1942–1948.

[21] Khajanchi, S. *Stability analysis of a mathematical model for glioma-immune interaction under optimal therapy*, Int. J. Nonlin. Sci. Numer. Simul., 20(3-4) (2019), 269–285.

[22] Khalil, H.K. and Grizzle, J.W. *Nonlinear systems*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 2002.

[23] Kopp, R.E. *Pontryagin maximum principle*, in: Mathematics in Science and Engineering, Vol. 5, Amsterdam: Elsevier, (1962), 255–279.

[24] Kurzweil, J. *Ordinary differential equations: Introduction to the theory in the real domain*, Amsterdam: Elsevier, 2014.

[25] La Salle, J.P. *The Stability of dynamical systems*, Philadelphia: SIAM, 1976.

[26] Ledzewicz, U., and Schättler, H. *Analysis of a mathematical model for low-grade gliomas under chemotherapy as a dynamical system*, Nonlinear Anal. Real World Appl., 85 (2025), 104344.

[27] Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G. and Soffietti, R. *The 2021 WHO classification of tumors of the central nervous system: a summary*, Neuro Oncol., 23(8) (2021), 1231–1251.

[28] McAsey, M., Mou, L. and Han, W. *Convergence of the forward-backward sweep method in optimal control*, Comput. Optim. Appl., 53(1) (2012), 207–226.

[29] McDuff, S.G., Dietrich, J., Atkins, K.M., Oh, K.S., Loeffler, J.S. and Shih, H.A. *Radiation and chemotherapy for high-risk lower grade gliomas: choosing between temozolomide and PCV*, Cancer Med., 9(1) (2020), 3–11.

[30] Mellinghoff, I.K., Van Den Bent, M.J., Blumenthal, D.T., Touat, M., Peters, K.B., Clarke, J., Mendez, J., Yust-Katz, S., Welsh, L., Mason, W.P. and Ducray, F. *Vorasidenib in IDH1- or IDH2-mutant low-grade glioma*, N. Engl. J. Med., 389(7) (2023), 589–601.

[31] Minniti, G., Paolini, S., Antonelli, M., Gianno, F., Tini, P., Lanzetta, G., Arcella, A., De Pietro, R., Giraffa, M., Capone, L. and Romano, A. *Long-term treatment outcomes of temozolomide-based chemoradiation in patients with adult-type diffuse IDH-mutant grade 2 astrocytoma*, J. Neurooncol., 164(2) (2023), 331–339.

[32] Nagumo, M. *Über die lage der integralkurven gewöhnlicher differentialgleichungen*, Proc. Phys Math. Soc. Jpn., 24 (1942), 551–559.

[33] Pace, A., Vidiri, A., Galie, E., Carosi, M., Telera, S., Cianciulli, A.M., Canalini, P., Giannarelli, D., Jandolo, B. and Carapella, C.M. *Temozolomide chemotherapy for progressive low-grade glioma: clinical benefits and radiological response*, Ann. Oncol., 14(12) (2003), 1722–1726.

[34] Patel, S.H., Bansal, A.G., Young, E.B., Batchala, P.P., Patrie, J.T., Lopes, M.B., Jain, R., Fadul, C.E. and Schiff, D. *Extent of surgical resection in lower-grade gliomas: differential impact based on molecular subtype*, AJNR Am. J. Neuroradiol., 40(7) (2019), 1149–1155.

[35] Reuvers, M.J., Burgers, V.W., Koekkoek, J.A., Brandsma, D., Compter, A., Kouwenhoven, M.C., van den Bent, M.J., Frissen, S.A., Husson, O. and van der Graaf, W.T. *Aiming at a moving target the daily life experiences of adolescents and young adults with a low-grade glioma*, Neurooncol. Pract., 12(1) (2025), 113–121.

[36] Ribba, B., Kaloshi, G., Peyre, M., Ricard, D., Calvez, V., Tod, M., Čajavec-Bernard, B., Idbaih, A., Psimaras, D., Dainese, L. and Pallud, J. *A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy*, Clin. Cancer Res., 18(18) (2012), 5071–5080.

[37] Shaw, E.G., Wang, M., Coons, S.W., Brachman, D.G., Buckner, J.C., Stelzer, K.J., Barger, G.R., Brown, P.D., Gilbert, M.R. and Mehta, M.P. *Randomized trial of radiation therapy plus procarbazine, lomustine, and vincristine chemotherapy for supratentorial adult low-grade glioma: initial results of RTOG 9802*, J. Clin. Oncol., 30(25) (2012), 3065–3070.

[38] Spitaels, J., Devriendt, D., Sadeghi, N., Luce, S., De Witte, O., Goldman, S., Mélot, C. and Lefranc, F. *Management of supratentorial recurrent low-grade glioma: a multidisciplinary experience in 35 adult patients*, Oncol. Lett., 14(3) (2017), 2789–2795.

[39] Teng, C., Zhu, Y., Li, Y., Dai, L., Pan, Z., Wanggou, S. and Li, X. *Recurrence- and malignant progression-associated biomarkers in low-grade gliomas and their roles in immunotherapy*, Front. Immunol., 13 (2022), 899710.

[40] Van den Driessche, P., and Watmough, J. *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci., 180(1-2) (2002), 29–48.

[41] Wan, Q., Wu, X., Zhou, J., Wu, W., Cao, Y., Sun, C., Li, Z., Gong, Z., Tang, H., Li, Q. and Chu, J. *The hypoxia-associated high-risk cell subpopulation distinctly enhances the progression of glioma*, Adv. Sci., 12(17) (2025), 2416231.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1271–1294
https://doi.org/10.22067/ijnao.2026.97551.1828
https://ijnao.um.ac.ir/



Research Article



Traveling-wave patterns and bifurcation dynamics in a spatial predator-prey model

S. Hussain* and M. Sen

Abstract

We study a diffusive predator-prey model with nonlinear predator-prey interactions and prey dispersal. Using bifurcation analysis combined with numerical simulations, we reveal a rich variety of dynamical behaviors, including the stability of steady states, wave fronts, wave pulses, periodic and multi-periodic wave trains, and chaotic spatiotemporal patterns. The results show that diffusion and interaction strength play a critical role in shaping the traveling-wave dynamics through saddle-node, Hopf, and period-doubling bifurcations. In particular, increasing diffusion can destabilize steady states and induce oscillatory dynamics, ultimately leading to chaotic population fluctuations. These findings provide ecological insight into how spatial dispersal and nonlinear interactions shape complex population patterns in predator-prey systems.

AMS subject classifications (2020): Primary 37N25; Secondary 92-08, 92D25.

Keywords: Predator-Prey model; Traveling waves; Stability analysis; Bifurcation.

*Corresponding author

Received 30 January 2026; revised 14 June 2026; accepted 22 June 2026

Saddam Hussain
Department of Mathematics and Computing Technology, NIT Patna, India 800005. e-mail: saddamh.phd20.ma@nitp.ac.in

Moitri Sen
Department of Mathematics and Computing Technology, NIT Patna, India 800005. e-mail: moitri300784@gmail.com

How to cite this article
Hussain, S. and Sen, M., Traveling-wave patterns and bifurcation dynamics in a spatial predator-prey model. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1271-1294. https://doi.org/10.22067/ijnao.2026.97551.1828

1 Introduction

The prey-predator model, commonly referred to as the Lotka–Volterra model, is a cornerstone of mathematical ecology and provides a fundamental framework for describing the dynamic interaction between two interdependent populations: prey, which constitute a food resource, and predators, which rely on prey for survival. This classical formulation has played a crucial role in advancing our understanding of population oscillations, species coexistence, and ecological balance. Over the years, the basic model has been extended in various directions to incorporate biological mechanisms observed in real ecosystems, thereby improving its ecological realism and predictive capacity. One important refinement of the classical predator-prey framework is the incorporation of cooperative hunting behavior among predators. In many natural systems, predators do not act independently but instead hunt in groups, thereby increasing capture efficiency and altering functional responses. Such cooperative strategies can significantly modify population dynamics by amplifying predation pressure and influencing stability properties of equilibria. The ecological relevance of cooperative hunting and its mathematical formulation were first systematically explored by Alves and Hilker [4], who demonstrated that predator cooperation can lead to qualitatively new dynamical behaviors not present in classical models.

While most early predator-prey models focused on temporal dynamics alone, spatial effects are now recognized as an essential component of ecological systems. Species movement, dispersal, and spatial heterogeneity strongly influence persistence, invasion, and pattern formation [25]. To capture these effects, predator-prey models are naturally extended to reaction-diffusion systems by incorporating diffusion terms that describe random movement in space. The inclusion of spatial dynamics is motivated by the seminal work of Turing on the chemical basis of morphogenesis [21], which demonstrated that diffusion, when coupled with nonlinear interactions, can destabilize homogeneous steady states and generate spatial patterns. From an ecological standpoint, reaction-diffusion predator-prey models provide a powerful framework for studying the combined effects of biological interactions and spatial dispersal [1]. In plankton communities, forest insect populations, and spatially structured food webs, population densities are rarely uniform across space. Instead, ecosystems often exhibit traveling invasion fronts, localized outbreaks, sustained oscillatory waves, and complex spatiotemporal patterns [8, 16]. Dispersal, commonly represented through diffusion, plays a dual role in such systems: it promotes persistence and spatial spread, yet can also destabilize homogeneous equilibria when coupled with nonlinear predation mechanisms, leading to oscillatory or even chaotic dynamics [13]. Consequently, the emergence of wave fronts, wave pulses, and wave trains in reaction-diffusion predator-prey models offers a mechanistic explanation for invasion processes, recurrent outbreaks, and cyclic population behavior observed in nature.

A particularly important mathematical tool for understanding these spatial phenomena is invariant manifold theory, which plays a central role in the analysis of traveling-wave solutions in reaction-diffusion systems. Traveling waves provide a mathematical description of spatial propagation processes occurring in diverse contexts, including population dynamics, nutrient transport, chemical reactions, and biological signal transmission [16, 15]. By introducing a moving coordinate frame, reaction-diffusion equations can be reduced to systems of ordinary differential equations, where traveling waves correspond to special trajectories in phase space. In this formulation, wave fronts, wave pulses, and wave trains arise as heteroclinic, homoclinic, and periodic orbits, respectively [13, 19, 24]. The existence and stability of these solutions are fundamentally governed by the geometry of invariant manifolds associated with equilibria and periodic orbits. Despite their conceptual importance, the rigorous construction of homoclinic and heteroclinic orbits remains a challenging problem, particularly in systems of dimension three or higher. Except for a few special cases, explicit analytical expressions for such connecting orbits are rarely available [8, 20]. As a result, a wide range of analytical and numerical techniques has been developed, including Poincaré maps, Lin's method, center manifold reductions, and numerical continuation, to establish the existence and continuation of global connections [9, 10, 14, 26]. In recent years, numerical bifurcation software such as `AUTO` [6] and `MATCONT` [5] has become indispensable for computing equilibria, periodic orbits, and their invariant manifolds with high accuracy, enabling systematic investigations of traveling waves in reaction-diffusion systems [18, 2, 11, 12].

Although traveling-wave solutions and diffusion-driven dynamics have been extensively investigated in predator-prey reaction-diffusion systems [7, 25, 22, 14], the combined effects of prey dispersal, a strong Allee effect, and cooperative hunting on traveling-wave dynamics remain largely unexplored. In contrast to [7, 14], where the emphasis is primarily on the existence of traveling waves and heteroclinic structures, the present work investigates how cooperative predation and prey diffusion interact to generate a rich bifurcation scenario involving wave fronts, pulse waves, periodic wave trains, multi-periodic oscillations, and chaotic traveling waves. Moreover, unlike [25] and [22], which focus on diffusion-driven instability and multiple wave solutions in more general reaction-diffusion settings, the present study reveals how hunting cooperation and prey dispersal jointly induce complex spatiotemporal behaviour through Hopf and period-doubling bifurcations. To the best of our knowledge, the emergence of chaotic traveling waves in a predator-prey system incorporating both cooperative hunting and Allee effects has not been previously reported.

Motivated by these developments, the objective of the present study is to investigate the existence and structure of traveling-wave solutions in a diffusive predator-prey model with nonlinear interaction terms. Building upon the predator-prey model proposed in [23], we focus on

the effects of prey dispersal on traveling-wave dynamics and the associated bifurcation structure. Using bifurcation theory, invariant manifold analysis, and numerical continuation techniques, we investigate how variations in diffusion and interaction parameters influence the qualitative behavior of traveling-wave solutions and lead to transitions between different dynamical regimes. The remainder of the paper is organized as follows. Section 2 introduces the reaction-diffusion model, analyzes the existence and stability of its equilibria, derives the traveling-wave equations, and investigates their phase-space structure. Section 3 employs numerical continuation methods to compute homoclinic and heteroclinic orbits and classify the resulting traveling-wave solutions. Finally, Section 4 discusses the ecological implications of our findings and summarizes the main conclusions of the study.

2 Diffusive model

We consider a prey-predator reaction-diffusion model obtained by introducing one-dimensional spatial diffusion only in the prey population, extending the nonspatial system (4) proposed in the paper [23]. The primary objective of the present work is to investigate the existence of traveling-wave solutions for the following system:

$$\begin{aligned} \frac{\partial N}{\partial t} &= N\Big(\sigma(1-N)(N-\beta) - (1+\alpha P)P\Big) + d\frac{\partial^2 N}{\partial x^2}, \\ \frac{\partial P}{\partial t} &= P\delta(1-P) + e(1+\alpha P)NP, \end{aligned} \tag{1}$$

where $N(x,t)$ and $P(x,t)$ denote the densities of the prey and predator populations at position x and time t, respectively. The parameter $d > 0$ represents the diffusion coefficient of the prey species, modeling its random movement in space, while the predator population is assumed to be immobile. This asymmetric diffusion assumption reflects situations in which prey species exhibit higher mobility than predators and has been widely adopted in ecological and epidemiological modeling frameworks.

The prey growth is governed by the Allee effect, represented by the term $\sigma(1-N)(N-\beta)$, where $\sigma > 0$ is the intrinsic growth rate of the prey and $\beta \in (-1,1)$ denotes the Allee threshold below, which the prey population experiences negative growth [3]. Predator consumption of prey follows a Holling type-I functional response, given by $(1+\alpha P)P$, where $\alpha \geq 0$ represents the strength of cooperative hunting among predators. It is assumed that predators hunt prey in groups, thereby enhancing their predation efficiency and increasing the prey capture rate. The predator population grows logistically in the absence of prey with intrinsic growth rate $\delta > 0$ and

carrying capacity normalized to unity. The term $e(1+\alpha P)NP$ represents the gain in predator biomass due to predation, where $e>0$ denotes the conversion efficiency of consumed prey into predator reproduction.

System (1) is inspired by the predator-prey framework introduced in [22], incorporating both the Allee effect in prey dynamics and a Holling type-I functional response. The inclusion of diffusion only in the prey equation allows the model to capture spatial propagation phenomena driven by prey dispersal, which plays a crucial role in the formation of traveling-wave solutions connecting biologically relevant equilibria. Such spatially structured dynamics are essential for understanding invasion processes, range expansion, and pattern formation in ecological systems.

2.1 Traveling-wave reduction

In order to study the traveling waves, we introduce the coordinate, $z=x-ct$, where $c>0$ denotes the wave speed, and seek solutions of the form $N(x,t)=U(z), \quad P(x,t)=V(z)$. Substitution into (1) yields

$$\begin{aligned} -cU' &= dU'' + U\Big(\sigma(1-U)(U-\beta)-(1+\alpha V)V\Big), \\ -cV' &= V\delta(1-V)+e(1+\alpha V)UV, \end{aligned} \tag{2}$$

where primes denote differentiation with respect to z. Introducing the auxiliary variable $W=U'$, system (2) can be rewritten as the following first-order autonomous system:

$$\begin{aligned} U' &= W, \\ W' &= \frac{1}{d}\Big(-cW-\sigma U(1-U)(U-\beta)+(1+\alpha V)UV\Big), \\ V' &= -\frac{1}{c}\Big(V\delta(1-V)+e(1+\alpha V)UV\Big). \end{aligned} \tag{3}$$

System (3) is three-dimensional, in contrast to the four-dimensional traveling-wave system obtained when both species diffuse. This reduction in dimension plays a crucial role in the qualitative analysis of wave solutions. Motivated by the biological context of (1), we confine our analysis of system (3) to the domain $\Omega=\{(U,W,V)\in\mathbb{R}^3 : U>0, V>0\}$. A traveling-wave solution of (1) is defined as any bounded solution of system (3) that lies entirely within the domain Ω. Such solutions represent wave profiles that move through space while maintaining their shape over time [15, 16, 8].

To obtain bounded solutions of system (3), we consider three principal classes of traveling waves: wave pulses, wave fronts, and wave trains. A wave pulse is associated with a homoclinic orbit of the traveling-wave system, in which a trajectory leaves an equilibrium point E_1 along its unstable manifold and returns to the same equilibrium along its stable manifold, so that the orbit lies in $W^u(E_1) \cap W^s(E_1)$. Such solutions display a pronounced excursion in the amplitudes of the state variables followed by convergence back to the resting state, and in the original reaction-diffusion system they correspond to localized pulses propagating through space while connecting a spatially homogeneous steady state to itself. A wave front, on the other hand, arises from a heteroclinic connection between two distinct equilibria, where the orbit departs from E_1 along $W^u(E_1)$ and approaches another equilibrium E_2 along its stable manifold $W^s(E_2)$, that is, it lies in $W^u(E_1) \cap W^s(E_2)$; this represents a traveling interface that connects two different spatially homogeneous steady states. Finally, wave trains correspond to periodic traveling-wave solutions of the reduced system, characterized by closed orbits in phase space. These solutions are spatially periodic and temporally steady in the moving frame, representing a continuous sequence of oscillations that propagate with constant speed without decay, and they play a key role in describing sustained rhythmic patterns in the underlying reaction-diffusion model [22].

2.2 Stability analysis

An equilibrium of (3) is given by $E = (U^*, 0, V^*)$, where (U^*, V^*) satisfies

$$\begin{aligned} U^*\Big(\sigma(1-U^*)(U^*-\beta) - (1+\alpha V^*)V^*\Big) &= 0, \\ V^*\Big(\delta(1-V^*) + e(1+\alpha V^*)U^*\Big) &= 0. \end{aligned} \tag{4}$$

From (4), we identify the following classes of equilibria.

(i) Trivial equilibrium.

$$E_0 = (0, 0, 0).$$

(ii) Boundary equilibria.

$$E_\beta = (\beta, 0, 0), \qquad E_1 = (1, 0, 0) \qquad E_2 = (0, 0, 1).$$

The equilibrium E_β, E_1 corresponds to extinction of the P-population and persistence of N and E_2 correspond to extinction of the N-population and persistence of P at its intrinsic steady states.

(iii) Coexistence equilibria. If $U^* > 0$ and $V^* > 0$, then (U^*, V^*) satisfies

$$\begin{aligned} \sigma(1-U^*)(U^*-\beta) &= (1+\alpha V^*)V^*, \\ \delta(1-V^*) + e(1+\alpha V^*)U^* &= 0. \end{aligned} \tag{5}$$

Due to the nonlinear coupling between U^* and V^*, explicit closed-form expressions for these quantities are generally difficult to obtain. Consequently, the coexistence equilibrium is characterized implicitly by system (5). For suitable parameter values, system (5) admits positive solutions, yielding the biologically relevant coexistence equilibrium

$$E_* = (U^*, 0, V^*).$$

2.2.1 Linearization and Jacobian matrix

The Jacobian matrix of (3) evaluated at an equilibrium $E = (U^*, 0, V^*)$ is given by

$$J(E) = \begin{pmatrix} 0 & 1 & 0 \\ -a_{11} & -\frac{c}{d} & -a_{13} \\ -a_{31} & 0 & -a_{33} \end{pmatrix},$$

where

$$\begin{aligned} a_{11} &= \frac{1}{d}\Big(\sigma(-3U^{*2} + 2(1+\beta)U^* - \beta) - (1+\alpha V^*)V^*\Big), \\ a_{13} &= -\frac{1}{d}\Big(1+2\alpha V^*\Big)U^*, \\ a_{31} &= \frac{e}{c}\Big(1+\alpha V^*\Big)V^*, \\ a_{33} &= \frac{1}{c}\Big(\delta(1-2V^*) + e(1+2\alpha V)U^*\Big). \end{aligned}$$

2.2.2 Characteristic equation

The characteristic polynomial of $J(E)$ takes the form

$$\lambda^3 + \left(\frac{c}{d} + a_{33}\right)\lambda^2 + \left(a_{11} + \frac{c}{d}a_{33}\right)\lambda + (a_{11}a_{33} - a_{13}a_{31}) = 0. \tag{6}$$

The signs of the coefficients in (6) determine the local stability properties of the equilibrium [17].

2.2.3 Stability of boundary equilibria

Lemma 1. The equilibrium point E_0 is a stable equilibrium whenever $-1 < \beta < 0$ with $dim^s(E_0) = 3$ and saddle whenever $0 < \beta < 1$ with $dim^s(E_0) = 2$, $dim^u(E_0) = 1$.

Proof. At E_0, the Jacobian reduces to

$$J(E_0) = \begin{pmatrix} 0 & 1 & 0 \\ \frac{\sigma\beta}{d} & -\frac{c}{d} & 0 \\ 0 & 0 & -\frac{\delta}{c} \end{pmatrix}.$$

The corresponding eigenvalues are

$$\lambda_{1,2} = \frac{-c \pm \sqrt{c^2 + 4\sigma\beta}}{2d}, \qquad \lambda_3 = -\frac{\delta}{c} < 0.$$

Hence, E_0 is a stable equilibrium whenever $-1 < \beta < 0$ on a three dimensional manifold and saddle whenever $0 < \beta < 1$ in two dimensional stable and one dimensional unstable manifold. □

Lemma 2. The equilibrium point E_β exist for $0 < \beta < 1$ and is a saddle with $dim^s(E_\beta) = 2$, $dim^u(E_\beta) = 1$

Proof. The Jacobian matrix at E_β is therefore

$$J(E_\beta) = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{\sigma\beta(1-\beta)}{d} & -\frac{c}{d} & \frac{\beta}{d} \\ 0 & 0 & -\frac{\delta+e\beta}{c} \end{pmatrix}.$$

The eigenvalues satisfy

$$\lambda_{1,2} = \frac{-c \pm \sqrt{c^2 + 4\sigma\beta(1-\beta)}}{2d}, \qquad \lambda_3 = -\frac{\delta + e\beta}{c} < 0.$$

Thus, equilibrium E_β is a saddle point with a one-dimensional unstable manifold and a two-dimensional stable manifold. □

Lemma 3. The equilibrium point E_1 is a saddle equilibrium with $dim^s(E_1) = 2$, $dim^u(E_1) = 1$.

Proof. The Jacobian matrix at E_1 is

$$J(E_1) = \begin{pmatrix} 0 & 1 & 0 \\ \frac{\sigma(1-\beta)}{d} & -\frac{c}{d} & \frac{1}{d} \\ 0 & 0 & -\frac{\delta+e}{c} \end{pmatrix}.$$

The eigenvalues are

$$\lambda_{1,2} = \frac{-c \pm \sqrt{c^2 + 4\sigma(1-\beta)}}{2d}, \qquad \lambda_3 = -\frac{\delta+e}{c} < 0.$$

Hence, E_1 is also a saddle equilibrium with a two-dimensional stable manifold and a one-dimensional unstable manifold. □

Lemma 4. The equilibrium point E_2 is also a saddle equilibrium with $dim^s(E_2) = 1$, $dim^u(E_2) = 2$.

Proof. The Jacobian matrix at E_2 is given by

$$J(E_2) = \begin{pmatrix} 0 & 1 & 0 \\ \frac{\sigma\beta+1+\alpha}{d} & -\frac{c}{d} & 0 \\ -\frac{e(1+\alpha)}{c} & 0 & \frac{\delta}{c} \end{pmatrix}.$$

The characteristic polynomial of $J(E_2)$ is

$$(\lambda - \tfrac{\delta}{c})\left(d\lambda^2 + c\lambda - (\sigma\beta + 1 + \alpha)\right) = 0.$$

Hence, the eigenvalues are

$$\lambda_{1,2} = \frac{-c \pm \sqrt{c^2 + 4\big(\sigma\beta + 1 + \alpha\big)}}{2}, \qquad \lambda_3 = \frac{\delta}{c} > 0.$$

Since $\lambda_3 > 0$ and the quadratic term admits one positive and one negative root, equilibrium E_2 possesses a two-dimensional unstable manifold and a one-dimensional stable manifold. Therefore, E_2 is a saddle equilibrium of index two. □

2.2.4 Stability of coexistence equilibrium

If a coexistence equilibrium $E_* = (U^*, 0, V^*)$ exists, its local stability is determined by the eigenvalues of the Jacobian matrix evaluated at E_*. The corresponding characteristic equation is

$$\lambda^3 + \left(\frac{c}{d} + b_{33}\right)\lambda^2 + \frac{1}{d}\Big(b_{11} + cb_{33}\Big)\lambda + \frac{1}{d}\Big(b_{11}b_{33} - b_{13}b_{31}\Big) = 0, \tag{7}$$

where

$$\begin{aligned} b_{11} &= \sigma U^*(-2U^* + \beta + 1), \qquad & b_{13} &= -(1 + 2\alpha V^*)U^*, \\ b_{31} &= \tfrac{e}{c}\Big(1 + \alpha V^*\Big)V^*, \qquad & b_{33} &= \tfrac{1}{c}\Big(-\delta + e\alpha U^*\Big)V^*. \end{aligned}$$

By the Routh–Hurwitz criterion for cubic polynomials, the equilibrium E_* is locally asymptotically stable if and only if

$$\frac{c}{d} + b_{33} > 0, \qquad b_{11} + cb_{33} > 0, \qquad b_{11}b_{33} - b_{13}b_{31} > 0,$$

and

$$\left(\frac{c}{d} + b_{33}\right)(b_{11} + cb_{33}) > (b_{11}b_{33} - b_{13}b_{31}).$$

Although the stability conditions can be stated in terms of the coefficients of the characteristic polynomial, these coefficients depend nonlinearly on the model parameters and the equilibrium values U^* and V^*. Consequently, determining explicit parameter regions that guarantee stability or instability is analytically cumbersome. To gain further insight into the dynamics of the system and to investigate the influence of key parameters on the stability of the coexistence equilibrium E_*, we therefore complement the analytical results with numerical simulations. The numerical investigations presented in the next section illustrate the local behavior of solutions near E_* and identify parameter regimes associated with stable and unstable dynamics.

2.2.5 Hopf bifurcation analysis

The occurrence of Hopf bifurcation at the equilibrium point $E_* = (U^*, 0, V^*)$ is determined by the characteristic equation (7). According to the Hopf bifurcation theorem, a Hopf bifurcation occurs when a pair of complex conjugate eigenvalues of the Jacobian matrix $J(E_*)$ crosses the imaginary axis as a system parameter varies, while the remaining eigenvalue has a nonzero real part.

Let d be chosen as the bifurcation parameter. For (7) to admit a pair of purely imaginary roots $\lambda = \pm i\omega$ ($\omega > 0$), the following necessary conditions must be satisfied:

$$\frac{c}{d} + b_{33} > 0, \tag{8}$$

$$b_{13}b_{31} - b_{11}b_{33} > 0, \tag{9}$$

and the Hopf-bifurcation threshold is obtained by

$$d_H = -\frac{c(b_{11} + cb_{33})}{b_{13}b_{31} + cb_{33}^2}, \tag{10}$$

where

$$\omega^2 = \frac{1}{d}\Big(b_{11} + cb_{33}\Big). \tag{11}$$

In addition, the transversality condition must hold, namely,

$$\frac{d}{dd}\Re(\lambda(d))\Big|_{d=d_H} \neq 0. \tag{12}$$

This condition ensures that the eigenvalues cross the imaginary axis with nonzero speed as the bifurcation parameter d passes through the critical value d_H.

When all the above conditions are satisfied, the equilibrium point $E_* = (U^*, 0, V^*)$ undergoes a Hopf bifurcation at $d = d_H$. As a consequence, a family of small-amplitude periodic solutions emerges from E, leading to sustained oscillations in the population densities. These temporal oscillations play a crucial role in the formation of wave trains in the corresponding reaction-diffusion system, as will be demonstrated in the subsequent sections.

3 Numerical simulation

In this section, the traveling-wave system (3) was solved numerically using the classical fourth-order Runge–Kutta (RK4) method. Since traveling-wave solutions are defined on the unbounded domain $z \in (-\infty, \infty)$, the computational domain was truncated to the finite interval $[-L, L]$ with $L = 250$. Unless otherwise stated, the numerical simulations were performed using a uniform step size $h = 0.001$.

The numerical integration was initiated at $z = -L$ using initial data chosen in a sufficiently small neighborhood of the equilibrium state from which the traveling orbit originates. The solution was then integrated forward up to $z = L$. Since a fixed-step RK4 scheme was employed, no adaptive numerical tolerance parameter was required.

To verify the accuracy and robustness of the numerical computations, convergence tests were performed using different step sizes and computational domains. The resulting traveling-wave

profiles exhibited no significant qualitative or quantitative differences, confirming that the reported numerical solutions are independent of the discretization parameters and domain truncation. The parameter values were fixed according to the ecological model proposed in [23] and are summarized in Table 1.

Table 1: Parameter values used in numerical simulations.

s	β	e	δ	c
8	-0.108	0.17	0.2298	0.6

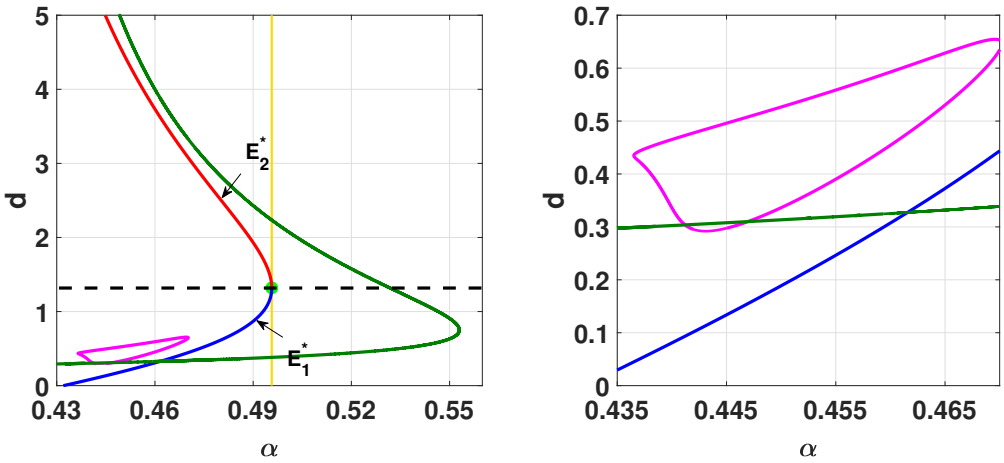
For these fixed parameters, numerical continuation and bifurcation analysis were carried out using the software package MATCONT. The primary objective of this study is to examine how variations in the hunting cooperation parameter α and the diffusion coefficient d influence the qualitative dynamics of the traveling-wave system (3). In particular, we investigate the stability of equilibria and the associated bifurcation structure, including saddle-node, Hopf, and period-doubling bifurcations.

The resulting bifurcation diagrams reveal transitions between different traveling-wave behaviors, including wave fronts, wave pulses, periodic and multi-periodic wave trains, and chaotic traveling patterns, highlighting the important roles of hunting cooperation and prey diffusion in shaping the spatiotemporal dynamics of the predator-prey system.

3.1 Bifurcation

For the chosen set of parameters, the boundary equilibrium points E_0, E_1, and E_2, together with the interior equilibrium E^*, coexist. Numerical bifurcation analysis reveals the occurrence of a saddle-node bifurcation at $\alpha = \alpha_{ZH}$, where the number of interior equilibrium points changes from two to zero as the parameter α is varied from left to right in Figure 1(a). In the parameter region where two interior equilibria exist, they are denoted by E_1^* and E_2^*. The blue curve in Figure 1(a) represents the Hopf bifurcation curve associated with the equilibrium E_1^*. Below this curve, E_1^* is locally asymptotically stable, whereas it loses stability upon crossing the curve, giving rise to periodic oscillations. In the unstable region of E_1^*, a period-doubling bifurcation occurs, indicated by the magenta curve. Inside the resulting period-doubling loop, solutions with higher-order periodicity, including 2-periodic, 4-periodic, and chaotic dynamics, are observed, indicating a classical route to chaos. The red curve corresponds to the Hopf bifurcation associated with the equilibrium E_2^*. Below this curve, E_2^* is a saddle-focus with one positive real eigenvalue and the

remaining eigenvalues having negative real parts, whereas above the curve, the real parts of the complex conjugate eigenvalues become positive, leading to oscillatory instability.



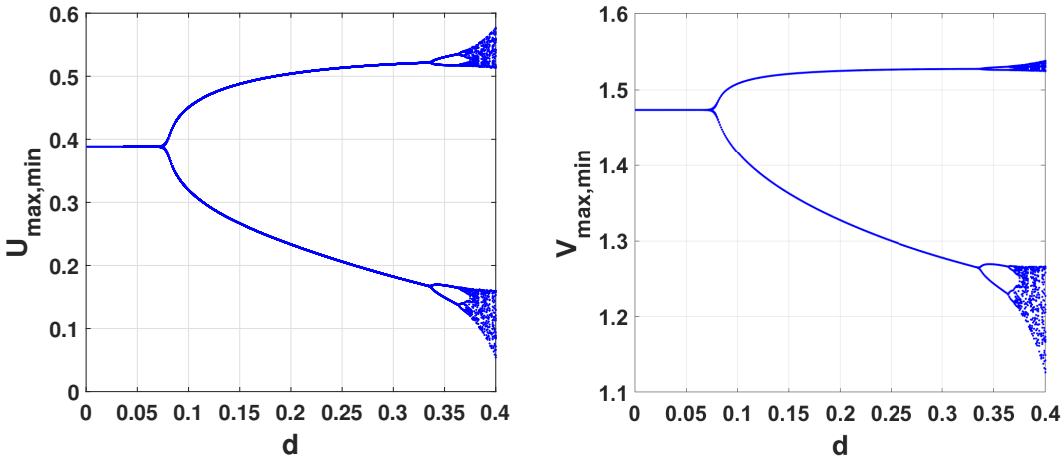
(a) (b)

Figure 1: (a) Two-dimensional bifurcation diagram with respect to α and d. The yellow vertical line shows the saddle-node bifurcation curve, the red curve shows the Hopf bifurcation curve for the equilibrium E_2^*, the blue curve shows the Hopf-bifurcation for the equilibrium E_1^* and the magenta curve represents the period-doubling bifurcation curve.The green curve represents the homoclinic bifurcation curve. The green dot represents the zero-Hopf point. (b) The zoomed version of Figure 1(a).

Figure 2 presents the one-dimensional bifurcation diagrams obtained by plotting the local maxima and minima of the prey and predator densities with respect to the diffusion coefficient d, while keeping the parameter $\alpha = 0.44$ fixed. For small values of d, both species converge to a constant equilibrium state, which is reflected in the diagram by a single branch for each variable. This indicates that the equilibrium is locally asymptotically stable in this parameter regime. As d increases and crosses a critical threshold, the equilibrium loses stability through a Hopf bifurcation, leading to the emergence of sustained oscillations. Beyond this point, the bifurcation diagrams split into two distinct branches corresponding to the maximum and minimum values attained by the prey and predator populations, thereby signaling periodic oscillatory dynamics.

With a further increase in the diffusion coefficient d, the system undergoes a sequence of secondary bifurcations. The appearance of multiple discrete branches in the bifurcation diagrams corresponds to period-doubling bifurcations, where oscillations with twice the original period are generated. This cascade continues as d increases, resulting in increasingly complex oscillatory patterns, including 2-periodic, 4-periodic, and higher-order periodic solutions. Eventually, the accumulation of these bifurcations gives rise to chaotic dynamics, which are manifested in the diagrams as densely populated bands of maxima and minima. This behavior is observed consis-

tently for both prey and predator species, demonstrating that diffusion plays a crucial role in destabilizing the equilibrium and driving the system through a transition from steady states to periodic oscillations and ultimately to chaotic spatiotemporal dynamics.



(a) (b)

Figure 2: One dimensional bifurcation diagram with respect to the diffusion coefficient d for fixed $\alpha = 0.44$, (a) prey species and (b) predator species.

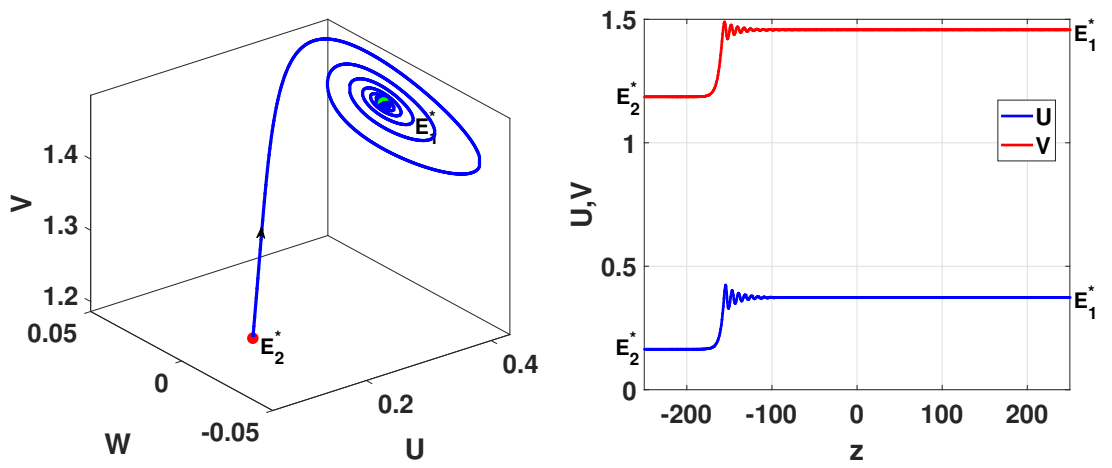
To determine the nature of the Hopf bifurcation, the first Lyapunov coefficient was computed numerically using MATCONT. For the parameter value $\alpha = 0.44$, the Hopf bifurcation occurs at

$$d_H = 0.080745985,$$

and the corresponding first Lyapunov coefficient is

$$l_1 = -5.39211921 < 0.$$

Therefore, the Hopf bifurcation is supercritical, and a branch of stable periodic solutions emerges from the coexistence equilibrium as the diffusion coefficient d passes through the critical value d_H. This observation is consistent with the numerical bifurcation diagrams presented in the subsequent section.



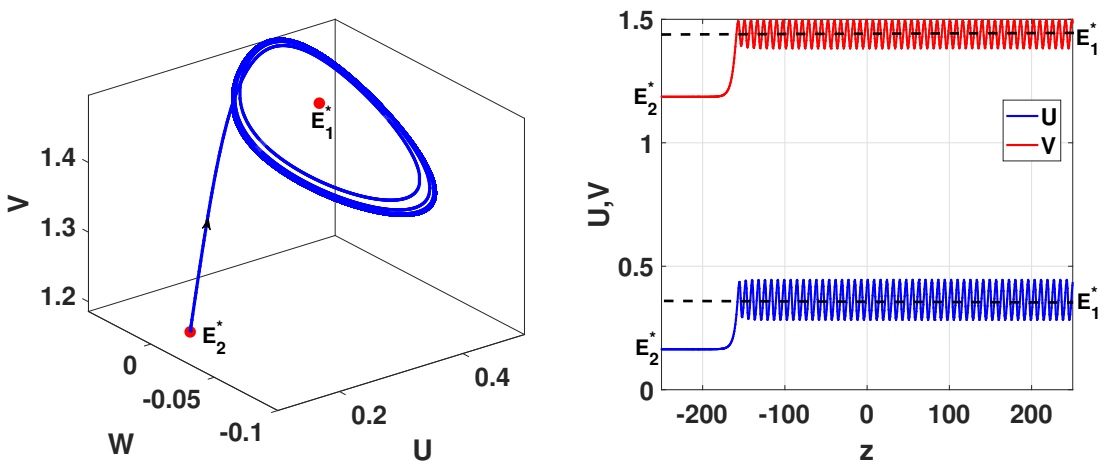
(a) (b)

Figure 3: (a) Phase portrait of the traveling-wave system showing convergence to a stable spiral equilibrium. (b) Corresponding spatial profiles of the prey (U) and predator (V) populations. Parameter values are $\alpha = 0.45$ and $d = 0.1$.

3.2 Phase portrait

To validate the bifurcation diagram shown in Figure 1, we present the corresponding phase portrait and spatial profiles of the prey and predator populations. For parameter values chosen from the regime in which the equilibrium E_1^* is stable, namely $\alpha = 0.45$ and $d = 0.1$, the traveling-wave system admits a heteroclinic connection between two distinct equilibria. As shown in Figure 3, the phase portrait indicates that the trajectory originates in the neighborhood of the equilibrium E_2^* and departs along its unstable manifold. After a transient evolution, the orbit approaches and eventually settles at the equilibrium E_1^* along its stable manifold, confirming the existence of a heteroclinic orbit lying in the intersection $W^u(E_2^*) \cap W^s(E_1^*)$. The corresponding spatial profiles of the state variables, depicted in Figure 3(b), clearly illustrate a transition from the state associated with E_2^* to that of E_1^*, with small oscillations observed as the solution converges to the final steady state. In the context of the original reaction-diffusion system, this heteroclinic connection represents a wave front, describing a traveling interface that propagates with constant speed and connects two distinct spatially homogeneous equilibrium states.

For parameter values taken from the regime in which the equilibrium E_1^* loses stability through a Hopf bifurcation, specifically $\alpha = 0.45$ and $d = 0.25$, the traveling-wave system exhibits a heteroclinic-type connection from the equilibrium E_2^* to a periodic orbit generated at E_1^*. As illustrated by the phase portrait in Figure 4(a), the trajectory departs from the equilibrium E_2^*

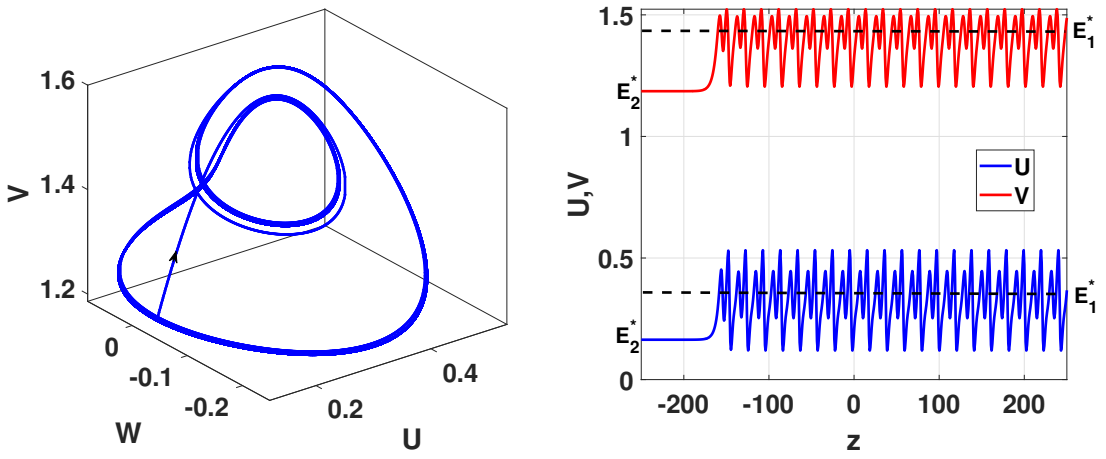


(a) (b)

Figure 4: (a) Phase portrait of the traveling-wave system showing convergence to a stable limit cycle. (b) Corresponding spatial profiles of the prey (U) and predator (V) populations. Parameter values are $\alpha = 0.45$ and $d = 0.25$.

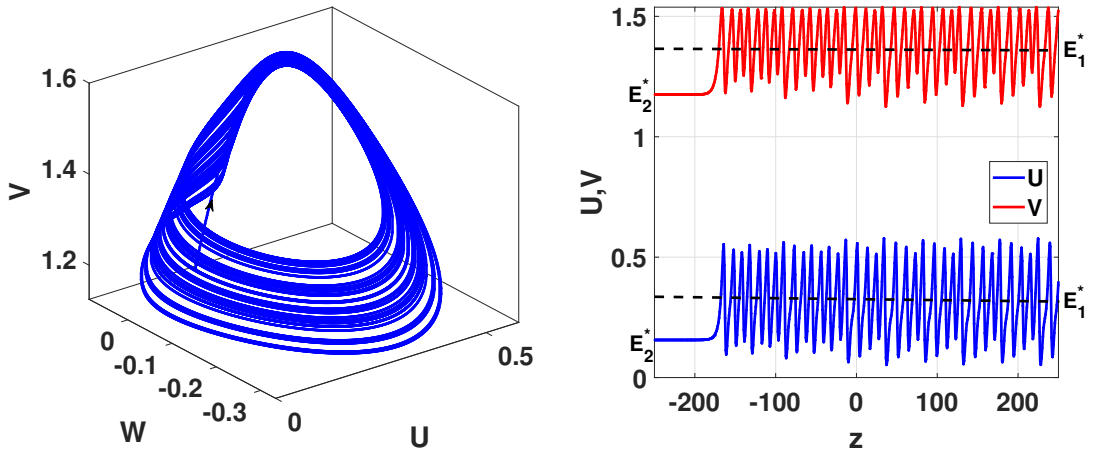
along its unstable manifold and, instead of converging to a steady equilibrium, it is attracted to a stable limit cycle that emerges due to the Hopf bifurcation at E_1^*. This indicates that the orbit lies in the intersection of the unstable manifold of E_2^* and the stable manifold of the periodic orbit associated with E_1^*. The corresponding spatial profiles of the state variables, shown in Figure 4(b), display a transition from the homogeneous state corresponding to E_2^* to sustained oscillations with constant amplitude as the traveling coordinate z increases. In the original reaction-diffusion system, this behavior corresponds to a wave train solution, characterized by spatially periodic oscillations that propagate with a constant wave speed, representing a traveling pattern that connects a steady state to a stable periodic wave.

Continuing from the previous wave-train scenario, when the parameters are chosen from the region inside the magenta loop in Figure 1, specifically $\alpha = 0.45$ and $d = 0.4$, the traveling-wave system exhibits a further qualitative change in its dynamics. In this case, the equilibrium E_1^* gives rise to a stable 2-periodic orbit, as illustrated by the phase portrait in Figure 5(a). The corresponding spatial profiles of the prey and predator variables, shown in Figure 5(b), display sustained oscillations with two distinct peaks per period along the traveling coordinate. In the original reaction-diffusion framework, this behavior corresponds to a 2-periodic wave train, representing a more complex traveling pattern that emerges as a continuation of the wave-train solutions observed for smaller values of d.



(a) (b)

Figure 5: (a) Phase portrait of the traveling-wave system showing 2-period oscillations. (b) Corresponding spatial profiles of the prey (U) and predator (V) populations. Parameter values are $\alpha = 0.45$ and $d = 0.4$.



(a) (b)

Figure 6: (a) Phase portrait of the traveling-wave system showing chaotic behavior. (b) Corresponding spatial profiles of the prey (U) and predator (V) populations. Parameter values are $\alpha = 0.44$ and $d = 0.4$.

As a further continuation of the wave-train dynamics discussed above, a more complex behavior is observed when the parameters are chosen as $\alpha = 0.44$ and $d = 0.4$. In this regime, the equilibrium E_2^* no longer supports periodic motion but instead gives rise to chaotic dynamics, as evidenced by the irregular trajectory in the phase portrait shown in Figure 6(a). The corresponding spatial profiles of the prey and predator variables, depicted in Figure 6(b), exhibit aperiodic oscillations with varying amplitudes along the traveling coordinate. In the original reaction-diffusion system, this behavior corresponds to a chaotic wave train, representing a breakdown of regular periodic wave trains and highlighting a transition from periodic to chaotic traveling patterns as parameters vary.

To further verify the presence of chaos, the largest Lyapunov exponent was computed using the standard Benettin algorithm. Let $\delta U(0)$ denote an infinitesimal perturbation of the trajectory $U(z)$. The largest Lyapunov exponent is defined by

$$\lambda_{\max} = \lim_{z\to\infty} \frac{1}{z} \ln \left| \frac{\delta U(z)}{\delta U(0)} \right|. \tag{13}$$

A positive value of $\lambda_{\max}$ indicates exponential divergence of nearby trajectories and therefore provides a quantitative signature of chaotic dynamics, whereas $\lambda_{\max} < 0$ corresponds to stable dynamics. Figure 7 depicts the variation of the largest Lyapunov exponent with respect to the diffusion coefficient d. It can be observed that $\lambda_{\max}$ remains negative for small values of d, indicating stable dynamics. As d increases, $\lambda_{\max}$ approaches zero, corresponding to the onset of oscillatory behavior through Hopf bifurcation. Furthermore, for the parameter value used in Figure 6, the largest Lyapunov exponent is found to be

$$\lambda_{\max} = 0.0243034466 > 0,$$

which confirms that the irregular traveling-wave pattern shown in Figure 6 is indeed chaotic. Thus, Figure 7 provides quantitative evidence supporting the chaotic traveling-wave dynamics observed in Figure 6.

Figure 8 illustrates the emergence of a pulse-type traveling wave for the parameter values $\alpha = 0.439$ and $d = 0.30151634$ taken from the bifurcation diagram in Figure 1. Panel (a) shows the phase portrait of the traveling-wave system in the (U, V)-plane. The trajectory originates from the saddle equilibrium point E_2, makes a large excursion in phase space, and eventually returns to the same equilibrium. This closed connecting orbit represents a homoclinic orbit to E_2, indicating the occurrence of a homoclinic bifurcation. Such a global bifurcation arises when the limit cycle generated through a Hopf bifurcation grows in amplitude and collides with the saddle equilibrium as the diffusion parameter d is increased.

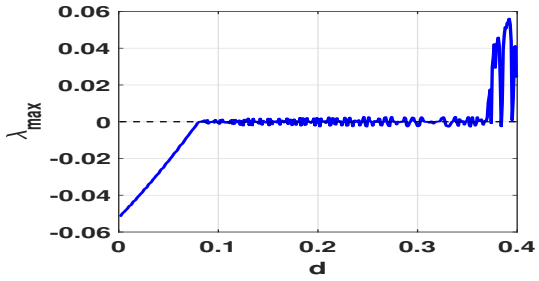


Figure 7: Variation of the largest Lyapunov exponent $\lambda_{\max}$ with respect to the diffusion coefficient d for fixed $\alpha = 0.44$.

The corresponding spatial profiles of the prey and predator densities are displayed in panel (b), respectively. The prey density $U(z)$ exhibits a sharp localized peak followed by a rapid decay to the equilibrium level, while the predator density $V(z)$ shows a pronounced overshoot before relaxing back to its steady state. These localized structures confirm the formation of a traveling pulse, characterized by a large deviation from the background equilibrium over a finite spatial region and convergence to the same equilibrium state as $z \to \pm\infty$.

Ecologically, the pulse wave corresponds to a transient but intense outbreak of prey followed by a delayed predator response, after which both populations return to their background densities. Such pulse-like invasion events are commonly observed in ecological systems where dispersal and strong predator-prey interactions interact to produce localized population bursts.

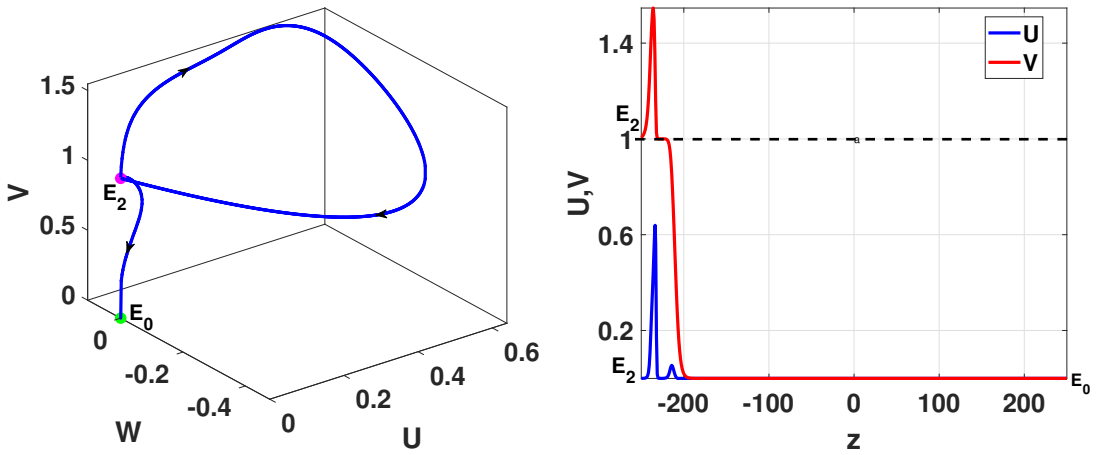


Figure 8: (a) Phase portrait of the traveling-wave system showing a pulse wave. (b) Corresponding spatial profiles of the prey (U) and predator (V) populations. Parameter values are $\alpha = 0.439$ and $d = 0.30151634$.

4 Discussion and Conclusion

In this work, we investigated a diffusive predator-prey model governed by the system (1) with particular emphasis on the role of diffusion and nonlinear predator-prey interactions in shaping spatiotemporal population dynamics. Through a combination of analytical bifurcation analysis and extensive numerical simulations, we revealed a rich spectrum of dynamical behaviors, including steady states, wave fronts, wave pulses, periodic wave trains, multi-periodic oscillations, and chaotic traveling patterns.

The bifurcation analysis demonstrated that the system admits multiple boundary and interior equilibria, whose existence and stability strongly depend on the interaction parameter α and the diffusion coefficient d. A saddle-node bifurcation was identified at a critical value $\alpha = \alpha_{ZH}$, where the number of interior equilibria changes, marking a fundamental reorganization of the system's ecological structure. Hopf bifurcation curves associated with the interior equilibria delineate parameter regions where uniform population distributions lose stability and give rise to oscillatory dynamics. In particular, one interior equilibrium undergoes a Hopf bifurcation leading to stable periodic oscillations, while further parameter variation results in a cascade of period-doubling bifurcations culminating in chaotic behavior.

The one-dimensional bifurcation diagrams obtained by plotting the maxima and minima of prey and predator densities with respect to the diffusion coefficient clearly illustrate how increasing diffusion destabilizes spatially homogeneous equilibria. For small diffusion rates, both species settle at stable equilibrium densities, corresponding to ecological coexistence. As diffusion increases beyond a critical threshold, oscillatory dynamics emerge, followed by multi-periodic and chaotic fluctuations. This indicates that diffusion, often interpreted ecologically as dispersal or movement of species, can act as a destabilizing mechanism that promotes complex population cycles rather than spatial homogenization.

The traveling-wave analysis further revealed the ecological mechanisms underlying spatial pattern formation. Homoclinic orbits correspond to wave pulses, representing localized population outbreaks that propagate through space and eventually decay, while heteroclinic orbits give rise to wave fronts connecting distinct ecological states, such as transitions from low-density to high-density population regimes. Moreover, periodic and multi-periodic wave trains were shown to arise through Hopf and period-doubling bifurcations, representing sustained spatial oscillations in prey and predator densities. The emergence of chaotic wave trains highlights the possibility of irregular, unpredictable population patterns driven purely by intrinsic nonlinear interactions and dispersal.

From an ecological perspective, the cooperation parameter α quantifies the intensity of hunting cooperation among predators. Larger values of α correspond to stronger cooperative effects,

whereby predators benefit from coordinated hunting strategies, information sharing, or collective attacks, leading to enhanced predation efficiency. Such cooperative behavior has been documented in several social predator populations, including wolves, lions, African wild dogs, and dolphins. Consequently, variations in α may significantly influence species coexistence, population persistence, and the stability of ecological communities.

The results further demonstrate that prey dispersal can act as an important ecological mechanism capable of generating complex spatiotemporal dynamics. Although diffusion is commonly associated with spatial mixing and homogenization, the present analysis shows that increasing the prey diffusion coefficient can destabilize stable coexistence states and induce periodic, multi-periodic, and chaotic oscillations. Ecologically, diffusion-induced chaos implies that small changes in the spatial distribution of prey may lead to large and unpredictable fluctuations in both prey and predator populations, thereby affecting ecosystem stability and the predictability of population outbreaks.

The traveling-wave solutions obtained in this study also possess clear ecological interpretations. Wave fronts correspond to invasion processes in which one ecological state gradually replaces another as populations spread through space, whereas pulse waves represent localized population outbreaks that propagate through a finite spatial region before returning to the background equilibrium state. Similar pulse-like propagation phenomena have been reported in biological invasions, forest insect outbreaks, locust population expansions, and epidemic spreading processes. Furthermore, the chaotic traveling waves observed in the present model describe irregular invasion fronts whose amplitudes and spatial structures vary unpredictably in space. Comparable spatiotemporal complexity has been reported in plankton ecosystems, vegetation dynamics in semi-arid environments, and predator-prey communities exhibiting strong nonlinear interactions.

Overall, the present study highlights the important roles of prey dispersal and cooperative hunting in shaping ecological dynamics. The interplay between diffusion, nonlinear interactions, and cooperative predation generates a rich variety of traveling-wave structures, ranging from wave fronts and pulse waves to periodic and chaotic wave trains. These findings provide valuable insight into the mechanisms responsible for spatial population patterns, invasion phenomena, recurrent outbreaks, and irregular spatiotemporal fluctuations observed in ecological systems.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare that they have no conflicts of interest. All authors jointly worked on the results, read and approved the final manuscript, and declare that they have no competing interests.

References

[1] Aguirre, P., Flores, J.D. and González-Olivares, E. *Bifurcations and global dynamics in a predator–prey model with a strong Allee effect on the prey, and a ratio-dependent functional response*, Nonlinear Anal. Real World Appl. 16 (2014), 235–249.

[2] Aguirre, P., Krauskopf, B. and Osinga, H.M. *Global invariant manifolds near a Shilnikov homoclinic bifurcation*, J. Comput. Dyn. 1(1) (2014) 1–38.

[3] Allee, W.C. *Animal aggregations: A study in general sociology*, University of Chicago Press, Chicago, 1931.

[4] Alves, M.T. and Hilker, F.M. *Hunting cooperation and Allee effects in predators*, J. Theor. Biol. 419 (2017), 13–22.

[5] Dhooge, A., Govaerts, W., Kuznetsov, Y.A., Meijer, H.G.E. and Sautois, B. *New features of the software MatCont for bifurcation analysis of dynamical systems*, Math. Comput. Model. Dyn. Syst. 14(2) (2008), 147–175.

[6] Doedel, E.J., Champneys, A.R., Dercole, F., Fairgrieve, T.F., Kuznetsov, Y.A., Oldeman, B., Paffenroth, R.C., Sandstede, B., Wang, X.J. and Zhang, C.H. *AUTO-07P: Continuation and bifurcation software for ordinary differential equations*, Concordia University, 2007.

[7] Dunbar, S.R. *Traveling waves in diffusive predator–prey equations: periodic orbits and point-to-periodic heteroclinic orbits*, SIAM J. Appl. Math. 46(6) (1986), 1057–1078.

[8] Homburg, A.J. and Sandstede, B. *Homoclinic and heteroclinic bifurcations in vector fields*, in: Handbook of Dynamical Systems, Vol. 3, Elsevier, 2010, 379–524.

[9] Hsu, C.H., Yang, C.R., Yang, T.H. and Yang, T.S., *Existence of traveling wave solutions for diffusive predator–prey type systems*, J. Differ. Equ. 252(4) (2012), 3040–3075.

[10] Huang, J., Lu, G. and Ruan, S. *Existence of traveling wave solutions in a diffusive predator–prey model*, J. Math. Biol. 46(2) (2003), 132–152.

[11] Krauskopf, B. and Osinga, H.M. *Computing invariant manifolds via the continuation of orbit segments*, in: Numerical Continuation Methods for Dynamical Systems, Springer, 2007, 117–154.

[12] Krauskopf, B., Osinga, H.M., Doedel, E.J., Henderson, M.E., Guckenheimer, J., Vladimirsky, A., Dellnitz, M. and Junge, O. *A survey of methods for computing (un)stable manifolds of vector fields*, Int. J. Bifurc. Chaos, 15(3) (2005), 763–791.

[13] Li H. and Xiao, H. *Traveling wave solutions for diffusive predator–prey type systems with nonlinear density dependence*, Comput. Math. Appl. 74(10) (2017), 2221–2230.

[14] Manna, K., Pal, S. and Banerjee, M. *Analytical and numerical detection of traveling wave and wave-train solutions in a prey–predator model with weak Allee effect*, Nonlinear Dyn. 100(3) (2020), 2989–3006.

[15] Murray, J.D. *Mathematical biology*, Springer-Verlag, Berlin, 1989.

[16] Murray, J.D. *Spatial models and biomedical applications*, Mathematical Biology, Springer, 2003.

[17] Perko, L. *Differential equations and dynamical systems*, Vol. 7, Springer Science & Business Media, 2013.

[18] Sandstede, B. *Stability of N-fronts bifurcating from a twisted heteroclinic loop and an application to the FitzHugh–Nagumo equation*, SIAM J. Math. Anal., 29(1) (1998), 183–207.

[19] Sherratt, J.A. and Smith, M.J. *Periodic traveling waves in cyclic populations: field studies and reaction–diffusion models*, J. R. Soc. Interface. 5(22) (2008), 483–505.

[20] Takahashi, L.T., Maidana, N.A., Ferreira Jr, W.C., Pulino, P. and Yang, H.M. *Mathematical models for the Aedes aegypti dispersal dynamics: Traveling waves by wing and wind*, Bull. Math. Biol. 67(3) (2005), 509–528.

[21] Turing, A.M. *The chemical basis of morphogenesis*, Bull. Math. Biol. 52(1) (1990), 153–197.

[22] Villar-Sepúlveda, E., Aguirre, P. and Breña-Medina, V.F. *A case study of multiple wave solutions in a reaction-diffusion system using invariant manifolds and global bifurcations*, SIAM J. Appl. Dyn. Syst. 22(2) (2023), 918–950.

[23] Vishwakarma, K. and Sen, M. *Role of Allee effect in prey and hunting cooperation in a generalist predator*, Math. Comput. Simul. 190 (2021), 622–640.

[24] Wu, X., Luo, Y. and Hu, Y. *Traveling waves in a diffusive predator–prey model incorporating a prey refuge*, Abstract. Appl. Anal. 2014 (2014), Article ID 679131.

[25] Yadav, R., Sen, M. and Pal, S. *Impact of diffusion-driven instability on traveling wave solutions*, Phys. Rev. E. 111(2) (2025), 024202.

[26] Yamashita, W.M., Takahashi, L.T. and Chapiro, G. *Traveling wave solutions for the dispersive models describing population dynamics of Aedes aegypti*, Math. Comput. Simul. 146 (2018), 90–99.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1295–1318
https://doi.org/10.22067/ijnao.2026.96710.1788
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Practical early stopping for adaptive barrier SGD: Balancing stochastic speed with validation accuracy

A. Fillali*, S. Addoune and R. Zeghdane

Abstract

Stochastic gradient descent (SGD) algorithms based on adaptive barrier functions are highly efficient for large-scale constrained optimization. However, their inherent stochastic nature leads to the critical "last-iterate problem", where update noise causes significant oscillations, making reliance on the final solution a high-stakes gamble. To address this instability, we introduce the early-stopped barrier SGD (EB-SGD) algorithm, a practical approach that replaces reliance on the noisy last iterate with a robust "best-iterate tracking" mechanism. Our method

*Corresponding author

Received 26 November 2025; revised 12 February 2026; accepted 14 February 2026

Abdelouakil Fillali
Laboratoire d'Analyse Mathématiques et ses Applications (LAMA), Department of Mathematics, University Mohamed El Bachir El Ibrahimi, El Anasser, Bordj Bou Arreridj, 34030, Algeria. e-mail: abdelouakil.fillali@univ-bba.dz

Smail Addoune
Laboratoire d'Analyse Mathématiques et ses Applications (LAMA), Department of Mathematics, University Mohamed El Bachir El Ibrahimi, El Anasser, Bordj Bou Arreridj, 34030, Algeria. e-mail: smail.addoune@univ-bba.dz

Rebiha Zeghdane
Laboratoire d'Analyse Mathématiques et ses Applications (LAMA), Department of Mathematics, University Mohamed El Bachir El Ibrahimi, El Anasser, Bordj Bou Arreridj, 34030, Algeria. e-mail: rebiha.zeghdane@univ-bba.dz

How to cite this article
Fillali, A., Addoune, S. and Zeghdane, R., Practical early stopping for adaptive barrier SGD: Balancing stochastic speed with validation accuracy. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1295-1318. https://doi.org/10.22067/ijnao.2026.96710.1788

employs periodic validation, where the cheap stochastic process is temporarily halted to compute a high-quality, deterministic "Validation Score." This score is defined as a combination of the full-batch objective function and a large penalty for constraint violation. We maintain the iterate that achieves the best score recorded so far and utilize a "patience"-based stopping criterion to filter out minor stochastic oscillations and halt the algorithm only upon persistent stagnation. This strategy introduces a deliberate computational trade-off: Exchanging many inexpensive stochastic steps for fewer, more expensive validation steps. We prove theoretically that the sequence of best iterates returned by EB-SGD converges almost surely to the unique optimal solution. Furthermore, our numerical experiments confirm the practical superiority of this approach, demonstrating significantly reduced variance (higher stability) and finding a final solution that is substantially more accurate than the baseline last-iterate method.

AMS subject classifications (2020): Primary 90C15, 90C25; Secondary 49M30, 65K05, 68T05.

Keywords: Constrained optimization, Stochastic gradient descent, Early stopping, Best-iterate tracking.

1 Introduction

Stochastic gradient descent (SGD) and its variants have become cornerstone algorithms [3, 10, 11, 14, 24] in modern large-scale optimization. While exceptionally efficient for unconstrained problems, its application to constrained optimization remains an "active area of research" [19]. Many classical approaches require "computationally expensive projection steps" [21], which become prohibitive when the number of constraints is very large.

To overcome this challenge, recent methods have been developed that rely on sampling only a subset of constraints, such as those based on penalty functions [12, 22] or primal-dual approaches [23]. A "particularly promising direction," however, is the Adaptive Relaxed Barrier SGD algorithm proposed by Dimitrieski, Cao, and Ebenbauer [7]. This approach avoids expensive projections [8, 9] and allows for sampling both the objective function and the constraints.

While this baseline algorithm is "guaranteed to converge almost surely" [7], its practical application faces a critical limitation due to "inherent stochastic noise." Because of this noise, the validation error curve does not decrease monotonically; rather, as Prechelt [16] demonstrated, real-world validation curves are "ugly" and exhibit "multiple local minima." This leads to what we term the "last-iterate problem": Relying on the final iterate x_k is a gamble, as the stochastic process oscillates significantly rather than settling on a single point [6].

In this paper, we address this practical limitation by proposing the "Early-stopped barrier SGD" (EB-SGD) algorithm. Instead of relying on the noisy last iterate, we integrate the baseline algorithm [7] with a robust "best-iterate tracking" mechanism. Our method periodically halts the stochastic process to compute a deterministic, high-quality "Validation Score." We then apply a formal stopping criterion, such as the "Successive Increases" (UP_s) rule formalized by Prechelt [16], to this deterministic validation path. This "patience" parameter s makes the algorithm robust to small, random oscillations.

This strategy introduces a deliberate computational trade-off: Exchanging many inexpensive stochastic steps for fewer, more expensive validation steps. By doing so, we aim to filter out the high-variance oscillations inherent to the barrier method and recover a high-precision solution.

We provide a rigorous theoretical analysis (Theorem 1) establishing that this approach preserves the "strong almost sure convergence guarantees" of the underlying relaxed barrier method [7, 13]. Our numerical experiments provide a "clear validation" of these practical claims. The results demonstrate that EB-SGD decisively overcomes the "random peak" problem, achieves significantly reduced variance, and finds a final solution of "substantially higher precision" compared to the baseline algorithm.

2 The proposed methodology

In this section, we introduce our algorithm, which integrates the adaptive relaxed barrier SGD from Dimitrieski, Cao, and Ebenbauer [7] with a robust best-iterate tracking and early stopping mechanism. Instead of relying on the noisy last iterate, our method uses periodic validation to find a practically superior solution.

2.1 Problem formulation and assumptions

The objective is to solve the following constrained optimization problem:

$$\min_{x\in\mathbb{R}^d} f(x) := \frac{1}{n}\sum_{i=1}^{n} f_i(x) \quad \text{subject to} \quad g_j(x) \le 0, \quad \text{for all } j \in \{1,\dots,m\}, \tag{1}$$

where we assume that the objective functions f_i and constraint functions g_j are continuously differentiable ($\mathcal{C}^1$) functions. We adopt the following standard assumptions, with mathematical formulations that are common in this field

[4, 5, 18].

Assumption 1 (Objective Function Properties)**.** The objective function f and its components f_i are assumed to possess the following properties for constants $L \geq \mu > 0$ and for any $x, y \in \mathbb{R}^d$:

1. **L-Smoothness:** Each component $f_i(x)$ is L-smooth, meaning its gradient is Lipschitz continuous with constant L:

$$f_i(y) \leq f_i(x) + \nabla f_i(x)^T(y - x) + \frac{L}{2}||y - x||^2.$$

2. **Convexity:** Each component $f_i(x)$ is convex:

$$f_i(y) \geq f_i(x) + \nabla f_i(x)^T(y - x).$$

3. **Strong Convexity:** The total function $f(x)$ is μ-strongly convex:

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2}||y - x||^2.$$

These assumptions, although restrictive, are not uncommon in convergence proofs for gradient-based optimization algorithms.

[4, 7].

Assumption 2 (Constraint Functions)**.** The constraint functions $g_j(x)$ are affine, defined as follows:

$$g_j(x) := a_j^T x + b_j,$$

where $a_j \in \mathbb{R}^d$ and $b_j \in \mathbb{R}$. The feasible set $\mathbb{X} := \{x \in \mathbb{R}^d : g_j(x) \leq 0, \text{ for all } j\}$ is assumed to have a nonempty interior, which is a fundamental condition for applying barrier-based methods [4].

Definition 1 (Relaxed Logarithmic Barrier)**.** The function $B : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ utilized in this work is defined as follows:

$$B(z, \delta) := \begin{cases} -\delta \log(-z), & z < -\delta, \\ \frac{1}{2}\left(\frac{(z+2\delta)^2}{\delta} - \delta\right) - \delta \log(\delta), & z \geq -\delta. \end{cases} \tag{2}$$

This specific relaxation is globally defined, continuously differentiable (C^2), and convex with respect to its first argument z [7].

Remark 1 (Convexity of the Relaxed Barrier)**.** It is crucial to note that the specific relaxed barrier function $B(z, \delta)$ employed in the underlying algorithm [7] is convex in its first argument

z. Consequently, since the constraint functions $g_j(x)$ are affine (Assumption 2), the composite barrier term $B(g_j(x), \delta)$ preserves convexity with respect to x. This follows directly from the property that the composition of a convex function with an affine mapping is convex. This property is fundamental for guaranteeing the uniqueness of the solution in the relaxed formulation used in Theorem 1.

Assumption 3 (Stochastic Sampling)**.** At each iteration k, an objective index $i_k \sim \mathcal{U}\{1, \ldots, n\}$ and a constraint index $j_k \sim \mathcal{U}\{1, \ldots, m\}$ are sampled independently and uniformly at random. This uniform distribution is chosen to ensure that the stochastic gradients are unbiased estimators of the true gradients, thereby satisfying the standard assumptions for SGD algorithms [17]. This is demonstrated by the following equations:

1. **For the objective function:** The expected value of the sampled stochastic gradient is equal to the true gradient of the full function [7, 17]:

$$\mathbb{E}[\nabla f_{i_k}(x)] = \frac{1}{n} \sum_{i=1}^{n} \nabla f_i(x) = \nabla f(x).$$

2. **For the barrier function:** The expected value of the sampled stochastic gradient of the barrier function equals the average effect of the gradients over all constraints [7]:

$$\mathbb{E}[\nabla B(g_{j_k}(x), \delta)] = \frac{1}{m} \sum_{j=1}^{m} \nabla B(g_j(x), \delta).$$

2.2 The "last-iterate" problem in stochastic optimization

The foundational algorithm by Dimitrieski, Cao, and Ebenbauer [7] proposes the following stochastic update rule, which samples both an objective component and a constraint:

$$x_{k+1} = x_k - \gamma_k (\nabla f_{i_k}(x_k) + \nabla B(g_{j_k}(x_k), \delta_k)). \tag{3}$$

While this baseline algorithm is guaranteed to converge almost surely asymptotically [7], its practical behavior in finite time differs. Even with a diminishing step-size schedule ($\gamma_k \propto k^{-0.8}$), the error curve exhibits significant nonmonotonicity. This instability arises primarily from the **adaptive nature of the barrier function**. As the barrier parameter δ_k decreases, the barrier becomes increasingly steep and the Hessian ill-conditioned near the constraint boundaries [9].

Consequently, stochastic gradient updates that push iterates towards the boundary result in large repulsive gradients, causing the sequence to oscillate rather than settling smoothly.

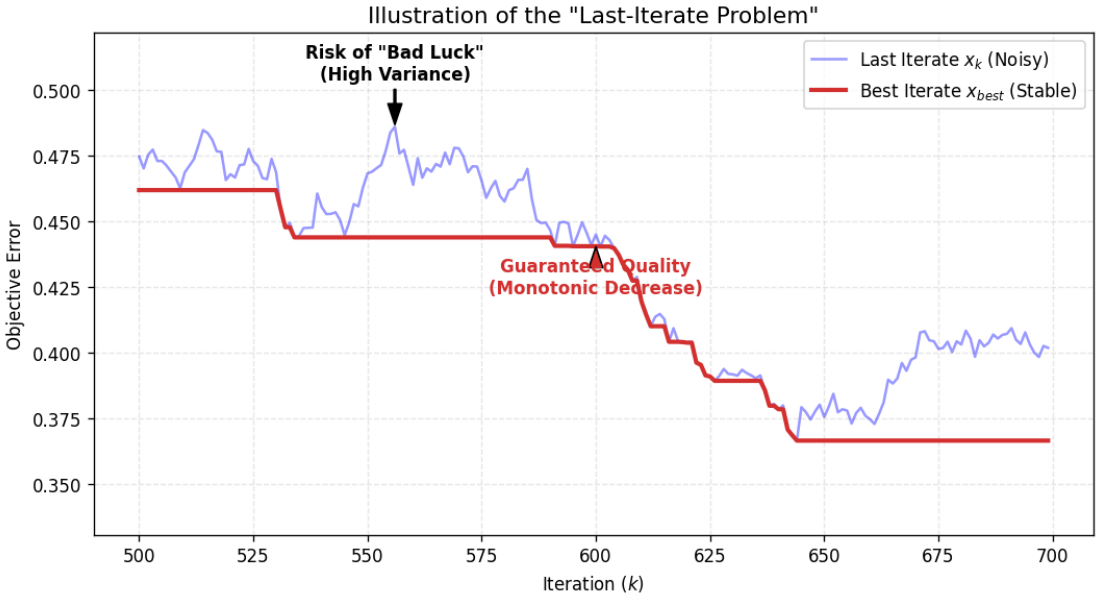


Figure 1: Illustration of the "last-iterate problem" (Zoom-in view, iterations 500–700). Using the standard step-size schedule from Dimitrieski, Cao, and Ebenbauer [7] ($\gamma_k \propto k^{-0.8}$), the inherent stochastic noise causes the current iterate x_k (blue) to oscillate significantly. Relying on the final point is a "gamble," as the error may stop at a random peak (indicated by the black arrow). In contrast, our proposed best-iterate tracking mechanism (red) guarantees monotonic improvement and algorithmic stability.

This phenomenon renders reliance on the "last iterate" x_k a gamble, as illustrated in Figure 1. As empirically demonstrated by Prechelt [16], real-world validation curves are rarely smooth; they are often "ugly" and exhibit multiple local minima. It is worth noting that while Polyak–Ruppert averaging [15] is a standard technique to mitigate such oscillations by averaging all iterates, our proposed approach differs by explicitly selecting the *best* feasible iterate based on validation, rather than indiscriminately averaging the trajectory. This leads to the proposed solution for the "last-iterate problem".

2.3 Best-iterate tracking mechanism

To overcome the instability of the last iterate, we propose a best-iterate tracking mechanism based on periodic validation, a standard practice in early stopping literature.

Instead of only tracking the last iterate x_k, we periodically halt the stochastic process (e.g., every V iterations) to compute a deterministic, high-quality "Validation Score" for the current iterate. This score is defined as a combination of the full-batch objective function and a penalty for the maximum constraint violation:

$$\text{Score}(x_k) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} f_i(x_k)}_{\text{Full Objective}} + \underbrace{P \cdot \max_{j\in\{1,\ldots,m\}}(0, g_j(x_k))}_{\text{Max Violation Penalty}}, \tag{4}$$

where P is a large penalty coefficient (e.g., 10^6). From a theoretical standpoint, exact penalty methods require P to exceed the maximum Lagrange multiplier ($P > ||\lambda^*||_\infty$) [1, 9]. In this work, we do not derive a formal bound for λ^*; instead, we adopt a practical heuristic approach inspired by exact penalty methods. We choose P sufficiently large to empirically ensure that any infeasible point incurs a prohibitive cost, effectively prioritizing feasible solutions during validation. Crucially, our sensitivity analysis (Section 4.5) demonstrates that EB-SGD is remarkably robust to the choice of P, maintaining strict feasibility and high accuracy across varying orders of magnitude.

We then maintain two tracking variables throughout the training process:

1. `best_score`: The lowest (best) validation score recorded so far (analogous to $E_{opt}(t)$ in Prechelt [16]).

2. `x_best`: The specific iterate x_k that achieved the `best_score`.

2.4 Early stopping and the computational trade-off

This validation step introduces a new computational trade-off. The original algorithm [7] is fast because its per-iteration cost is independent of n and m. Our validation step is computationally "expensive" because it requires iterating over all n objective components and all m constraints to compute the score in (4).

However, the benefit is that we no longer need to run the algorithm for an arbitrarily large, fixed number of iterations (K). We can stop intelligently once the algorithm is no longer improving. This is achieved by adopting a formal stopping criterion, such as the "Successive Increases" (UP_s) rule formalized by Prechelt [16]. This "patience" parameter s makes the algorithm robust to small, random oscillations, exchanging many cheap stochastic steps for fewer, more expensive validation steps to ensure high solution quality.

2.5 Algorithm specification: EB-SGD

We consolidate these concepts into our proposed algorithm, EB-SGD, detailed in Algorithm 1. This algorithm uses the update rule from (3) as its "engine" but uses the validation and tracking mechanism from Section 2.3 to select the final output.

Algorithm 1 Early-Stopped Barrier SGD (EB-SGD)

Require: Initial point x_0, validation frequency V, patience s.

$x_{\text{best}} \leftarrow x_0$
best_score $\leftarrow$ Score(x_0) ▷ Using (4)
patience_counter $\leftarrow 0$
$x_k \leftarrow x_0$
for $k = 0, 1, 2, \ldots$ **do** ▷ Stochastic Update Step (from [7])
    Sample $i_k \sim \mathcal{U}\{1, \ldots, n\}$ and $j_k \sim \mathcal{U}\{1, \ldots, m\}$
    Compute schedules γ_k and δ_k
    $x_{k+1} \leftarrow x_k - \gamma_k(\nabla f_{i_k}(x_k) + \nabla B(g_{j_k}(x_k), \delta_k))$
    $x_k \leftarrow x_{k+1}$
    if $(k+1) \pmod V = 0$ **then** ▷ Time for periodic validation
        current_score $\leftarrow$ Score(x_k)
        if current_score $<$ best_score **then**
            best_score $\leftarrow$ current_score
            $x_{\text{best}} \leftarrow x_k$ ▷ Save the best iterate
            patience_counter $\leftarrow 0$
        else
            patience_counter $\leftarrow$ patience_counter $+ 1$
        end if
    end if
    if patience_counter $\geq s$ **then** ▷ Stopping criterion met
        break
    end if
end for
return x_{best} ▷ Return the best solution found, not the last

We now provide the formal convergence analysis for the proposed EB-SGD algorithm (Algorithm 1) to establish that its output, the sequence of best iterates x_{best}, converges almost surely to the unique optimal solution.

Theorem 1 (Convergence of the Best-Iterate Sequence)**.** Let Assumptions 1, 2, and 3 hold. Let the step-size $\{\gamma_k\}_{k\geq 0}$ and barrier sequence $\{\delta_k\}_{k\geq 0}$ (defined as $\delta_k = \delta_\infty + \epsilon_k$) satisfy the following standard conditions:

$$\sum_{k=0}^{\infty} \gamma_k = \infty, \quad \sum_{k=0}^{\infty} \gamma_k^2 < \infty, \quad \text{and} \quad \sum_{k=0}^{\infty} \gamma_k \epsilon_k < \infty.$$

Then, the infinite sequence of best iterates $\{x_{best,k}\}_{k\geq 0}$, generated by the tracking mechanism of Algorithm 1 (considered without the early stopping criterion), converges almost surely (a.s.) to the unique, feasible optimal solution $x^*(\delta_\infty)$.

Proof. The proof relies on the convergence of the underlying "engine" sequence X_k and the continuity of the validation score function.

Step 1: Convergence of Validation Points

The foundational result from Dimitrieski, Cao, and Ebenbauer [7] states that the full sequence X_k converges almost surely to $x^*(\delta_\infty)$. The sequence of iterates checked during periodic validation, X_{k_v} (where $k_v = V, 2V, 3V, \dots$), is an infinite subsequence of X_k. Therefore, this subsequence must also converge to the same limit:

$$\lim_{v\to\infty} X_{k_v} = x^*(\delta_\infty) \quad (a.s.).$$

Step 2: Convergence of the Score Function

The validation score function, $Score(x) = f(x) + P \cdot \max_j(0, g_j(x))$, is continuous. This follows because $f(x)$ (Assumption 1), $g_j(x)$ (Assumption 2), and the max function are all continuous. By the Continuous Mapping Theorem (see e.g., [20], and for application to almost sure convergence, see [2]), since $X_{k_v} \to x^*(\delta_\infty)$ (a.s.) and $Score(\cdot)$ is continuous, the sequence of scores must also converge:

$$\lim_{v\to\infty} Score(X_{k_v}) = Score(x^*(\delta_\infty)) \quad (a.s.).$$

Step 3: Convergence of the Best Score

Let $S^* = Score(x^*(\delta_\infty))$. The variable *best_score* at validation step v is the running infimum of the observed scores: $best\_score = \inf_{i\leq v}\{Score(X_{k_i})\}$. This sequence $\{best\_score_v\}$ is "monotonically nonincreasing" by the definition of the running infimum. Furthermore, since $X_{k_v} \to x^*(\delta_\infty)$ (a.s.) and $x^*(\delta_\infty)$ is the unique minimizer of $Score(x)$ on the approximately feasible set, the sequence of scores $Score(X_{k_v})$ is bounded below by S^*. Following standard results from analysis (e.g., [20] and [2]), since a nonincreasing sequence that is bounded below must converge, the sequence of best scores converges to the infimum S^*:

$$lim_{v\to\infty}\text{best\_score} = S^* = Score(x^*(\delta_\infty)) \text{ (a.s.)}.$$

Step 4: Establishing Feasibility and Uniqueness. Let x_c be any cluster point of the sequence x_{best}. By the definition of x_{best} and the continuity of $Score(\cdot)$, we have $Score(x_c) = S^* = Score(x^*(\delta_\infty))$. It is worth noting that while the penalty term is flat (zero) over the feasible region, the uniqueness of the minimizer is guaranteed by the objective function itself. Recall that $Score(x) = f(x)$ for all feasible x. Since $f(x)$ is μ-strongly convex (Assumption 1), the function $Score(x)$ inherits this strong convexity (as the sum of a strongly convex function and a convex penalty). Consequently, the score function admits a unique minimizer, ensuring that convergence in score values implies convergence of the iterates to the unique optimizer. Regarding feasibility: Assume for contradiction that x_c is infeasible. For a sufficiently large penalty P (acting as an exact penalty), we would have $Score(x_c) > S^*$, which contradicts the convergence of scores. Thus, x_c must be feasible. Combining feasibility with the strict convexity of $f(x)$, we conclude that x_c must coincide uniquely with the theoretical minimizer $x^*(\delta_\infty)$.

Step 5: Strong Convergence (Uniqueness)

From Step 4, we consider the properties of the barrier-augmented objective function $F(x) = f(x)+\frac{1}{m}\sum_{j=1}^{m} B(g_j(x), \delta_\infty)$. Recall that $f(x)$ is μ-strongly convex (Assumption 1). Furthermore, as highlighted in Remark 1, the relaxed barrier terms $B(g_j(x), \delta_\infty)$ are convex functions (as established in [7]). By standard optimization theory, the sum of a μ-strongly convex function and a convex function results in a strictly μ-strongly convex function. Consequently, the total objective $F(x)$ admits a *unique* minimizer. Since any cluster point x_c of the sequence has been shown to minimize this objective, x_c must coincide with the unique optimal solution $x^*(\delta_\infty)$. Therefore, the entire sequence converges almost surely to $x^*(\delta_\infty)$. □

Remark 2 (Asymptotic Convergence vs. Finite Stopping)**.** It is important to clarify the distinction between the theoretical guarantee of Theorem 1 and the practical execution of Algorithm 1. Theorem 1 establishes that the infinite sequence of best iterates $\{x_{best,k}\}_{k\geq 0}$ converges almost surely to the optimal solution $x^*(\delta_\infty)$. This asymptotic property is crucial because it proves that the "best-iterate" tracking mechanism acts as a stable estimator, filtering out the oscillatory noise of the last iterate. In practice, the "patience" parameter s (Line 17 of Algorithm 1) enforces a finite stopping time. While the finite algorithm cannot strictly "converge" in the asymptotic sense, Theorem 1 provides the theoretical justification that the sequence being tracked is indeed directing towards the unique optimum, validating the quality of the solution returned upon stopping.

3 Numerical experiments and results

To validate the practical performance of our proposed **EB-SGD** algorithm, we conduct a series of numerical experiments. The primary objective is to compare its convergence, stability, and efficiency against the original **Base SGD** algorithm proposed by Dimitrieski, Cao, and Ebenbauer [7].

3.1 Experimental setup

To ensure a fair and direct comparison, we replicate the academic benchmark problem described in [7, Section III].

- **Problem Dimensions:** We use a problem dimension of $d = 50$.

- **Objective Function:** The objective $f(x)$ is a finite sum of $n = 10$ components. Following the benchmark problem in [7], the objective function is defined as follows:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}(\alpha_{i,j}x_j + \log(1 + e^{-\alpha_{i,j}x_j}) + (x_j - \beta)^2).$$

  Each component $f_i(x)$ is composed of a logistic regression term and a strongly convex quadratic term [7]. We intentionally place the unconstrained minimizer outside the feasible set.

  The objective function coefficients $\alpha_{i,j}$ are sampled uniformly from $[0.1, 1.1]$. The parameter β determines the location of the unconstrained minimizer. In our setup, we tuned $\beta \approx 2.365$ specifically to set the Euclidean norm of the unconstrained minimizer to $||x_f^*||_2 = 15$. This configuration follows the benchmark problem design in [7], ensuring that the unconstrained optimum lies well outside the feasible region, thereby forcing the algorithm to actively engage with the constraints.

- **Constraints:** The feasible set is defined by $m = 10,000$ affine inequality constraints. Following the setup in [7], these constraints are generated to form an outer approximation of an ellipsoidal set $\mathcal{E}$. First, the set is defined as $\mathcal{E} := \{x \in \mathbb{R}^d : x^\top Qx \leq 100\}$, where Q is a diagonal matrix with entries $q_i \sim \mathcal{U}(1, 1.5)$. Then, m points (y_j) are sampled from the boundary of this set, and the constraints are defined as the supporting halfspaces $g_j(x) =$

$y_j^\top Qx - 100 \leq 0$. This construction ensures the problem is feasible and Assumption 2 is satisfied [7].

- **Implementation and Environment:** The simulations are conducted in Python. We run $R = 1000$ independent simulations, each with a different random seed, to gather statistics on performance. The experiments were conducted on the Google Colab platform with the following specifications: Intel(R) Xeon(R) CPU @ 2.00GHz, 13.61 GB of RAM, and an NVIDIA Tesla T4 GPU (15 GB VRAM), running on Linux (6.6.97+) with Python 3.12.11.

3.2 Algorithms for comparison

To isolate the effect of our tracking mechanism, we run the same underlying stochastic update (3) for a fixed $K = 10,000$ iteration budget. We compare the performance of two different output strategies running on this identical engine:

- **Base SGD (last-iterate) [7]:** This is the original strategy, which uses the final iterate x_k as its output. The blue line in our plot (Figure 2) represents the error $||x_k - x^*||_2$ at each iteration k.

- **Best-Iterate Tracking (Ours):** This strategy implements the validation and tracking mechanism from Section 2.3. The red line (Figure 2) represents the error $||x_{\text{best}} - x^*||_2$ at each iteration k, where x_{best} is the best-performing iterate recorded *up to that point.*

This experimental design allows for a direct, step-by-step comparison of the stability and quality of the two output strategies.

It is important to note that for this comparative experiment, the `break` condition (early stopping) from Algorithm 1 is *not* used. We must run the simulation for the full 10,000 iterations to generate the complete error curves for comparison. The resulting stability and superior accuracy of the best-iterate tracking curve (red) is precisely what *justifies* the use of an early stopping criterion in our final, practical EB-SGD algorithm (Algorithm 1).

Both algorithms use the *exact same* hyperparameter schedules as defined in [7] to ensure a fair comparison:

- **Step-size schedule:** $\gamma_k = 0.3k^{-0.8}$.

- **Barrier schedule:** $\delta_k = \delta_\infty + \epsilon_k$, with $\delta_\infty = 10^{-6}$ and $\epsilon_k = 5k^{-0.3}$.

3.3 Evaluation metrics

We evaluate the algorithms based on the following metrics:

1. **Convergence Error:** We measure the l_2-norm of the error, $||x_k - x^*||_2$, where x_k is the current iterate and x^* is the true (pre-computed) constrained optimizer.

2. **Statistical Performance:** We plot the **mean** error averaged over all $R = 1000$ runs. We also plot the **standard deviation** as a shaded region around the mean, which visually represents the algorithm's stability and robustness to different stochastic paths.

4 Results and discussion

It is well-established that barrier-based SGD significantly outperforms projection-based methods in terms of execution time for large-scale constrained problems, as projection steps become prohibitive [7]. Therefore, our comparative analysis focuses on evaluating the stability improvements of EB-SGD against the standard last-iterate baseline [7] and Polyak–Ruppert averaging [15]. The results are analyzed in terms of convergence speed, final solution accuracy, and statistical robustness.

4.1 Convergence and stability analysis

Figure 2 (Left) presents the mean convergence error over $R = 1000$ independent runs. The visual results highlight the distinct behaviors of the three strategies:

- **Base SGD (Blue):** Exhibits the characteristic "last-iterate problem," stagnating at a higher error level with significant variance due to the ill-conditioned barrier near the boundary.

- **Polyak–Ruppert Averaging (Green):** While averaging successfully reduces the variance compared to Base SGD, it converges slowly. The error curve shows a persistent "lag." This underperformance arises because Polyak averaging accumulates the entire history of iterates, including the **sub-optimal initial solutions** and the transient iterates generated while the barrier parameter δ_k was still large. This "memory effect" biases the average away from the fine-grained solution required at the final stages.

- **EB-SGD (Red-Ours):** Our best-iterate tracking mechanism demonstrates superior performance. By explicitly selecting the iterate that minimizes the validation score, it effectively discards the poor initial steps and instantaneously locks onto the high-quality solutions as soon as they appear, avoiding the lag associated with averaging.

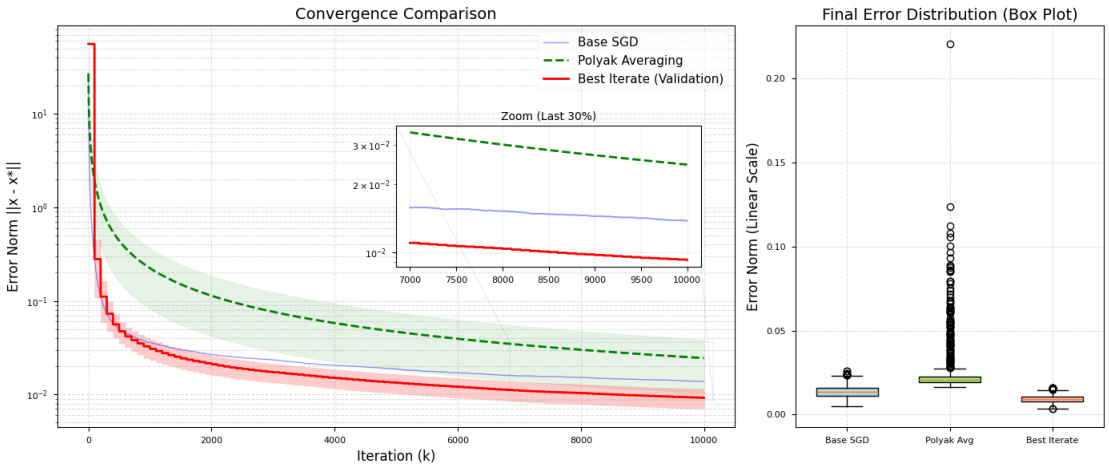


Figure 2: Performance Comparison: (Left) Mean convergence error over 1000 runs. EB-SGD (Red) converges faster and deeper than Polyak Averaging (Green), which suffers from initialization bias. (Right) Box plot of the final error distribution. EB-SGD demonstrates the lowest median error and tightest spread.

4.2 Solution quality and robustness

Table 1 provides a quantitative summary of the final solutions. The "best-iterate" strategy achieves a final error norm of 9.20×10^{-3}, which is substantially lower than both Base SGD (1.38×10^{-2}) and Polyak Averaging (2.45×10^{-2}).

The poor performance of Polyak averaging in this specific context—constrained optimization with adaptive barriers—confirms that indiscriminately averaging all iterates is not ideal when the "target" (the central path) is shifting. The box plot in Figure 2 (Right) further reinforces this, showing that EB-SGD offers the most robust distribution with minimal outliers.

Table 1: Quantitative comparison of final solutions (Mean ± Std. Dev. over 1000 runs).

Metric	**Base SGD**	**Polyak Avg**	**EB-SGD (Ours)**
Final Error Norm ($\|\|x - x^*\|\|_2$)	$1.38 \times 10^{-2} \pm 3.5 \times 10^{-3}$	$2.45 \times 10^{-2} \pm 1.4 \times 10^{-2}$	$\mathbf{9.20 \times 10^{-3} \pm 2.2 \times 10^{-3}}$
Final Objective ($f(x)$)	81.7414	81.7420	**81.7412**
Max Violation	0.00	0.00	0.00

4.3 Computational trade-off analysis for validation frequency (V)

To quantitatively analyze the "computational trade-off" introduced in Section 2.4, we conducted a sensitivity analysis on the validation frequency hyperparameter, V. This parameter directly controls the balance between the low computational cost of stochastic steps and the 'expensive' cost of full-batch validation steps.

Experimental Setup

For this analysis, we used the same problem setup described in Section 3.1 ($d = 50, n = 10, m = 10000$), following the benchmark problem in [7]. We fixed the total iteration budget at $K = 10,000$ and ran the best-iterate tracking mechanism (Algorithm 1) with different values for V, ranging from $V = 20$ (very frequent validation) to $V = 1000$ (infrequent validation). We measured two key metrics:

1. **Mean Wall-Clock Time:** The total real-world time (in seconds) to complete $10,000$ iterations, including the overhead of all validation computations.
2. **Mean Final Error:** The l_2-norm of the error ($||x_{\text{best}} - x^*||_2$) of the returned solution x_{best} after the full budget was exhausted.

Results and Analysis

The results of this analysis are presented in Figure 3 and Table 2.

The data clearly illustrates the trade-off:

- **Frequent Validation (Small V):** When V is small (e.g., $V = 20$), the expensive validation step is executed frequently (500 times), resulting in the highest wall-clock time (1.02s). However, this frequent checking allows the algorithm to effectively "capture" high-quality iterates, yielding the **lowest final error** (8.54×10^{-3}).

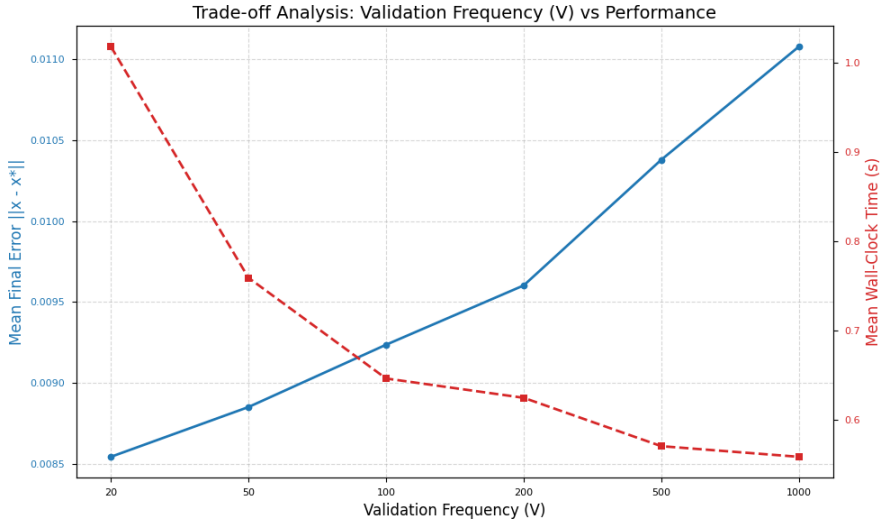


Figure 3: Sensitivity analysis for the validation frequency (V). The left axis (blue) shows the mean final error, while the right axis (red) shows the mean wall-clock time (seconds). The plot demonstrates a clear trade-off between solution accuracy and computational cost.

Table 2: Quantitative comparison of the impact of validation frequency (V) on computational cost and solution accuracy (Mean ± Std. Dev. over 1000 runs).

Freq. (V)	**Mean Time (s)**	**Mean Final Error ($\|\|x - x^*\|\|_2$)**	**Mean Max Violation**
$V = 20$	1.02 ± 0.33	$8.54 \times 10^{-3} \pm 2.07 \times 10^{-3}$	0.0
$V = 50$	0.76 ± 0.22	$8.85 \times 10^{-3} \pm 2.15 \times 10^{-3}$	0.0
$V = 100$	0.65 ± 0.16	$9.24 \times 10^{-3} \pm 2.13 \times 10^{-3}$	0.0
$V = 200$	0.62 ± 0.17	$9.60 \times 10^{-3} \pm 2.26 \times 10^{-3}$	0.0
$V = 500$	0.57 ± 0.14	$1.04 \times 10^{-2} \pm 2.30 \times 10^{-3}$	0.0
$V = 1000$	0.56 ± 0.13	$1.11 \times 10^{-2} \pm 2.43 \times 10^{-3}$	0.0

- **Infrequent Validation (Large V):** Conversely, when V is large (e.g., $V = 1000$), the computational overhead is reduced. The execution time drops to (0.56s), approaching the cost of Base SGD. However, the algorithm "flies blind" for 1000 steps at a time, increasing the likelihood of "missing" an optimal iterate. This results in the **highest final error** (1.11×10^{-2}).

This experiment confirms that the choice of V is a practical decision that balances the available computational budget against the required solution precision.

4.4 Impact of the early stopping patience criterion (s)

In our final experiment, we analyze the practical effect of the early stopping mechanism itself, governed by the 'patience' parameter, s, introduced in Section 2.4. This parameter defines the "Successive Increases" (UP_s) criterion and is critical for making the algorithm robust to the "ugly," oscillating validation curves described by Prechelt [16].

Experimental Setup

We test the full EB-SGD algorithm (Algorithm 1) using the same benchmark problem. Instead of a fixed budget, we allow the algorithm to run until the stopping criterion (Line 18) is met, using a validation frequency of $V = 100$. We test a broad range of patience values, from $s = 5$ (relatively impatient) to $s = 200$ (very patient). We measure the resulting trade-off between

1. **Computational Cost:** The mean wall-clock time (in seconds) the algorithm ran before stopping.
2. **Solution Quality:** The mean final error ($||x_{\text{best}} - x^*||_2$) of the solution returned upon termination.

Results and Analysis

The results, shown in Figure 4 and Table 3, demonstrate the essential role of patience in stochastic optimization.

Table 3: Impact of patience parameter (s) on computational cost and solution accuracy (Mean ± Std. Dev. over 1000 runs).

Patience (s)	**Mean Time (s)**	**Mean Final Error ($||x - x^*||_2$)**	**Mean Max Violation**
$s = 5$	0.12 ± 0.05	$2.44 \times 10^{-2} \pm 7.73 \times 10^{-3}$	0.0
$s = 10$	0.20 ± 0.09	$1.79 \times 10^{-2} \pm 5.79 \times 10^{-3}$	0.0
$s = 20$	0.36 ± 0.16	$1.32 \times 10^{-2} \pm 4.05 \times 10^{-3}$	0.0
$s = 50$	0.62 ± 0.17	$9.53 \times 10^{-3} \pm 2.28 \times 10^{-3}$	0.0
$s = 100$	0.65 ± 0.17	$9.23 \times 10^{-3} \pm 2.18 \times 10^{-3}$	0.0
$s = 200$	0.72 ± 0.25	$9.32 \times 10^{-3} \pm 2.20 \times 10^{-3}$	0.0

We observe a clear and direct trade-off:

- **Impatience (Small s):** With small patience values (e.g., $s = 5$), the algorithm stops very quickly (mean time ≈ 0.12s). However, this rapid termination often mistakes early

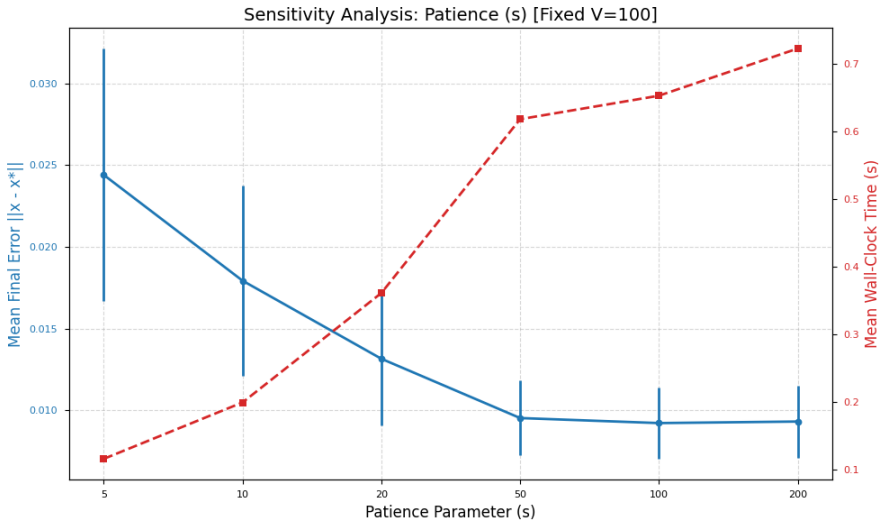


Figure 4: Sensitivity analysis for the patience parameter (s). The left axis (blue) shows the mean final error, while the right axis (red) shows the mean wall-clock time (computational cost). This illustrates the classic trade-off between computational budget and solution accuracy.

stochastic oscillations for permanent stagnation, resulting in a **higher final error** (2.44×10^{-2}).

- **Patience (Large s):** As s increases, the algorithm is forced to be more "patient," effectively filtering out the stochastic noise. By running longer (e.g., $s = 200$, mean time ≈ 0.72s), it waits out these oscillations and finds a **significantly more accurate solution** (9.32×10^{-3}).

This experiment validates the necessity of the s parameter. It provides a direct, tunable "knob" for the user to balance their desired solution quality against their available computational budget. Notably, while increasing patience from $s = 5$ to $s = 100$ yields substantial gains in accuracy, extending it further to $s = 200$ yields diminishing returns, suggesting that an optimal patience range exists for a given problem scale.

4.5 Sensitivity to penalty parameter P

To address the theoretical concerns regarding the exact penalty condition and justify the choice of $P = 10^6$ in the validation score (4), we conducted a comprehensive sensitivity analysis. We

evaluated the algorithm's performance with P varying across varying orders of magnitude, $P \in [10^3, 10^8]$, while maintaining all other hyperparameters fixed.

As illustrated in Figure 5, the EB-SGD algorithm demonstrates remarkable robustness to the magnitude of P. Both the mean final error and the constraint violation remained statistically invariant across the tested range. This empirical evidence suggests that because the underlying adaptive barrier mechanism naturally enforces feasibility, the specific value of P becomes less critical, provided it is sufficiently large to filter out infeasible outliers during the transient phase. Thus, $P = 10^6$ is a safe and effective choice.

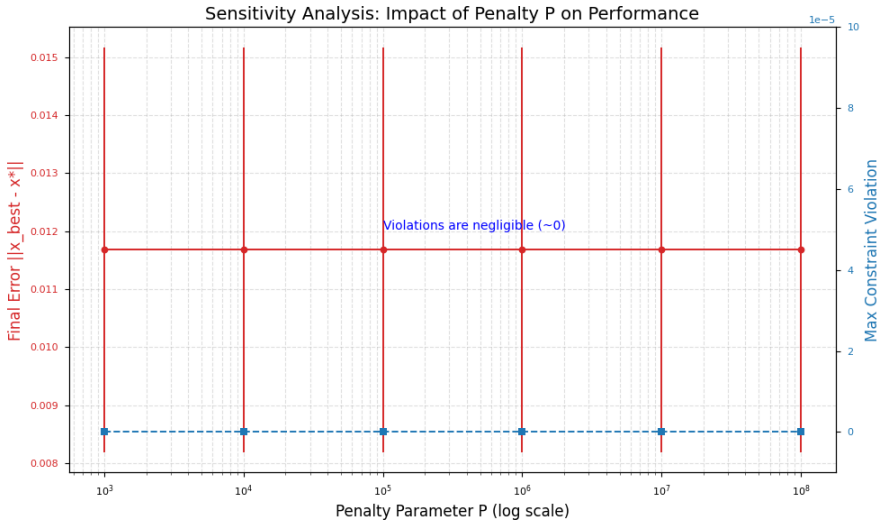


Figure 5: Sensitivity Analysis: The impact of varying the penalty parameter P on the final error (red) and maximum constraint violation (blue). The performance is consistently stable, confirming the robustness of the method to the penalty magnitude.

4.6 Ablation study: Choice of penalty metric

To validate the design choice of using the maximum constraint violation in the validation score ($S(x) = f(x) + P \cdot \max_j g_j(x)^+$), we conducted an ablation study comparing it against an alternative sum-based metric ($S_{alt}(x) = f(x) + P \cdot \sum_j g_j(x)^+$).

The results, visualized in Figure 6, demonstrate that both strategies achieve perfect feasibility with mean maximum violations converging to zero (≈ 0). Furthermore, both metrics converged to the identical objective value of 81.7412. This confirms that the algorithm's feasibility guarantee is primarily driven by the adaptive barrier mechanism rather than the specific aggregation of

the penalty in the validation step. However, we retain the *Max-Penalty* formulation as it aligns theoretically with the worst-case guarantee required for strict feasibility.

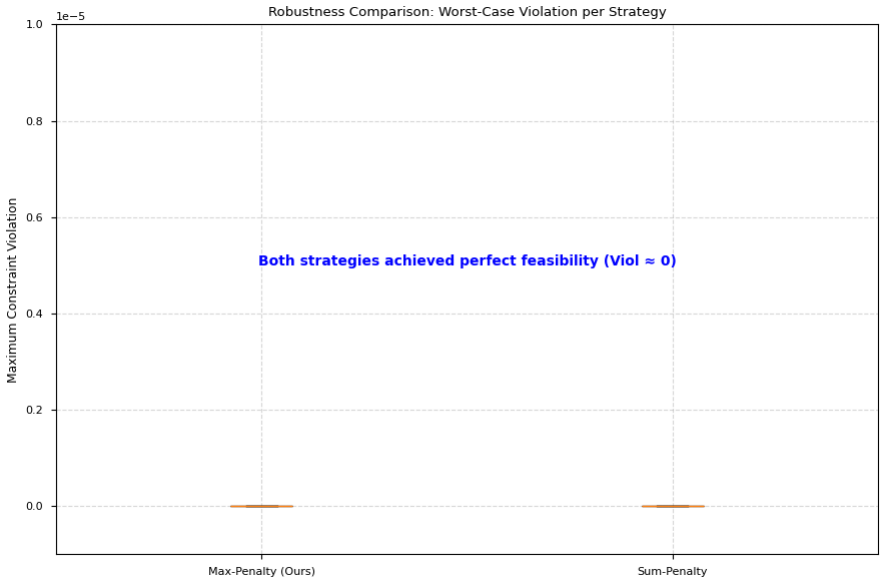


Figure 6: Robustness Comparison: Box plot showing the worst-case violations for Max-Penalty vs. Sum-Penalty strategies. Both methods consistently achieve zero violations, confirming the algorithm's inherent robustness.

5 Conclusion

In this paper, we introduced the **EB-SGD** algorithm, a practical and effective modification for stochastic constrained optimization. We began by identifying a fundamental limitation in baseline algorithms [7], which we term the **"last-iterate problem"**. We demonstrated that relying on the final, noise-ridden iterate is a "gamble", as real-world validation curves are "ugly" and oscillatory.

Our approach directly addresses this limitation by replacing reliance on the last iterate with a robust **"best-iterate tracking" mechanism**. Through "periodic validation" using a deterministic "Validation Score", and a "patient" (s) stopping criterion, our algorithm is able to "filter out stochastic noise" and reliably identify a superior solution.

We proved theoretically (Theorem 1) that the sequence of best iterates (x_{best}) returned by our algorithm **converges almost surely** to the unique, optimal solution. More importantly, our numerical experiments (Section 4) demonstrated the practical benefits of this approach:

- **Superior Accuracy:** EB-SGD returns a solution that is "significantly more accurate" (lower final error norm) than the baseline last-iterate method.

- **Robust Stability:** The most significant finding is the "significantly reduced variance". This confirms our tracking mechanism is "more stable and robust" to stochastic noise, filtering out the "random peaks" that plague the last-iterate method.

These findings confirm that the **"computational trade-off"**—exchanging cheap stochastic steps for "computationally expensive" validation checks —is highly beneficial, leading to a final solution that is not only more accurate but also certifiably feasible and "significantly more stable". Finally, our sensitivity analyses of the validation frequency (V) and patience (s) provide practical "knobs" for users to balance their required solution precision against their available computational budget. It is important to note that our experimental validation in this work was focused on convex optimization problems with affine constraints. While the proposed EB-SGD framework is theoretically grounded, extending its empirical validation to nonconvex or nonaffine settings remains a subject for future investigation.

Future work

The promising results of our EB-SGD algorithm open several interesting avenues for future research. Based on our findings, we suggest the following directions:

- **Adaptive and Autonomous Hyperparameter Tuning:** Developing a self-tuning mechanism for the validation frequency V and the patience parameter s to automatically detect the optimal balance between computational budget and solution quality, thereby reducing user dependence.

- **Variance-Based Validation Frequency:** Designing a heuristic to dynamically adjust the validation frequency V based on the observed variance in the deterministic Validation Score, ensuring frequent checks only when stochastic noise is high.

- **Trend Analysis for Patience (s):** Employing advanced statistical monitoring or time-series analysis techniques to robustly differentiate between minor stochastic oscillation and persistent performance stagnation, allowing for an automatic adjustment of s.

- **Scalable Stochastic Validation Schemes:** Investigating the replacement of the expensive full-batch validation step with a low-cost, statistically sound mini-batch validation

scheme, leveraging concentration inequalities to guarantee high-confidence early stopping decisions.

- **Extension to Nonaffine Constraints:** Investigating the performance, stability, and theoretical guarantees of the EB-SGD approach when applied to problems with general nonlinear (nonaffine) convex constraints, moving beyond the current focus on affine constraints.

- **Generalization to Nonconvex Optimization:** Exploring the empirical performance and potential convergence guarantees of the EB-SGD algorithm within the domain of nonconvex optimization, specifically for large-scale training of deep neural networks.

- **Tail-Averaging Variants:** Since our experiments demonstrated that standard Polyak–Ruppert averaging suffers from "memory lag" due to the shifting central path, future work should investigate "Tail-Averaging" (or windowed averaging) schemes. These methods average only the most recent iterates (e.g., the last suffix of the trajectory), potentially combining the variance-reduction benefits of averaging with the dynamic tracking capability required for adaptive barrier methods.

- **High-Dimensional Scalability:** While this study established the stability and robustness of EB-SGD on the standard benchmark ($d = 50$), investigating the algorithm's scalability to significantly higher-dimensional problems (e.g., $d \geq 100$) remains a critical direction. Future work will focus on analyzing the convergence rates and computational trade-offs in such high-dimensional regimes under varying constraint geometries.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest. article.

References

[1] Bertsekas, D.P. *Nonlinear programming*, 2nd ed., Athena Scientific, 1999.

[2] Billingsley, P. *Convergence of probability measures*, John Wiley & Sons, 1968.

[3] Bottou, L., Curtis, F.E. and Nocedal, J. *Optimization methods for large-scale machine learning,* SIAM Rev. 60(2) (2018), 223–311.

[4] Boyd, S. and Vandenberghe, L. *Convex optimization*, Cambridge University Press, 2004.

[5] Bubeck, S. *Convex optimization: Algorithms and complexity,* Found. Trends Mach. Learn. 8(3–4) (2015), 231–358.

[6] Dieuleveut, A., Durmus, A. and Bach, F. *Bridging the gap between constant step size stochastic gradient descent and Markov chains,* arXiv preprint arXiv:1707.06386 (2018).

[7] Dimitrieski, N., Cao, J. and Ebenbauer, C. *Stochastic gradient descent for constrained optimization based on adaptive relaxed barrier functions,* Int. J. Robust. Nonlinear Control. (2025).

[8] Feller, C. and Ebenbauer, C. *A stabilizing iteration scheme for model predictive control based on relaxed barrier functions,* Automatica 80 (2017), 328–339.

[9] Fiacco, A.V. and McCormick, G.P. *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley & Sons, 1968.

[10] Garrigos, G. and Gower, R.M. *Handbook of convergence theorems for (stochastic) gradient methods,* arXiv preprint arXiv:2301.11235 (2023).

[11] Gower, R.M., Loizou, N., Qian, X., Sailanbayev, A., Shulgin, E. and Richtárik, P. *SGD: General analysis and improved rates,* arXiv preprint arXiv:1901.09401 (2019).

[12] Li, M., Grigas, P. and Atamtürk, A. *New penalized stochastic gradient methods for linearly constrained strongly convex optimization,* SIAM J. Optim. 32(2) (2022), 859–887.

[13] Liu, J. and Yuan, Y. *On almost sure convergence rates of stochastic gradient methods*, In Proceedings of the 35th Annual Conference on Learning Theory, PMLR, 178 (2022), 1–21.

[14] Nocedal, J. and Wright, S.J. *Numerical optimization*, 2nd ed., Springer, 2006.

[15] Polyak, B.T. and Juditsky, A.B. *Acceleration of stochastic approximation by averaging,* SIAM J. Control Optim. 30(4) (1992), 838–855.

[16] Prechelt, L. Early stopping but when?, In: Orr, G.B., Müller, K.R. (eds.) *Neural networks: Tricks of the trade*, Lecture Notes in Computer Science, vol 1524, Springer, Berlin, Heidelberg, 1998, 55–69.

[17] Robbins, H. and Monro, S. *A stochastic approximation method,* Ann. Math. Stat. 22(3) (1951), 400–407.

[18] Rockafellar, R.T. *Convex analysis*, Princeton University Press, 1970.

[19] Roy, S.K. and Harandi, M. *Constrained stochastic gradient descent: The good practice*, In Int. Conf. Digit. Image Comput. Tech. Appl. (DICTA), IEEE, 2017, 1–8.

[20] Rudin, W. *Principles of mathematical analysis*, 3rd ed., McGraw-Hill, 1976.

[21] Wang, M. and Bertsekas, D.P. *Incremental constraint projection methods for variational inequalities,* Math. Program. 150 (2015), 321–363.

[22] Wang, X., Ma, S. and Yuan, Y. *Penalty methods with stochastic approximation for stochastic nonlinear programming,* arXiv preprint arXiv:1605.05609 (2016).

[23] Yan, Y. and Xu, Y. *Adaptive primal-dual stochastic gradient method for expectation-constrained convex stochastic programs,* Math. Program. Comput. 14(2) (2022), 319–363.

[24] Zhang, T. *Solving large scale linear prediction problems using stochastic gradient descent algorithms,* In Proc. 21st Int. Conf. Mach. Learn. (ICML), 2004, 116.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1319–1340
https://doi.org/10.22067/ijnao.2026.97725.1837
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A novel collocation-based numerical method for higher-order linear ODE systems with pharmacokinetic applications

F. Kaci*

Abstract

This paper introduces a collocation method founded on the independence polynomial of the path graph to solve systems of linear differential equations subject to mixed conditions. The approach converts the original differential system and its associated conditions into an equivalent matrix representation, resulting in a linear algebraic system whose unknowns are the independence coefficients. Several test problems are conducted to assess the numerical efficiency and accuracy of the method. As a practical application, the technique is applied to the three-compartment pharmacokinetic model, illustrating its robustness, numerical stability, and computational simplicity. The results demonstrate the proposed method's effectiveness and reliability in solving complex systems of differential equations.

AMS subject classifications (2020): 65L05; 65L10; 65N35; 05C31

Keywords: Systems of differential equations; Graph; Independence polynomial; Collocation method; Compartment pharmacokinetic model.

*Corresponding author

Received 6 February 2026; revised 28 May 2026; accepted 7 July 2026

Fatma Kaci
Department of Mathematics, Faculty of Exact Sciences, Mohamed Khider University, Biskra, Algeria. e-mail: fatma.kaci@univ-biskra.dz

How to cite this article
Kaci, F., A novel collocation-based numerical method for higher-order linear ODE systems with pharmacokinetic applications. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1319-1340. https://doi.org/10.22067/ijnao.2026.97725.1837

1 Introduction

The numerical solution of systems of ordinary differential equations (ODEs) plays an important role in many scientific and technical disciplines, including physics, engineering, biology, and economics [5, 9, 13, 17]. These systems often model complex phenomena involving several interdependent variables and higher-order dynamic relationships. However, their exact solution remains, in the majority of cases, inaccessible, which justifies the use of numerical methods.

Over the past few decades, several numerical approaches have been developed to solve systems of linear differential equations, particularly those of high order. Classical methods include generalized Runge–Kutta methods, finite difference schemes, and spectral methods, as well as wavelet and collocation techniques, which are widely used for their flexibility and accuracy. In particular, spectral, wavelet, and collocation methods rely on polynomial or piecewise polynomial approximations to represent the solution through suitable basis functions. For illustration, Akyüz and Sezer [1] and Ozturk [12] presented Chebyshev collocation methods for solving high-order linear problems, where the original equations are transformed into equivalent matrix forms. In a related study, Çetin, Sezer, and Güler [4] employed Lucas polynomials to solve high-order linear problems with variable coefficients, along with residual error estimation. More recently, Shiralashetti and Kumbinarasaiah [15] introduced a numerical algorithm based on an exact Parseval frame based on Laguerre wavelets. For a class of non-linear Lane–Emden-type equations, Krithikka and Hariharan [10] presented algorithms using Hermite wavelet and Fermat polynomials. Gümgüm, Özdek, and Özaltun [7] proposed a Legendre wavelet method for solving high-order nonlinear equations with variable and proportional delays, while Mall and Chakraverty [11] developed a single-layer Legendre neural network model to tackle nonlinear systems. Basit et al. [3] used a symmetric Bernstein approximation to solve a system of fourth-order nonlinear ODEs on any finite interval $[a, b]$. The technique is based on the operational matrices of Bernstein polynomials using Chebyshev collocation nodes. Although these approaches have shown good performance in several situations, some may encounter difficulties with high-order systems due to issues of numerical stability or implementation complexity.

In this paper, we employ a novel basis function to numerically solve systems of high-order linear differential equations. We propose a collocation method based on independence polynomials of an even path. This basis is derived from graph theory. Although it lacks orthogonality, which can sometimes lead to less favorable numerical conditioning, the practical results indicate that it is nonetheless effective for the problems addressed. A comparison with standard orthogonal polynomials is also provided to highlight the effectiveness and potential of the proposed approach.

A graph $G = (V(G), E(G))$ consists of a finite nonempty set $V(G)$ of n vertices together with a prescribed set $E(G)$ of m unordered pairs of distinct vertices of $V(G)$. Each pair $e = (u, v)$ of vertices in $E(G)$ is an edge of G. We say that the vertices are adjacent if they are joined by an edge. A subset A of $V(G)$ is an *independent set* of G if no two vertices of A are adjacent. The *independence polynomial* of a graph G is the polynomial

$$I(G, x) = \sum_{k=0}^{\alpha(G)} a_k x^k,$$

where a_k is the number of independent set of cardinality k of G and $\alpha(G)$ is the cardinality of the largest independent set of G.

A *path* of order n, denoted P_n, is a graph on n vertices $v_1, \ldots, v_n$ where v_i is adjacent to v_{i+1}, for all $i = 1, 2, \ldots, n-1$. The length of a path is the number of edges in it. We define $P_0 = (\emptyset, \emptyset)$. See Figure 1.

Arocha [2] deduced that the independence polynomial of a path P_n can be computed as follows:

$$I(P_n, x) = \sum_{j=0}^{\lfloor (n+1)/2 \rfloor} \binom{n+1-j}{j} x^j.$$

In particular,

$$I(P_0, x) = 1, \quad I(P_1, x) = 1 + x, \quad I(P_2, x) = 1 + 2x, \quad I(P_3, x) = 1 + 3x + x^2.$$



Figure 1: Paths P_1, P_2, and P_3.

Let us consider the following systems of ordinary differential equations with variable coefficients:

$$\sum_{n=0}^{m} \sum_{j=1}^{k} P_n^{i,j}(x)\, u_j^{(n)}(x) = f_i(x), \quad i = 1, 2, \ldots, k, \tag{1}$$

with the mixed conditions

$$\sum_{n=0}^{m-1} \left(\beta_n^{i,j}\, u_j^{(n)}(a) + \gamma_n^{i,j}\, u_j^{(n)}(b) \right) = \alpha_{i,j}, \quad i = 0, 1, \ldots, m-1, \quad j = 1, 2, \ldots, k. \tag{2}$$

Here, $j = 1, 2, \ldots, k$. The function $u_j^{(0)}(x) = u_j(x)$ is an unknown function and $u \in C^m([a, b])$, while $P_n^{i,j}(x)$ and $f_i(x)$ are known continuous functions defined on the interval $a \leq x \leq b$. The coefficients $\beta_n^{i,j}$, $\gamma_n^{i,j}$, and $\alpha_{n,i}$ are appropriate real constants.

2 Operational matrices for the independence polynomial

The purpose of this paper is to develop an approximate representation of the solution to (1) under conditions (2) by applying the independence polynomial of even paths in the form:

$$u_j(x) \simeq u_{j,N}(x) = \sum_{n=0}^{N} c_{j,n} I(P_{2n}, x), \quad j = 1, 2, \ldots, k, \tag{3}$$

where $c_{j,n}$, $n = 0, \ldots, N$, $j = 1, 2, \ldots, k$, are unknown independence coefficients.

For every $N \in \mathbb{N}$, the independence polynomial vector is given by

$$I_N(x) = [I(P_0, x) \;\; I(P_2, x) \ldots \;\; I(P_l, x)], \qquad N = \lceil \frac{l+1}{2} \rceil.$$

The vector $I_N(x)$ can be written in the matrix form:

$$I_N(x) = X(x)A, \tag{4}$$

where $X(x) = [1 \; x \ldots \; x^N]$ and A is an $(N+1) \times (N+1)$ matrix in which

$$A_{ij} = \begin{cases} \binom{2j-i}{i-1}, & i \leq j, \\ \\ 0, & i > j. \end{cases}$$

Consider the matrix form of the solution $u_{j,N}$:

$$u_{j,N}(x) = I_N(x) \; C_j, \quad C_j = [c_{j,0} \; c_{j,1} \; \ldots \; c_{j,N}]^T \tag{5}$$

and its derivatives $u_{j,N}^{(n)}$:

$$u_{j,N}^{(n)}(x) = I_N^{(n)}(x) C_j, \quad j = 1, \ldots, k, \quad \text{and } n = 1, 2, \ldots, m. \tag{6}$$

From (3), (5), and (6), we obtain

$$u_{j,N}(x) = X(x)\ A\ C_j,$$

$$u_{j,N}^{(n)}(x) = X^{(n)}(x)\ A\ C_j.$$

The matrix $X^{(n)}(x)$ can be found in terms of $X(x)$ as follows:

$$X^{(1)}(x) = X(x)\ L,$$

$$X^{(2)}(x) = X^{(1)}(x)\ L = X(x)\ L^2, \ldots$$

$$X^{(n)}(x) = X(x)L^n,$$

where L is the following matrix:

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \ldots & 0 \\ 0 & 0 & 2 & 0 & 0 \ldots & 0 \\ 0 & 0 & 0 & 3 & 0 \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ 0 \ldots & 0 \ldots & 0 & 0 & 0 & N \\ 0 \ldots & 0 & \ldots & 0 & 0 & 0 \end{bmatrix},$$

which gives

$$u_{j,N}^{(n)}(x) = X(x)\ L^n A\ C_j. \tag{7}$$

We also write the following Kronecker product:

$$\mathbf{u}^{(n)}(x) = \mathbf{X}(x)\ \mathbf{L}^n \mathbf{A}\ \mathbf{C}, \tag{8}$$

where

$$\mathbf{u}^{(n)}(x) = \begin{bmatrix} u_{1,N}^{(n)}(x) \\ u_{2,N}^{(n)}(x) \\ \vdots \\ u_{k,N}^{(n)}(x) \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix},$$

and

$$\mathbf{X}(x) = \begin{bmatrix} X(x) & 0 & \ldots & 0 \\ 0 & X(x) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X(x) \end{bmatrix}_{k \times k}, \quad \mathbf{L} = \begin{bmatrix} L & 0 & \ldots & 0 \\ 0 & L & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & L \end{bmatrix}_{k \times k},$$

$$\mathbf{A} = \begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A \end{bmatrix}_{k\times k}.$$

The system (1) can consequently be expressed in the matrix form:

$$\sum_{n=0}^{m} \mathbf{P}_n(x)\mathbf{u}^{(n)}(x) = \mathbf{f}(x), \tag{9}$$

where

$$\mathbf{P}_n(x) = \begin{bmatrix} P_n^{1,1}(x) & P_n^{1,2}(x) & \dots & P_n^{1,k}(x) \\ P_n^{2,1}(x) & P_n^{2,2}(x) & \dots & P_n^{2,k}(x) \\ \vdots & \vdots & \ddots & \vdots \\ P_n^{k,1}(x) & P_n^{k,2}(x) & \dots & P_n^{k,k}(x) \end{bmatrix} \quad \text{and} \quad \mathbf{f}(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{bmatrix}.$$

3 Independence polynomial method (IPM)

To construct the numerical scheme, we introduce the collocation points

$$x_i = a + \frac{b-a}{N}\, i, \quad i = 0, 1, \dots N \text{ and } 0 \le a \le x_i \le b < \infty.$$

Substituting these points into (9) transforms it into a system of matrix equations that forms the basis of our method:

$$\sum_{n=0}^{m} \mathbf{P}_n(x_i)\mathbf{u}^{(n)}(x_i) = \mathbf{f}(x_i), \qquad i = 0, 1, \dots, N. \tag{10}$$

This can be expressed with the following matrix equation:

$$\sum_{n=0}^{m} \mathbf{P}_n\mathbf{U}^{(n)} = \mathbf{F}, \tag{11}$$

where

$$\mathbf{P}_n = \begin{bmatrix} \mathbf{P}_n(x_0) & 0 & \dots & 0 \\ 0 & \mathbf{P}_n(x_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{P}_n(x_N) \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}(x_0) \\ \mathbf{f}(x_1) \\ \vdots \\ \mathbf{f}(x_N) \end{bmatrix}, \quad \text{and} \quad \mathbf{U}^{(n)} = \begin{bmatrix} \mathbf{u}^{(n)}(x_0) \\ \mathbf{u}^{(n)}(x_1) \\ \vdots \\ \mathbf{u}^{(n)}(x_N) \end{bmatrix}.$$

From (8), we have

$$\mathbf{u}^{(n)}(x_i) = \mathbf{X}(x_i)\mathbf{L}^n\mathbf{A}\mathbf{C},$$

and thus $\mathbf{U}^{(n)}$ can be written in compact form as follows:

$$\mathbf{U}^{(n)} = \tilde{\mathbf{X}}\mathbf{L}^n\mathbf{A}\mathbf{C}, \tag{12}$$

where

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}(x_0) \\ \mathbf{X}(x_1) \\ \vdots \\ \mathbf{X}(x_N) \end{bmatrix}$$

in which

$$\mathbf{X}(x_i) = \begin{bmatrix} X(x_i) & 0 & \dots & 0 \\ 0 & X(x_i) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X(x_i) \end{bmatrix}_{k\times k}, \qquad i = 0, 1, \dots, N.$$

Substituting the expression given in (12) into (11) leads to the fundamental matrix equation:

$$\left\{ \sum_{n=0}^{m} \mathbf{P}_n\tilde{\mathbf{X}}\mathbf{L}^n\mathbf{A} \right\} \mathbf{C} = \mathbf{F}. \tag{13}$$

The matrices $\mathbf{P}_n, \tilde{\mathbf{X}}$, $\mathbf{L}$, and $\mathbf{A}$ are square block matrices of dimension $k(N+1)$, whereas $\mathbf{C}$ and $\mathbf{F}$ are block vectors of size $k(N+1)$.

Equation (13) can be written as follows:

$$\mathbf{V}\mathbf{C} = \mathbf{F} \text{ or } [\mathbf{V}; \mathbf{F}], \tag{14}$$

where

$$\mathbf{V} = \sum_{n=0}^{m} \mathbf{P}_n\tilde{\mathbf{X}}\mathbf{L}^n\mathbf{A} = [v_{s,t}].$$

Equation (14) gives rise to a linear system containing $k(N+1)$ algebraic equations for the $k(N+1)$ unknown coefficients.

Imposition of conditions

Starting from (8) and the condition (2), the corresponding matrix representation is obtained as follows:

$$\sum_{n=0}^{m-1} \left[\beta_n X(a) + \gamma_n X(b)\right] \mathbf{L}^n \mathbf{A}\ \mathbf{C} = \tilde{\alpha}, \tag{15}$$

where

$$\beta_n = \begin{bmatrix} \beta_n^1 & 0 & \dots & 0 \\ 0 & \beta_n^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_n^k \end{bmatrix}_{k\times k} \quad \text{with} \quad \beta_n^j = \begin{bmatrix} \beta_n^{0,j} \\ \beta_n^{1,j} \\ \vdots \\ \beta_n^{m-1,j} \end{bmatrix}, \qquad j = 1, 2, \dots, k,$$

$$\gamma_n = \begin{bmatrix} \gamma_n^1 & 0 & \dots & 0 \\ 0 & \gamma_n^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \gamma_n^k \end{bmatrix}_{k\times k} \quad \text{with} \quad \gamma_n^j = \begin{bmatrix} \gamma_n^{0,j} \\ \gamma_n^{1,j} \\ \vdots \\ \gamma_n^{m-1,j} \end{bmatrix}, \qquad j = 1, 2, \dots, k,$$

$$\tilde{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}, \quad \text{with} \quad \alpha_j = \begin{bmatrix} \alpha_{0,j} \\ \alpha_{1,j} \\ \vdots \\ \alpha_{m-1,j} \end{bmatrix}, \qquad j = 1, 2, \dots, k.$$

The matrix form (15) can be expressed as

$$\mathbf{WC} = \tilde{\alpha} \text{ or } [\mathbf{W}; \boldsymbol{\alpha}], \tag{16}$$

where

$$\mathbf{W} = \sum_{n=0}^{m-1} \left[\beta_n X(a) + \gamma_n X(b)\right] \mathbf{L}^n \mathbf{A}.$$

Now, we will replace the rows of the augmented matrix $[\mathbf{W}; \boldsymbol{\alpha}]$ by the row mk of the augmented matrix $[\mathbf{V}; \mathbf{F}]$, to get the new augmented matrix

$$\mathbf{\mathbb{V}C} = \mathbb{F} \quad \text{or} \quad [\mathbb{V}; \mathbb{F}]. \tag{17}$$

If $rank\mathbb{V} = rank[\mathbb{V}; \mathbb{F}] = k(N+1)$, then the unknown coefficients vector $\mathbf{C}$ is determined by solving this linear system.

The determined coefficients $c_{j,0}, c_{j,1}, \ldots, c_{j,N}$, for $j = 1, 2, \ldots, k$, are substituted into (3). We get the independence solution:

$$u_{j,N}(x) = \sum_{n=0}^{N} c_{j,n} I(P_{2n}, x), \qquad j = 1, 2, \ldots, k. \tag{18}$$

Error analysis

In this section, we derive an error estimate for the proposed method. The analysis is based on the error bounds previously obtained in [8]. Before stating our main result, we recall the following interpolation theorem.

Theorem 1. [6] Let u be a function in $\mathcal{C}^{N+1}([a, b])$, and let P_N be a polynomial of degree $\leq N$ that interpolates the function u at $(N+1)$ distinct points $x_0, x_1, \ldots, x_N$ in $[a, b]$. Then to each $x \in [a, b]$ there exists a point $\xi_x \in [a, b]$ such that

$$u(x) - P_N(x) = \frac{u^{(N+1)}(\xi_x)}{(N+1)!} \prod_{i=0}^{N} (x - x_i).$$

Let $u = (u_1, u_2, \ldots, u_k)^T$ be the exact solution of (1) and let $P_{j,N}$ denote the interpolating polynomial of u_j for every $j = 1, 2, \ldots, k$. Set $P_N = \left(P_{1,N}, P_{2,N}, \ldots, P_{k,N}\right)^T$.

We can write for every $j = 1, \ldots, k$,

$$u_j(x) = P_{j,N}(x) + R_{j,N}(x),$$

where

$$R_{j,N}(x) = \frac{u_j^{(N+1)}(\xi_x^j)}{(N+1)!} \prod_{i=0}^{N} (x - x_i), \qquad \xi_x^j \in [a, b].$$

Now, let $u_N = \left(u_{1,N}, u_{2,N}, \ldots, u_{k,N}\right)^T$ be the independence solution vector of (1) given by (18); this means that u_N satisfies (2). Therefore u_N and P_N are solutions of $\mathbf{\mathbb{V}C} = \mathbb{F}$ and $\mathbf{\mathbb{V}C}^* = \mathbb{F} + \Delta(\mathbf{F})$, respectively, where $\Delta(\mathbf{F})$ is the $k(N+1)$ vector:

$$\Delta(\mathbf{F}) = \left(\Delta(\mathbf{f}_1), \Delta(\mathbf{f}_2), \ldots, \Delta(\mathbf{f}_k)\right)^T,$$

$$[\Delta(\mathbf{f}_j)]_i = -\sum_{n=0}^{m}\sum_{j=1}^{k} P_n^{i,j}(x_i)\, R_{j,N}^{(n)}(x_i).$$

The following theorem provides an upper bound for the error between the exact solution and the independence solution of the problem (1)–(2).

Theorem 2. Let u and u_N be the exact and independence solution vectors of (1)–(2), respectively, and let P_N be the interpolating polynomial vector of u. Let $\mathbf{V}$,$\mathbf{F}$, $\mathbf{C}$, $\mathbf{C}^*$, and $\Delta(\mathbf{F})$, be defined as above. If each u_j, for every $j = 1, 2, \ldots, k$, is sufficiently smooth, then

$$|u_j(x) - u_{j,N}(x)| \leq |R_{j,N}(x)| + \| X(x)A(C_j - C_j^*) \|_\infty .$$

Proof. We use the term $P_{j,N}$, for every $j = 1, 2, \ldots, k$, as follows:

$$\begin{aligned} |u_j(x) - u_{j,N}(x)| &= |(u_j(x) - P_{j,N}(x)) + (P_{j,N}(x) - u_{j,N}(x))| \\ &\leq |u_j(x) - P_{j,N}(x)| + |u_{j,N}(x) - P_{j,N}(x)| \\ &\leq |R_{j,N}(x)| + |u_{j,N}(x) - P_{j,N}(x)|. \end{aligned}$$

On the other hand, from (7) we have

$$\begin{aligned} |u_{j,N}(x) - P_{j,N}(x)| &= |X(x)AC_j - X(x)A\ C_j^*| \\ &\leq |X(x)A(C_j - C_j^*)| \\ &\leq \|X(x)A(C_j - C_j^*)\|_\infty. \end{aligned}$$

□

4 Numerical examples and discussion

In this paper, several test problems are solved numerically using the proposed method. The numerical results obtained are presented in order to highlight the diversity of the cases treated, in terms of orders, coefficients, types of equations and solutions. This variety illustrates both the generality and the computational efficiency of the proposed method for solving systems of linear ordinary differential equations. Note that all the numerical simulations related to the proposed method have been performed using MATLAB. The absolute errors $|e_{j,N}(x_i)|$ in tables are the values of $|u_j(x_i) - u_{j,N}(x_i)|$ at selected points x_i.

Example 1. Consider the second-order linear differential equation system:

$$\begin{cases} u_1''(x) + xu_1(x) + xu_2(x) = 2, \\ \\ u_2''(x) + 2xu_2(x) + 2xu_1(x) = -2, \end{cases} \qquad 0 \le x \le 1, \tag{19}$$

with the boundary conditions $u_1(0) = 0,\ u_1'(0) = -1,\ u_2(1) = 0$, and $u_2'(1) = -1$. Here, we have

$$\mathbf{P}_0(x) = \begin{bmatrix} x & x \\ 2x & 2x \end{bmatrix}, \quad \mathbf{P}_1(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{P}_2(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{f}(x) = \begin{bmatrix} 2 \\ -2 \end{bmatrix}.$$

The set of the collocation points for $N = 2$ is

$$\{x_0 = 0, x_1 = \frac{1}{2}, x_2 = 1\}.$$

Hence, the fundamental matrix equation of the problem from (13) is written as

$$\left\{ \mathbf{P}_0\tilde{\mathbf{X}}\mathbf{L}^0\mathbf{A} + \mathbf{P}_1\tilde{\mathbf{X}}\mathbf{L}^1\mathbf{A} + \mathbf{P}_2\tilde{\mathbf{X}}\mathbf{L}^2\mathbf{A} \right\} \mathbf{C} = \mathbf{F},$$

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{P}_0(0) & 0 & 0 \\ 0 & \mathbf{P}_0(\frac{1}{2}) & 0 \\ 0 & 0 & \mathbf{P}_0(1) \end{bmatrix}, \quad \mathbf{P}_1 = \begin{bmatrix} \mathbf{P}_1(0) & 0 & 0 \\ 0 & \mathbf{P}_1(\frac{1}{2}) & 0 \\ 0 & 0 & \mathbf{P}_1(1) \end{bmatrix},$$

$$\mathbf{P}_2 = \begin{bmatrix} \mathbf{P}_2(0) & 0 & 0 \\ 0 & \mathbf{P}_2(\frac{1}{2}) & 0 \\ 0 & 0 & \mathbf{P}_2(1) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{f}(0) \\ \mathbf{f}(\frac{1}{2}) \\ \mathbf{f}(1) \end{bmatrix}, \quad \tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}(0) \\ \mathbf{X}(\frac{1}{2}) \\ \mathbf{X}(1) \end{bmatrix},$$

$$\mathbf{X}(0) = \begin{bmatrix} X(0) & 0 \\ 0 & X(0) \end{bmatrix}, \quad \mathbf{X}(\frac{1}{2}) = \begin{bmatrix} X(\frac{1}{2}) & 0 \\ 0 & X(\frac{1}{2}) \end{bmatrix},$$

$$\mathbf{X}(1) = \begin{bmatrix} X(1) & 0 \\ 0 & X(1) \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} L & 0 \\ 0 & L \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \text{ where } C_1 = \begin{bmatrix} c_{1,0} \\ c_{1,1} \\ c_{1,2} \end{bmatrix} \text{ and } C_2 = \begin{bmatrix} c_{2,0} \\ c_{2,1} \\ c_{2,2} \end{bmatrix}.$$

Thus, the augmented matrix is calculated as

$$[\mathbf{V};\mathbf{F}] = \begin{bmatrix} 0 & 0 & 6 & 0 & 0 & 0 & ; & 2 \\ 0 & 0 & 0 & 0 & 0 & 6 & ; & -2 \\ \frac{1}{2} & 1 & 63/8 & \frac{1}{2} & 1 & 15/8 & ; & 2 \\ 1 & 2 & 15/4 & 1 & 2 & 39/4 & ; & -2 \\ 1 & 3 & 14 & 1 & 3 & 8 & ; & 2 \\ 2 & 6 & 16 & 2 & 6 & 22 & ; & -2 \end{bmatrix}.$$

On the other hand, from (16), the matrix form of conditions is the following one:

$$[\mathbf{W};\alpha] = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & ; & 0 \\ 0 & 2 & 4 & 0 & 0 & 0 & ; & -1 \\ 0 & 0 & 0 & 1 & 3 & 8 & ; & 0 \\ 0 & 0 & 0 & 0 & 2 & 10 & ; & -1 \end{bmatrix}.$$

The new augmented matrix based on the conditions is given as

$$[\bar{\mathbf{V}};\bar{\mathbf{F}}] = \begin{bmatrix} 0 & 0 & 6 & 0 & 0 & 0 & ; & 2 \\ 0 & 0 & 0 & 0 & 0 & 6 & ; & -2 \\ 1 & 1 & 1 & 0 & 0 & 0 & ; & 0 \\ 0 & 2 & 4 & 0 & 0 & 0 & ; & -1 \\ 0 & 0 & 0 & 1 & 3 & 8 & ; & 0 \\ 0 & 0 & 0 & 0 & 2 & 10 & ; & -1 \end{bmatrix}.$$

Consequently, the system is solved, yielding the following independence coefficient vector:

$$\mathbf{C} = (\bar{\mathbf{V}})^{-1}\bar{\mathbf{F}} = \left[\tfrac{5}{6} \ \tfrac{-7}{6} \ \tfrac{1}{3} \ \tfrac{-5}{6} \ \tfrac{7}{6} \ \tfrac{-1}{3}\right]^T.$$

Finally, the vector C is substituted into (5), then

$$u_{1,N}(x) = I_N(x)\ C_1 \quad \text{and} \quad u_{2,N}(x) = I_N(x)\ C_2,$$

where

$$C_1 = \left[\tfrac{5}{6} \ \tfrac{-7}{6} \ \tfrac{1}{3}\right]^T \quad \text{and} \quad C_2 = \left[\tfrac{-5}{6} \ \tfrac{7}{6} \ \tfrac{-1}{3}\right]^T.$$

We obtain the approximate solutions as follows:

$$u_{1,N}(x) = x^2 - x \quad \text{and} \quad u_{2,N}(x) = -x^2 + x,$$

which are the exact solutions of the system.

Example 2. Consider the fourth-order linear differential equation system [18]:

$$\begin{cases} u_1^{(4)}(x) - \cos(x)u_2^{(2)}(x) + xu_3^{(1)}(x) - u_1(x) = xe^x + \cos^2(x), \\ u_2^{(4)}(x) + \sin(x)u_1^{(3)}(x) + \cos(x)u_1(x) - \cos(x)u_3(x) = (1-e^x)\cos(x), \\ e^{(-x)}u_3^{(4)}(x) + u_2^{(2)}(x) - \cos(x)u_1^{(1)}(x) + u_2(x) = \sin^2(x), \\ 0 \le x \le 1, \end{cases} \tag{20}$$

with the initial conditions $u_1(0) = 0, u_1^{(1)}(0) = 1, u_1^{(2)}(0) = 0, u_1^{(3)}(0) = -1, u_2(0) = 1, u_2^{(1)}(0) = 0, u_2^{(2)}(0) = -1, u_2^{(3)}(0) = 0, u_3(0) = u_3^{(1)}(0) = u_3^{(2)}(0) = u_3^{(3)}(0) = 1$.

The exact solution of the problem is given by $u_1(x) = \sin(x)$, $u_2(x) = \cos(x)$ and $u_3(x) = e^x$.

As observed in Table 1, the accuracy of the computed solution tends to decrease with increasing N. The same table further compares the performance of the proposed method with that of the exponential approximation method reported in [18], used to solve problem (20).

Example 3. Consider the following system of linear differential equations with variable coefficients [12, 15]:

$$\begin{cases} u_1^{(2)}(x) - u_2^{(1)}(x) + u_3^{(1)}(x) - e^{-x}u_1(x) + e^x u_3(x) = e^x - 2e^{-x} + xe^{-x}, \\ u_2^{(2)}(x) + u_1^{(1)}(x) - u_3^{(1)}(x) - u_1(x) = xe^{-x} - e^{-x}, \qquad 0 \le x \le 1, \\ u_3^{(2)}(x) - u_1^{(1)}(x) + u_2^{(1)}(x) + e^x u_2(x) + u_1(x) = 2e^{-x} - xe^{-x} + x, \end{cases} \tag{21}$$

with initial conditions $u_1(0) = 1, u_2(0) = 0, u_3(0) = 1, u_1'(0) = 1, u_2'(0) = 1, u_3'(0) = -1$ and the exact solution $u_1(x) = e^x, u_2(x) = xe^{-x}$ and $u_3(x) = e^{-x}$.

Table 2 presents a comparison of the absolute errors obtained by the proposed method and the Chebyshev operational method [12]. It can be seen that the current method provides a slightly more accurate approximation with slightly lower errors.

In terms of numerical performance, our method also stands out for its significantly reduced computational time. Indeed, it requires approximately 1 second to produce the results, compared to approximately 7 seconds for the method in [12], as indicated in [15]. This difference highlights

Table 1: Absolute errors for the IPM method and the method used in [18] for Example 2.

x_i	$\lvert e_{1,5}(x_i)\rvert$	$\lvert e_{1,5}(x_i)\rvert$ [18]	$\lvert e_{1,10}(x_i)\rvert$	$\lvert e_{1,10}(x_i)\rvert$ [18]
0	2.8623e-17	5.9999e-12	4.4938e-11	1.0221e-9
0.2	1.6081e-08	3.3672e-5	4.3323e-11	3.6854e-6
0.4	2.7146e-07	5.2758e-4	2.3235e-11	3.8405e-5
0.6	1.0021e-06	1.6690e-3	2.0988e-10	1.4109e-4
0.8	2.2176e-05	1.6112e-3	5.9634e-10	3.5199e-4
1	1.3750e-04	3.8807e-3	1.4827e-09	7.6995e-4
x_i	$\lvert e_{2,5}(x_i)\rvert$	$\lvert e_{2,5}(x_i)\rvert$ [18]	$\lvert e_{2,10}(x_i)\rvert$	$\lvert e_{2,10}(x_i)\rvert$ [18]
0	2.2204e-16	3.3999e-11	1.3980e-10	1.8999e-10
0.2	1.7676e-07	5.3148e-5	2.2176e-10	6.7587e-7
0.4	2.8260e-06	8.1712e-4	3.5811e-10	6.7557e-6
0.6	1.5149e-07	2.6665e-3	5.7150e-10	2.3055e-5
0.8	8.8001e-05	3.5885e-3	8.8584e-10	5.0852e-5
1	5.3442e-04	4.1086e-4	1.3963e-09	8.9223e-5
x_i	$\lvert e_{3,5}(x_i)\rvert$	$\lvert e_{3,5}(x_i)\rvert$ [18]	$\lvert e_{3,10}(x_i)\rvert$	$\lvert e_{3,10}(x_i)\rvert$ [18]
0	0	4.4409e-16	2.3906e-11	9.7000e-9
0.2	1.9495e-07	2.4145e-4	7.4221e-11	3.1086e-5
0.4	3.1353e-06	3.7515e-3	1.6324e-10	3.2342e-4
0.6	1.1942e-06	1.1835e-2	2.9924e-10	1.1828e-3
0.8	1.1694e-04	1.1622e-2	4.8362e-10	2.9217e-3
1	7.2002e-04	2.6565e-2	4.1498e-10	6.3328e-3

the computational efficiency of the proposed method, which manages to improve accuracy while significantly reducing computational costs. These results confirm the value of the developed method, both in terms of accuracy and numerical efficiency.

Example 4. Consider the system of fourth-order linear differential equations with variable coefficients, subject to mixed conditions:

$$\begin{cases} u_1^{(4)}(x) + u_2^{(2)}(x) = xe^x - 4e^{-x} - \sin(x), \\ \qquad\qquad\qquad\qquad\qquad\qquad 0 \le x \le 2, \\ u_2^{(4)}(x) + u_1^{(2)}(x) = \sin(x) + xe^{-x} - 2e^{-x}, \end{cases} \tag{22}$$

with

$$u_1(0) + u_1(2) = 2e^{-2}, \quad u_1^{(1)}(0) + u_1^{(2)}(2) = 1 - e^{-2}, \quad u_1^{(2)}(0) = -2,$$

$$u_1^{(3)}(0) = 3, \quad u_2(0) = 0, \quad u_2^{(1)}(0) = 1, \quad u_2^{(2)}(0) = 0, \quad u_2^{(3)}(0) = -1,$$

and the exact solutions $u_1(x) = xe^{-x},\ u_2(x) = \sin(x)$.

Table 2: Comparison of absolute errors and CPU Times for Example 3

	$\lvert e_{1,7}\rvert$		$\lvert e_{2,7}\rvert$			$\lvert e_{3,7}\rvert$	
x_i	[12] CPU time: 7.04 s	IPM CPU time: 0.08 s	[12] CPU time: 7.04 s	IPM CPU time: 0.08 s	x_i	[12] CPU time: 7.04 s	IPM CPU time: 0.08 s
0	2.00e-12	1.02e-12	4.59e-13	1.22e-12	0	0	4.45e-13
0.2	1.61e-07	1.35e-09	1.00e-07	1.33e-08	0.2	2.51e-08	3.61e-09
0.4	1.52e-06	1.47e-09	1.72e-07	2.69e-08	0.4	6.90e-07	1.12e-08
0.6	1.58e-06	9.59e-09	1.08e0-7	3.77e-08	0.6	9.87e-07	2.20e-08
0.8	1.93e-07	2.44e-08	1.70e-07	4.41e-08	0.8	2.25e-07	3.63e-08
1	1.86e-08	4.98e-07	1.01e-08	1.25e-06	1	2.21e-10	7.18e-08

Table 3: Absolute errors obtained using IPM for Example 4

x_i	$\lvert e_{1,8}(x_i)\rvert$	$\lvert e_{2,8}(x_i)\rvert$	$\lvert e_{1,10}(x_i)\rvert$	$\lvert e_{2,10}(x_i)\rvert$
0	7.7201e-04	2.7756e-15	8.9828e-06	1.4778e-12
0.2	5.8858e-04	1.0296e-09	6.9495e-06	1.3847e-11
0.4	4.0525e-04	1.7632e-08	4.9170e-06	1.5365e-10
0.6	2.2220e-04	7.1867e-08	2.8867e-06	6.3313e-10
0.8	3.9576e-05	1.7061e-07	8.5960e-07	1.7425e-09
1	1.4246e-04	3.0795e-07	1.1629e-06	3.9044e-09
1.2	3.2380e-04	4.4658e-07	3.1796e-06	7.7289e-09
1.4	5.0359e-04	6.4648e-07	5.1891e-06	1.4047e-08
1.6	6.7437e-04	2.1068e-06	7.1729e-06	2.6467e-08
1.8	8.0147e-04	1.0797e-05	8.9045e-06	7.9137e-08
2	7.7201e-04	4.6495e-05	8.9827e-06	3.7759e-07

As reported in Table 3, the errors obtained by IPM for $N = 8$ and $N = 10$ are of order 10^{-k} over the domain, indicating that the method provides a good approximation of the exact solution. The decrease of the error when passing from $N = 8$ to $N = 10$ illustrates the expected improvement in accuracy with larger truncation parameters.

Example 5. Consider the initial value problem:

$$\begin{cases} u_1'(x) + u_2'(x) + u_2(x) = x - e^{-x}, \\ \\ u_1'(x) + 4u_2'(x) + u_1(x) = 1 + 2e^{-x}, \end{cases} \qquad 0 \le x \le 1, \tag{23}$$

with $u_1(0) = 1$, $u_2(0) = 0$, and exact solutions

Table 4: Comparison of absolute errors for Example 5

x_i	$\lvert e_{1,5}(x_i)\rvert$ [1]	$\lvert e_{1,5}(x_i)\rvert$ [14]	$\lvert e_{1,5}(x_i)\rvert$ IPM
0.1	4.510522e-05	6.2951e-05	5.9187e-07
0.2	7.985043e-05	1.5714e-04	1.0110e-06
0.5	9.719089e-05	4.3862e-04	1.0978e-06
0.8	8.006002e-05	8.0076e-04	9.0094e-07
1	1.067677e-04	1.0598e-03	1.3659e-05
x_i	$\lvert e_{2,5}(x_i)\rvert$ [1]	$\lvert e_{2,5}(x_i)\rvert$ [14]	$\lvert e_{2,5}(x_i)\rvert$ IPM
0.1	2.247723e-05	4.5250e-04	2.9327e-07
0.2	3.984701e-05	6.5580e-04	5.0134e-07
0.5	4.890662e-05	6.7891e-04	5.4596e-07
0.8	4.064222e-05	7.8896e-04	4.5116e-07
1	5.390356e-05	1.2735e-03	6.7555e-06

$$u_1(x) = e^{-x} + 3\ e^{x/3} - 3, \quad u_2(x) = -\frac{1}{2}e^{-x} + \frac{3}{2}\ e^{-x/3} + x - 1.$$

Table 4 illustrates a comparison of the absolute errors obtained by the presented method with those by the Chebyshev collocation method [1] and the rational Chebyshev collocation method [14] for $N = 5$. The numerical results show that the proposed method provides a noticeable improvement compared to the other methods considered.

Example 6. We consider the sixth-order boundary-value ordinary differential equation: [16]

$$u^{(6)}(x) + u(x) = 6(2x\cos(x) + 5\sin(x)), \qquad -1 \le x \le 1,$$

with $u(-1) = u(1) = 0$, $u'(-1) = u'(1) = 2\sin(1)$ and $u^{(2)}(-1) = -u^{(2)}(1) = -4\cos(1) - 2\sin(1)$. The exact solution is $u(x) = (x^2 - 1)\sin(x)$.

We define the maximum and mean absolute errors, $E_{\max}$ and E_{mean}, respectively, by

$$E_{\max} = \max_{1\le i\le M} |e_N(x_i)|, \qquad E_{\text{mean}} = \frac{1}{M}\sum_{i=1}^{M} |e_N(x_i)|,$$

where M denotes the number of spatial grid points. Table 5 lists the maximum absolute error and the mean error obtained by IPM for various choices of N. We observe that, for each value of N, the maximum absolute error and the mean error are of the same order of magnitude. This indicates that the numerical error is distributed rather uniformly over the spatial grid, with no significant local peaks.

Table 5: $E_{\max}$ and E_{mean} for Example 6

N	$E_{\max}$	E_{mean}
9	1.2227e-03	5.5342e-04
10	3.6091e-04	1.6416e-04
11	2.6185e-05	1.1938e-05
12	6.4901e-06	2.9727e-06
13	2.8248e-07	1.3046e-07
14	1.8677e-06	6.1177e-07

5 Three-compartment pharmacokinetic models

Consider the three-compartment pharmacokinetic model, commonly used to model the kinetics of a drug within the body. This model divides the body into three interconnected compartments: the central compartment $C1$, the site of drug elimination, and two peripheral compartments $C2$ and $C3$, representing fast and slow exchange tissues, respectively. Drug transfers between compartments are modeled by first-order transfer constants: k_{12}, k_{21} between C_1 and C_2 and k_{13}, k_{31} between C_1 and C_3. The elimination occurs from the central compartment with rate constant k_{10}. The system is governed by the following set of ordinary differential equations:

$$\begin{cases} \dfrac{dC_1}{dt} = -(k_{10} + k_{12} + k_{13})C_1 + k_{21}C_2 + k_{31}C_3, \\ \dfrac{dC_2}{dt} = k_{12}C_1 - k_{21}C_2, \\ \dfrac{dC_3}{dt} = k_{13}C_1 - k_{31}C_3. \end{cases} \tag{24}$$

The analytical solution of this system is, in general, difficult to obtain due to the coupling of the equations and the presence of several parameters. Therefore, the use of numerical methods is essential to effectively approximate the concentrations in the different compartments over time.

We provide the numerical solution of the system (24) for $t > 0$ using IPM with the initial conditions: $C_1(0) = 5.0000\mu g \cdot kg^{-1}$ (Dose), $C_2(0) = 0$ and $C_3(0) = 0$. The parameter values for the model are $k_{12} = 0.105\ min^{-1}$, $k_{21} = 0.064\ min^{-1}$, $k_{13} = 0.022\ min^{-1}$, $k_{31} = 0.0034\ min^{-1}$, $k_{10} = 0.0827\ min^{-1}$.

Figure 2 shows the concentration of the drug in each compartment obtained using the proposed IPM method for $N = 10$ and the standard `ode23` solver, illustrating a close agreement between both approaches.

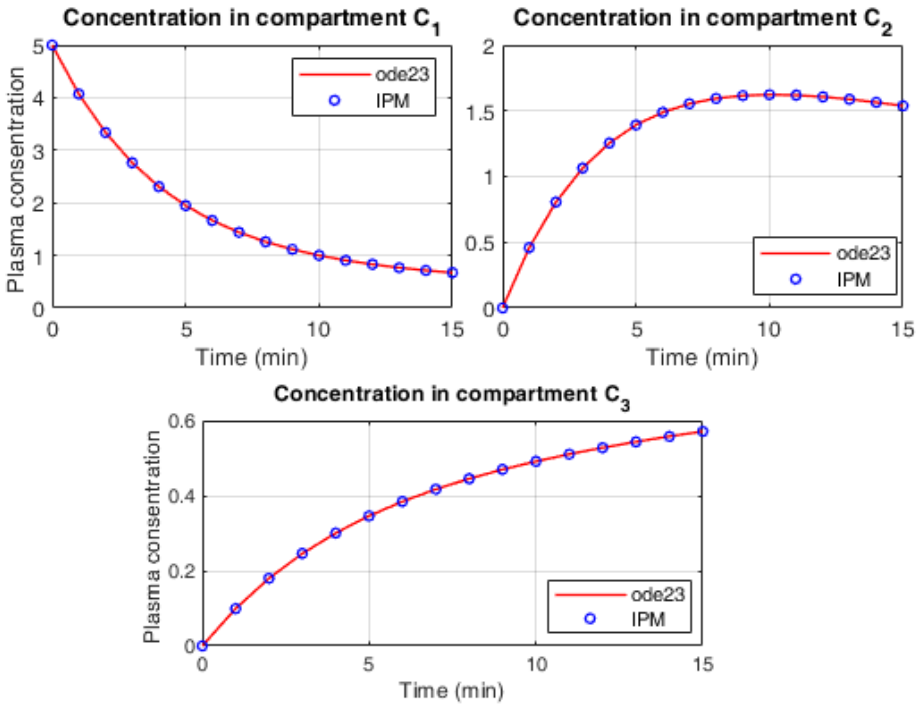


Figure 2: Concentration in the three compartments for $N = 10$

To measure the performance of the proposed approach, we rely on the estimated error function $\tilde{E}_N(x)$, which can also be used when no exact solution is available. If $C_{1,N}(x)$, $C_{2,N}(x)$, and $C_{3,N}(x)$ serve as approximations to the solution of system (24), then inserting these functions and their derivatives into (24) should lead to relations that are satisfied only approximately. For $x_k \in [a, b]$, the associated error terms $\tilde{E}_{i,N}(x_k),\ i = 1, 2, 3$ are defined by

$$\begin{aligned}\tilde{E}_{1,N}(x_k) =& \Big| C'_{1,N}(x_k) + (k_{10} + k_{12} + k_{13})\, C_{1,N}(x_k) - k_{21} C_{2,N}(x_k) \\ & - k_{31} C_{3,N}(x_k) \Big| \approx 0,\end{aligned}$$

$$\tilde{E}_{2,N}(x_k) = \left| C'_{2,N}(x_k) - k_{12} C_{1,N}(x_k) + k_{21} C_{2,N}(x_k) \right| \cong 0,$$

$$\tilde{E}_{3,N}(x_k) = \left| C'_{3,N}(x_k) - k_{13} C_{1,N}(x_k) + k_{31} C_{3,N}(x_k) \right| \cong 0,$$

and $\tilde{E}_{i,N}(x) < 10^{-s_k}$, $i = 1, 2, 3$, is satisfied for a positive constant s_k. If $\max 10^{-s_k} = 10^{-s}$ is specified, then the truncation limit N is increased until $\tilde{E}_{i,N}(x_k)$, $i = 1, 2, 3$, at all points, fall below the prescribed 10^{-s}.

The estimated error functions for $N = 6, 8, 10$ are given in Tables 6, 7, and 8. The graphics of estimated the error functions for $N = 10$ are presented in Figure 3.

Table 6: Estimated error function for $N = 6, 8, 10$ for C_1

x_i	$\tilde{E}_{1,6}$	$\tilde{E}_{1,8}$	$\tilde{E}_{1,10}$
1	3.1036e-04	5.6536e-06	5.6669e-08
3	6.9027e-05	1.1956e-06	5.1010e-12
5	3.4417e-15	5.0910e-07	2.7003e-09
7	3.9963e-05	3.7212e-07	2.2108e-09
9	9.4458e-05	4.4541e-07	4.0115e-10
11	2.2646e-04	8.8744e-07	1.0229e-08
13	6.9269e-04	3.2310e-06	1.1929e-07
15	1.3516e-02	4.6249e-04	9.9699e-06

Table 7: Estimated error function for $N = 6, 8, 10$ for C_2

x_i	$\tilde{E}_{2,6}$	$\tilde{E}_{2,8}$	$\tilde{E}_{2,10}$
1	1.7830e-04	3.2479e-06	3.2555e-08
3	3.9654e-05	6.8684e-07	3.2266e-13
5	1.2670e-14	2.9247e-07	1.5697e-09
7	2.2958e-05	2.1377e-07	1.1885e-09
9	5.4264e-05	2.5588e-07	3.8696e-11
11	1.3009e-04	5.0981e-07	6.6038e-09
13	3.9794e-04	1.8561e-06	6.6827e-08
15	7.7644e-03	2.6569e-04	5.7311e-06

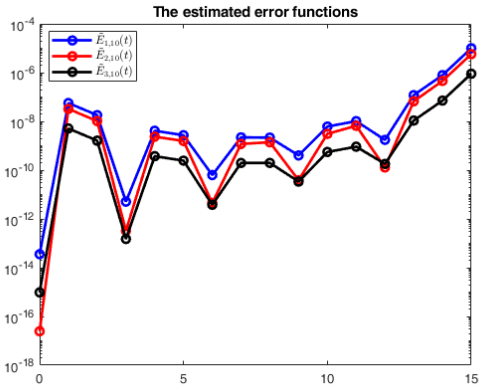


Figure 3: Estimated error functions for $N = 10$

Table 8: Estimated error function for $N = 6, 8, 10$ for C_3

x_i	$\tilde{E}_{3,6}$	$\tilde{E}_{3,8}$	$\tilde{E}_{3,10}$
1	2.8055e-05	5.1107e-07	5.1226e-09
3	6.2397e-06	1.0808e-07	1.5015e-13
5	3.5384e-15	4.6021e-08	2.4530e-10
7	3.6125e-06	3.3638e-08	1.9725e-10
9	8.5386e-06	4.0263e-08	3.3955e-11
11	2.0471e-05	8.0220e-08	9.1691e-10
13	6.2616e-05	2.9207e-07	1.0833e-08
15	1.2218e-03	4.1807e-05	9.0107e-07

6 Conclusion

In this work, a collocation method based on the independence polynomial of the even path has been developed for solving systems of higher-order linear differential equations with mixed conditions. An error analysis was carried out to assess the accuracy of the proposed method, and upper bounds for the approximation error were derived. Furthermore, numerical experiments conducted on various benchmark systems demonstrate the accuracy and efficiency of the proposed method, which often yields superior results compared to existing approaches. Moreover, it was observed that as the parameter N increases, the approximation error decreases, confirming the convergence of the method. Finally, the application of the proposed approach to the three-compartment pharmacokinetic model further validated its robustness and reliability, as the estimated error functions remained very small, indicating excellent numerical performance.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Akyüz, A. and Sezer, M. *Chebyshev polynomial solutions of systems of high-order linear differential equations with variable coefficients*, Appl. Math. Comput., 144(2) (2003), 237–247.

[2] Arocha, J.L. *Propiedades del polinomio independiente de un grafo*, Rev. Cienc. Mat., 5 (1984), 103–110.

[3] Basit, M., Shahnaz, K., Malik, R., Karim, S.A.A. and Khan, F. *An effective approach based on generalized Bernstein basis functions for the system of fourth-order initial value problems for an arbitrary interval*, Mathematics, 11(14) (2023), 3076.

[4] Çetin, M., Sezer, M. and Güler, C. *Lucas polynomial approach for system of high-order linear differential equations and residual error estimation*, Math. Probl. Eng., 2015 (2015), 625984.

[5] Cieślik, S. *Mathematical modeling of the dynamics of linear electrical systems with parallel calculations*, Energies, 14(10) (2021), 2930.

[6] Davis, P.J. and Rabinowitz, P. *Methods of numerical integration*, Academic Press, 2nd ed., 1984.

[7] Gümgüm, S., Özdek, D. and Özaltun, G. *Legendre wavelet solution of high-order nonlinear ordinary delay differential equations*, Turkish J. Math., 43(3) (2019), 1339–1352.

[8] Işık, O.R., Sezer, M. and Güney, M. *A rational approximation based on Bernstein polynomials for high order initial and boundary value problems*, Appl. Math. Comput., 217 (2011), 9438–9450.

[9] Johnston, M.D., Edwards, C.M., Bodmer, W.F., Maini, P.K. and Chapman, S.J. *Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer*, Proc. Natl. Acad. Sci. U.S.A., 104(10) (2007), 4008–4013.

[10] Krithikka, S. and Hariharan, G. *Two robust operational matrix algorithms for solving nonlinear Lane–Emden-type equations in astrophysics using Hermite and Fermat polynomials*, NRIAG J. Astron. Geophys., 14(1) (2025), 1–15.

[11] Mall, S. and Chakraverty, S. *Application of Legendre neural network for solving ordinary differential equations*, Appl. Soft Comput., 43 (2016), 347–356.

[12] Ozturk, Y. *Numerical solution of systems of differential equations using operational matrix method with Chebyshev polynomials*, J. Taibah Univ. Sci., 12(2) (2018), 155–162.

[13] Rafikov, M. and Limeira, E.D.H. *Mathematical modelling of the biological pest control of the sugarcane borer*, Int. J. Comput. Math., 89(3) (2012), 390–401.

[14] Ramadan, M.A., Raslan, K.R. and Nassar, M.A. *An approximate solution of systems of high-order linear differential equations with variable coefficients by means of a rational Chebyshev collocation method*, Electron. J. Math. Anal. Appl., 4(1) (2016), 86–98.

[15] Shiralashetti, S.C. and Kumbinarasaiah, S. *Laguerre wavelets exact Parseval frame-based numerical method for the solution of system of differential equations*, Int. J. Appl. Comput. Math., 6(4) (2020), 1–16.

[16] Siddiqi, S.S. and Twizell, E.H. *Spline solutions of linear sixth order boundary value problems*, Int. J. Comput. Math., 60 (1996), 295–304.

[17] Tzafriri, A.R. *Michaelis–Menten kinetics at high enzyme concentrations*, Bull. Math. Biol., 65(6) (2003), 1111–1129.

[18] Yüzbaşi, Ş. and Sezer, M. *An exponential matrix method for solving systems of linear differential equations*, Math. Methods Appl. Sci., 36(3) (2013), 336–348.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1341–1365
https://doi.org/10.22067/ijnao.2026.96531.1779
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Conformable power series method for solving Fisher–Kolmogorov–Petrovskii–Piskunov equation

D.L. Suthar, F. Desalew, M. Aychluh*, and G. Tirite

Abstract

This work presents the application of the conformable power series method to approximate the nonlinear time-fractional Fisher–Kolmogorov–Petrovskii–Piskunov equation, which models tumor growth and invasion. Using the conformable derivative, the method transforms the fractional equation into a recursive series expansion, yielding accurate semi-analytical approximations. Three numerical examples with different nonlinear powers are solved. Convergence and error analyses of the series solution are established. Results demonstrate excellent agreement with exact and existing numerical solutions, with small absolute errors in all test examples.

AMS subject classifications (2020): 65L05; 35R11, 44A99.

*Corresponding author

Received 18 November 2025; revised 14 February 2026; accepted 15 February 2026

Daya Lal Suthar
Department of Mathematics, Wollo University, Dessie, Ethiopia. e-mail: dlsuthar@gmail.com

Fasil Desalew
Department of Mathematics, Wollo University, Dessie, Ethiopia. e-mail: fasildesalew22@gmail.com

Mulualem Aychluh
Department of Mathematics, Samara University, Samara, Ethiopia. e-mail: muleaychluh62@gmail.com

Gizachew Tirite Gellow
Department of Mathematics, Debre Tabor University, Debre Tabor, Ethiopia. e-mail: gellow-gizachewtirite@gmail.com

How to cite this article
Suthar, D.L., Desalew, S., Aychluh, M. and Tirite, G. , Conformable power series method for solving Fisher–Kolmogorov–Petrovskii–Piskunov equation . *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1341-1365. https://doi.org/10.22067/ijnao.2026.96531.1779

Keywords: Fisher–KPP equation; Power series method; conformable derivative.

1 Introduction

Fractional calculus generalizes classical differentiation and integration to noninteger orders, enabling the mathematical description of systems with memory effects and nonlocal interactions. Since its conceptual origins in a 1695 correspondence between L'Hôpital and Leibniz, fractional calculus has evolved through contributions from Euler, Laplace, Riemann, Liouville, and others, becoming a vital tool in fields such as viscoelasticity, control theory, and transport in porous media [9, 21, 2]. This mathematical framework finds diverse applications in scientific disciplines, particularly for modeling complex systems with memory effects and nonlocal dynamics that cannot be adequately described by classical calculus. Notably, Lacroix became the first to formally mention a fractional derivative in his 1819 publication [34].

Partial differential equations play a central role in modeling dynamical systems, particularly those involving diffusion and reaction processes. Among these, the Fisher–Kolmogorov–Petrovskii–Piskunov (Fisher–KPP) equation, a reaction–diffusion model introduced independently in the 1930s by Fisher and by Kolmogorov, Petrovsky, and Piskunov, has become foundational in mathematical biology [31, 13, 19]. Originally derived from population genetics to model microbial population dynamics, it describes how spatial dispersal affects population densities over time [20]. Contemporary applications extend to analyzing micro-morphogenesis in modern microbiology [32, 37]. This study employs a conformable power series method (CFPSM) to approximate solutions for the nonlinear time-fractional Fisher–KPP equation [37].

Solving nonlinear fractional equations remains challenging, which has prompted the development of various analytical and numerical techniques. Several researchers have investigated the time-fractional Fisher–KPP equation. Angstmann and Henry [3] analyzed both the time-fractional Fitzhugh–Nagumo and Fisher–KPP equations, while Khaled [27] applied the homotopy perturbation method to approximate solutions for the time-fractional nonlinear Fisher–KPP equation. Other studies have examined related aspects: The authors of [10] investigated the time-asymptotic behavior of level sets in Fisher–KPP equations with fractional diffusion in periodic media; Cabré and Roquejoffre [11] studied front propagation in fractional diffusion Fisher–KPP equations; and Cabré and Roquejoffre [12] analyzed the effects of fractional diffusion in these models. Hamel and Roques [24] focused on fast propagation in KPP equations with slowly decaying initial conditions, and Hariharan [25] employed the homotopy analysis method to solve both classical and fractional KPP equations.

Berkovich [8] employed the factorization method to derive analytical invariant solutions of the Fisher–KPP equation. Griffiths and Schiesser [22] analyzed solutions to the Fisher–KPP equation using MATLAB® and Maple™software. Fasano, Herrero, and Rodrigo [18] examined both fast and slow invasion waves in acid-mediated tumor growth models.

Semi-analytical series expansion methods, including the fractional power series (FPS) method, have gained popularity for solving fractional differential equations. These methods provide systematically improvable approximations without spatial discretization. Since nonlinear equations present significant analytical challenges, researchers have developed numerous numerical solution techniques. The FPS method, a well-established semi-analytical technique covered in computational analysis literature, has become fundamental for solving both linear and nonlinear time-fractional differential equations. This approach is particularly valuable in computational mathematics for efficiently deriving functional approximations [4]. Aychluh and Ayalew [6] applied the FPS method to the time-fractional Kuramoto-Sivashinsky equation, while El-Ajou et al. [16] implemented FPS as a generalization of classical power series. Habenom, Suthar, and Mulualem [23] solved the fractional Fokker-Planck equation using the FPS method, while Cui and Hu [14] successfully applied the same approach to fractional differential equations. The current method offers several advantages for solving differential equations, particularly the fractional Fisher–KPP equation. This technique provides semi-analytical approximations through convergent series representations, achieving high accuracy when rapid convergence occurs. The approach is straightforward to implement without requiring complex numerical algorithms or specialized software, making it accessible for various applications. This study implements a conformable fractional power series method (CFPSM) to solve a nonlinear fractional Fisher–KPP model.

Khalil et al. [28] introduced the conformable derivative, a definition that preserves fundamental calculus rules such as the product, quotient, and chain rules. This formulation has attracted substantial attention from researchers, with multiple studies establishing its fundamental characteristics [26, 1]. Silva, Moreira, and Moret [35] investigated the conformable Laplace transform for fractional differential equations. Zhao and Luo [39] examined the physical interpretation of the general conformable operator, while Eroglu, Avci, and Ozdemir [17] applied conformable derivatives to optimal control problems in heat conduction. Özkan and Kurt [33] derived exact solutions for conformable integro-differential equations using the conformable Laplace transform method. The clear properties of conformable derivative simplify the extension of classical series techniques to fractional frameworks, making the CFPSM computationally efficient. While CFPSM has been applied to several fractional models, to our knowledge, it has not yet been employed for the nonlinear time-fractional Fisher–KPP equation.

The proposed noninteger order power series method exhibits a significant limitation in its dependence on initial function selection. The technique's effectiveness hinges on appropriate initial conditions, as poor initial assumptions may lead to slow convergence or inaccurate approximations. Furthermore, the method faces challenges in multi-dimensional applications where managing series convergence becomes increasingly difficult. While the approach provides solutions within a stable convergence radius, highly nonlinear systems often require numerous series terms for adequate approximation, increasing computational demands compared to finite difference, finite element, or finite volume methods. Truncation errors present another limitation, as series termination introduces approximation errors that become particularly significant when computational constraints prevent inclusion of sufficient terms. The fractional nature of the equations introduces additional complexities, especially in computing coefficients for higher-order nonlinear time fractional derivatives. Despite these limitations, the power series method offers distinct advantages over traditional analytical and numerical approaches for fractional differential equations. It maintains applicability across linear and nonlinear problems while achieving high accuracy through series expansion. The method avoids discretization-related numerical instabilities and round-off errors by working with continuous expressions, making it a valuable approach for solving fractional differential equations when appropriate initial conditions are available and computational resources permit sufficient term inclusion.

This research is motivated by the need for a direct and easy method to solve fractional Fisher–KPP equations. The CFPS approach offers an accurate solution due to its simplicity and computational efficiency. To the best of our knowledge, this represents the first application of the CFPSM to solve the time-fractional Fisher–KPP equation. Because the conformable derivative preserves the classical calculus rules (chain rule, product rule, and quotient rule), integer-order power series techniques can be easily extended to fractional domains, offering mathematical simplicity without compromising analytical rigour. CFPSM produces semi-analytical solutions where accuracy can be systematically improved by increasing the series terms.

The paper is organized as follows: Section 2 outlines essential definitions and theorems of conformable calculus and the FPS method. Section 3 details the solution methodology for the fractional Fisher–KPP equation, including numerical examples, graphical and tabular results, and discussion. Section 4 presents concluding remarks.

2 Fundamental concepts

This section presents the fundamental principles of conformable calculus and the essential concepts of the FPS method formulated within the conformable framework.

Definition 1 ([36]). Let $\upsilon(x,\ t) : \mathbb{R} \times (0,\infty) \to \mathbb{R}$ and let $t > 0$. The conformable time fractional partial derivative of order $\kappa \in (0,1]$ is defined as follows:

$$T_t^{\kappa}\upsilon(x,\ t) = \lim_{\varepsilon \to 0} \frac{\upsilon(x,\ t + \varepsilon t^{1-\kappa}) - \upsilon(x,\ t)}{\varepsilon}, \qquad t > 0. \tag{1}$$

The above definition has several important properties. For $\kappa \in (0,1]$, if $\upsilon(x,\ t)$ and $\vartheta(x,\ t)$ are κ-differentiable at a point $t > 0$, then [36]

1. **Linearity:** $T_t^{\kappa}(a\upsilon(x,\ t) \pm b\vartheta(x,\ t)) = aT_t^{\kappa}\upsilon(x,\ t) \pm bT_t^{\kappa}\vartheta(x,\ t)$, for all $a, b \in \mathbb{R}$.
2. **Power rule:** $T_t^{\kappa}(t^{\sigma}) = \sigma t^{\sigma-\kappa}$ for all $\sigma \in \mathbb{R}$.
3. **Constant rule:** $T_t^{\kappa}(\eta) = 0$, for all constant functions $\upsilon(x,\ t) = \eta$.
4. **Quotient rule:** $T_t^{\kappa}(\upsilon/\vartheta) = \dfrac{\vartheta T_t^{\kappa}\upsilon - \upsilon T_t^{\kappa}\vartheta}{\vartheta^2}$.
5. **Product rule:** $T_t^{\kappa}(\upsilon\vartheta) = \upsilon T_t^{\kappa}\vartheta + \vartheta T_t^{\kappa}\upsilon$.

If $\upsilon(x,\ t)$ is differentiable in the classical sense with respect to t, then

$$T_t^{\kappa}\upsilon(x,\ t) = t^{1-\kappa}\frac{\partial \upsilon(x,\ t)}{\partial t}. \tag{2}$$

Definition 2 ([5]). For $0 < \kappa \leq 1$, a power series of the form

$$\upsilon(x,t) = \sum_{i=0}^{\infty} q_i(x)t^{i\kappa} = q_0(x) + q_1(x)t^{\kappa} + q_2(x)t^{2\kappa} + q_3(x)t^{3\kappa} + \cdots, \tag{3}$$

where $x \in I$, $0 \leq t < R^{1/\kappa}$, and $R > 0$, is called a *multiple FPS*, with t as the variable and q_i as the coefficient functions.

All terms in the multiple FPS, except the first, vanish as in classical power series, indicating convergence. The series converges absolutely for

$$0 \leq t < R^{1/\kappa}, \quad (R > 0),$$

where $R^{1/\kappa}$ is the radius of convergence.

Theorem 1 (**Conformable Differentiation of Series**). If $\upsilon(x,t) = \sum_{i=0}^{\infty} q_i(x)t^{i\kappa}$ converges for $0 \leq t < R$, then

$$T_t^{\kappa}\upsilon(x,t) = \kappa\sum_{i=1}^{\infty} iq_i(x)t^{(i-1)\kappa}.$$

Proof. Apply linearity and the power rule of the conformable derivative:

$$\begin{aligned} T_t^\kappa v(x,t) &= T_t^\kappa \left(\sum_{i=0}^{\infty} q_i(x) t^{i\kappa} \right) \\ &= T_t^\kappa \left(q_0(x) + q_1(x) t^\kappa + q_2(x) t^{2\kappa} + q_3(x) t^{3\kappa} + \cdots \right) \\ &= q_0(x) T_t^\kappa(1) + q_1(x) T_t^\kappa(t^\kappa) + q_2(x) T_t^\kappa(t^{2\kappa}) + q_3(x) T_t^\kappa(t^{3\kappa}) + \cdots \\ &= \sum_{i=1}^{\infty} q_i(x) T_t^\kappa \left(t^{i\kappa} \right). \end{aligned}$$

Using the property $T_t^\kappa(t^{i\kappa}) = i\kappa t^{(i-1)\kappa}$, we obtain

$$T_t^\kappa v(x,t) = \kappa \sum_{i=1}^{\infty} i q_i(x) t^{(i-1)\kappa}.$$

□

3 Solution for the fractional Fisher–KPP equation

In recent years, the representation of physical problems using fractional operators has yielded significant results. A major challenge in this field involves obtaining solutions for nonlinear equations. Given the practical importance of these solutions, researchers have developed numerous analytical and numerical methods. Notable techniques include the homotopy perturbation method [30], variational iteration method [38], Laplace perturbation method [30], Adomian decomposition method [15], and homotopy decomposition method [7], among others. This study employs the CFPSM to approximate solutions for the nonlinear time-fractional Fisher–KPP equation, which takes the following form:

$$T_t^\kappa v(x,t) = c \frac{\partial^2 v}{\partial x^2} + av + bv^m, \qquad t > 0,\ \kappa \in (0,1], \tag{4}$$

where a, b, $m \neq 1$, and c are arbitrary constants. For the integer-order case ($\kappa = 1$), an analytical solution is given by [22, p. 173]:

$$v(x,t) = \left[\beta + \exp\left(\lambda t + \frac{\mu x}{\sqrt{c}} \right) \right]^{\frac{2}{1-m}}, \tag{5}$$

where

$$\lambda = \frac{a(1-m)(m+3)}{2(m+1)}, \quad \mu = \sqrt{\frac{a(1-m)^2}{2(m+1)}}, \quad \beta = \sqrt{\frac{-b}{a}}.$$

In this section, we present three numerical examples ($m = 2, 3, 4$), which demonstrate how the proposed approach can be applied to highly nonlinear problems.

Example 1. Consider the following Fisher–KPP equation:

$$T_t^{\kappa} \upsilon(x,t) = c\frac{\partial^2 \upsilon}{\partial x^2} + a\upsilon + b\upsilon^2, \tag{6}$$

with the initial condition:

$$\upsilon\,(x,\ 0) = \left[\beta + \exp\left(\frac{\mu x}{\sqrt{c}}\right)\right]^{-2}.$$

We assume the solution can be expressed as an FPS:

$$\upsilon(x,\ t) = \sum_{i=0}^{\infty} q_i t^{i\kappa} = q_0 + q_1 t^{\kappa} + q_2 t^{2\kappa} + q_3 t^{3\kappa} + \ldots. \tag{7}$$

Applying the conformable derivative to (7) and using the property in (2), we obtain:

$$\begin{aligned} T_t^{\kappa}\upsilon(x,\ t) &= \sum_{i=1}^{\infty} q_i t^{1-\kappa}\frac{d}{dt}t^{i\kappa} = \kappa\sum_{i=1}^{\infty} i q_i t^{(i-1)\kappa} \\ &= \kappa\begin{pmatrix} q_1 + 2q_2 t^{\kappa} + 3q_3 t^{2\kappa} + 4q_4 t^{3\kappa} + 5q_5 t^{4\kappa} \\ + 6q_6 t^{5\kappa} + 7q_7 t^{6\kappa} + 8q_8 t^{7\kappa} + \ldots \end{pmatrix}. \end{aligned} \tag{8}$$

The spatial derivative term is:

$$\begin{aligned} \frac{\partial^2 \upsilon}{\partial x^2} &= \frac{\partial^2}{\partial x^2}\left[\sum_{i=0}^{\infty} q_i t^{i\kappa}\right] \\ &= \frac{\partial^2}{\partial x^2}q_0 + \frac{\partial^2}{\partial x^2}q_1 t^{\kappa} + \frac{\partial^2}{\partial x^2}q_2 t^{2\kappa} + \frac{\partial^2}{\partial x^2}q_3 t^{3\kappa} + \ldots. \end{aligned} \tag{9}$$

The nonlinear term $v^2(x,t)$ expands to

$$\begin{aligned} v^2(x,t) = {} & q_0^2 + 2q_0 q_1 t^{\kappa} + \left(q_1^2 + 2q_0 q_2\right) t^{2\kappa} + 2\left(q_0 q_3 + q_1 q_2\right) t^{3\kappa} \\ & + \left(q_2^2 + 2q_0 q_4 + 2q_1 q_3\right) t^{4\kappa} + 2\left(q_0 q_5 + q_1 q_4 + q_2 q_3\right) t^{5\kappa} \\ & + \left(q_3^2 + 2q_0 q_6 + 2q_1 q_5 + 2q_2 q_4\right) t^{6\kappa} \\ & + 2\left(q_0 q_7 + q_1 q_6 + q_2 q_5 + q_3 q_4\right) t^{7\kappa} + \cdots. \end{aligned} \tag{10}$$

Substituting (7)–(10) into (6) and matching coefficients of like powers of t^{κ} yields the recurrence relations:

$$\begin{cases} q_0 = \upsilon_0, \\ q_1 = \dfrac{c\dfrac{\partial^2 q_0}{\partial x^2} + aq_0 + bq_0^2}{\kappa}, \\ q_2 = \dfrac{c\dfrac{\partial^2 q_1}{\partial x^2} + aq_1 + 2bq_0q_1}{2\kappa}, \\ q_3 = \dfrac{c\dfrac{\partial^2 q_2}{\partial x^2} + aq_2 + b\left[2q_0q_2 + q_1^2\right]}{3\kappa}, \\ q_4 = \dfrac{c\dfrac{\partial^2 q_3}{\partial x^2} + aq_3 + 2b\left[q_0q_3 + q_1q_2\right]}{4\kappa}, \\ q_5 = \dfrac{c\dfrac{\partial^2 q_4}{\partial x^2} + aq_4 + b\left[2q_0q_4 + 2q_1q_3 + q_2^2\right]}{5\kappa}, \\ q_6 = \dfrac{c\dfrac{\partial^2}{\partial x^2}q_5 + aq_5 + 2b\left[q_0q_5 + q_1q_4 + q_2q_3\right]}{6\kappa}, \\ q_7 = \dfrac{c\dfrac{\partial^2}{\partial x^2}q_6 + aq_6 + b\left[2q_0q_6 + 2q_1q_5 + 2q_2q_4 + q_3^2\right]}{7\kappa}, \\ q_8 = \dfrac{c\dfrac{\partial^2}{\partial x^2}q_7 + aq_7 + 2b\left[q_0q_7 + q_1q_6 + q_2q_5 + q_3q_4\right]}{8\kappa}. \\ \vdots \end{cases} \tag{11}$$

Higher-order terms follow similarly. Thus, the series solution is

$$\begin{aligned} \upsilon(x,t) = q_0 &+ q_1t^{\kappa} + q_2t^{2\kappa} + q_3t^{3\kappa} + q_4t^{4\kappa} \\ &+ q_5t^{5\kappa} + q_6t^{6\kappa} + q_7t^{7\kappa} + q_8t^{8\kappa} + \ldots. \end{aligned}$$

Convergence Analysis

Theorem 2 (Convergence of the FPS Solution)**.** Let $\upsilon(x,t)$ be defined in a Banach space $\mathfrak{B}$ with norm $\|\cdot\|$. Consider the series solution of the fractional Fisher–KPP equation given by

$$\upsilon(x,t) = \sum_{n=0}^{\infty} \upsilon_n(x,t),$$

where $\upsilon_0(x,t) = q_0(x)$ is the initial approximation and subsequent terms $\upsilon_n(x,t)$ are determined recursively via the CFPSM procedure. If there exists a constant $\delta \in (0,1)$ such that

$$\|\upsilon_{n+1}(x,t)\| \leq \delta \|\upsilon_n(x,t)\| \quad \text{for all } n \in \mathbb{N}, \tag{12}$$

then the series solution converges uniformly in $\mathfrak{B}$.

Proof. Let $w_n(x,t) = \sum_{i=0}^{n} \upsilon_i(x,t)$ denote the partial sum of the series. We prove that $\{w_n\}_{n=0}^{\infty}$ is a Cauchy sequence in $\mathfrak{B}$. First, from condition (12), we have

$$\|\upsilon_n(x,t)\| \leq \delta^n \|\upsilon_0(x,t)\| \quad \text{for all } n \in \mathbb{N}.$$

Now, for any $n > m \geq 0$, consider

$$\begin{aligned}
\|w_n - w_m\| &= \left\| \sum_{i=0}^{n} \upsilon_i\,(x,\ t) - \sum_{i=0}^{m} \upsilon_i\,(x,\ t) \right\| \\
&= \left\| \sum_{i=m+1}^{n} \upsilon_i\,(x,\ t) \right\| \\
&\leq \sum_{i=m+1}^{n} \|\upsilon_i\,(x,\ t)\| \quad \text{(by the triangle inequality)} \\
&\leq \sum_{i=m+1}^{n} \delta^i \|\upsilon_0\,(x,\ t)\|.
\end{aligned}$$

Since $0 < \delta < 1$, the geometric series converges

$$\sum_{i=m+1}^{\infty} \delta^i = \frac{\delta^{m+1}}{1-\delta}.$$

Thus, for all $n > m$:

$$\|w_n - w_m\| \leq \frac{\delta^{m+1}}{1-\delta} \|\upsilon_0\,(x,\ t)\|.$$

Given $\|\upsilon_0\| < \infty$ and $\delta \in (0,1)$, we have

$$\lim_{m\to\infty} \frac{\delta^{m+1}}{1-\delta} \|\upsilon_0\,(x,\ t)\| = 0.$$

Therefore, for any $\epsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that for all $n, m \geq N_0$,

$$\|w_n - w_m\| < \epsilon.$$

Since $\mathfrak{B}$ is a Banach space, $\{w_n\}$ converges to some $v^* \in \mathfrak{B}$. Each w_n satisfies the CFPSM recurrence derived from the original equation. Therefore, the series converges to the exact solution. □

Theorem 3. Let $\upsilon(x,t) = \sum_{n=0}^{\infty} \upsilon_n(x,t)$ be the exact series solution of the fractional Fisher–KPP equation obtained by CFPSM, and let $w_n(x,t) = \sum_{i=0}^{n} \upsilon_n(x,t)$ be the n-term partial sum approximation. If there exists $\delta \in (0,1)$ such that

$$\|\upsilon_{n+1}(x,t)\| \leq \delta \|\upsilon_n(x,t)\| \quad \text{for all } n \geq 0,$$

then the truncation error $e_n(x,t) = |\upsilon(x,t) - w_n(x,t)|$ satisfies

$$e_n(x,t) \leq \frac{\delta^{n+1}}{1-\delta} \|\upsilon_0(x,t)\|. \tag{13}$$

Proof. The exact solution can be expressed as follows:

$$\upsilon(x,t) = w_n(x,t) + \sum_{i=n+1}^{\infty} \upsilon_i(x,t).$$

Therefore,

$$e_n(x,t) = \left| \sum_{i=n+1}^{\infty} \upsilon_i(x,t) \right| \leq \sum_{i=n+1}^{\infty} |\upsilon_i(x,t)| \leq \sum_{i=n+1}^{\infty} \delta^i \|\upsilon_0(x,t)\|.$$

The last inequality follows from the contraction condition $\|\upsilon_n\| \leq \delta^n \|\upsilon_0\|$. Evaluate the geometric series:

$$\sum_{i=n+1}^{\infty} \delta^i = \delta^{n+1}(1 + \delta + \delta^2 + \cdots) = \frac{\delta^{n+1}}{1-\delta}.$$

Thus,

$$e_n(x,t) \leq \frac{\delta^{n+1}}{1-\delta} \|\upsilon_0(x,t)\|.$$

□

For numerical Example 1, setting $a = 1$, $b = -1$, $c = 0.1$, and $m = 2$, we compute the first few coefficients:

$$q_0(x) = \frac{1}{\left(\exp\left(\frac{\sqrt{15}x}{3}\right) + 1\right)^2},$$

$$q_1(x) = \frac{5 \exp\left(\frac{\sqrt{15}x}{3}\right)}{3\kappa \left(\exp\left(\frac{\sqrt{15}x}{3}\right) + 1\right)^3},$$

$$q_2(x) = \frac{25\exp\left(\frac{\sqrt{15}x}{3}\right)\left[2\exp\left(\frac{\sqrt{15}x}{3}\right)-1\right]}{36\kappa^2\left(\exp\left(\frac{\sqrt{15}x}{3}\right)+1\right)^4},$$

$$q_3(x) = \frac{125\exp\left(\frac{\sqrt{15}x}{3}\right)\left[4\exp\left(\frac{2\sqrt{15}x}{3}\right)-7\exp\left(\frac{\sqrt{15}x}{3}\right)+1\right]}{648\kappa^3\left(\exp\left(\frac{\sqrt{15}x}{3}\right)+1\right)^5},$$

$$q_4(x) = \frac{625\exp\left(\frac{\sqrt{15}x}{3}\right)\left[\begin{array}{r}8\exp\left(\sqrt{15}x\right)+18\exp\left(\frac{\sqrt{15}x}{3}\right)\\ -33\exp\left(\frac{2\sqrt{15}x}{3}\right)-1\end{array}\right]}{15552\kappa^4\left(\exp\left(\frac{\sqrt{15}x}{3}\right)+1\right)^6},$$

$$q_5(x) = \frac{625\exp\left(\frac{\sqrt{15}x}{3}\right)\left(\begin{array}{l}171\exp\left(\frac{\sqrt{215}x}{3}\right)-41\exp\left(\frac{\sqrt{15}x}{3}\right)\\ -131\exp\left(\sqrt{15}x\right)+16\exp\left(\frac{\sqrt{415}x}{3}\right)+1\end{array}\right)}{93312\kappa^5(\exp\left(\frac{\sqrt{15}x}{3}\right)+1)^7},$$

$$q_6(x) = \frac{3125\exp\left(\frac{\sqrt{15}x}{3}\right)\left(\begin{array}{l}1208\exp\left(\sqrt{15}x\right)+88\exp\left(\frac{\sqrt{15}x}{3}\right)\\ -718\exp\left(\frac{\sqrt{215}x}{3}\right)-473\exp\left(\frac{\sqrt{415}x}{3}\right)\\ +32\exp\left(\frac{\sqrt{515}x}{3}\right)-1\end{array}\right)}{3359232\kappa^6(\exp\left(\frac{\sqrt{15}x}{3}\right)+1)^8},$$

$$q_7(x) = \frac{15625\exp\left(\frac{\sqrt{15}x}{3}\right)\left(\begin{array}{l}1+64\exp\left(\sqrt{215}x\right)-8422\exp\left(\sqrt{15}x\right)\\-183\exp\left(\frac{\sqrt{15}x}{3}\right)+2682\exp\left(\frac{\sqrt{215}x}{3}\right)\\+7197\exp\left(\frac{\sqrt{415}x}{3}\right)-1611\exp\left(\frac{\sqrt{515}x}{3}\right)\end{array}\right)}{141087744\kappa^7\left(\exp\left(\frac{\sqrt{15}x}{3}\right)+1\right)^9},$$

$$q_8(x) = \frac{78125\exp\left(\frac{\sqrt{15}x}{3}\right)\left(\begin{array}{l}49780\exp\left(\sqrt{15}x\right)-5281\exp\left(\sqrt{215}x\right)\\+374\exp\left(\frac{\sqrt{15}x}{3}\right)-9327\exp\left(\frac{\sqrt{215}x}{3}\right)\\-78095\exp\left(\frac{\sqrt{415}x}{3}\right)+38454\exp\left(\frac{\sqrt{515}x}{3}\right)\\+128\exp\left(\frac{\sqrt{715}x}{3}\right)-1\end{array}\right)}{6772211712\kappa^8\left(\exp\left(\frac{\sqrt{15}x}{3}\right)+1\right)^{10}}.$$

This section employs the CFPSM to obtain numerical solutions for the time-fractional Fisher–KPP equation. The results presented in the reference tables show that the fractional order κ significantly influences the solution behavior. Both tabular data and graphical representations confirm that the proposed method accurately captures the key features of the Fisher–KPP model across different nonlinearities and fractional orders. The tables provide a quantitative assessment of the CFPSM's performance by comparing the approximate solutions with exact analytical values and, where applicable, with results from existing methods. The close agreement observed validates the accuracy and reliability of the proposed approach.

Figure 1(a) displays the approximate solution $\upsilon(x,t)$ obtained via the CFPSM for the Fisher–KPP equation under the initial condition specified in Equation (6), while Figure 1(b) illustrates the corresponding exact solution. To evaluate the accuracy of the proposed method, error analysis is presented in Table 1 for fixed x and varying t, confirming the effectiveness of the approach in solving the Fisher–KPP equation with the initial condition from Example 1. Figure 2 shows a surface plot of the absolute error for Example 1, revealing a maximum absolute error of 2.5670×10^{-8}.

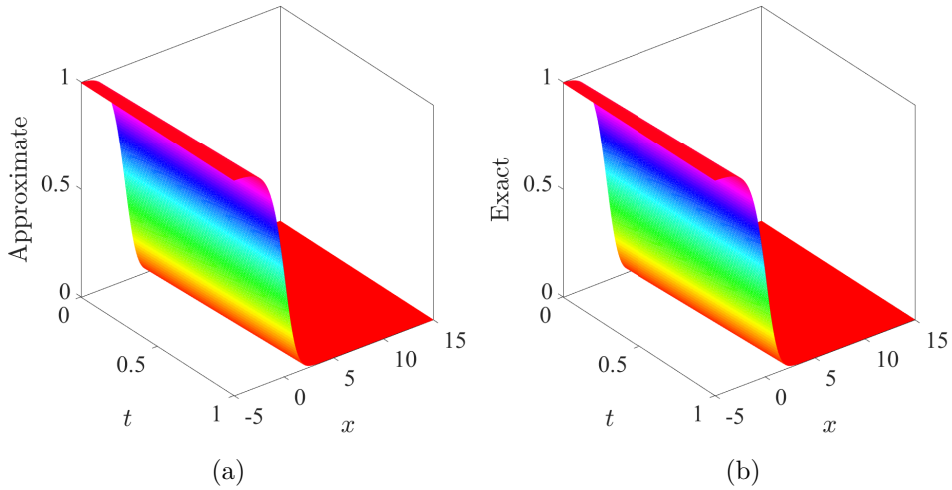


Figure 1: Comparison between exact solution and approximate results for Example 1 with $a = 1,\ b = -1,\ c = 0.1, \kappa = 1$

Table 1: Comparisons of the present results against analytical values for Example 1 at $\kappa = 1$

x	t	Exact	CFPSM	Abs. Error
0.2	0.2	0.22765581	0.22765581	3.3928×10^{-12}
	0.4	0.26912727	0.26912727	1.5652×10^{-09}
	0.6	0.31377638	0.31377643	5.3170×10^{-08}
	0.8	0.36086503	0.36086564	6.1156×10^{-07}
	1.0	0.40953226	0.40953609	3.8266×10^{-06}

Figure 3(a) illustrates the solution profile of $\upsilon(x,t)$ for Example 1 at $\kappa = 0.85$ and different time levels. Figure 3(b) shows the absolute difference between the exact and CFPSM results for $\kappa = 1$ and $t = 0.1$. Figures 4(a)–4(b) display the solution behavior of $\upsilon(x,t)$ for Example 1 at $\kappa = 0.25$ and $\kappa = 0.5$, respectively.

Table 2: Absolute errors for Example 1 at different values of κ

$(x,\ t)$	κ	Absolute error
(2.0, 0.1)	0.99	2.9952×10^{-05}
	0.97	9.3705×10^{-05}
	0.95	1.6304×10^{-04}
	0.93	2.3856×10^{-04}
	0.91	3.2094×10^{-04}

Tables 2, 4, and 8 present the absolute errors for different fractional orders κ of the Fisher–KPP equation in Examples 1–3. These results demonstrate that as the fractional order κ in-

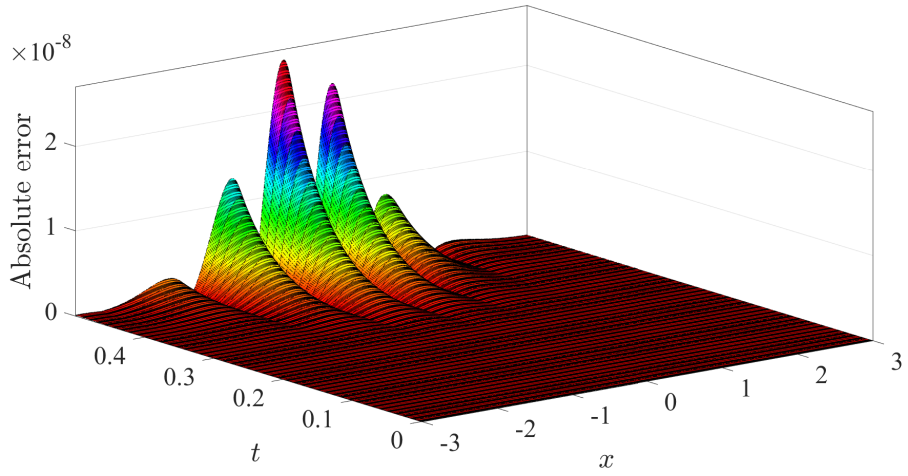


Figure 2: Absolute Error max $= 2.5670 \times 10^{-08}$, for Example 1 with $a = 1,\ b = -1,\ c = 0.1, \kappa = 1$

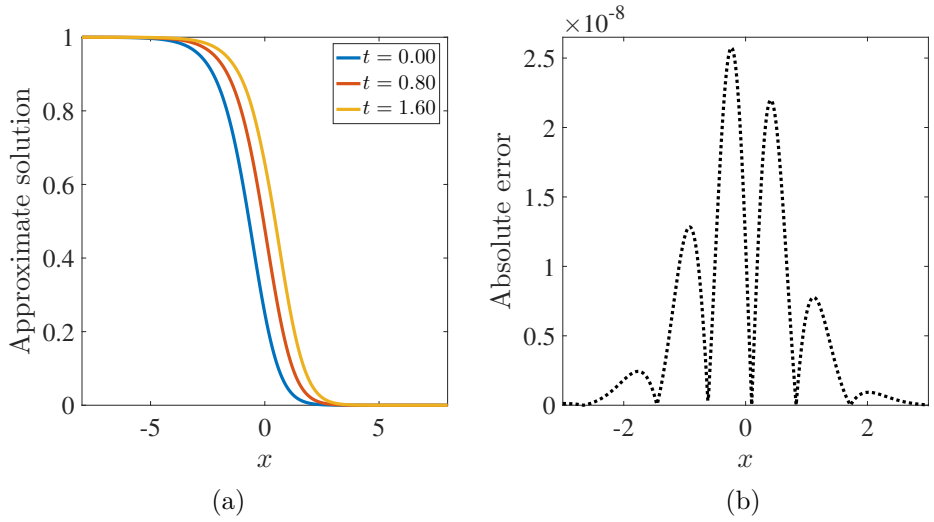


Figure 3: a. Approximate solution for $\kappa = 0.85$; b. Absolute error plot for Example 1 at $\kappa = 1,\ t = 0.1$

creases, the numerical solutions obtained using the proposed technique converge toward the exact solution.

Example 2. Consider the time fractional Fisher–KPP equation:

$$T_t^{\kappa} \upsilon(x,t) = c\frac{\partial^2 \upsilon}{\partial x^2} + a\upsilon + b\upsilon^3,\ t \geq 0,\ 0 < \kappa \leq 1, \tag{14}$$

with the initial condition

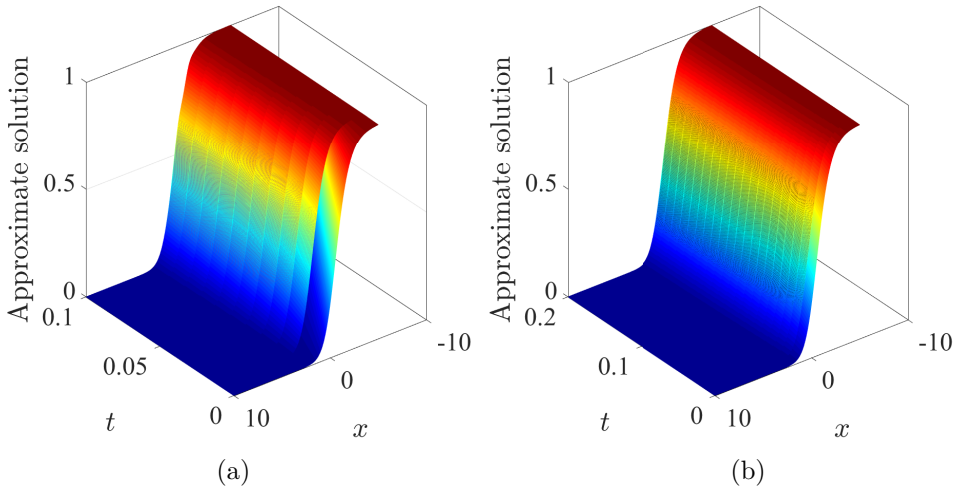


Figure 4: (a) Surf plot of Example 1 for $\kappa = 0.25$. (b) Surf plot of Example 1 at $\kappa = 0.5$

$$\upsilon\left(x,\ 0\right) = \left[\beta + \exp\left(\frac{\mu x}{\sqrt{c}}\right)\right]^{-1}. \tag{15}$$

The exact solution for the nonfractional case at $\kappa = 1$ is

$$\upsilon(x,t) = \left[\beta + \exp\left(\lambda t + \frac{\mu x}{\sqrt{c}}\right)\right]^{-1}. \tag{16}$$

The nonlinear term $\upsilon^3\left(x,\ t\right)$ can be computed as below:

$$\begin{aligned} &\upsilon^3(x) \\ &= \begin{pmatrix} q_0^3 + 3q_1q_0t^{\kappa} + 3\left(q_0q_1^2 + q_0^2q_2\right)t^{2\kappa} + \left(q_1^3 + 3q_0^2q_3 + 6q_0q_1q_2\right)t^{3\kappa} \\ + \left(3q_0q_2^2 + 3q_1^2q_2 + 3q_0^2q_4 + 6q_0q_1q_3\right)t^{4\kappa} \\ + \left(3q_1q_2^2 + 3q_1^2q_3 + 3q_0^2q_5 + 6q_0q_1q_4 + 6q_0q_2q_3\right)t^{5\kappa} \\ + \left(q_2^3 + 3\left(q_0q_3^2 + q_1^2q_4 + q_0^2q_6 + 2q_0q_1q_5 + 2q_0q_2q_4 + 2q_1q_2q_3\right)\right)t^{6\kappa} \\ + 3\begin{pmatrix} q_1q_3^2 + q_2^2q_3 + q_1^2q_5 + q_0^2q_7 \\ + 2\left(q_0q_1q_6 + q_0q_2q_5 + q_0q_3q_4 + q_1q_2q_4\right) \end{pmatrix}t^{7\kappa} + \ldots \end{pmatrix}. \end{aligned} \tag{17}$$

Substituting (7)–(9) and (17) into (14) and looking at the coefficients of t^{κ}, hence, we obtain

$$
\begin{cases}
q_1 = \dfrac{c\dfrac{\partial^2 q_0}{\partial x^2} + aq_0 + bq_0^3}{\kappa}, \quad q_2 = \dfrac{c\dfrac{\partial^2 q_1}{\partial x^2} + aq_1 + 3bq_0^2 q_1}{2\kappa}, \\
q_3 = \dfrac{c\dfrac{\partial^2 q_2}{\partial x^2} + aq_2 + 3b\left(q_0 q_1^2 + q_0^2 q_2\right)}{3\kappa}, \quad q_4 = \dfrac{c\dfrac{\partial^2 q_3}{\partial x^2} + aq_3 + b\left(q_1^3 + 3q_0^2 q_3 + 6q_0 q_1 q_2\right)}{4\kappa} \\
q_5 = \dfrac{c\dfrac{\partial^2 q_4}{\partial x^2} + aq_4 + 3b\left(q_0 q_2^2 + q_1^2 q_2 + q_0^2 q_4 + 2q_0 q_1 q_3\right)}{5\kappa}, \\
q_6 = \dfrac{c\dfrac{\partial^2 q_5}{\partial x^2} + aq_5 + 3b\left(q_1 q_2^2 + q_1^2 q_3 + q_0^2 q_5 + 2q_0 q_1 q_4 + 2q_0 q_2 q_3\right)}{6\kappa}, \\
q_7 = \dfrac{c\dfrac{\partial^2 q_6}{\partial x^2} + aq_6 + b\left[q_2^3 + 3\left(\begin{array}{l} q_0\left(q_3^2 + q_0 q_6 + 2q_1 q_5 + 2q_2 q_4\right) \\ + q_1^2 q_4 + 2q_1 q_2 q_3 \end{array}\right)\right]}{7\kappa}, \\
q_8 = \dfrac{c\dfrac{\partial^2 q_7}{\partial x^2} + aq_7 + 3b\left[\begin{array}{l} q_1 q_3^2 + q_2^2 q_3 + q_1^2 q_5 + q_0^2 q_7 \\ + 2\left(q_0\left(q_1 q_6 + q_2 q_5 + q_3 q_4\right) + q_1 q_2 q_4\right) \end{array}\right]}{8\kappa}.
\end{cases} \tag{18}
$$

Following the same procedure, the remaining iterative terms can be obtained. The coefficients

Table 3: Comparisons of the present results against analytical values for Example 2 at $\kappa = 1$

x	t	Exact	CFPSM	Absolute Error
2.0	0.2	0.015185146	0.015185146	2.4356×10^{-10}
	0.4	0.020389481	0.020389463	1.7384×10^{-08}
	0.6	0.027327977	0.027327771	2.0605×10^{-07}
	0.8	0.036539559	0.036538489	1.0698×10^{-06}
	1.0	0.048700671	0.048697711	2.9600×10^{-06}

are as follows:

$$
q_1 = \frac{3\exp\left(\sqrt{5}x\right)}{2\kappa(\exp\left(\sqrt{5}x\right)+1)^2},
$$

$$
q_2 = \frac{9\exp\left(\sqrt{5}x\right)\left(\exp\left(\sqrt{5}x\right)-1\right)}{8\kappa^2(\exp\left(\sqrt{5}x\right)+1)^3},
$$

$$
q_3(x) = \frac{9\exp\left(\sqrt{5}x\right)\left(\exp\left(\sqrt{25}x\right)-4\exp\left(\sqrt{5}x\right)+1\right)}{16\kappa^3(\exp\left(\sqrt{5}x\right)+1)^4},
$$

$$
q_4(x) = \frac{27\exp\left(\sqrt{5}x\right)\left[\begin{array}{l} 11\exp\left(\sqrt{5}x\right) - 11\exp\left(\sqrt{25}x\right) \\ + \exp\left(\sqrt{35}x\right) - 1 \end{array}\right]}{128\kappa^4(\exp\left(\sqrt{5}x\right)+1)^5},
$$

$$q_5(x) = \frac{81\exp\left(\sqrt{5}x\right)\begin{bmatrix}66\exp\left(\sqrt{25}x\right) - 26\exp\left(\sqrt{5}x\right)\\ -26\exp\left(\sqrt{35}x\right) + \exp\left(\sqrt{45}x\right) + 1\end{bmatrix}}{1280\kappa^5(\exp\left(\sqrt{5}x\right)+1)^6},$$

$$q_6(x) = \frac{81\exp\left(\sqrt{5}x\right)\begin{bmatrix}168\exp\left(\sqrt{5}x\right) - 897\exp\left(\sqrt{25}x\right)\\ +752\exp\left(\sqrt{35}x\right) + 83\exp\left(\sqrt{45}x\right)\\ -56\exp\left(\sqrt{55}x\right) + \exp\left(\sqrt{65}x\right) - 3\end{bmatrix}}{5120\kappa^6(\exp\left(\sqrt{5}x\right)+1)^8},$$

$$q_7(x) = \frac{243\exp\left(\sqrt{5}x\right)\begin{bmatrix}11662\exp\left(\sqrt{35}x\right) - 1214\exp\left(\sqrt{25}x\right)\\ -106\exp\left(\sqrt{5}x\right) - 18476\exp\left(\sqrt{45}x\right)\\ +6194\exp\left(\sqrt{55}x\right) + 358\exp\left(\sqrt{65}x\right)\\ -118\exp\left(\sqrt{75}x\right) + \exp\left(\sqrt{85}x\right) + 3\end{bmatrix}}{71680\kappa^7(\exp\left(\sqrt{5}x\right)+1)^{10}},$$

$$q_8(x) = \frac{243\exp\left(\sqrt{5}x\right)\begin{bmatrix}852\exp\left(\sqrt{5}x\right) - 31025\exp\left(\sqrt{25}x\right)\\ +302352\exp\left(\sqrt{35}x\right) - 1037658\exp\left(\sqrt{45}x\right)\\ +1500040\exp\left(\sqrt{55}x\right) - 844434\exp\left(\sqrt{65}x\right)\\ +132336\exp\left(\sqrt{75}x\right) + 4131\exp\left(\sqrt{85}x\right)\\ -732\exp\left(\sqrt{95}x\right) + 3\exp\left(\sqrt{105}x\right) - 9\end{bmatrix}}{1146880\kappa^8(\exp\left(\sqrt{5}x\right)+1)^{12}}.$$

Figure 5(a) shows the absolute error for the Fisher–KPP solution in Example 2, while Figure 5(b) displays the solution computed using the CFPSM at $\kappa = 0.75$. Figure 6 presents a mesh plot of the absolute error for Example 2 at $\kappa = 1$. Table 3 compares the exact and CFPSM results for Example 2 at $\kappa = 1$, demonstrating near-perfect agreement.

Table 4: Absolute errors for Example 2 at different values of κ

$(x,\ t)$	κ	Absolute error for $\upsilon(x,\ t)$
$(2.0,\ 0.1)$	0.99	6.5365×10^{-05}
	0.98	1.3348×10^{-04}
	0.97	2.0448×10^{-04}
	0.96	2.7853×10^{-04}
	0.95	3.5577×10^{-04}

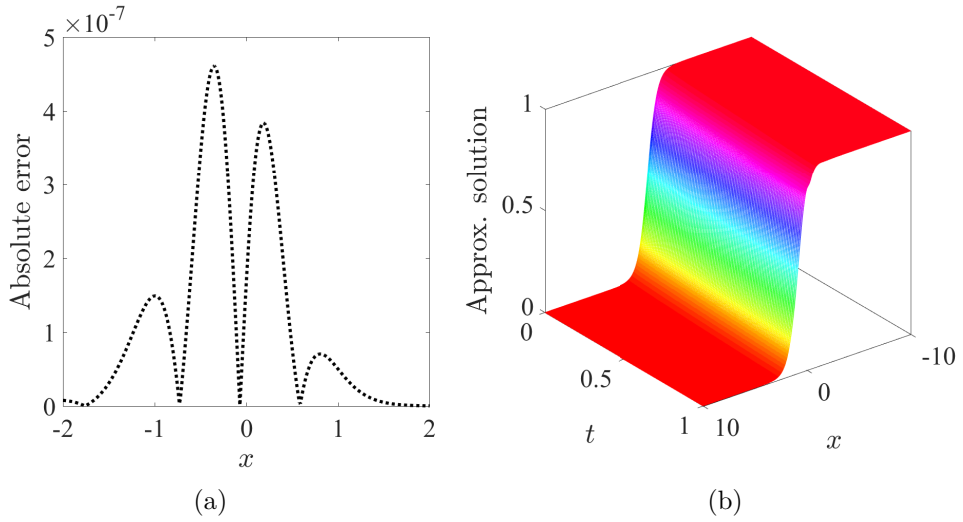


Figure 5: (a) Absolute error plot for Example 2 at $\kappa = 1,\ t = 0.2$. (b) Approximate solution of Example 2 for $\kappa = 0.75$

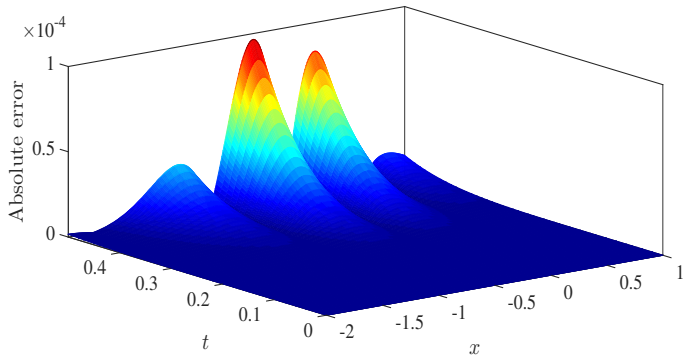


Figure 6: Absolute error for Example 2 with $\kappa = 1$

Table 5: Method comparison for Example 1 with $a = -1,\ b = 0,\ c = 1$ and $q_0(x) = x + \exp(-x)$

x	t	$\kappa = 0.98$	$\kappa = 0.98$
		Method in [29]	Present method
0.5	0.1	1.05647673	1.05586096
	0.2	1.01266710	1.01151583
	0.3	0.97347374	0.97194245
	0.4	0.93809566	0.93646575
	0.5	0.90587041	0.90458680

Tables 5 and 6 compare the CFPSM with the method presented in [29]. These results confirm that the solutions obtained by the proposed method for the Fisher–KPP equation are in excellent agreement with existing numerical results.

Table 6: Method comparison for Example 1 with $a=-1,\ b=0,\ c=1$ and $q_0(x)=x+\exp(-x)$

x	t	$\kappa=1.0$	$\kappa=1.0$
		Method in [29]	Present method
0.5	0.1	1.0589494100	1.058949370
	0.3	0.9769494100	0.976939770
	0.5	0.9099160766	0.909795992
	0.7	0.8554494100	0.854823364
	0.9	0.7940306600	0.809815979

Example 3. Consider the time fractional Fisher–KPP equation:

$$T_t^{\kappa}\upsilon(x,t)=c\frac{\partial^2\upsilon}{\partial x^2}+a\upsilon+b\upsilon^4, \tag{19}$$

with the initial condition

$$\upsilon(x,\ 0)=\left[\beta+\exp\left(\frac{\mu x}{\sqrt{c}}\right)\right]^{-2/3}. \tag{20}$$

The exact solution for the nonfractional case at $\kappa=1$ is

$$\upsilon(x,t)=\left[\beta+\exp\left(\lambda t+\frac{\mu x}{\sqrt{c}}\right)\right]^{-2/3}. \tag{21}$$

The nonlinear term $\upsilon^4(x,\ t)$ can be computed as follows:

$$\begin{aligned}&\upsilon^4(x,t)\\&=\begin{pmatrix}4q_0^3q_1t^{\kappa}+q_0^2\left(4q_0q_2+6q_1^2\right)t^{2\kappa}+4q_0\left(q_0^2q_3+3q_0q_1q_2+q_1^3\right)t^{3\kappa}\\+q_0^4+\left(q_1^4+4q_0^3q_4+6q_0^2q_2^2+12q_0q_1^2q_2+12q_0^2q_1q_3\right)t^{4\kappa}\\+4\left(q_1^3q_2+q_0^3q_5+3q_0q_1q_2^2+3q_0q_1^2q_3+3q_0^2q_1q_4+3q_0^2q_2q_3\right)t^{5\kappa}\\+4\begin{pmatrix}q_0q_2^3+q_1^3q_3+q_0^3q_6+3q_0q_1^2q_4+3q_0^2q_1q_5+3q_0^2q_2q_4\\+1.5q_0^2q_3^2+1.5q_1^2q_2^2+6q_0q_1q_2q_3\end{pmatrix}t^{6\kappa}+\dots\end{pmatrix}.\end{aligned}$$

The coefficients are as follow:

$$q_0=\frac{1}{(\exp(3x)+1)^{2/3}},$$

$$q_1=\frac{7\exp(3x)}{5\kappa(\exp(3x)+1)^{5/3}},$$

$$q_2 = \frac{49\exp(3x)(2\exp(3x)-3)}{100\kappa^2(\exp(3x)+1)^{8/3}},$$

$$q_3 = \frac{343\exp(3x)(4\exp(6x)-27\exp(3x)+9)}{3000\kappa^3(\exp(3x)+1)^{11/3}},$$

$$q_4 = \frac{2401\exp(3x)(234\exp(3x)-171\exp(6x)+8\exp(9x)-27)}{120000\kappa^4(\exp(3x)+1)^{14/3}},$$

$$\vdots$$

Table 7: Comparisons of the present results against analytical values for Example 3 at $\kappa = 1$

x	t	Exact	CFPSM	Abs. Error
0.2	0.02	0.50981094	0.50981094	1.2606×10^{-10}
	0.04	0.51889992	0.51889991	4.0763×10^{-09}
	0.06	0.583909237	0.583903284	3.1268×10^{-08}
	0.08	0.53712165	0.53712152	1.3305×10^{-07}
	0.1	0.54624177	0.54624136	4.0988×10^{-07}

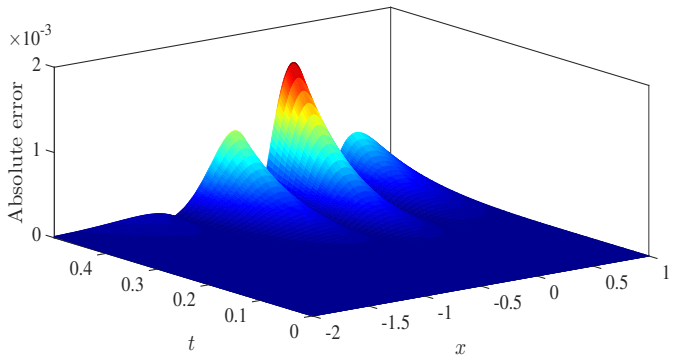


Figure 7: Absolute error for Example 3 with $\kappa = 1$

For Example 3, Table 7 compares the exact and CFPSM results at $\kappa = 1$, showing excellent agreement. Figure 7 displays the error function for Example 3, further illustrating the method's accuracy and applicability.

The consistency of these results underscores the reliability of the proposed method for solving highly nonlinear equations. Both the excellent agreement with exact solutions and the behavior of the error plots confirm the accuracy and robustness of the CFPSM.

Figure 8(a) displays the absolute error, while Figure 8(b) shows the approximate solution for Example 3 at $\kappa = 0.85$. In summary, the accuracy of the CFPSM has been validated

Table 8: Absolute errors for Example 3 at different values of κ

$(x,\ t)$	κ	Absolute error for $\upsilon\,(x,\ t)$
$(0.2,\ 0.08)$	0.99	$2.05900637 \times 10^{-3}$
	0.98	$4.21096997 \times 10^{-3}$
	0.97	$6.43844304 \times 10^{-3}$
	0.96	$8.74387351 \times 10^{-3}$
	0.95	$1.11297450 \times 10^{-2}$

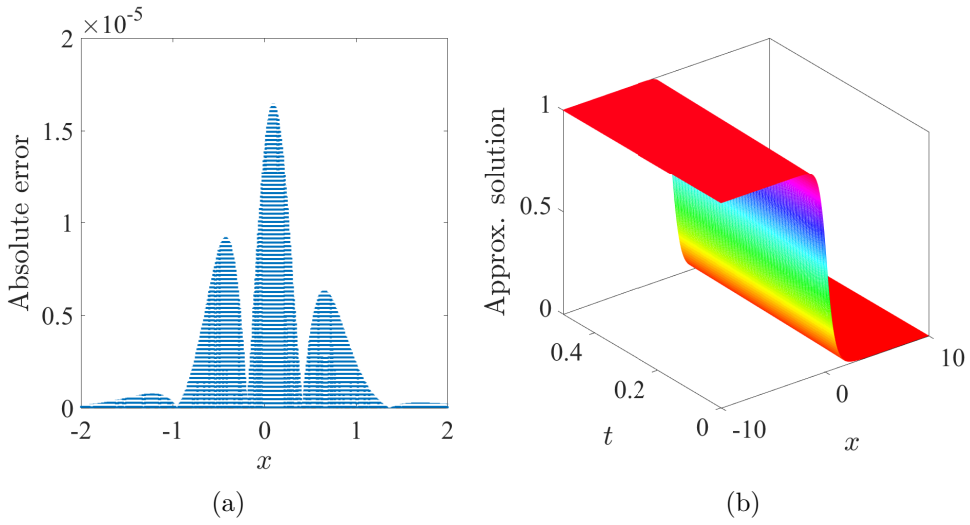


Figure 8: (a) Absolute error plot for Example 3 at $\kappa = 1,\ t = 0.2$. (b) Approximate solution of Example 3 for $\kappa = 0.85$

by comparing the computed results with exact solutions and existing numerical results from the literature. The near-perfect agreement demonstrates the applicability and reliability of the proposed technique for approximating nonlinear fractional differential equations.

4 Conclusion

This work has successfully applied the CFPSM to solve the nonlinear time-fractional Fisher–KPP equation, a widely used model for biological invasion and tumor growth. By leveraging the mathematical simplicity and rule-preserving properties of the conformable derivative, we derived accurate semi-analytical approximations without the need for spatial discretization or complex numerical algorithms. The method was implemented for three representative cases with varying nonlinear powers $(m = 2, 3, 4)$ and fractional orders $\kappa \in (0, 1]$. In all examples, the CFPSM

solutions show excellent agreement with the corresponding exact solutions. In addition, convergence and error analyses of the series solution were established. The graphical and tabular results consistently illustrated that as $\kappa \to 1$, the fractional solutions converge smoothly to the classical integer-order solutions. All computations, including the derivation of series coefficients, and graphical visualizations, were performed using MATLAB R2016a. In conclusion, the CF-PSM shows excellent approximation behavior for solving nonlinear time-fractional Fisher–KPP equations. Future work may extend the method to multi-dimensional systems, variable-order fractional operators, or coupled nonlinear models.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Abdeljawad, T. *On conformable fractional calculus,* J. Comput. Appl. Math. 279 (2015), 57–66. https://doi.org/10.1016/j.cam.2014.10.016

[2] Albalawi, W., Shah, R., Shah, N.A., Chung, J.D., Ismaeel, S.M. and El-Tantawy, S.A. *Analyzing both fractional porous media and heat transfer equations via some novel techniques,* Mathematics 11 (2023), 1350. https://doi.org/10.3390/math11061350

[3] Angstmann, C.N. and Henry, B.I. *Time fractional Fisher–KPP and FitzHugh–Nagumo equations,* Entropy 22(9) (2020), 1035. https://doi.org/10.3390/e22091035

[4] Apostol, T.M. *Calculus*, Vol. 1, 2nd ed., Blaisdell Publishing Company, Waltham, MA, 1967.

[5] Arqub, O.A., Edwan, R., Al-Smadi, M. and Momani, S. *Solving space-fractional Cauchy problem by modified finite-difference discretization scheme,* Alexandria Eng. J. 59 (2020), 2409–2417. https://doi.org/10.1016/j.aej.2020.01.041

[6] Aychluh, M. and Ayalew, M. *The fractional power series method for solving the nonlinear Kuramoto–Sivashinsky equation,* Int. J. Appl. Comput. Math. 11 (2025), 29. https://doi.org/10.1007/s40819-025-01850-9

[7] Atangana, A. *A note on the triple Laplace transform and its applications to some kind of third-order differential equation,* Abstr. Appl. Anal. 2013 (2013), Article ID 769102, 1–10. https://doi.org/10.1155/2013/769102

[8] Berkovich, L.M. *Factorization as a method of finding exact invariant solutions of the Kolmogorov–Petrovskii–Piskunov equation and the related Semenov and Zeldovich equations,* Sov. Math. Dokl. 45 (1992), 162–167.

[9] Boulaaras, S., Jan, R. and Pham, V.T. *Recent advancement of fractional calculus and its applications in physical systems,* Eur. Phys. J. Spec. Top. 232 (2023), 2347–2350. https://doi.org/10.1140/epjs/s11734-023-01002-4

[10] Cabré, X., Coulon, A.C. and Roquejoffre, J.-M. *Propagation in Fisher–KPP type equations with fractional diffusion in periodic media,* C. R. Math. 350(19–20) (2012), 885–890. https://doi.org/10.1016/j.crma.2012.10.007

[11] Cabré, X. and Roquejoffre, J.-M. *Front propagation in Fisher–KPP equations with fractional diffusion,* C. R. Acad. Sci. Paris Ser. I 347 (2009), 1361–1366.

[12] Cabré, X. and Roquejoffre, J.-M. *The influence of fractional diffusion in Fisher–KPP equations,* Commun. Math. Phys., submitted for publication, 2012.

[13] Chellaboina, V., Bhat, S.P., Haddad, W.M. and Bernstein, D.S. *Modeling and analysis of mass-action kinetics,* IEEE Control Syst. Mag. 29 (2009), 60–78.

[14] Cui, R. and Hu, Y. *Fractional power series method for solving fractional differential equation,* J. Adv. Math. 14(4) (2016), 6156–6159.

[15] Duan, J.S., Rach, R., Baleanu, D. and Wazwaz, A.M. *A review of the Adomian decomposition method and its applications to fractional differential equations,* Commun. Fract. Calc. 3 (2012), 73–99.

[16] El-Ajou, A., Arqub, O.A., Zhour, Z.A. and Momani, S. *New results on fractional power series: theories and applications,* Entropy 15 (2013), 5305–5323.https://doi.org/10.3390/e15125305

[17] Eroğlu, B.İ., Avci, D. and Özdemir, N. *Optimal control problem for a conformable heat conduction equation,* Acta Phys. Pol. A 132 (2017), 658–662. https://doi.org/10.12693/APhysPolA.132.658

[18] Fasano, A., Herrero, M.A. and Rodrigo, M.R. *Slow and fast invasion waves in a model of acid-mediated tumour growth,* Math. Biosci. 220 (2009), 45–56. https://doi.org/10.1016/j.mbs.2009.05.003

[19] Fisher, R.A. The wave of advance of advantageous genes, Ann. Eugen. 7 (1937), 353–369.

[20] Fuentes, M.A., Kuperman, M.N. and Kenkre, V.M. *Nonlocal interaction effects on pattern formation in population dynamics,* Phys. Rev. Lett. 91 (2003), 158104.

[21] Giusti, A., Colombaro, I., Garra, R., Garrappa, R. and Mentrelli, A. *On variable-order fractional linear viscoelasticity,* Fract. Calc. Appl. Anal. 27 (2024), 1564–1578. https://doi.org/10.1007/s13540-024-00288-y

[22] Griffiths, G. and Schiesser, W.E. *Traveling Wave Analysis of Partial Differential Equations: Numerical and Analytical Methods with MATLAB and Maple,* Academic Press, Amsterdam, 2010. https://doi.org/10.1016/B978-0-12-384652-5.00010-8

[23] Habenom, H., Suthar, D.L. and Mulualem, A. *Solution of fractional Fokker–Planck equation using fractional power series method,* J. Sci. Arts 3(48) (2019), 593–600.

[24] Hamel, F. and Roques, L. *Fast propagation for KPP equations with slowly decaying initial conditions,* J. Differential Equations 249 (2010), 1726–1745. https://doi.org/10.1016/j.jde.2010.06.005

[25] Hariharan, G. *The homotopy analysis method applied to the Kolmogorov–Petrovskii–Piskunov (KPP) and fractional KPP equations,* J. Math. Chem. 51 (2013), 992–1000. https://doi.org/10.1007/s10910-012-0132-5

[26] Jarad, F., Uğurlu, E., Abdeljawad, T. and Baleanu, D. *On a new class of fractional operators,* Adv. Difference Equ. 2017(1) (2017), 247. https://doi.org/10.1186/s13662-017-1306-z

[27] Khaled, A.G. *The homotopy perturbation method applied to the nonlinear fractional Kolmogorov–Petrovskii–Piskunov equations,* Appl. Math. Lett. 24 (2011), 1428–1434. https://doi.org/10.1016/j.aml.2011.03.025

[28] Khalil, R., Al Horani, M., Yousef, A. and Sababheh, M. *A new definition of fractional derivative,* J. Comput. Appl. Math. 264 (2014), 65–70. https://doi.org/10.1016/j.cam.2014.01.002

[29] Khan, S.Y. and Ahmed, S.A. *An approximate solution of fractional Kolmogorov–Petrovskii–Piskunov equations,* MATEMATIKA 35(3) (2019), 377–385. https://doi.org/10.11113/matematika.v35.n3.1170

[30] Khan, Y. and Wu, Q. *Homotopy perturbation transform method for nonlinear equations using He's polynomials,* Comput. Math. Appl. 61 (2011), 1963–1967. https://doi.org/10.1016/j.camwa.2010.08.016

[31] Kolmogorov, A.N., Petrovskii, I. and Piskunov, N. *A study of the diffusion equation with increase in the amount of substance and its application to a biology problem,* Byul. Moskovskogo Gos. Univ. 1 (1937), 1–25.

[32] Murray, J.D. *Mathematical Biology I: An Introduction,* 3rd ed., Springer, New York, 2001.

[33] Özkan, O. and Kurt, A. *The analytical solutions for conformable integral equations and integro-differential equations by conformable Laplace transform,* Opt. Quantum Electron. 50 (2018), 81. https://doi.org/10.1007/s11082-018-1342-2

[34] Ross, B. *Fractional Calculus and Its Applications,* Proceedings of the International Conference Held at the University of New Haven, June 1974, Springer-Verlag, Berlin, 1975.

[35] Silva, F.S., Moreira, D.M. and Moret, M.A. *Conformable Laplace transform of fractional differential equations,* Axioms 7 (2018), 55. https://doi.org/10.3390/axioms7030055

[36] Thabet, H., Kendre, S. and Peters, J. *Analytical solutions for nonlinear systems of conformable space-time fractional partial differential equations via generalized fractional differential transform,* Vietnam J. Math. 47 (2019), 487–507. https://doi.org/10.1007/s10013-019-00340-y

[37] Unal, A.O. *On the Kolmogorov-Petrovskii-Piskunov equation,* Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat. 62 (2013), 1–10.

[38] Wu, G.C. *New trends in the variational iteration method,* Commun. Fract. Calc. 2 (2011), 59–75.

[39] Zhao, D. and Luo, M. *General conformable derivative and its physical interpretation,* Calcolo 54 (2017), 903–917. https://doi.org/10.1007/s10092-017-0213-8

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1366–1400
https://doi.org/10.22067/ijnao.2026.97704.1834
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Impact of lifestyle interventions and adequate medical infrastructure on diabetes prevention: Dynamics analysis and optimal control

M. Kalita[1,2], Z. Mondal[3], B. Bhuyan[4], B.J. Nath*[,3], H.K. Sarmah[1]

Abstract

Diabetes mellitus is a chronic metabolic disorder that affects the ability to regulate blood glucose levels in the human body. In recent decades, the rapid increase in the number of people living with diabetes around the world has made it one of the most pressing public health concerns and has placed a growing burden on healthcare systems. The onset of diabetes is driven by a combination of genetic predisposition and lifestyle factors, including poor diet, physical inactivity, and obesity, emphasizing the importance of early prevention and effects lifestyle interventions. In this study, we formulate and analyze a nonlinear mathematical model to

*Corresponding author

Received 5 February 2026; revised 26 April 2026; accepted 10 May 2026

Manash Kalita, (manashkalita077@gmail.com)
Zenith Mondal, (znth.mndl96@gmail.com)
Barlin Bhuyan, (barlinbhuyan@gmail.com)
Bhagya Jyoti Nath, (bhagyajyotinath13@gmail.com)
Hemanta Kumar Sarmah, (hsarmah@gauhati.ac.in)

[1] Department of Mathematics, Gauhati University, Guwahati 781014, Assam, India.
[2] Department of Mathematics, Lakhimpur Girls' College, Lakhimpur 787031, Assam, India.
[3] Department of Mathematics, Rabindranath Thakur Vishwavidyalaya, Hojai 782436, Assam, India.
[4] Department of Mathematics, Mahapurusha Srimanta Sankaradeva Viswavidyalaya, Nagaon 782001, Assam, India.

How to cite this article
Kalita, M., Mondal, Z., Bhuyan, B., Nath, B.J. and Sarmah, H.K., Impact of lifestyle interventions and adequate medical infrastructure on diabetes prevention: Dynamics analysis and optimal control. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1366-1400. https://doi.org/10.22067/ijnao.2026.97704.1834

study the effect of lifestyle intervention and proper medical infrastructure on diabetes prevention. The model incorporates the interactions between five state variables related to diabetes mellitus, namely, healthy population, pre-diabetic population due to unhealthy lifestyle, pre-diabetic population due to genetic factors, diabetic population without complications, and diabetic population with complications. We study the basic properties of the model system, such as existence, nonnegativity, and boundedness of solutions. Furthermore, a nontrivial equilibrium point for the model system is identified, and local and global stability of the equilibrium points are established. Additionally, we propose an optimal control problem by introducing two control variables that prevent healthy people from becoming pre-diabetic due to poor lifestyle through awareness campaigns promoting healthy lifestyle and clinical interventions to improve treatment rates for the diabetic population with complications. The optimal control problem is further solved to reduce the burden of diabetes with and without complications. Finally, our findings suggest that lifestyle interventions aimed at raising awareness among the healthy population about unhealthy lifestyles, together with adequate medical infrastructure to improve treatment rates among diabetic individuals with complications, can substantially ease the growing strain imposed by diabetes on healthcare systems.

AMS subject classifications (2020): Primary 37M05, 34D23, 37N25; Secondary 92B05, 49K15.

Keywords: Diabetes; Awareness; Lifestyle intervention; Medical infrastructure; Simulation; Optimal control.

1 Introduction

Diabetes mellitus, popularly known as diabetes, is a chronic and progressive disease that has become a major challenge to the world's healthcare system [23]. It has become one of the leading causes of mortality and morbidity around the world, contributing to the increasing cost of the healthcare system worldwide [14]. Diabetes-related health costs were projected to reach USD 1,028 billion by 2030, up from an anticipated USD 966 billion in 2021 [14]. Diabetes can arise due to a variety of factors, including genetic predisposition and irregular lifestyle habits such as an unhealthy diet and physical inactivity. Another common cause is the inability of the pancreas to produce sufficient insulin, a hormone that enables glucose from food to enter the body's cells and be used for energy [36]. Type 1 diabetes occurs when the immune system repeatedly attacks and damages the pancreatic cells responsible for regulating blood glucose levels, resulting in a build-up of glucose in the bloodstream, whereas type 2 diabetes occurs when the body may not produce sufficient insulin or cannot effectively use the insulin it produces, resulting in higher than normal blood glucose levels. Pre-diabetes is a health condition when the blood sugar level

is elevated but not high enough to be considered diabetes [36]. Individuals with pre-diabetes are more likely to acquire type 2 diabetes if the condition is not treated. Early diagnosis is crucial as untreated pre-diabetes can progress to type 2 diabetes, which in turn can lead to serious health complications such as heart disease and even death. Blood tests are available to detect diabetes and pre-diabetes, facilitating timely intervention and management.

In 2021, a study on global health found that 529 million people around the world were living with diabetes [30], indicating that approximately 6.1% of the population had diabetes. The highest rates were in the Middle East and North Africa (9.3%), and in Oceania (12.3%). Qatar had the highest rate for people aged $75-79$, at 76.1%. Total diabetes prevalence, particularly among older adults, predominantly reflects type 2 diabetes, which in 2021 accounted for 96.0% of all diabetes cases and 95.4% of diabetes-related DALYs worldwide. In the year 2021, more than half of the health problems from type 2 diabetes were linked to high body mass index (BMI), which has increased significantly since 1990. By 2051, it is expected that over 1.31 billion people will have diabetes, with rates above 10% in regions like the Middle East, North Africa (16.8%), and Latin America (11.3%). Out of 204 countries, 89 are projected to have diabetes rates over 10%.

As per the report of the World Health Organization (WHO), the incidence of diabetes is not stable but is rising each year [37]. This increase is largely attributed to changes in lifestyle among the global population. Moreover, social interaction is an important factor influencing lifestyle, and it plays a significant role in lifestyle changes that may increase an individual's risk of developing diabetes. Furthermore, genetic factors are responsible for the incidence of diabetes, particularly in individuals with a family history of the disease [15]. Therefore, the rise in diabetes cases is linked to both genetic factors and unhealthy lifestyles, including an unhealthy level of inactivity, poor dietary choices, and other destructive routines. Currently, countries around the world are facing a significant increase in the number of people with diabetes. If not properly treated, all forms of diabetes can result in severe complications. When individuals with diabetes learn to effectively control their blood glucose levels, this awareness is crucial in reducing the risk of severe complications. Another well-established method to mitigate the long-term effects of diabetes by controlling blood glucose levels is to perform [13]. Recent studies show that lifestyle intervention is an effective method for diabetes remission [12, 18]. Diabetes commonly presents with symptoms such as excessive thirst, frequent urination, heightened hunger, exhaustion, weight loss, and blurred vision. Some complications of diabetes are heart disease, kidney failure, nerve damage, coma and eventually, death. On the other hand, diabetes without complications made it difficult to lead a healthy lifestyle, especially when it remains undiagnosed.

In recent years, mathematical modeling has been widely used to study the dynamics of various diseases and extensive research has been carried out by many researchers to better un-

derstand their transmission, progression, and control [28, 29, 1, 2, 38, 4, 27]. In recent decades, numerous mathematicians have proposed and studied different mathematical models related to population dynamics [5, 33, 8, 16, 25, 11, 21, 22, 20, 9, 7, 31, 32, 35] and glucose-insulin dynamics [24, 19, 17, 3] of diabetes. In the majority of diabetes studies, researchers have primarily concentrated on continuous and discrete time models, which are governed by ordinary differential equations. Also, mathematical analyses such as stability around the equilibrium points, the formulation and solution of optimal control problems to find different optimal strategies are performed in these studies, like other disease-related works [6, 7, 28, 29]. A basic mathematical model to illustrate the dynamics of a population affected by diabetes, which was proposed by Boutayeb et al. [5] in 2016, can be expressed by the following system of differential equations:

$$\begin{aligned}
\frac{dP}{dt} &= n - (I_1 + I_2 + I_3 + \mu)P + \gamma_1 E,\\
\frac{dE}{dt} &= I_1 P - (\gamma_1 + \beta_1 + \beta_3 + \mu)E,\\
\frac{dD}{dt} &= I_2 P + \beta_1 E + \gamma_2 C - (\beta_2 + \mu)D,\\
\frac{dC}{dt} &= I_3 P + \beta_2 D + \beta_3 E - (\gamma_2 + \mu + \delta)C,
\end{aligned}$$

where P, E, D, and C represent healthy, pre-diabetic, diabetic with and without complications, respectively. The parameters n, I_1, I_2, I_3, μ, β_1, β_2, β_3, γ_1, γ_2, δ recruitment rate of adult population, rate of healthy people become pre-diabetic, rate of healthy people become diabetic, rate of healthy people developing complication, probability of pre-diabetic people become diabetic, probability of diabetic people developing complication, probability of pre-diabetic people developing complication, rate of pre-diabetic people become healthy, rate of complication that are cured, and mortality rate due to complications, respectively. The authors introduced an optimal control approach with an objective to reduce the burden of diabetes. In 2016, Derouich et al. [8] published a paper, where they suggested an optimal control technique modeling the progression from pre-diabetes to diabetes with and without complications and found that utilizing that optimal control, the number of diabetes cases with and without complications may be greatly decreased over 10 years. In 2018, Permatasari, Tjahjana, and Udjiani [33] constructed an optimal control model with diabetes caused disabled population to prevent adults from diabetes by minimizing the number of pre-diabetics at minimum cost. In 2020, Kouidere et al. [16] proposed an optimal control model to find a strategy that can minimize treatable diabetics through awareness campaigns about the complications of diabetes, the bad impacts of an irregular lifestyle and the surrounding environment, along with treatment. Expanding the work by W. Boutayeb et al. [5], A. Kouidere et al. [15] proposed a novel model in 2021 by introducing two

new state variables: pre-diabetic individuals due to genetic factors and those affected by negative socio-environmental influences. They also formulated an optimal control problem to reduce the number of diabetics without complications and the vaccination cost by conducting awareness campaigns about the complications of diabetes, the bad impact of an irregular lifestyle, along with the surrounding environment. In 2022, Nasir [25] developed a five-compartmental model of the diabetes population and also proposed an optimal control problem to minimize the pre-diabetic population with minimum cost. It was found that the total number of people with diabetes can be suppressed effectively by the awareness of a healthy diet and regular lifestyle. In 2021, Gupta and Kumar [11] proposed a mathematical model considering that pre-diabetic individuals went into remission due to lifestyle changes, including a healthy diet and physical exercise.

Using both deterministic and stochastic modeling approaches, Mollah and Biswas [21] assessed the effectiveness of diabetes awareness programs and found that increased population-level awareness significantly lowers the risk of diabetes. Furthermore, Mollah and Biswas [22] proposed a population model for Type 2 diabetes to study the combined effects of awareness programs and treatment. Also, they proposed an optimal control problem and found that both awareness programs and treatment are effective in reducing the prevalence of Type 2 diabetes. Mishra and Kumari [20] first proposed a model incorporating the effect of awareness through word-of-mouth communication only and then extended the model to incorporate the effect of awareness through social media and TV advertisements. Their study suggested that social media and television advertisements are more effective in increasing awareness among individuals and in reducing the risk of diabetes mellitus. In 2025, Pal et al. [31] proposed a mathematical model to study the prevalence of diabetes and kidney infection, and then formulated and solved an optimal control problem with two controls, representing the impacts of awareness at the local and global levels, by enhancing information through media. In another study, Pal et al. [32] proposed a mathematical model to study the effects of awareness campaigns on the prevention of diabetes and viral infections. Then, they proposed an optimal control problem considering the same controls as in [31]. Their findings indicate that interventions delivered through social media and television advertisements are significantly more effective in enhancing awareness and reducing the propagation of diabetes risk compared to reliance on word-of-mouth communication alone.

From the above studies, it has been established that awareness among the population plays a crucial role in the management of diabetes. Nowadays, the widespread availability of mobile phones and internet-based platforms such as YouTube has made health-related information accessible to a large segment of the population. Even illiterate individuals can acquire knowledge through audio-visual content, while those without direct access to mobile phones often become informed through interpersonal communication within families and communities, as younger

members typically have access to internet-enabled devices. This indirect diffusion of information ensures that awareness regarding the benefits of lifestyle interventions reaches almost the entire population. In addition, government-led awareness initiatives and public health campaigns further strengthen the dissemination of information related to diabetes prevention. Along with awareness, the availability of adequate and efficient treatment facilities plays a crucial role in diabetes control, as early diagnosis, timely medical intervention, and proper management help prevent disease progression and reduce the severity of diabetes-related complications. Keeping these in mind, we propose a nonlinear model incorporating the effect of lifestyle interventions among the healthy population through awareness, along with the role of proper treatment facilities for timely diagnosis of the complicated diabetic population.

The rest of this article is structured as follows: In Section 2, we present the continuous differential equations model that describes the dynamics of a population of diabetics. In Section 3, we establish some basic properties of the model, including the existence, nonnegativity, and boundedness of solutions. In this section, the stability analysis is conducted around the model's equilibrium point. We discuss the optimal control problem for the proposed model in Section 4, providing findings indicating the existence of the optimal control and characterizing it using Pontryagin's minimum principle. Section 5 contains numerical simulations which were done using MATLAB software. Finally, in Section 6, we summarize the findings of our study.

2 Model formulation

To formulate the model, we first introduce the notation and nomenclature associated with diabetes mellitus. The total population at any given time $t > 0$ is denoted by $N(t)$, and is divided into five subclasses based on health status, genetic predisposition, lifestyle, and medical condition. Individuals who are currently healthy but may develop pre-diabetes or diabetes over time due to genetic factors or unhealthy lifestyle habits are included in the healthy class, denoted by $H(t)$. The pre-diabetic population is further divided into two subclasses: $G(t)$, representing individuals who become pre-diabetic primarily due to genetic factors, and $B(t)$, representing individuals who become pre-diabetic mainly due to unhealthy lifestyle behaviors. Similarly, the diabetic population is divided into two subclasses: $D(t)$, denoting diabetic individuals without complications, and $C(t)$, denoting diabetic individuals with complications. The progression to complications is assumed to occur because of poor glycemic control, delayed diagnosis, or inadequate treatment. Therefore, the total population $N(t)$ at any time t is given by $N(t) = H(t) + G(t) + B(t) + D(t) + C(t)$. The formulation of our model is based on the following assumptions:

- Let n be the constant growth rate of the healthy adult population, who may later become pre-diabetic or diabetic [5]. We assume that technological advancement has increased public awareness regarding the role of lifestyle interventions in the prevention of diabetes.

- We assume that healthy adults may progress to a pre-diabetic state either due to genetic predisposition or unhealthy lifestyle behaviors, and may eventually develop diabetes if timely preventive measures are not taken [16]. For the model formulation, we assume that healthy individuals enter the pre-diabetic stage due to genetic factors at the rate v and due to an unhealthy lifestyle at the rate λ. Furthermore, γ_1 and γ_2 denote the rates at which pre-diabetic individuals caused by genetic factors progress to diabetes without complications and with complications, respectively. Similarly, δ represents the rate at which pre-diabetes caused by an unhealthy lifestyle progresses to diabetes due to the continued effects of irregular and unhealthy habits.

- We assume that a fraction of healthy individuals may progress directly to diabetes, with or without complications, without passing through the pre-diabetic stage [16]. This may occur due to factors such as late diagnosis, genetic susceptibility, lack of regular glucose monitoring, and other related causes. In the model, the parameters α_1 and α_2 represent the rates at which healthy individuals become diabetic without complications and with complications, respectively.

- We assume that individuals in the pre-diabetic class caused by an unhealthy lifestyle may return to the healthy class through lifestyle modification, regular physical activity, and improved dietary habits. It is also assumed that diabetic individuals without complications may achieve remission and return to the healthy class through early diagnosis, sustained lifestyle intervention, and appropriate medical treatment [11]. In the model, r_1 and r_2 represent the remission rates of the pre-diabetic class due to unhealthy lifestyle and the diabetic class without complications, respectively.

- In the absence of adequate treatment or regular monitoring, diabetic individuals without complications are assumed to progress to diabetes-related complications at the rate η.

- We assume that diabetic individuals with complications may return to the diabetic class without complications through effective treatment, whose success depends on the availability and quality of medical infrastructure, such as timely access to specialists, diagnostic facilities, and continuous care. Therefore, a saturated treatment function $\frac{\kappa C}{1+lC}$ [39, 26] is used to represent the recovery rate from complications. Here, κ denotes the

treatment rate, while l is a nonnegative parameter representing limitations in medical resources. Thus, when healthcare resources are limited, the treatment rate initially increases with the number of diabetic individuals with complications, but eventually approaches a maximum saturation level as healthcare capacity becomes fully utilized. Furthermore, β represents the mortality rate due to diabetes-related complications.

The above assumptions form the basis for the differential equation model, which captures the dynamics of healthy, pre-diabetic and diabetic populations with the impact of lifestyle intervention through awareness and the adequate medical facilities. The flowchart in Figure 1 illustrates the interactions between different populations of the model. Hence, the governing system of differential equations for the model is derived as follows:

$$\begin{aligned}
\frac{dH(t)}{dt} &= n - (\mu + \alpha_1 + \alpha_2 + v + \lambda)H(t) + r_1 B(t) + r_2 D(t),\\
\frac{dG(t)}{dt} &= vH(t) - (\gamma_1 + \gamma_2 + \mu)G(t),\\
\frac{dB(t)}{dt} &= \lambda H(t) - (\delta + \mu + r_1)B(t),\\
\frac{dD(t)}{dt} &= \alpha_1 H(t) + \gamma_1 G(t) + \delta B(t) - (\mu + \eta + r_2)D(t) + \frac{\kappa C}{1 + lC},\\
\frac{dC(t)}{dt} &= \alpha_2 H + \gamma_2 G(t) + \eta D(t) - \frac{\kappa C}{1 + lC} - (\mu + \beta)C(t).
\end{aligned} \tag{1}$$

We analyse the system with the initial conditions

$$H(0) \geq 0, \quad G(0) \geq 0, \quad B(0) \geq 0, \quad D(0) \geq 0, \quad C(0) \geq 0. \tag{2}$$

Moreover, all the parameters used in the model (1) are considered to be positive, and their

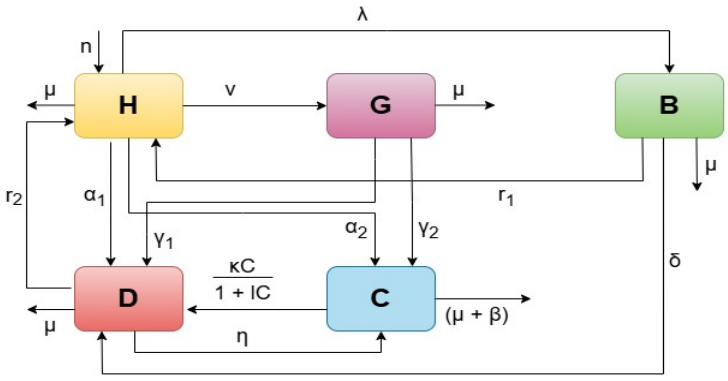


Figure 1: The schematic diagram for the model (1).

detailed descriptions and values are given in Table 1.

Table 1: Meanings of parameters for the model (1) along with their values.

Parameter(s)	Description	Units	Values
n	Constant growth rate of Healthy population	person day^{-1}	600000
μ	Natural death rate of healthy population	day^{-1}	0.02
α_1	Rate at which the healthy population becomes diabetic without complications	day^{-1}	0.3
α_2	Rate at which the healthy population becomes diabetic with complications.	day^{-1}	0.1
v	Rate at which the healthy population enters the pre-diabetic stage due to genetic factors.	day^{-1}	0.1
λ	Rate at which the healthy population enters the pre-diabetic stage due to a bad lifestyle	day^{-1}	0.2
γ_1	Rate at which the pre-diabetic population due to genetic factors become diabetic patients without complications	day^{-1}	0.2
γ_2	Rate at which the pre-diabetic population due to genetic factors become diabetic patients with complications	day^{-1}	0.1
δ	Rate at which the pre-diabetic population, due to a bad lifestyle, becomes diabetic patients without complications	day^{-1}	0.8
η	Rate at which the diabetic population without complication propagates to the complicated stage	day^{-1}	0.5
β	Represents the mortality rate of the diabetic population with complications.	day^{-1}	0.05
r_1	Rate at which the pre-diabetic population, due to a bad lifestyle, moves to a healthy population	day^{-1}	0.08
r_2	Rate at which the diabetic population without complications enters into a healthy category	day^{-1}	0.08
κ	The treatment rate	day^{-1}	0.65
l	The limitations of medical resources	person^{-1}	0.0003

3 Mathematical analysis of the model (1)

3.1 Basic properties of the model (1)

In this part, we demonstrate many fundamental aspects of the model (1), such as existence, nonnegativity, and boundedness of the model solutions. The verification of these basic properties plays a significant role in establishing the biological feasibility of the model (1).

Theorem 1. The dynamical system (1) with the initial condition (2) has a unique solution.

The proof of the theorem is given in *Appendix A*.

Theorem 2. Under the initial conditions (2), the solutions $H(t)$, $G(t)$, $B(t)$, $D(t)$ and $C(t)$ of the system (1) are nonnegative for all $t \geq 0$.

The proof of the theorem is given in *Appendix B*.

Theorem 3. The set

$$\Omega = \Big\{(H(t), G(t), B(t), D(t), C(t)) \in \mathbb{R}^5_+ \\ |H(t) + G(t) + B(t) + D(t) + C(t) \leq \frac{n}{\mu}\Big\},$$

is positively invariant under system (1) with initial conditions (2).

The proof of the theorem is given in *Appendix C*.

3.2 Calculation of equilibrium point

Generally, the solutions of $\frac{dH(t)}{dt} = \frac{dG(t)}{dt} = \frac{dB(t)}{dt} = \frac{dD(t)}{dt} = \frac{dC(t)}{dt} = 0$ are the equilibrium points of the model (1). Let the equilibrium point $T^*(H^*, G^*, B^*, D^*, C^*)$ be a solution of the following system:

$$n - (\mu + \alpha_1 + \alpha_2 + v + \lambda)H^*(t) + r_1 B^*(t) + r_2 D^*(t) = 0, \tag{3}$$

$$vH^*(t) - (\gamma_1 + \gamma_2 + \mu)G^*(t) = 0, \tag{4}$$

$$\lambda H^*(t) - (\delta + \mu + r_1)B^*(t) = 0, \tag{5}$$

$$\alpha_1 H^*(t) + \gamma_1 G^*(t) + \delta B^*(t) - (\mu + \eta + r_2)D^*(t) + \frac{\kappa C^*}{1 + lC^*} = 0, \tag{6}$$

$$\alpha_2 H^* + \gamma_2 G^*(t) + \eta D^*(t) - \frac{\kappa C^*}{1 + lC^*} - (\mu + \beta)C^*(t) = 0. \tag{7}$$

Since the state variable represents the individual population, only a positive solution is the biologically feasible equilibrium point. From (3), (4) and (5) we have

$$G^* = \frac{vH^*}{\gamma_1 + \gamma_2 + \mu}, \tag{8}$$

$$B^* = \frac{\lambda H^*}{\delta + \mu + r_1}, \tag{9}$$

$$D^* = \frac{-n}{r_2} + \frac{(\mu + \alpha_1 + \alpha_2 + v + \lambda)H^*}{r_2} - \frac{r_1 \lambda H^*}{r_2(\delta + \mu + r_1)}. \tag{10}$$

By substituting the expression of G^*, B^*, and D^* in (6) we obtain

$$(\mu + \beta)l(C^*)^2 - (Xl - \kappa - (\mu + \beta))C^* - X = 0, \tag{11}$$

where

$$\begin{aligned} X =& \alpha_2 H^* + \frac{\gamma_2 v}{\gamma_1 + \gamma_2 + \mu} H^* + \frac{\eta(\mu + \alpha_1 + \alpha_2 + v + \lambda)}{r_2} H^* \\ & - \frac{\lambda r_1}{r_2(\delta + \mu + r_1)} H^* - \frac{\eta n}{r_2}. \end{aligned}$$

The roots of the equation are as follows:

$$C^* = \frac{(Xl - \kappa - \mu - \beta) \pm \sqrt{(Xl - \kappa - \mu - \beta)^2 + 4X(\mu + \beta)l}}{2l(\mu + \beta)}. \tag{12}$$

Since

$$\sqrt{(Xl-\kappa-\mu-\beta)^2+4X(\mu+\beta)l} > \sqrt{(Xl-\kappa-\mu-\beta)^2} = |Xl-\kappa-\mu-\beta|,$$

therefore, irrespective of the sign of the term $(Xl-\kappa-\mu-\beta)$, the root

$$C^* = \frac{(Xl-\kappa-\mu-\beta)+\sqrt{(Xl-\kappa-\mu-\beta)^2+4X(\mu+\beta)l}}{2l(\mu+\beta)},$$

is always positive.

Again, from (6) we obtain

$$H^* = -\frac{n(\mu+\eta+r_2)}{A} - \frac{\kappa C^*}{A(1+lC^*)}, \tag{13}$$

where

$$A = \alpha_1 + \frac{\gamma_1 v}{\gamma_1+\gamma_2+\mu} + \frac{\delta\lambda}{\delta+\mu+r_1} + \frac{\lambda r_1(\mu+\eta+r_2)}{r_2(\delta+\mu+r_1)} - \frac{(\mu+\eta+r_2)(\mu+\alpha_1+\alpha_2+v+\lambda)}{r_2}.$$

If

$$A = \alpha_1 + \frac{\gamma_1 v}{\gamma_1+\gamma_2+\mu} + \frac{\delta\lambda}{\delta+\mu+r_1} + \frac{\lambda r_1(\mu+\eta+r_2)}{r_2(\delta+\mu+r_1)} < \frac{(\mu+\eta+r_2)(\mu+\alpha_1+\alpha_2+v+\lambda)}{r_2},$$

then H^* is positive, along with G^* and B^*. From (10), if

$$\frac{(\mu+\alpha_1+\alpha_2+v+\lambda)H^*}{r_2} > \frac{n}{r_2} + \frac{\lambda r_1 H^*}{r_2(\delta+\mu+r_1)},$$

then D^* is positive.

3.3 Local stability of the equilibrium point T^*

Theorem 4. The equilibrium point T^* is locally asymptotically stable under the following conditions:

$$r_1 + r_2 < \mu + \alpha_1 + \alpha_2 + v + \lambda,$$

$$
\begin{aligned}
&v < \gamma_1 + \gamma_2 + \mu,\\
&\lambda < \delta + \mu + r_1,\\
&\alpha_1 + \gamma_1 + \delta + \frac{\kappa}{(1+lC)^2} < \mu + \eta + r_2,\\
&\alpha_2 + \gamma_2 + \eta < \mu + \beta + \frac{\kappa}{(1+lC)^2}.
\end{aligned}
$$

Proof. The Jacobian matrix around the interior equilibrium point T^* for the model (1) is given by

$$
J(T^*) = \begin{bmatrix}
-(\mu+\alpha_1+\alpha_2+v+\lambda) & 0 & r_1 & r_2 & 0\\
v & -(\gamma_1+\gamma_2+\mu) & 0 & 0 & 0\\
\lambda & 0 & -(\delta+\mu+r_1) & 0 & 0\\
\alpha_1 & \gamma_1 & \delta & -(\mu+\eta+r_2) & \frac{\kappa}{(1+lC)^2}\\
\alpha_2 & \gamma_2 & \eta & -\frac{\kappa}{(1+lC)^2} & -(\mu+\beta)
\end{bmatrix}.
$$

Here, all the diagonal entries of the matrix $J(T^*)$ are negative. Also, it is not always easy to find eigenvalues for a large matrix. Thus, by Gersgorin's theorem, the interior equilibrium T^* is locally asymptotically stable provided the following condition holds:

$$
\begin{aligned}
&r_1 + r_2 < \mu + \alpha_1 + \alpha_2 + v + \lambda,\\
&v < \gamma_1 + \gamma_2 + \mu,\\
&\lambda < \delta + \mu + r_1,\\
&\alpha_1 + \gamma_1 + \delta + \frac{\kappa}{(1+lC)^2} < \mu + \eta + r_2,\\
&\alpha_2 + \gamma_2 + \eta < \mu + \beta + \frac{\kappa}{(1+lC)^2}.
\end{aligned}
$$

□

3.4 Global Stability of the equilibrium point T^*

In this section, we have shown the global stability of the model (1) around the equilibrium point T^* with the help of suitable Lyapunov function. Consider the following Lyapunov positive definite function:

$$
L = \frac{1}{2}(H-H^*)^2 + \frac{1}{2}(G-G^*)^2 + \frac{1}{2}(B-B^*)^2 + \frac{1}{2}(D-D^*)^2 + \frac{1}{2}(C-C^*)^2. \tag{14}
$$

By differentiating L with respect to t, we have

$$\frac{dL}{dt} =(H-H^*)\frac{dH}{dt}+(G-G^*)\frac{dG}{dt}+(B-B^*)\frac{dB}{dt}+(D-D^*)\frac{dD}{dt} \tag{15}$$

$$+(C-C^*)\frac{dC}{dt}. \tag{16}$$

Using (3)–(7), we have

$$\begin{aligned}\frac{dL}{dt} =& -(\mu+\alpha_1+\alpha_2+v+\lambda)(H-H^*)^2-(\gamma_1+\gamma_2+\mu)(G-G^*)^2\\ &-(\delta+\mu+r_1)(B-B^*)^2-(\mu+\eta+r_2)(D-D^*)^2\\ &-\left(\frac{\kappa}{(1+lC^*)^2}+\mu+\beta\right)(C-C^*)^2\\ &+(r_1+\lambda)(B-B^*)(H-H^*)\\ &+(r_2+\alpha_1)(D-D^*)(H-H^*)+v(G-G^*)(H-H^*)\\ &+\gamma_1(D-D^*)(B-B^*)+\alpha_2(C-C^*)(H-H^*)\\ &+\gamma_2(C-C^*)(G-G^*)+\left(\frac{\kappa}{(1+lC^*)^2}+\eta\right)(D-D^*)(C-C^*)\\ &+\frac{\kappa C^*}{1+lC^*}[(D-D^*)-(C-C^*)].\end{aligned}$$

From the above equation, using Young's inequality, we have $\frac{dL}{dt}$ is negative definite provided the following conditions hold:

$$(\delta+\mu+r_1)(\mu+\alpha_1+\alpha_2+v+\lambda)>\left(\frac{r_1+\lambda}{2}\right)^2, \tag{17}$$

$$(\mu+\eta+r_2)(\mu+\alpha_1+\alpha_2+v+\lambda)>\left(\frac{r_2+\lambda}{2}\right)^2, \tag{18}$$

$$(\gamma_1+\gamma_2+\mu)(\mu+\alpha_1+\alpha_2+v+\lambda)>\frac{v^2}{4}, \tag{19}$$

$$(\mu+\eta+r_2)(\gamma_1+\gamma_2+\mu)>\frac{\gamma^2}{4}, \tag{20}$$

$$(\mu+\eta+r_2)(\delta+\mu+r_1)>\frac{\delta^2}{4}, \tag{21}$$

$$(\mu+\eta+r_2)\left(\frac{\kappa}{(1+lC^*)^2}+\mu+\beta\right)>\frac{\left(\frac{\kappa}{(1+lC^*)^2}+\eta\right)^2}{4}, \tag{22}$$

$$\left(\frac{\kappa}{(1+lC^*)^2}+\mu+\beta\right)(\mu+\alpha_1+\alpha_2+v+\lambda)>\frac{\alpha_2^2}{4}, \tag{23}$$

$$\left(\frac{\kappa}{(1+lC^*)^2}+\mu+\beta\right)(\gamma_1+\gamma_2+\mu)>\frac{\gamma_2^2}{4}, \tag{24}$$

$$(C - C^*) > (D - D^*). \tag{25}$$

So, by Lyapunov's Direct method, we can conclude that our model (1) shows global stability around the interior equilibrium point T^* under the above conditions. Thus, we have the following theorem.

Theorem 5. The equilibrium point T^* is globally asymptotically stable under the conditions (17)–(25)

4 Optimal control problem

The optimal control problem aims to identify time-dependent intervention strategies that minimize the diabetic population and related complications at the population level, subject to the underlying disease transmission dynamics and control costs. In general, an optimal control problem for a controlled model consists of three main parts: an objective function, a set of state variables and controls. To formulate the optimal control problem associated with system (1), we introduce two time-dependent controls $u_1(t)$ and $u_2(t)$, representing the lifestyle intervention due to awareness about healthy lifestyle and the proper medical infrastructure, respectively. The detailed description and impact of these controls are given below:

1. **Lifestyle intervention control:** The control $u_1(t)$ represents the intensity of the behavioral modification as a result of awareness to maintain a proactive and healthy lifestyle, which includes performing regular physical activities, maintaining a healthy weight, and following a balanced diet, where $0 \leq u_1(t) \leq 1$. An increase in $u_1(t)$ reduces the rate at which individuals transition from the healthy class to pre-diabetic and diabetic classes by discouraging unhealthy lifestyle choices. Moreover, the control $u_1(t)$ enhances behavioral change among individuals in the pre-diabetic class due to bad lifestyle, thereby increasing the remission rate from this class back to the healthy population as $u_1(t)$ increases. Similarly, higher levels of $u_1(t)$ improve self-management and lifestyle adherence among diabetic individuals without complications, leading to an increased remission rate from the diabetic without complications class to the healthy class.

2. **Enhancing medical infrastructure:** The control $u_2(t)$ represents the intensity of clinical interventions aimed at improving the treatment outcomes of diabetic individuals with complications, where $0 \leq u_2(t) \leq 1$. This control captures the effect of expanding and strengthening medical resources, including improved healthcare infrastructure, availability of specialized medical personnel, access to diagnostic facilities, and timely thera-

peutic management. An increase in $u_2(t)$ enhances the treatment rate of individuals in the diabetic-with-complications compartment, thereby facilitating their transition to the diabetic-without-complications class and reducing the overall burden of severe diabetes related outcomes in the population.

Now, the controlled model based on (1) is stated below:

$$\begin{aligned}
\frac{dH(t)}{dt} =& n-(\mu+\alpha_2+v)H(t)-\alpha_1(1-u_1(t))H(t)\\
&-\lambda(1-u_1(t))H(t)+r_1u_1(t)B(t)+r_2u_1(t)D(t),\\
\frac{dG(t)}{dt} =& vH(t)-(\gamma_1+\gamma_2+\mu)G(t),\\
\frac{dB(t)}{dt} =& \lambda(1-u_1(t))H(t)-\delta B(t)-(\mu+r_1u_1(t))B(t),\\
\frac{dD(t)}{dt} =& \alpha_1(1-u_1(t))H(t)+\gamma_1G(t)+\delta B(t)+\frac{\kappa u_2(t)C(t)}{1+lC(t)}\\
&-\eta D(t)-(\mu+r_2u_1(t))D(t),\\
\frac{dC(t)}{dt} =& \alpha_2H(t)+\gamma_2G(t)-\frac{\kappa u_2(t)C(t)}{1+lC(t)}-(\mu+\beta)C(t)+\eta D(t),
\end{aligned} \tag{26}$$

Moreover, the objective functional is

$$\mathcal{J}(B,u)=\int_0^T\{C(t)+D(t)+\frac{1}{2}A_1u_1^2(t)+\frac{1}{2}A_2u_2^2(t)\}dt, \tag{27}$$

which has to be minimized. Here, the positive constants A_1 and A_2, referred to as weighting constants, are related to costs of following a healthy diet, engaging in regular physical activities, organizing awareness programs, expanding medical resources, etc. It will also be kept in mind that these costs should also be minimized. Hence, our main objective is to find an optimal control set (u_1^*,u_2^*) such that

$$\mathcal{J}(u_1^*,u_2^*)=\min_{(u_1^*,u_2^*)\in\Gamma}J(u_1(t),u_2(t)), \tag{28}$$

subject to the state constraints (26) and initial conditions (2). The control set Γ is defined as follows:

$$\Gamma=\{(u_1,u_2)|u_i\text{'s are measurable},0\le u_i\le 1,\ \text{for all } i=1,2 \text{ and } t\in[0,t_f]\}. \tag{29}$$

4.1 Existence of the optimal control problem

Now, the first target is to show the existence of an optimal control set (u_1^*, u_2^*) that can minimize the objective function (27). In order to establish the existence of optimal control, methodologies from several previous research works [10, 31, 32, 28, 29] were applied, resulting in the following results.

Theorem 6. For the control model (26) with the controls (u_1, u_2), there exists an optimal control set (u_1^*, u_2^*) such that

$$\mathcal{J}(u_1^*, u_2^*) = \min_{(u_1^*, u_2^*) \in \Gamma} J(u_1(t), u_2(t)).$$

Proof. It can be easily shown that the set of state variables under the controls (u_1, u_2) in (26) is bounded, which confirms the existence of solutions. Hence, the set Γ is nonempty. From the definition, the control space Γ is convex and closed. Consider the integrand of the objective functional as

$$\mathcal{L} = C(t) + D(t) + \frac{1}{2}A_1 u_1^2(t) + \frac{1}{2}A_2 u_2^2(t). \tag{30}$$

The Hessian matrix for the integrand $\mathcal{L}$ is given by

$$H_{\mathcal{L}} = \begin{bmatrix} \mathcal{L}_{u_1,u_1} & \mathcal{L}_{u_1,u_2} \\ \mathcal{L}_{u_2,u_1} & \mathcal{L}_{u_2,u_2} \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}. \tag{31}$$

Clearly, the Hessian matrix $H_{\mathcal{L}}$ is positive definite, which indicates that the integrand function $\mathcal{L}$ is convex for all $u_i \in \Gamma$, $i = 1, 2$. Since the state variables $D(t)$ and $C(t)$ are bounded, so by order completeness property, they have supremum, let $\xi_1 = \sup_{t\in[0,t_f]} D(t)$ and $\xi_2 = \sup_{t\in[0,t_f]} C(t)$, then we have

$$C(t) + D(t) + A_1 u_1^2(t) + A_2 u_2^2(t) \geq \xi_3 |u_1|^\alpha + \xi_4 |u_2|^\alpha - \xi_1 - \xi_2,$$

choosing $\xi_3 = A_1$, $\xi_4 = A_2$, and $\alpha = 2$. So, the integrand $\mathcal{L}$ is bounded below.

From Fleming and Rishel [10], thus, the existence of the optimal control set (u_1, u_2) can be concluded, and for this, the objective functional $\mathcal{J}(u_1^*, u_2^*)$ is minimized. □

4.2 Characterization of the optimal control

Now, we apply Pontryagin's minimum principle [34] to find the necessary conditions for the optimal control problem. For this, the Hamiltonian $\mathcal{H}$ related to state system (26) is considered as below:

$$\mathcal{H} = C(t) + D(t) + \frac{1}{2}A_1u_1^2(t) + \frac{1}{2}A_2u_2^2(t) + \sum_{i=1}^{5}\theta_i f_i(H, G, B, D, C), \tag{32}$$

where

$$\begin{aligned}
f_1(H,G,B,D,C) &= n - (\mu+\alpha_2+v)H(t) - \alpha_1(1-u_1(t))H(t) - \lambda(1-u_1(t))H(t)\\
&\quad + r_1u_1(t)B(t) + r_2u_1(t)D(t),\\
f_2(H,G,B,D,C) &= vH(t) - (\gamma_1+\gamma_2+\mu)G(t),\\
f_3(H,G,B,D,C) &= \lambda(1-u_1(t))H(t) - \delta B(t) - (\mu + r_1u_1(t))B(t),\\
f_4(H,G,B,D,C) &= \alpha_1(1-u_1(t))H(t) + \gamma_1G(t) + \delta B(t) + \frac{\kappa u_2(t)C(t)}{1+lC(t)}\\
&\quad - \eta D(t) - (\mu + r_2u_1(t))D(t),\\
f_5(H,G,B,D,C) &= \alpha_2H(t) + \gamma_2G(t) - \frac{\kappa u_2(t)C(t)}{1+lC(t)} - (\mu+\beta)C(t) + \eta D(t),
\end{aligned}$$

and θ_1, θ_2, θ_3, θ_4, and θ_5 are adjoint functions that need to be determined. Applying Pontryagin's maximum principle [34], we have the following theorem.

Theorem 7. For a given set of optimal controls (u_1^*, u_2^*) with associated optimal solutions H^*, G^*, B^*, D^*, and C^* to the model (26), there exist adjoint variables θ_1, θ_2, θ_3, θ_4, and θ_5 which satisfy the equations

$$\begin{aligned}
\theta_1'(t) &= \theta_1\mu + (\theta_1-\theta_2)v + (1-u_1)(\theta_1-\theta_4)\alpha_1 + (\theta_1-\theta_5)\alpha_2\\
&\quad + (1-u_1)(\theta_1-\theta_3)\lambda,\\
\theta_2'(t) &= \theta_2\mu + (\theta_2-\theta_4)\gamma_1 + (\theta_2-\theta_5)\gamma_2,\\
\theta_3'(t) &= \theta_3\mu + (\theta_3-\theta_1)u_1r_1 + (\theta_3-\theta_4)\delta,\\
\theta_4'(t) &= -1 + \theta_4\mu + (\theta_4-\theta_5)\eta + (\theta_4-\theta_1)u_2r_2,\\
\theta_5'(t) &= -1 + (\theta_5-\theta_4)\frac{\kappa u_2}{(1+lC)^2} + \theta_5(\mu+\beta),
\end{aligned}$$

with the tranversality conditions

$$\theta_1(t_f) = 0, \quad \theta_2(t_f) = 0, \quad \theta_3(t_f) = 0, \quad \theta_4(t_f) = 0, \quad \theta(t_f) = 0. \tag{33}$$

The optimal controls are given by

$$u_1^* = \min\left\{\max\left\{\frac{(\theta_3-\theta_1)(\lambda H + r_1 B) + (\theta_4-\theta_1)(\alpha_1 H + r_2 D)}{A_1}, 0\right\}, 1\right\}, \tag{34}$$

$$u_2^* = \min\left\{\max\left\{\frac{(\theta_5-\theta_4)\kappa C}{A_2(1+lC)}, 0\right\}, 1\right\}. \tag{35}$$

Proof. We can find the adjoint functions using the following equations:

$$\begin{aligned}
\theta_1'(t) &= -\frac{\partial \mathcal{H}}{\partial H} \\
&= \theta_1\mu + (\theta_1-\theta_2)v + (1-u_1)(\theta_1-\theta_4)\alpha_1 \\
&\quad + (\theta_1-\theta_5)\alpha_2 + (1-u_1)(\theta_1-\theta_3)\lambda, \\
\theta_2'(t) &= -\frac{\partial \mathcal{H}}{\partial G} \\
&= \theta_2\mu + (\theta_2-\theta_4)\gamma_1 + (\theta_2-\theta_5)\gamma_2, \\
\theta_3'(t) &= -\frac{\partial \mathcal{H}}{\partial B} \\
&= \theta_3\mu + (\theta_3-\theta_1)u_1 r_1 + (\theta_3-\theta_4)\delta, \\
\theta_4'(t) &= -\frac{\partial \mathcal{H}}{\partial D} \\
&= -1 + \theta_4\mu + (\theta_4-\theta_5)\eta + (\theta_4-\theta_1)u_2 r_2, \\
\theta_5'(t) &= -\frac{\partial \mathcal{H}}{\partial C} \\
&= -1 + (\theta_5-\theta_4)\frac{\kappa u_2}{(1+lC)^2} + \theta_5(\mu+\beta).
\end{aligned}$$

The transversality conditions are

$$\theta_1(t_f) = 0, \;\; \theta_2(t_f) = 0, \;\; \theta_3(t_f) = 0, \;\; \theta_4(t_f) = 0, \;\; \theta(t_f) = 0.$$

Moreover, the optimal controls u_1^* and u_2^* can be calculated using the following equations:

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial u_1} &= A_1 u_1 + (\theta_1-\theta_2)(\lambda H + r_1 B) + (\theta_1-\theta_4)(\alpha_1 H + r_2 D) = 0, \\
\frac{\partial \mathcal{H}}{\partial u_2} &= A_2 u_2 + (\theta_4-\theta_5)\frac{\kappa C}{1+lC} = 0.
\end{aligned}$$

which give

$$u_1 = \frac{(\theta_3-\theta_1)(\lambda H + r_1 B) + (\theta_4-\theta_1)(\alpha_1 H + r_2 D)}{A_1},$$

$$u_2 = \frac{(\theta_5 - \theta_4)\kappa C}{A_2(1 + lC)}.$$

Using the bounds of the controls u_1 and u_2, we get the optimal controls as follows:

$$u_1^* = \min\left\{\max\left\{\frac{(\theta_3 - \theta_1)(\lambda H + r_1 B) + (\theta_4 - \theta_1)(\alpha_1 H + r_2 D)}{A_1}, 0\right\}, 1\right\},$$
$$u_2^* = \min\left\{\max\left\{\frac{(\theta_5 - \theta_4)\kappa C}{A_2(1 + lC)}, 0\right\}, 1\right\}.$$

□

5 Numerical simulations

In this section, we perform a series of numerical simulations for the model (1) to validate the analytical results and study the effects of different model parameters on the model's dynamics using MATLAB. To write and compile code in MATLAB, we used the values of the model parameters (1) as given in Table 2. For these parameter values, the components of the equilibrium point T^* are calculated as $H^* = 956600$, $G^* = 298964$, $B^* = 212580$, $D^* = 897372$, and $C^* = 7900630$. For the Jacobian matrix J^* calculated at the equilibrium point T^*, the eigenvalues are obtained as $-0.9002 + 0.1297i$, $-0.9002 - 0.1297i$, 0.0700, -0.4413 and -0.2983, although one real part of eigenvalue is reported as positive, indicating the local asymptotic stability of the equilibrium point T^*. Figures 2 and 3 demonstrate that the solution trajectories of the system (1) start from different initial points in $H-B-C$ and $H-D-C$ space, respectively. All trajectories are observed to converge to the equilibrium point (H^*, B^*, C^*) irrespective of the initial points, indicating the global stability of T^*. Since these parameter values satisfy conditions (17)–(25), it follows that diabetes persists whenever these conditions hold.

Table 2: Values of parameters for the model (1).

Parameters	Values	Source	Parameters	Values	Source
n	600000	[5]	δ	0.8	[15]
μ	0.02	[5]	η	0.5	[5]
α_1	0.3	[5]	β	0.05	[11]
α_2	0.1	[5]	r_1	0.08	[11]
v	0.1	[5]	r_2	0.08	[11]
λ	0.2	[16]	κ	0.65	[26]
γ_1	0.2	[15]	l	0.0003	[26]
γ_2	0.1	[15]			

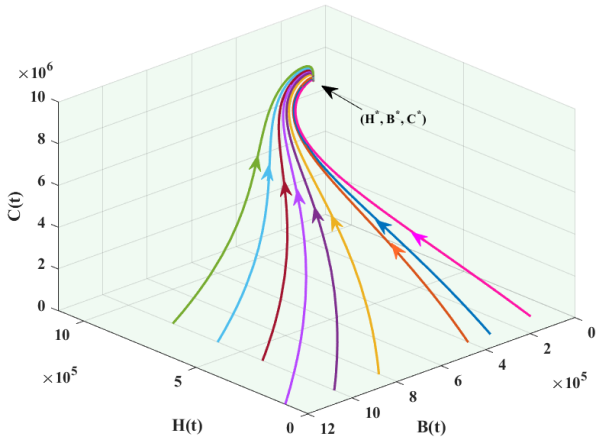


Figure 2: Global stability of (H^*, B^*, C^*) in $H - B - C$ plane

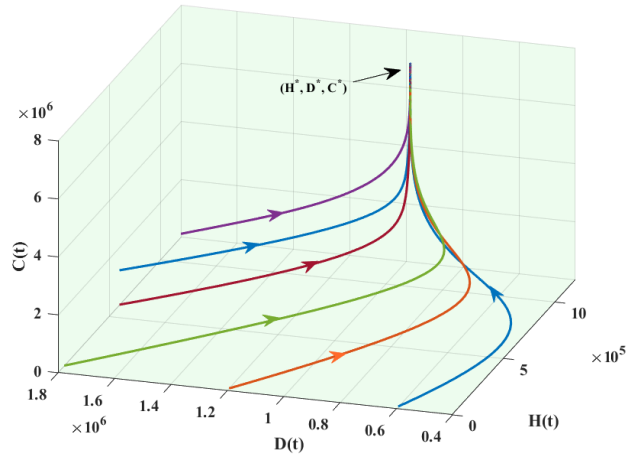


Figure 3: Global stability of (H^*, D^*, C^*) in $H - D - C$ plane

5.1 Effect of some key parameters on dynamics of the model (1)

We now examine the effects of key parameters α_1, λ, r_1, r_2, and κ on the dynamics of the model (1). Figure 4 shows the effects on the model system (1) as the parameter α_1 is varied, which denotes the rate at which the healthy population becomes diabetic without complications. On increasing the value of α_1, the healthy population (H), pre-diabetic population due to genetic

factors (G), and pre-diabetic population due to bad lifestyle (B) decrease, while the diabetic population without complications (D) and the diabetic population with complications increase (C). Thus, an increase in α_1 indicates a higher rate at which healthy individuals develop diabetes directly, bypassing the pre-diabetic stage. This leads to a decline in the healthy and pre-diabetic populations, while increasing the diabetic populations with and without complications. Biologically, this reflects poor early detection and prevention, resulting in faster disease progression and a higher overall diabetes burden.

In Figure 5, it can be seen that as the parameter value of λ increases, which represents the rate at which the healthy population become pre-diabetic due to bad lifestyles, the healthy population decreases. However, the pre-diabetic population due to a bad lifestyle, the diabetic population without complications and the diabetic population with complications increase. Biologically, an increase in λ signifies a higher influence of unhealthy lifestyle behaviors, causing more healthy individuals to transition into the pre-diabetic state due to poor lifestyle choices. As a result, the healthy population declines, while the pre-diabetic population due to bad lifestyle increases and subsequently contributes to higher diabetic populations with and without complications. This highlights the critical role of lifestyle-related risk factors in driving diabetes progression and complications.

Figure 6 shows the dynamics of the five population compartments of the model (1) for different values of the parameter r_1, which denotes the rate at which pre-diabetic individuals who adopt a healthy lifestyle return to the healthy class. The equilibrium level of the healthy population (H) is observed to increase, while the equilibrium level of the pre-diabetic population due to a poor lifestyle (B), the populations of diabetic individuals without complication (D) and diabetic individuals with complication (C) decrease with increasing value of r_1. Biologically, these results indicate that effective awareness-driven lifestyle interventions not only promote remission from pre-diabetes but also interrupt the disease progression pathway, increasing the healthy population and reducing diabetes-related complications.

The dynamics of the five population compartments of the model (1) for different values of the parameter r_2, representing the remission rate from diabetics without complication (D) to healthy (H), achieved through lifestyle interventions taking a high-protein diet and regular exercises, is shown in Figure 7. It is observed that the equilibrium level of the healthy class (H) increases, and the equilibrium levels of diabetics without complications (D) and with complications (C) decrease as the value of r_2 increases. The increase in the healthy class is more distinct compared to both diabetics classes. Furthermore, the equilibrium levels of pre-diabetic populations G and B increase with increasing value of r_2. This behavior suggests that an increase in r_2, which represents the rate of effective medical management or early detection, promotes recovery or stabilization of diabetic individuals, thereby reducing the burden of advanced disease. However,

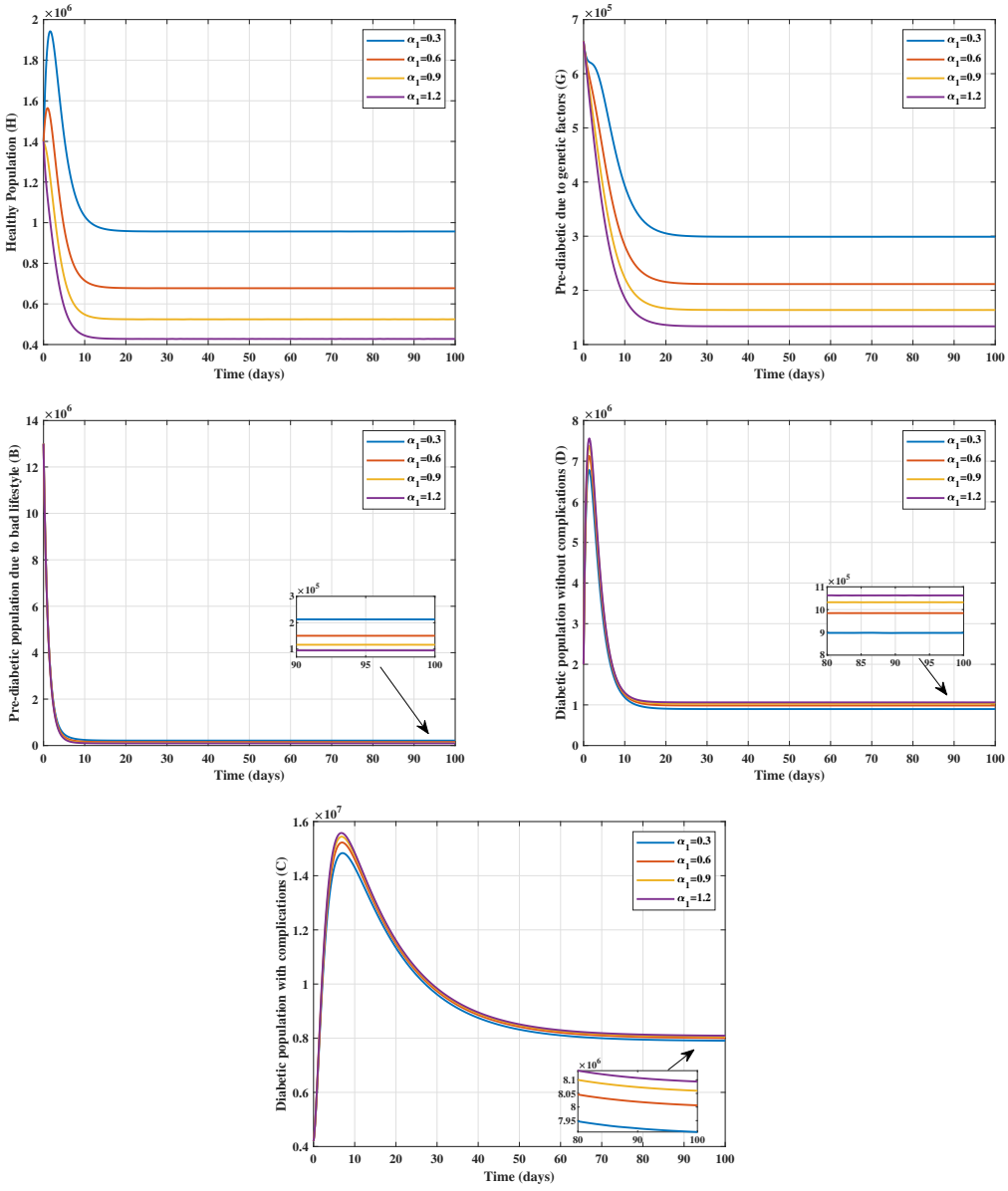


Figure 4: Dynamics of the populations of model (1) while varying α_1.

enhanced screening and prolonged survival may result in a temporary accumulation of individuals in the pre-diabetic classes G and B, even as the overall diabetic population declines.

Figure 8 shows the effect of the parameter κ, which represents the treatment rate for the diabetic population, on the model (1). As the parameter value κ increases, the healthy population (H), the pre-diabetic population due to a bad lifestyle (B), and the diabetic population without

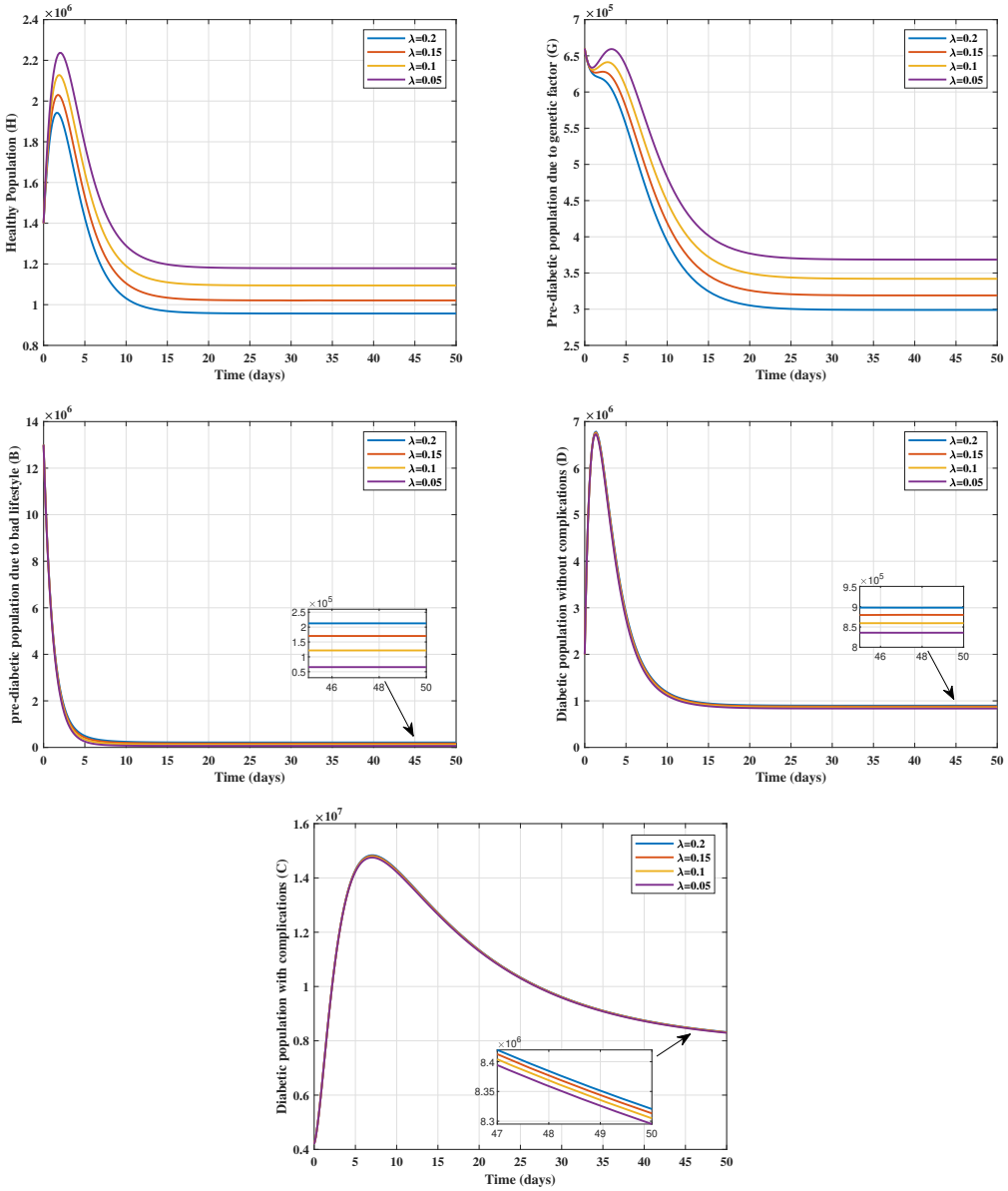


Figure 5: Dynamics of the populations of model (1) while varying λ.

complications (D) increase, while the diabetic population with complications (C) decreases significantly. Thus, an increase in the treatment rate κ reflects improved medical care and access to effective treatment for diabetic individuals. This leads to a significant reduction in the diabetic population with complications (C), as timely treatment prevents disease progression and promotes recovery. Consequently, the populations of healthy individuals (H), pre-diabetics due

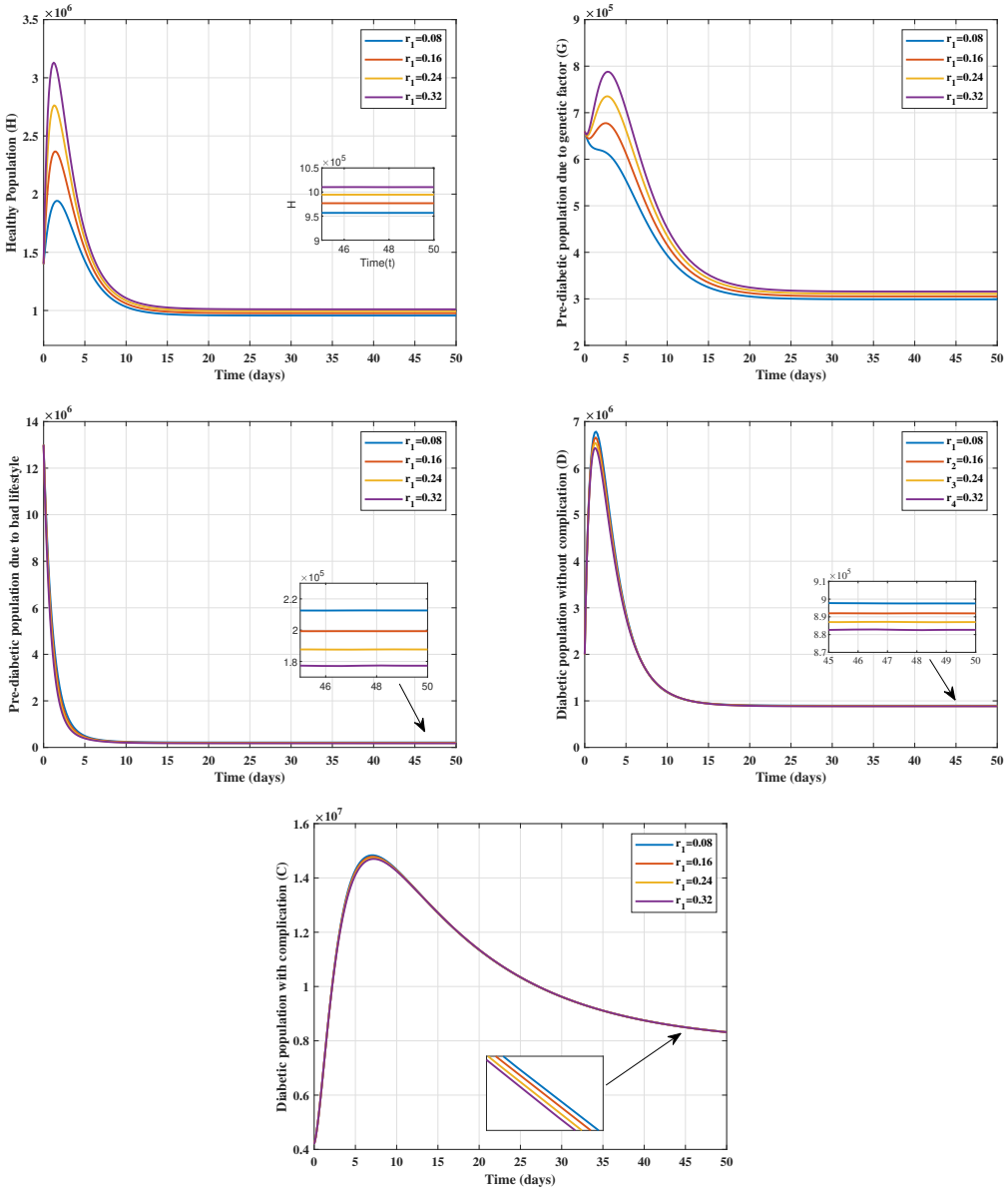


Figure 6: Dynamics of the populations of model (1) while varying r_1.

to bad lifestyle (B), and diabetics without complications (D) increase, indicating better disease management and overall improvement in population health.

Overall, the results indicate that unhealthy lifestyle adoption significantly alters the transient dynamics of the population and reduces the long-term healthy population, while increasing the disease burden at early stages. These findings highlight the critical biological role of lifestyle

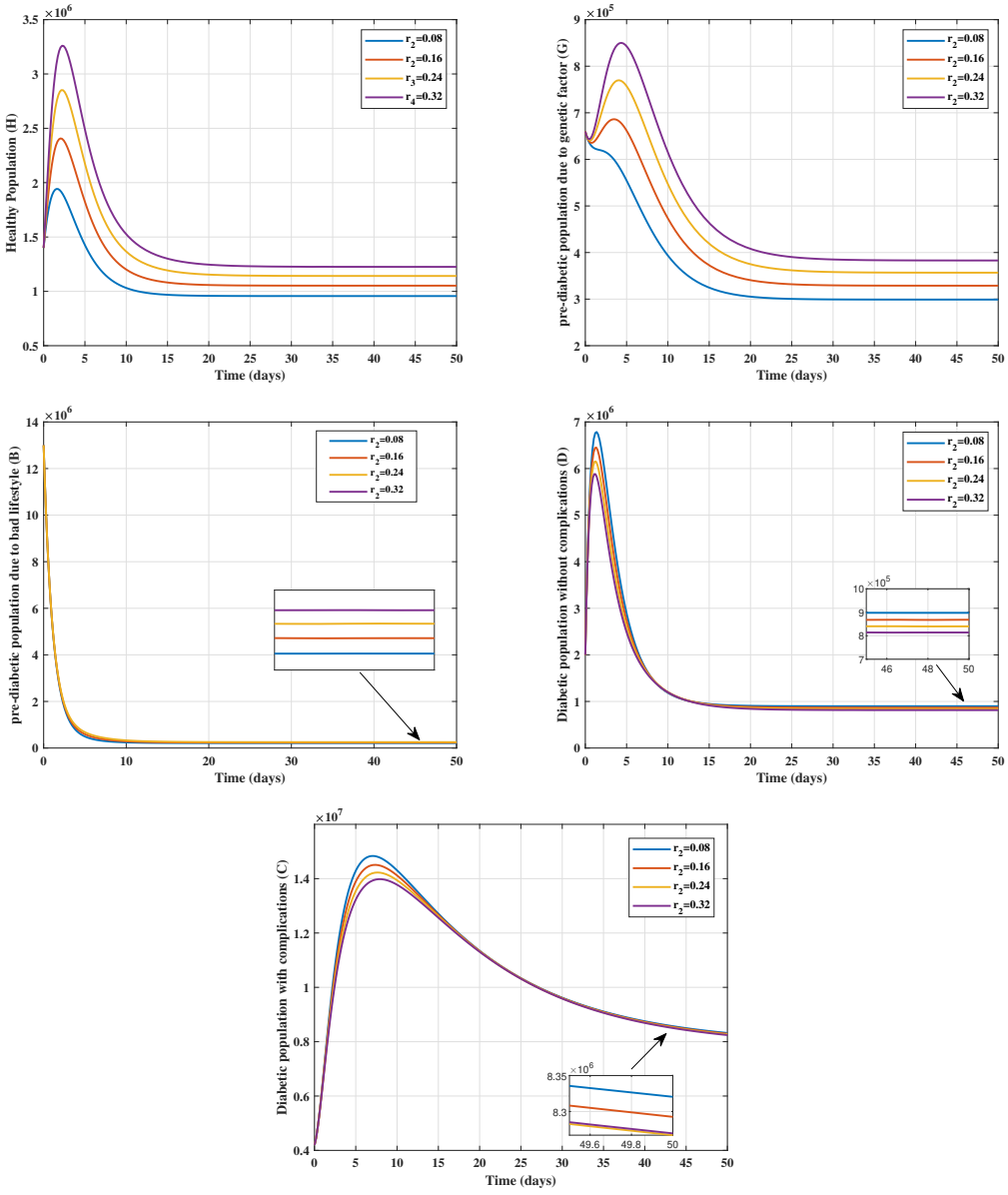


Figure 7: Dynamics of the populations of model (1) while varying r_2.

modification and preventive strategies in controlling the onset and progression of diabetes mellitus at the population level.

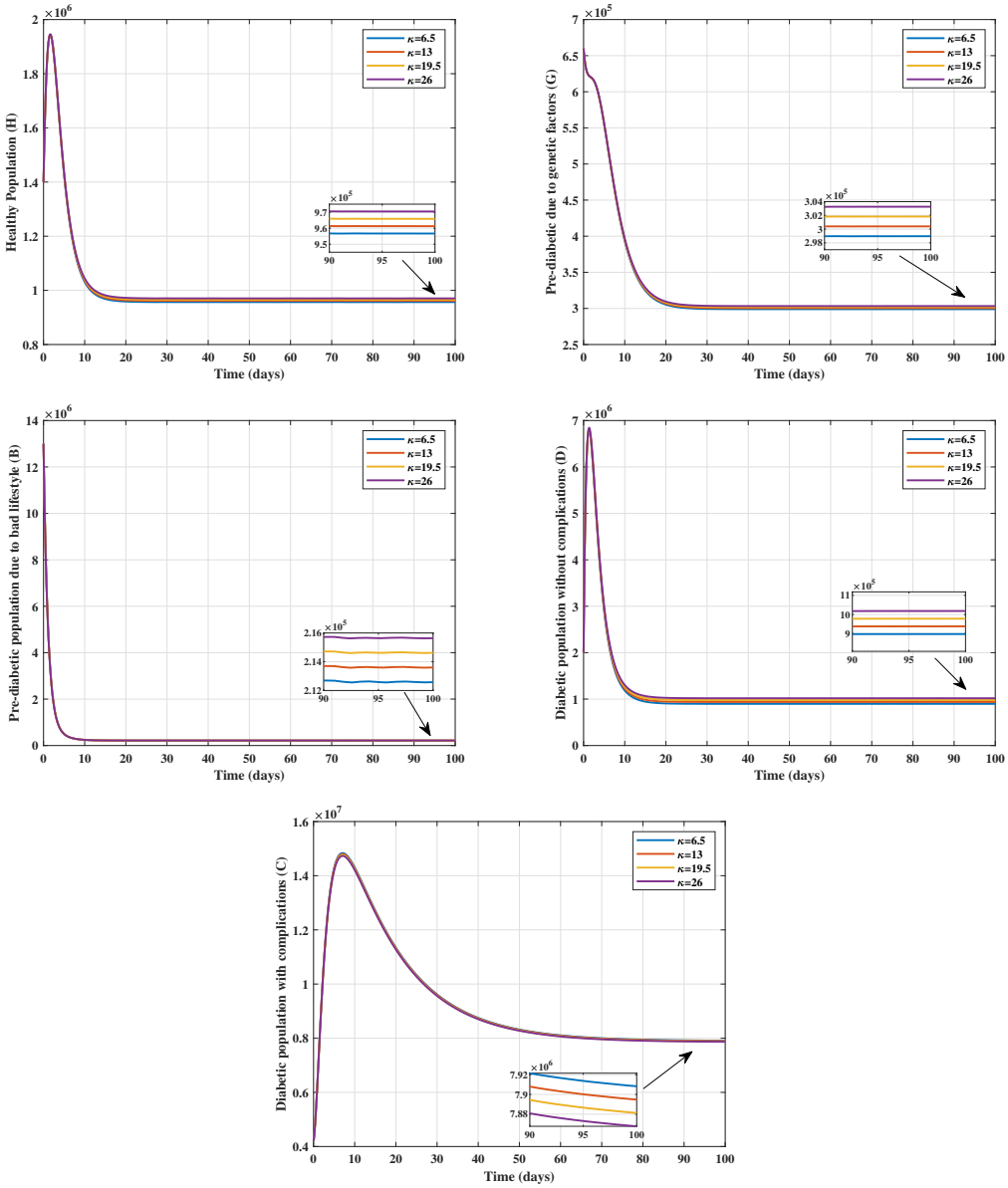


Figure 8: Dynamics of the populations of model (1) while varying κ.

5.2 Numerical simulations of the optimal control problem

In this part, we solve the optimality system by means of numerical simulations. For this purpose, we use an iterative method that involves forward solving the state system and backwards solving the adjoint system. In the first iteration, we make an initial guess for the controls. Before the

next iteration, we use the characterization to update the controls accordingly. This process is continued until successive iterations have converged. For solving the optimal control problem, we choose the weight functions as: $A_1 = 1$ and $A_2 = 400$ and the initial values for the model populations are considered as: $H(0) = 1400000$, $G(0) = 660000$, $B(0) = 13000000$, $D(0) = 2000000$, and $C(0) = 4200000$. The number of model populations after 100 months in both situations with and without controls is shown in Table 3. The values of all other parameters are considered as given in Table 2,

After 100 Months	Without control	With control
Healthy population(H)	956678	2835390
Pre-Diabetic due to bad lifestyle population(B)	212580	11.1205
Diabetics without complications(D)	897597	296860
Diabetics with complications(C)	7908450	7434080

Table 3: Evolution of the different diabetic populations after 100 months

Figure 9 illustrates the time evolution of the different population compartments of the diabetes model, comparing scenarios without control and with optimal control over the time interval $0 \leq t \leq 100$ days. It is clear from Figure 9(a) that the healthy population increases over time and eventually reaches a steady state. Under optimal control, the healthy population attains a higher equilibrium level compared to the scenario without any controlled measures. This suggests that awareness programs, lifestyle interventions, preventive strategies, and improving healthcare infrastructure are effective in sustaining individuals in the healthy class. Optimal control reduces the progression of healthy individuals toward pre-diabetes, thereby bolstering overall population health. From Figure 9(b), the results indicate that the pre-diabetic population is initially high and gradually declines as time progresses. Under optimal control, the population size reduced significantly throughout the simulation period. This indicates that interventions such as lifestyle modification, physical exercise, and dietary awareness effectively limit the growth of pre-diabetes caused by unhealthy behaviour. Optimal control reduces the burden of pre-diabetes, thereby preventing its progression to diabetes. Figure 9(c) shows that the diabetic population without complications rises quickly and reaches a stable level. Under optimal control, this compartment remains at a lower equilibrium compared to the uncontrolled scenario. This indicates that early awareness, lifestyle interventions, timely detection, continuous monitoring, improved healthcare and diagnostic facilities reduce the number of individuals progressing to the complicated diabetic stage. The implementation of the optimal control delays disease progression and improves glycaemic regulation, leading to a decrease in the diabetic population without complications. Figure 9(d) illustrates that, under optimal control, the population of diabetics with complications initially increases, reaches a peak, and then gradually declines, approaching a lower steady-state

level. Without any control measures, this population remains relatively higher. This highlights the effectiveness of clinical interventions, treatment compliance, and strengthened healthcare resources in reducing severe diabetic complications. Optimal control plays a crucial role in preventing the worsening of diabetes, reducing long-term health risks, mortality, and healthcare costs. The comparison clearly shows that optimal control strategies:

- Increase the number of healthy individuals,
- Reduce the number of pre-diabetic cases due to an unhealthy lifestyle,
- Limit the progression to diabetes,
- Effectively minimize diabetic-related complications.

6 Conclusions

In this manuscript, we have formulated a mathematical model of populations of Diabetes Mellitus with five compartments: Healthy population, pre-diabetics due to the genetic factors, pre-diabetics due to bad lifestyle, diabetic population without complications, and diabetics with complications. We established some basic properties of the model, like existence, nonnegativity and boundedness of the solutions, which are necessary for biological feasibility of the model. Furthermore, various conditions for local and global stability of a nonnegative equilibrium point have been derived. Then, to minimize the number of diabetic population and pre-diabetic populations due to a bad lifestyle, an optimal control problem is formulated using two control measures, which involve lifestyle intervention such as regular exercise, healthy diet, etc. and clinical intervention such as improving the healthcare infrastructure, enhanced diagnostic centre, etc. Then, we employed different results of control theory and obtained the characterizations of the optimal control. The problem is solved numerically using a forward-backwards iterative technique, resulting in an optimal control strategy which can maximize healthy population and minimizes the pre-diabetic population due to a bad lifestyle and the diabetic population to a significantly low level with proper diet, routine exercise, and enhanced medical facilities at minimum cost. The following conclusion can be drawn from this study:

- A mathematical model is formulated to study the dynamical relationship between the healthy population, the pre-diabetic population due to bad lifestyle and genetic factors, and the diabetic population with and without complications.

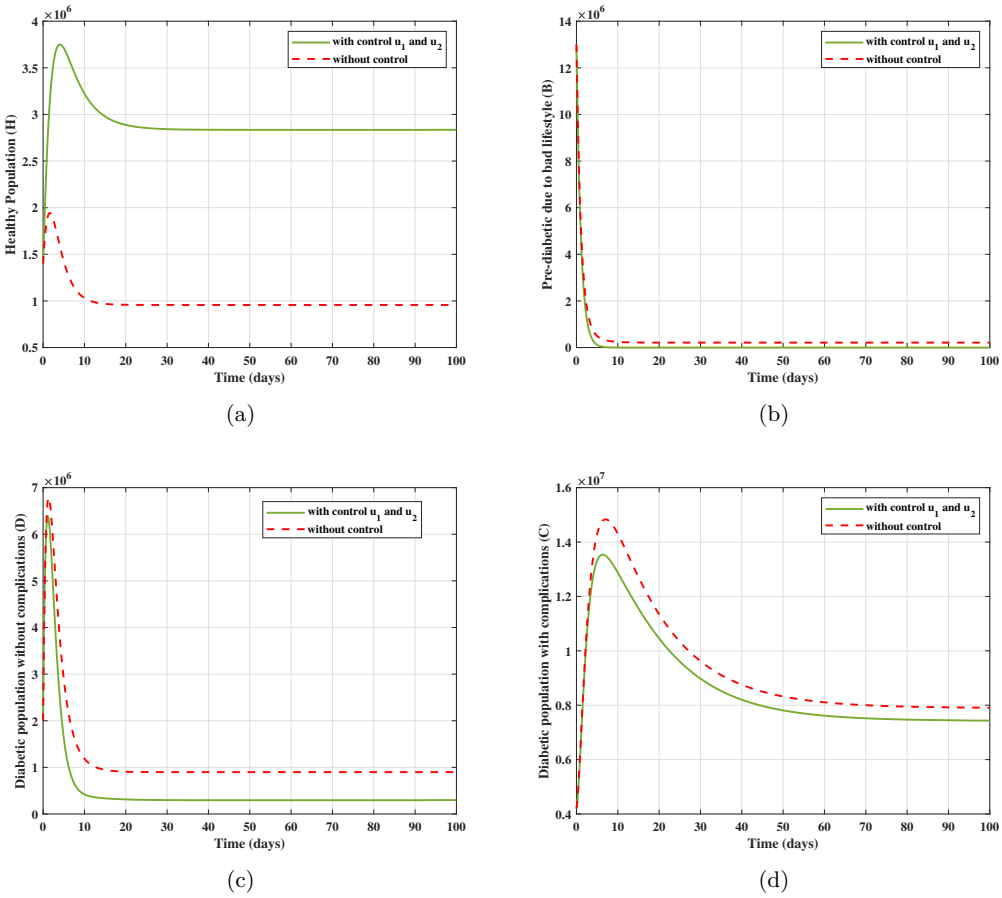


Figure 9: Evolution of model populations without and with optimal control u^*.

- Optimal control study reveals that increasing awareness among the pre-diabetic population, along with clinical interventions to improve the treatment rate of diabetics with complications, leads to an increase in the number of healthy individuals.

Although this study provides us with an optimal strategy to control the number of pre-diabetics due to a bad or irregular lifestyle via lifestyle intervention and adequate treatment facility, the model studied here can be further extended through the consideration of some other factors like delays caused when a healthy man becomes pre-diabetic, or the delay caused when a diabetic without complication becomes diabetic with complications, and so on, which will become the model more realistic. Also, this model can be extended by incorporating the memory effect through fractional order modeling. These improvements are left for future work.

Appendix A

Let $X = \begin{pmatrix} H(t) \\ G(t) \\ B(t) \\ D(t) \\ C(t) \end{pmatrix}$ and $\phi(X) = \begin{pmatrix} \frac{dH(t)}{dt} \\ \frac{dG(t)}{dt} \\ \frac{dB(t)}{dt} \\ \frac{dD(t)}{dt} \\ \frac{dC(t)}{dt} \end{pmatrix}$. The system (1) can be rewritten in the following form:

$$\phi(X) = \dot{A} = AX + B(X), \tag{36}$$

where

$$A = \begin{pmatrix} -(\mu+\alpha_1+\alpha_2+v+\lambda) & 0 & r_1 & r_2 & 0 \\ v & -(\gamma_1+\gamma_2+\mu) & 0 & 0 & 0 \\ \lambda & 0 & -(\delta+\mu+r_1) & 0 & 0 \\ \alpha_1 & \gamma_1 & \delta & -(\mu+\eta+r_2) & \beta_2 \\ \alpha_2 & \gamma_2 & 0 & \eta & -(\mu+\beta_1+\beta_2) \end{pmatrix}$$

and

$$B(X) = \begin{pmatrix} n \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then

$$||\phi(X_1) - \phi(X_2)|| \leq ||A||.||(X_1 - X_2||.$$

Thus, it is obtained that the function Φ is uniformly Lipschitz continuous with the restrictions $H(0) \geq 0$, $G(0) \geq 0$, $B(0) \geq 0$, $D(0) \geq 0$, and $C(0) \geq 0$. Hence, a unique solution of the system (1) exists.

Appendix B

From the first equation of the system (1), we obtain

$$\begin{aligned} \frac{dH(t)}{dt} &= n - (\mu+\alpha_1+\alpha_2+v+\lambda)H(t) + r_1B + r_2D, \\ \Longrightarrow \frac{dH(t)}{dt} &\geq -(\mu+\alpha_1+\alpha_2+v+\lambda)H(t), \end{aligned}$$

$$\implies \frac{dH(t)}{dt} + kH(t) \geq 0,$$

where $k = \mu + \alpha_1 + \alpha_2 + v + \lambda \geq 0$.

Multiplying by the integrating factor $e^{kt} \geq 0$ and solving the corresponding differential equation we have,

$$\begin{aligned} &\frac{d}{dt}\{H(t)e^{kt}\} = 0 \\ \implies &H(t) = H(0)e^{-kt}. \end{aligned}$$

Using the comparison theorem, we obtain $H(t) \geq 0$ for all t.

Similarly, we prove that $G(t) \geq 0$, $B(t) \geq 0$, $D(t) \geq 0$, and $C(t) \geq 0$ for all $t \geq 0$.

Appendix C

Since

$$N(t) = H(t) + G(t) + B(t) + D(t) + C(t),$$

by differentiating with respect to t, we get

$$\begin{aligned} &\frac{dN(t)}{dt} = \frac{dH(t)}{dt} + \frac{dG(t)}{dt} + \frac{dB(t)}{dt} + \frac{dD(t)}{dt} + \frac{dC(t)}{dt} \\ \implies &\frac{dN(t)}{dt} = n - \mu N(t) - \beta_1 C(t) \\ \implies &\frac{dN(t)}{dt} \leq n - \mu N(t) \\ \implies &N(t) \leq \frac{n}{\mu} - (\frac{n}{\mu} - N(0))e^{-\mu t}, \end{aligned}$$

where $N(0)$ represents the initial values of the population. Thus, $\lim\limits_{t\to\infty} \sup N(t) = \frac{n}{\mu}$, which implies that the region Ω is a positively invariant set for the system (1).

Funding

This research received no external funding.

Conflicts of Interest

All authors jointly worked on the results and they read and approved the final manuscript, and also declare that they have no competing interests.

References

[1] Aljohani, A., Shokri, A. and Mukalazi, H. *Analyzing the dynamic patterns of COVID-19 through nonstandard finite difference scheme*, Sci. Rep., 14(1) (2024), 8466.

[2] Aljohani, A., Hussain, A. and Shokri, A. *Mathematical modeling and the spreading of the cholera epidemic through numerical methods*, Comput. Syst. Oncol., 5(2) (2025), e70006.

[3] Banzi, W., Kambutse, I., Dusabejambo, V., Rutaganda, E., Minani, F., Niyobuhungiro, J., Mpinganzima, L. and Ntaganda, J.M. *Mathematical modelling of glucose–insulin system and test of abnormalities of type 2 diabetic patients*, Int. J. Math. Math. Sci., 2021(1) (2021), 6660177.

[4] Borah, P.B., Nath, B.J., Nath, K.C. and Sarmah, H.K. *Stability analysis of a nonlinear mathematical model for COVID-19 transmission dynamics*, Commun. Math. Biol. Neurosci., (2023), Article ID 26.

[5] Boutayeb, W., Lamlili, M.E., Boutayeb, A. and Derouich, M. *The dynamics of a population of healthy people, pre-diabetics and diabetics with and without complications with optimal control*, In: Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015 (MedCT 2015), Vol. 1, Springer Int. Publ., Cham (2016), 463–471.

[6] Chen, Y. and Zhang, J. *Dynamic analysis and optimal control of influenza model with discrete age structure*, Math. Methods Appl. Sci., 47(6) (2024), 4260–4282.

[7] Das, P., Das, S., Das, P., Rihan, F.A., Uzuntarla, M. and Ghosh, D. *Optimal control strategy for cancer remission using combinatorial therapy: a mathematical model-based approach*, Chaos Solitons Fractals, 145 (2021), 110789.

[8] Derouich, M., Boutayeb, A., Boutayeb, W. and Lamlili, M. *Optimal control approach to the dynamics of a population of diabetics*, Appl. Math. Sci., 8(56) (2014), 2773–2782.

[9] El Mansouri, A., Smouni, I., Khajji, B., Labzai, A., Belam, M. and Tidli, Y. *Optimal control in a mathematical model of a spread of the obesity epidemic and its impact on diabetes*, Commun. Math. Biol. Neurosci., (2023), Article ID 74.

[10] Fleming, W.H. and Rishel, R.W. *Deterministic and stochastic optimal control*, Springer Science & Business Media, 2012.

[11] Gupta, R. and Kumar, S. *Analysis of fractional-order population model of diabetes and effect of remission through lifestyle intervention*, Int. J. Appl. Comput. Math., 7(2) (2021), 53.

[12] Hallberg, S.J., Gershuni, V.M., Hazbun, T.L. and Athinarayanan, S.J. *Reversing type 2 diabetes: a narrative review of the evidence*, Nutrients, 11(4) (2019), 766.

[13] Hayes, C. and Kriska, A. *Role of physical activity in diabetes management and prevention*, J. Am. Diet. Assoc., 108(4) (2008), S19–S23.

[14] International Diabetes Federation. *IDF diabetes atlas*, 10th ed., International Diabetes Federation, Brussels (2021). Available at: https://www.idf.org.

[15] Kouidere, A., Khajji, B., Balatif, O. and Rachik, M. *A multi-age mathematical modeling of the dynamics of population diabetics with effect of lifestyle using optimal control*, J. Appl. Math. Comput., 67(1) (2021), 375–403.

[16] Kouidere, A., Labzai, A., Ferjouchia, H., Balatif, O. and Rachik, M. *A new mathematical modeling with optimal control strategy for the dynamics of population of diabetics and its complications with effect of behavioral factors*, J. Appl. Math., 2020(1) (2020), 1943410.

[17] Kumari, P., Singh, S. and Singh, H.P. *Bifurcation and stability analysis of glucose–insulin regulatory system in the presence of β-cells*, Iran. J. Sci. Technol. Trans. A Sci., 45(5) (2021), 1743–1756.

[18] Marín-Peñalver, J.J., Martín-Timón, I., Sevillano-Collantes, C. and del Cañizo-Gómez, F.J. *Update on the treatment of type 2 diabetes mellitus*, World J. Diabetes, 7(17) (2016), 354–395.

[19] Ma, M. and Li, J. *Dynamics of a glucose–insulin model*, J. Biol. Dyn., 16(1) (2022), 733–745.

[20] Misra, A.K. and Kumari, M. *Modeling the effects of TV and social media advertisements on diabetes*, J. Biol. Syst., 31(3) (2023), 1099–1124.

[21] Mollah, S. and Biswas, S. *Effect of awareness program on diabetes mellitus: deterministic and stochastic approach*, J. Appl. Math. Comput., 66(1) (2021), 61–86.

[22] Mollah, S. and Biswas, S. *Optimal control for the complication of type 2 diabetes: the role of awareness programs by media and treatment*, Int. J. Dyn. Control, 11(2) (2023), 877–891.

[23] Mutyambizi, C., Pavlova, M., Chola, L., Hongoro, C. and Groot, W. *Cost of diabetes mellitus in Africa: a systematic review of existing literature*, Glob. Health, 14(1) (2018), 3.

[24] Naik, P.A., Yeolekar, B.M., Owolabi, K.M. and Yeolekar, M. *Modeling and analyzing glucose–insulin interactions during diabetes through fractional dynamics in presence of glucagon*, J. Adv. Res. (2026), accepted.

[25] Nasir, H. *Stability analysis and optimal control of a five-state diabetic population model*, J. Stat. Manag. Syst., 25(1) (2022), 245–267.

[26] Nasir, H. *Hopf bifurcation analysis for a diabetic population model with two delays and saturated treatment*, Phys. Scr., 96(12) (2021), 125013.

[27] Nath, B.J., Dehingia, K., Mishra, V.N., Chu, Y.M. and Sarmah, H.K. *Mathematical analysis of a within-host model of SARS-CoV-2*, Adv. Differ. Equ., 2021(1) (2021), 113.

[28] Nath, B.J., Sarmah, H.K. and Maurer, H. *An optimal control strategy for antiretroviral treatment of HIV infection in presence of immunotherapy*, Qual. Theory Dyn. Syst., 21(2) (2022), 30.

[29] Nath, B.J., Sadri, K., Sarmah, H.K. and Hosseini, K. *An optimal combination of antiretroviral treatment and immunotherapy for controlling HIV infection*, Math. Comput. Simul., 217 (2024), 226–243.

[30] Ong, K.L., Stafford, L.K., McLaughlin, S.A., Boyko, E.J., Vollset, S.E., Smith, A.E., Dalton, B.E., Duprey, J., Cruz, J.A., Hagins, H. and Lindstedt, P.A. *Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021*, Lancet, 402 (2023), 203–234.

[31] Pal, K.K., Rai, R.K., Tiwari, P.K. and Misra, A.K. *Mathematical modeling of control measures for preventing kidney failure and managing diabetes: sensitivity analysis and optimal strategies*, J. Appl. Math. Comput., 71(5) (2025), 6455–6487.

[32] Pal, K.K., Rai, R.K., Tiwari, P.K. and Misra, A.K. *Role of awareness programs on diabetes prevention and control of viral infection: a study of optimal control*, Eur. Phys. J. Plus, 140(2) (2025), 150.

[33] Permatasari, A.H., Tjahjana, R.H. and Udjiani, T. *Existence and characterization of optimal control in mathematics model of diabetics population*, J. Phys. Conf. Ser., 983(1) (2018), 012069.

[34] Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mishchenko, E.F. *The mathematical theory of optimal processes*, John Wiley & Sons, London, UK (1962).

[35] Rai, R.K., Pal, K.K., Tiwari, P.K., Martcheva, M. and Misra, A.K. *Impact of social media and word-of-mouth on the transmission dynamics of communicable and non-communicable diseases*, Int. J. Biomath., 16 (2024), 2450094.

[36] Urbina, G., Riahi, D.N. and Bhatta, D. *Mathematical modeling of nonlinear blood glucose–insulin dynamics with beta cells effect*, Appl. Appl. Math., 15(1) (2020), 10.

[37] World Health Organization. *Report of a WHO/IDF consultation*, World Health Organization (2016).

[38] Xu, H.Y., Khalsaraei, M.M., Shokri, A., Ramos, H. and Molayi, M. *Nonstandard finite difference schemes with positivity property with application to a mathematical model of HIV infection*, Appl. Comput. Math., 23(4) (2024), 541–557.

[39] Zhonghua, Z. and Yaohong, S. *Qualitative analysis of a SIR epidemic model with saturated treatment rate*, J. Appl. Math. Comput., 34(1) (2010), 177–194.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1401–1431
https://doi.org/10.22067/ijnao.2026.96797.1789
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Mathematical modeling and control strategies for meningitis transmission

M. Baroudi*, B. Bettioui, M. Belam and A. Labzai

Abstract

Meningitis is one of the world's most alarming infectious diseases, characterized by rapid progression, high fatality, and severe long-term complications. Despite advances in vaccination and surveillance, recurrent epidemics—especially across Africa's "meningitis belt"—continue to threaten millions. This study introduces a novel compartmental model, $SEAIVHRQ$, integrating both medical and behavioral aspects of meningitis dynamics. The model divides

*Corresponding author

Received 2 December 2025; revised 6 May 2026; accepted 20 July 2026

Mohamed Baroudi
Laboratory of Advanced Research in Industrial and Logistic Engineering (LARILE), Applied Mathematics Team, Department of Mathematics and Computer Science, ENSEM, Hassan II University of Casablanca, Morocco. e-mail: m.mohamed.baroudi@gmail.com

Benyounes Bettioui
Centre Régional des Métiers de l'Éducation et de la Formation (CRMEF), 20340 Derb Ghalef, Casablanca, Maroc. e-mail: benyounes.bettioui@gmail.com

Mohamed Belam
Laboratory of Advanced Research in Industrial and Logistic Engineering (LARILE), Applied Mathematics Team, Department of Mathematics and Computer Science, ENSEM, Hassan II University of Casablanca, Morocco. e-mail: m.belam@gmail.com

Abderrahim Labzai
Laboratory of Analysis Modeling and Simulation, Department of Mathematics and Computer Science, Faculty of Sciences Ben M'sik, Hassan II University of Casablanca, Morocco. e-mail: labzaiabdo1977@gmail.com

How to cite this article
Baroudi, M., Bettioui, B., Belam, M. and Labzai, A., Mathematical modeling and control strategies for meningitis transmission. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1401-1431. https://doi.org/10.22067/ijnao.2026.96797.1789

the population into eight classes, incorporating vaccination and voluntary quarantine to reflect realistic public health conditions. Three time-dependent controls are applied to reduce transmission from asymptomatic and symptomatic individuals and to enhance awareness and treatment adherence. Using Pontryagin's Maximum Principle, we determine optimal strategies that minimize infection prevalence, hospitalization, and intervention costs. Numerical simulations show that combined preventive, treatment, and awareness measures substantially reduce infection peaks and accelerate recovery. This framework extends classical models and supports the WHO's "Defeating Meningitis by 2030" roadmap, offering practical insights for designing cost-effective and sustainable control strategies in resource-limited health systems.

AMS subject classifications (2020): 49J15, 93C10, 92B05, 93A30.

Keywords: Meningitis; Pontryagin's maximum principle; MATLAB; Bacterial meningitis.

1 Introduction

Meningitis is a severe infection of the meninges, the protective membranes surrounding the brain and the spinal cord. According to the World Health Organization (WHO), bacterial meningitis is among the most devastating infectious diseases, causing high morbidity and mortality worldwide [1, 44, 4]. The main bacterial agents include *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Group B Streptococcus*, *Listeria monocytogenes*, *Escherichia coli*, and *Mycobacterium tuberculosis* [3, 4, 5]. Transmission typically occurs from person to person through coughing, sneezing, or close contact with infected individuals or asymptomatic carriers [1, 6, 7]. The clinical symptoms are characterized by a sudden onset of fever, headache, stiff neck, nausea, vomiting, photophobia, and confusion [1, 2]. The case fatality rate varies from 3–10% in developed countries to as high as 20% in the African "meningitis belt," a region stretching from Senegal to Ethiopia [2, 7, 8]. Moreover, nearly 20% of survivors may develop long-term neurological sequelae such as deafness, epilepsy, cerebral palsy, speech disorders, limb loss, or intellectual disability [1, 2]. These alarming statistics highlight the serious threat of meningitis to human health and justify the need for advanced strategies of prevention and control [9, 4, 10].

The WHO launched the global roadmap *"Defeating Meningitis by 2030"*, which aims to eliminate bacterial meningitis epidemics, reduce vaccine-preventable cases by 50%, cut mortality by 70%, and improve quality of life for those affected [4]. While remarkable progress has been made through vaccination campaigns, surveillance, and case management, challenges remain [9, 10, 7]. Outbreaks continue to occur, particularly during the dry season (November–May) in sub-Saharan Africa, where climatic conditions, population movements, and poverty exacerbate

transmission dynamics [2, 7, 8]. In this context, mathematical modeling provides a rigorous framework to understand disease dynamics, to test hypothetical scenarios in the absence of sufficient real-world data, and to design optimal strategies for epidemic control [6, 11, 5, 7, 8, 12].

Several studies in the literature have applied mathematical and optimal control approaches to meningitis and other infectious diseases [6, 11, 13, 5, 20, 23, 14, 15, 12, 16]. Deterministic compartmental models have been formulated to analyze co-infections, such as tuberculosis-hepatitis B, malaria-meningitis, hepatitis B-HIV, and pneumonia-meningitis [17, 19, 11, 13, 21, 18]. Simulation results from these studies consistently demonstrated that implementing preventive measures, treatment, and vaccination together can significantly reduce the burden of infection [19, 11, 13, 21, 18, 20, 15, 12, 16]. For example, Tilahun [42] developed an optimal control model for pneumonia and meningitis co-infection, showing that combined preventive and treatment strategies achieved substantial reductions in both diseases. Similarly, other researchers have investigated hepatitis B and hepatitis C co-infection under fractional-order models, highlighting the complexity of overlapping epidemics [21, 20, 23, 14, 15]. These works collectively underline the importance of designing robust control strategies adapted to the biological, social, and economic contexts of infectious diseases [5, 20, 8, 12, 16, 22, 24, 25, 26, 27, 30, 28, 29].

In this study, we focus on a $SEAIVHRQ$ model tailored to meningitis transmission. This model subdivides the population into the following compartments: Susceptible (S), Exposed (E), Asymptomatic infected (A), Symptomatic infected (I), Vaccinated (V), Hospitalized (H), Recovered (R), and Quarantined (Q). The introduction of compartment Q is a key innovation in this work: It accounts for individuals who are infected but prefer to remain at home rather than seek hospital care, reflecting the reality of quarantine and self-isolation in resource-limited settings. Similarly, the compartment V represents individuals vaccinated at the onset, highlighting the role of proactive immunization. These additions make the model more realistic and allow for the assessment of both medical and behavioral interventions.

To improve epidemic control, we incorporate three time-dependent control functions:

- $c_1(t)$: reducing transmission from asymptomatic carriers (A), who often spread infection silently within communities.

- $c_2(t)$: reducing transmission from symptomatic individuals (I) through isolation, treatment, and limiting their contact with healthy individuals.

- $c_3(t)$: promoting awareness and responsibility, which includes ensuring access to medications at home, clear explanations of dosage, and stressing the importance of completing treatment.

These controls reflect practical strategies such as vaccination, isolation, quarantine, awareness campaigns, and logistic support for home-based care. By simulating different combinations of controls, we aim to identify the most effective strategies for minimizing infections, reducing mortality, and optimizing healthcare resource use.

The value of this study lies in extending classical meningitis models to include realistic features such as voluntary quarantine and proactive vaccination, and in analyzing their impacts under optimal control theory. By doing so, we provide insights into how to balance preventive measures, treatment interventions, and awareness programs in a cost-effective manner. This approach aligns with global health objectives, particularly the WHO's 2030 roadmap, and offers practical implications for policymakers in regions where meningitis remains a pressing public health challenge.

The remainder of this article is organized as follows. In Section 2, we present the formulation of the SEAIVHRQ model, defining state variables and parameters. Section 3 introduces the controlled mathematical model and discusses its qualitative properties. Section 4 addresses the challenge of identifying the optimal control strategies and outlines the algorithm employed. Section 5 provides numerical simulations to illustrate the effects of different control categories and compares their effectiveness. Finally, Section 6 offers concluding remarks and perspectives for future work.

2 Formulation of the model

In this work, we formulated the $SEAIVHRQ$ compartmental model to describe the transmission dynamics of meningitis within a human population.

For this purpose, the total population N is stratified into distinct epidemiological classes: Susceptible individuals S, exposed individuals E, asymptomatic infected individuals A, symptomatic infected individuals I, vaccinated individuals V, hospitalized individuals H, recovered individuals R, and quarantined individuals Q. The overall human population at time t is expressed as $N = S + E + A + I + V + H + R + Q$. A schematic diagram representing the flow between compartments in the proposed model is presented in Figure 1.

Compartment S: The $S(t)$ compartment represents the susceptible population who can contract the disease. The inflow comes from new recruits $\Sigma(1-\alpha)$, while the outflow occurs as susceptible individuals become exposed due to contact with symptomatic infected individuals $\beta_1 \frac{SI}{N}$ or asymptomatic infected individuals $\beta_2 \frac{SA}{N}$. In addition, natural mortality μ decreases this compartment, and some susceptibles move into the vaccinated class V at rate k_1.

Compartment V: The $V(t)$ compartment represents vaccinated individuals who are protected

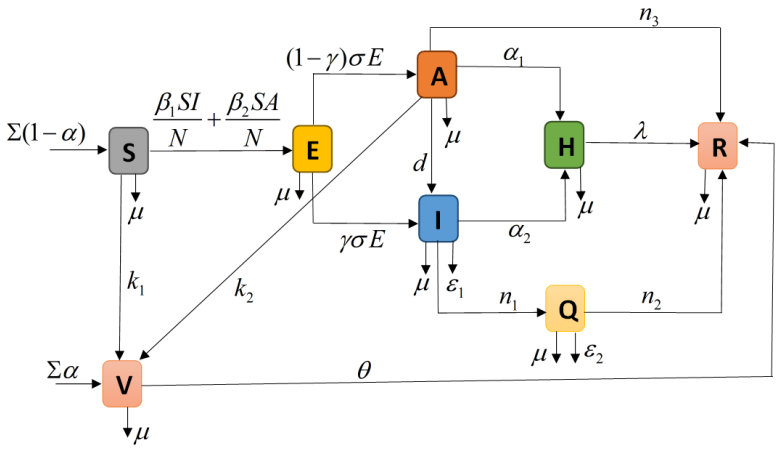


Figure 1: Model of population compartments.

from infection. This class increases through direct vaccination $\Sigma\alpha$, as well as vaccination of susceptible individuals at rate $k_1 S$ and asymptomatic individuals at rate $k_2 A$. The compartment decreases due to vaccine waning at rate θ and natural mortality μ.

Compartment E: The $E(t)$ compartment denotes the exposed individuals who have been infected but are not yet infectious. They are generated from susceptible individuals S through interactions with both symptomatic I and asymptomatic A infections. They progress to either the asymptomatic class A at rate $(1-\gamma)\sigma E$ or the symptomatic class I at rate $\gamma\sigma E$, and the compartment decreases due to natural mortality μ.

Compartment A: The $A(t)$ compartment represents asymptomatic infected individuals who can still transmit the infection. It increases as a fraction $(1-\gamma)\sigma E$ of exposed individuals become asymptomatic. The compartment decreases due to disease progression to the symptomatic class I at rate d, hospitalization at rate α_1, quarantine at rate n_3, and natural mortality μ.

Compartment I: The $I(t)$ compartment describes symptomatic infected individuals. It increases when a fraction $\gamma\sigma E$ of exposed individuals develop symptoms, and also when asymptomatic individuals A progress to symptomatic infection at rate d. The outflow occurs through hospitalization at rate α_2, disease-induced death ε_1, transfer to quarantine n_1, and natural mortality μ.

Compartment H: The $H(t)$ compartment represents hospitalized individuals. It receives inflow from asymptomatic individuals hospitalized at rate α_1 and symptomatic individuals hospitalized at rate α_2. The compartment decreases as patients either recover at rate λ or die from natural mortality μ.

Compartment R: The $R(t)$ compartment denotes recovered individuals who have acquired

immunity. Recovery comes from hospitalized individuals at rate λ, quarantined individuals at rate n_2, and asymptomatic individuals at rate n_3. The compartment decreases due to natural mortality μ.

Compartment Q: The $Q(t)$ compartment represents individuals in quarantine who are isolated at home. They enter from the symptomatic class I at rate n_1 and exit either by recovering at rate n_2, dying due to disease ε_2, or through natural mortality μ.

We perform an analysis of a system consisting of eight nonlinear differential equations:

$$\begin{cases} \frac{dS}{dt} = \Sigma(1-\alpha) - \beta_1 \frac{SA}{N} - \beta_2 \frac{SI}{N} - (k_1+\mu)S, \\ \frac{dV}{dt} = \Sigma\alpha + k_1 S + k_2 A - (\theta+\mu)V, \\ \frac{dE}{dt} = \beta_1 \frac{SA}{N} + \beta_2 \frac{SI}{N} - (\sigma+\mu)E, \\ \frac{dI}{dt} = \gamma\sigma E + dA - (\alpha_2 + n_1 + \varepsilon_1 + \mu)I, \\ \frac{dA}{dt} = (1-\gamma)\sigma E - (d + n_3 + \alpha_1 + \mu + k_2)A, \\ \frac{dH}{dt} = \alpha_1 A + \alpha_2 I - (\lambda+\mu)H, \\ \frac{dR}{dt} = \lambda H + n_3 A + n_2 Q + \theta V - \mu R, \\ \frac{dQ}{dt} = n_1 I - (n_2 + \varepsilon_2 + \mu)Q, \end{cases} \tag{1}$$

where $(S(0), V(0), E(0), I(0), A(0), H(0), R(0), Q(0)) \in \mathbb{R}^8_+$ represents the specified initial state.

2.1 Fundamental model characteristics

2.1.1 Nonnegativity of solutions

Theorem 1. If the initial conditions satisfy $S(0) \geq 0$, $V(0) \geq 0$, $E(0) \geq 0$, $I(0) \geq 0$, $A(0) \geq 0$, $H(0) \geq 0$, $R(0) \geq 0$ and $Q(0) \geq 0$, then the solutions to system (1) remain positive for all $t \geq 0$.

Proof. By examining the first equation in system (1), it becomes clear that

$$\begin{aligned} \frac{dS(t)}{dt} &= \Sigma(1-\alpha) + \left(-\frac{\beta_1 A}{N} - \frac{\beta_2 I}{N} - (k_1+\mu)\right) S(t) \\ &\geq -\left(\frac{\beta_1 A}{N} + \frac{\beta_2 I}{N} + k_1 + \mu\right) S(t), \end{aligned}$$

$$\frac{dS(t)}{dt} + \left(\frac{\beta_1 A}{N} + \frac{\beta_2 I}{N} + k_1 + \mu\right) S(t) \geq 0,$$

where $N(t) = \frac{\beta_1 A}{N} + \frac{\beta_2 I}{N} + k_1 + \mu$.

We obtain

$$\frac{dS(t)}{dt} + N(t)S(t) \geq 0,$$

$$\exp\left(\int_0^t N(s)ds\right)\frac{dS(t)}{dt} + N(t)\exp\left(\int_0^t N(s)ds\right)S(t) \geq 0,$$

$$\frac{d}{dt}\left(S(t)\exp\left(\int_0^t N(s)ds\right)\right) \geq 0,$$

$$\int_0^t \left(\frac{d}{ds}\left(S(s)\exp\left(\int_0^t N(s)ds\right)\right)\right)ds \geq 0,$$

$$S(t) \geq S(0)\exp\left(-\int_0^t N(s)ds\right),$$

$$S(t) \geq 0.$$

In the same manner, we have

$$\frac{dV}{dt} + (\theta + \mu).V = \Sigma\alpha + k_1 S + k_2 V,$$

$$\begin{aligned}\exp\left((\theta+\mu)t\right).\frac{dV}{dt} + (\theta+\mu).\exp\left((\theta+\mu)t\right).V \\ = \left(\Sigma\alpha + k_1 S + k_2 V\right).\exp\left((\theta+\mu)t\right),\end{aligned}$$

which can be rewritten as

$$\frac{d}{dt}\left(V(t).\exp\left((\theta+\mu)t\right)\right) = \left(\Sigma\alpha + k_1 S + k_2 V\right).\exp\left((\theta+\mu)t\right),$$

thus,

$$V(t).\exp\left((\theta+\mu)t\right)-V(0)=\int_0^{t_1}\left(\Sigma\alpha+k_1S(x)+k_2V(x)\right).\exp\left((\theta+\mu)t\right)dx.$$

Therefore,

$$\begin{aligned}V(t)=&V(0).\exp\left(-(\theta+\mu)t\right)\\&+\exp\left(-(\theta+\mu)t\right).\int_0^{t_1}\left(\Sigma\alpha+k_1S+k_2V\right).\exp\left((\theta+\mu)t\right)dx\\\geq&0\end{aligned}$$

Similarly, we illustrate $E(t)\geq 0,\ I(t)\geq 0,\ A(t)\geq 0,\ H(t)\geq 0,\ R(t)\geq 0$ *and* $Q(t)\geq 0$. □

2.2 Boundedness of the solutions

Theorem 2. The collection

$$\Omega_h=\left\{(S,V,E,I,A,H,R,Q)\in\mathbb{R}_+^8:0\leq S+V+E+I+A+H+R+Q\leq\frac{\Sigma}{\mu}\right\}$$

ensures positive invariance within system (1) based on the given initial conditions $S(t)\geq 0,\ V(t)\geq 0,\ E(t)\geq 0,\ I(t)\geq 0,\ A(t)\geq 0,\ H(t)\geq 0,\ R(t)\geq 0$ *and* $Q(t)\geq 0$.

Proof. As specified, $N=S+V+E+I+A+H+R+Q$.

Therefore

$$\frac{dN}{dt}=\Sigma-\varepsilon_1I-\varepsilon_2Q-\mu N,$$

$$\frac{dN}{dt}=\Sigma-\varepsilon_1I-\varepsilon_2Q-\mu N\leq\Sigma-\mu N,$$

$$\frac{dN}{dt}\leq\Sigma-\mu N,$$

$$N(t)\leq(1-e^{-\mu t})\frac{\Sigma}{\mu}+N(0)e^{-\mu t}.$$

As $t \to \infty$, it follows that limit $t \to \infty \sup N(t) = \frac{\Sigma}{\mu}$. Consequently, the region Ω is confirmed to be a positively invariant set for system (1),

$$N(t) \leq \frac{\Sigma}{\mu}.$$

□

2.3 Existence and uniqueness of solutions

Theorem 3. System (1), with the initial conditions

$$(S(0), V(0), E(0), I(0), A(0), H(0), R(0), Q(0))$$

has a unique solution. This ensures that, for any given set of initial values within the defined range, the system's behavior can be accurately predicted.

Proof. System (1) can be expressed as follows: $\Theta(Y) = KY + L(Y)$, where

$$Y = \begin{pmatrix} S \\ V \\ E \\ I \\ A \\ H \\ R \\ Q \end{pmatrix}, \qquad L(Y) = \begin{pmatrix} \frac{dS}{dt} \\ \frac{dV}{dt} \\ \frac{dE}{dt} \\ \frac{dI}{dt} \\ \frac{dA}{dt} \\ \frac{dH}{dt} \\ \frac{dR}{dt} \\ \frac{dQ}{dt} \end{pmatrix}$$

and

$$K = \begin{pmatrix} -(k_1+\mu) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_1 & -(\theta+\mu) & 0 & 0 & k_2 & 0 & 0 & 0 \\ 0 & & -(\sigma+\mu) & 0 & 0 & 0 & 0 & 0 \\ 0 & & \gamma\sigma & \omega_1 & d & 0 & 0 & 0 \\ 0 & 0 & (1-\gamma)\sigma & 0 & \omega_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 & \alpha_1 & -(\lambda+\mu) & 0 & 0 \\ 0 & \theta & 0 & 0 & n_3 & \lambda & -\mu & n_2 \\ 0 & 0 & 0 & n_1 & 0 & 0 & 0 & \omega_3 \end{pmatrix},$$

where $\omega_1 = -(\alpha_2 + n_1 + \varepsilon_1 + \mu)$, $\omega_2 = -(d + n_3 + \alpha_1 + \mu + k_2)$ *and* $\omega_3 = -(n_2 + \varepsilon_2 + \mu)$.

In addition to

$$L(Y) = \begin{pmatrix} \Sigma(1-\alpha) - \beta_1 \frac{SA}{N} - \beta_2 \frac{SI}{N} \\ \Sigma\alpha \\ \beta_1 \frac{SA}{N} + \beta_2 \frac{SI}{N} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

□

Assuming that Y_1 and Y_2 are solutions of (1), it follows that

$$\begin{aligned} |L(Y_1) - L(Y_2)| &\leq 2\left|\beta_1 \frac{S_1 A_1}{N} + \beta_2 \frac{S_1 I_1}{N} - \beta_2 \frac{S_2 A_2}{N} - \beta_2 \frac{S_2 I_2}{N}\right| \\ &\leq 2\left|\frac{\beta_1 S_1}{N}(A_1 - A_2) + \frac{\beta_1 A_2}{N}(S_1 - S_2) + \frac{\beta_2 S_1}{N}(I_1 - I_2)\right. \\ &\quad \left. + \frac{\beta_2 I_2}{N}(S_1 - S_2)\right| \\ &\leq 2\left(\frac{\beta_1 S_1}{N}|A_1 - A_2| + \frac{\beta_1 A_2}{N}|S_1 - S_2| + \frac{\beta_2 S_1}{N}|I_1 - I_2|\right. \\ &\quad \left. + \frac{\beta_2 I_2}{N}|S_1 - S_2|\right), \end{aligned}$$

where

$$|S| \leq \frac{\Sigma}{N}, \quad |I| \leq \frac{\Sigma}{N} \quad \textit{and} \quad |A| \leq \frac{\Sigma}{N}.$$

Therefore,

$$\begin{aligned} |L(Y_1) - L(Y_2)| \leq & \frac{2\Sigma}{N}\left(\beta_1 |A_1 - A_2| + \beta_1 |S_1 - S_2| + \beta_2 |I_1 - I_2|\right. \\ & \left. + \beta_2 |S_1 - S_2|\right). \end{aligned}$$

By factoring out the common term, we obtain

$$\|\Theta(Y_1) - \Theta(Y_2)\| \leq \left(\frac{2\beta_1\Sigma}{N} + \frac{2\beta_2\Sigma}{N}\right)|S_1 - S_2| + \frac{2\beta_1\Sigma}{N}|A_1 - A_2|$$

$$
\begin{aligned}
&+ \frac{2\beta_2\Sigma}{N} |I_1 - I_2| \\
&\leq Z \|Y_1 - Y_2\|,
\end{aligned}
$$

where

$$
Z = \max\left(\frac{2\beta_1\Sigma}{N} + \frac{2\beta_2\Sigma}{N}, \frac{2\beta_1\Sigma}{N}, \frac{2\beta_2\Sigma}{N}, \|K\|\right).
$$

Therefore, it follows that the mapping L possesses the property of being uniformly Lipschitz continuous. In addition, by enforcing the conditions $S(t) \geq 0$, $V(t) \geq 0$, $E(t) \geq 0$, $I(t) \geq 0$, $A(t) \geq 0$, $H(t) \geq 0$, $R(t) \geq 0$, and $Q(t) \geq 0$ within the positive cone $\mathbb{R}^8_+$, we guarantee the existence of at least one solution to the proposed system. This ensures that all the state variables remain nonnegative for any admissible initial data, thereby validating the consistency of the model under the prescribed assumptions. Consequently, the system is mathematically well-posed, which provides a rigorous foundation for further qualitative and numerical investigations [40].

3 The control mathematical model

The control strategies in this model are focused on three main variables aimed at reducing disease transmission and mitigating its health impact. The variable $c_1(t)$ represents preventive interventions directed towards asymptomatic individuals (A), such as enhancing protective awareness, maintaining social distancing, wearing masks, and adopting hygienic practices that limit the hidden spread of infection from this group. The variable $c_2(t)$ is associated with symptomatic individuals (I), through their rapid isolation, provision of appropriate treatment, and reduction of direct contact with others, which significantly lowers the transmission rate. In contrast, the variable $c_3(t)$ refers to the strategy that allows infected individuals to choose to remain at home while adhering to prescribed medications. This reduces the burden on hospitals and contributes to containing the spread of the disease by minimizing community-level interactions.

The integration of these three strategies ensures a balance between prevention, treatment, and home isolation, thereby providing greater effectiveness in controlling the disease while minimizing the associated health and social costs.

Category 1: The combination of personal protection measures (c_1), vaccination (c_2), and awareness raising program (c_3): $c_1 \neq 0,\ c_2 \neq 0,\ c_3 \neq 0$.

Category 2: The combination of personal protection measures (u_1) and awareness raising program (c_3): $c_1 \neq 0,\ c_3 \neq 0$.

Category 3: The use of only vaccination (c_2): $c_2 \neq 0$.

Category 4: The combination of personal protection measures (c_1) and vaccination (c_2): $c_1 \neq 0,\ c_2 \neq 0$.

Therefore

$$\begin{cases} \frac{dS}{dt} = \Sigma(1-\alpha) - \beta_1 \frac{SA}{N}(1-c_1(t)) - \beta_2 \frac{SI}{N}(1-c_2(t)) - (k_1+\mu)S, \\ \frac{dV}{dt} = \Sigma\alpha + k_1 S + k_2 A - (\theta+\mu)V, \\ \frac{dE}{dt} = \beta_1 \frac{SA}{N}(1-c_1(t)) + \beta_2 \frac{SI}{N}(1-c_2(t)) - (\sigma+\mu)E, \\ \frac{dI}{dt} = \gamma\sigma E + dA - (\alpha_2 + n_1 + \varepsilon_1 + \mu)I, \\ \frac{dA}{dt} = (1-\gamma)\sigma E - (d + n_3 + \alpha_1 + \mu + k_2)A, \\ \frac{dH}{dt} = \alpha_1 A + \alpha_2 I - (\lambda+\mu)H, \\ \frac{dR}{dt} = \lambda H + n_3 A + n_2 Q + \theta V - \mu R + c_3(t)Q, \\ \frac{dQ}{dt} = n_1 I - (n_2 + \varepsilon_2 + \mu)Q - c_3(t)Q. \end{cases} \tag{2}$$

Given $(S(0) \geq 0, V(0) \geq 0, E(0) \geq 0, I(0) \geq 0, A(0) \geq 0, H(0) \geq 0, R(0) \geq 0,$ $Q(0) \geq 0) \in \mathbb{R}^8_+$, these are the prescribed initial conditions.

4 The challenge of finding the best control strategy

The central difficulty in this framework is the minimization of the objective functional [26, 36, 37, 38, 39], which necessitates achieving an appropriate compromise between the application of control strategies and the expenses they incur. The overarching aim is to construct an optimal control policy that reduces the functional to its lowest possible value, while simultaneously accounting for the efficiency of the interventions as well as their economic burden. More precisely, the objective functional $J(c_1, c_2, c_3)$ captures the balance between the prevalence of infection and the financial cost associated with implementing control actions. The epidemiological dynamics are described through the infected classes $I(t)$, $A(t)$, and $Q(t)$ together with the recovered class $R(t)$. On the other hand, the costs of interventions are represented by the control variables $c_1(t)$, $c_2(t)$, and $c_3(t)$, weighted by the positive constants k_1, k_2, and k_3, respectively. Consequently, the formulation provides a rigorous quantitative tool to evaluate different strategies and to determine the most effective approach in terms of both public health outcomes and resource allocation.

$$J(c_1, c_2, c_3) = I(T) + A(T) + Q(T) - R(T)$$

$$+\int_0^t \left(I(t) + A(t) + Q(t) - R(t) + \frac{k_1}{2}c_1^2(t) + \frac{k_2}{2}c_2^2(t) + \frac{k_3}{2}c_3^2(t) \right) dt.$$

The coefficients $k_1 \geq 0$, $k_2 \geq 0$ and $k_3 \geq 0$ are cost factors, chosen to balance the relative importance of $c_1(t)$, $c_2(t)$ and $c_3(t)$ at any time t, where $t_f = T$ represents the terminal time.

Our primary objective is to find the optimal controls c_1^*, c_2^* and c_3^* such that

$$J\left(c_1^*, c_2^*, c_3^*\right) = \min_{(c_1,c_2,c_3)\in U_{ad}} J\left(c_1, c_2, c_3\right). \tag{3}$$

where U_{ad} refers to the set of allowable controls, defined by the following constraints:

$$\begin{aligned} U_{ad} = \{c_1, c_2, c_3 | \, & 0 \leq c_{1\min} \leq c_1(t) \leq c_{1\max} \leq 1, \\ & 0 \leq c_{2\min} \leq c_2(t) \leq c_{2\max} \leq 1 \\ & 0 \leq c_{3\min} \leq c_3(t) \leq c_{3\max} \leq 1; \qquad t \in [0, t_f]\}. \end{aligned}$$

4.1 Existence of optimal control

Theorem 4. Consider the regulation problem for system (2). There exists an optimal control vector $(c_1^*(t), c_2^*(t), c_3^*(t)) \in U_{ad}$ such that

$$J(c_1^*(t), c_2^*(t), c_3^*(t)) = \min_{(c_1(t),c_2(t),c_3(t))\in U_{ad}} J(c_1(t), c_2(t), c_3(t)).$$

This holds under the following conditions:

1. The control set U_{ad} and the corresponding state variables are nonempty.

2. The system's right-hand side is bounded by a linear combination of the state and control variables.

3. The control set U_{ad} is both convex and closed.

4. The integral $M(S, E, A, I, V, H, R, Q, c_1, c_2, c_3)$ within the objective function is convex over the admissible control set U_{ad}.

Proof. **Condition 1.**

To prove the nonemptiness of the control set and the associated state variables, we employ an existence theorem from Boyce et al. [12].

Consider the system described by $\dot{Y}_i = Y_{K_i}(t, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8)$ for $i \in \{1, \ldots, 8\}$, where $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8) = (S, E, A, I, V, H, R, Q)$. The functions $Y_S, Y_E, Y_A, Y_I, Y_V, Y_H, Y_R$ and Y_Q represent the right-hand side of (3). Given that all parameters are constants and $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7$ and Y_8 are continuous functions, it follows that $Y_S, Y_E, Y_A, Y_I, Y_V, Y_H, Y_R$ and Y_Q are also continuous. Additionally, the partial derivatives $\frac{\partial Y_{Y_i}}{\partial Y_i}$ for $i \in \{1, \ldots, 8\}$ are continuous.

Thus, a unique solution (S, E, A, I, V, H, R, Q) exists that satisfies the initial conditions, indicating that the control set and the corresponding state variables are nonempty, thus meeting the first condition.

Condition 2.

By definition, the admissible control set U_{ad} is closed. Consider any controls $(c_1, c_2, c_3) \in U_{ad}$ and $\xi \in [0, 1]$. We need to demonstrate that $0 \leq \xi c_1 + (1 - \xi)c_3 \leq 1$. Indeed, it follows that $\xi c_1 \leq \xi$ and $(1 - \xi)c_3 \leq (1 - \xi)$, which implies that $\xi c_1 + (1 - \xi)c_3 \leq \xi + (1 - \xi) = 1$. Therefore, we have $0 \leq \xi c_1 + (1 - \xi)c_3 \leq 1$ for all $(c_1, c_2, c_3) \in U_{ad}$ and $\xi \in [0, 1]$. This confirms that U_{ad} is both closed and convex, thereby satisfying the second condition.

Condition 3.

The right-hand sides of system (2) are continuous and bounded by a linear combination of the state variables and control inputs. Furthermore, they can be expressed as a linear function in terms of c_1, c_2 and c_3, with coefficients that depend on both the state variables and time.

Condition 4.

The integral within the objective function $I(t)+A(t)+Q(t)-R(t)+\frac{k_1}{2}c_1^2(t)+\frac{k_2}{2}c_2^2(t)+\frac{k_3}{2}c_3^2(t)$ is inherently convex over the set U_{ad}. Indeed, there exist constants Φ_1 and Φ_2, with $\beta > 1$, such that the integral in the objective functional satisfies

$$M(S, E, A, I, V, H, R, Q, c_1, c_2, c_3) \geq \Phi_1 + \Phi_2(|c_1|^2 + |c_2|^2 + |c_3|^2)^{\frac{\kappa}{2}},$$

where

$$\begin{aligned} &I(t) + A(t) + Q(t) - R(t) + \frac{k_1}{2}c_1^2(t) + \frac{k_2}{2}c_2^2(t) + \frac{k_3}{2}c_3^2(t) \\ &\quad \geq \Phi_1 + \Phi_2(|c_1|^2 + |c_2|^2 + |c_3|^2)^{\frac{\kappa}{2}}. \end{aligned}$$

Given that the state variables are bounded, let

$$\begin{aligned} \Phi_1 &= 4 \inf_{t \in [0,T]} (I(t) + A(t) + Q(t) - R(t)), \\ \Phi_2 &= \inf\left(\frac{k_1}{2}, \frac{k_2}{2}, \frac{k_3}{2}\right), \quad \textit{and} \quad \kappa = 2. \end{aligned}$$

□

Therefore, following the results presented by Fleming and Richel [19], and further elaborated in [35, 42, 43], we can confirm the existence of an optimal control.

4.2 Characterization of optimal control

In this part of the study, Pontryagin's maximum principle [37] is employed as a fundamental analytical framework for addressing the optimal control problem. The methodology begins with the introduction of the adjoint variables, which serve as an essential mathematical link between the system of differential equations describing the dynamics and the objective functional targeted for optimization. This relationship naturally leads to the construction of the Hamiltonian function, a central quantity that integrates both the system dynamics and the objectives of the control process.

By invoking Pontryagin's principle, the challenging task of determining an optimal control policy under given initial conditions is reformulated into the problem of continuously optimizing the Hamiltonian over time. This transformation not only simplifies the search for the optimal solution but also provides a systematic procedure to guide the system toward the desired outcome. Furthermore, this approach establishes a solid theoretical basis for exploring qualitative properties of the solutions and for developing efficient numerical algorithms capable of implementing the optimal strategies in practice.

We now define the Hamiltonian H at a given time t as follows:

$$H = I(t) + A(t) + Q(t) - R(t) + \frac{k_1}{2}c_1^2(t) + \frac{k_2}{2}c_2^2(t) + \frac{k_3}{2}c_3^2(t) + \sum_{i=1}^{8} \psi_i(t)\, u_i(S, E, A, I, V, H, R, Q).$$

In this scenario, u_i denotes the right-hand side of the differential equations system (2) for the i-th state variable, and ψ_i represents the corresponding adjoint functions. These adjoint functions are linked to the state variables and are used to reflect the impact of changes in the state on the cost functional.

Theorem 5. Suppose that $(c_1^*, c_2^*, c_3^*) \in U_{ad}$ denotes the optimal control variables, and let S^*, E^*, A^*, I^*, V^*, H^*, R^*, and Q^* represent the corresponding optimal trajectories for the state variables. Then, there exist adjoint functions ψ_1, ψ_2, ψ_3, ψ_4, ψ_5, ψ_6, ψ_7 and ψ_8 that satisfy the following conditions:

$$
\begin{aligned}
\psi_1' &= -\frac{\partial H}{\partial S} = \left(\beta_1 \frac{A}{N}(1-c_1) + \beta_2 \frac{I}{N}(1-c_2)\right)(\psi_1 - \psi_3) \\
&\quad + k_1(\psi_1 - \psi_2) + \mu\psi_1, \\
\psi_2' &= -\frac{\partial H}{\partial V} = \theta(\psi_2 - \psi_7) + \mu\psi_2, \\
\psi_3' &= -\frac{\partial H}{\partial E} = \sigma(\psi_3 - \psi_5) + \gamma\sigma(\psi_5 - \psi_4) + \mu\psi_3, \\
\psi_4' &= -\frac{\partial H}{\partial I} = -1 + \frac{\beta_2 S}{N}(1-c_2)(\psi_1 - \psi_3) \\
&\quad + \alpha_2(\psi_4 - \psi_6) + n_1(\psi_4 - \psi_8) + (\mu + \varepsilon_1)\psi_4, \\
\psi_5' &= -\frac{\partial H}{\partial A} = -1 + \frac{\beta_1 S}{N}(1-c_1)(\psi_1 - \psi_3) \\
&\quad + d(\psi_5 - \psi_4) + n_3(\psi_5 - \psi_7) + \alpha_1(\psi_5 - \psi_6) \\
&\quad + k_2(\psi_5 - \psi_2) + \mu\psi_5, \\
\psi_6' &= -\frac{\partial H}{\partial H} = \lambda(\psi_6 - \psi_7) + \mu\psi_6, \\
\psi_7' &= -\frac{\partial H}{\partial R} = 1 + \mu\psi_7, \\
\psi_8' &= -\frac{\partial H}{\partial Q} = -1 + (\psi_8 - \psi_7)(n_2 + c_3) + (\mu + \varepsilon_2)\psi_8.
\end{aligned}
$$

With the transversality conditions, ψ_1, ψ_2, ψ_5, and ψ_7 are all equal to zero at time T. On the other hand, ψ_3, ψ_4, and ψ_8 are set to minus one, and ψ_6 is set to one at the same time point.

The solution to the optimal control problem is provided by the control variables c_1^*, c_2^* and c_3^*. The optimal control values are obtained as follows:

$$
\begin{cases}
c_1^* = \min\left(1, \max\left(0, \frac{\beta_1 SA(\psi_3 - \psi_1)}{k_1 N}\right)\right), \\
c_2^* = \min\left(1, \max\left(0, \frac{\beta_2 SI(\psi_3 - \psi_1)}{k_2 N}\right)\right), \\
c_3^* = \min\left(1, \max\left(0, \frac{Q(\psi_8 - \psi_7)}{k_3}\right)\right).
\end{cases} \tag{4}
$$

Proof. The Hamiltonian H at any time t can be represented as follows:

$$
\begin{aligned}
H =& I(t) + A(t) + Q(t) - R(t) + \frac{k_1}{2}c_1^2(t) + \frac{k_2}{2}c_2^2(t) + \frac{k_3}{2}c_3^2(t) \\
&+ \psi_1 \left\{ \Sigma(1-\alpha) - \beta_1 \frac{S(t)A(t)}{N}(1 - c_1(t)) \right. \\
&\left. - \beta_2 \frac{S(t)I(t)}{N}(1 - c_2(t)) - (k_1 + \mu)S(t) \right\} \\
&+ \psi_2 \left\{ \Sigma\alpha + k_1 S(t) + k_2 A(t) - (\theta + \mu)V(t) \right\}
\end{aligned}
$$

$$
\begin{aligned}
&+ \psi_3 \left\{ \beta_1 \frac{S(t)A(t)}{N} \left(1 - c_1(t)\right) + \beta_2 \frac{S(t)I(t)}{N} \left(1 - c_2(t)\right) - (\sigma + \mu)E(t) \right\} \\
&+ \psi_4 \left\{ \gamma\sigma E(t) + dA(t) - (\alpha_2 + n_1 + \varepsilon_1 + \mu)I(t) \right\} \\
&+ \psi_5 \left\{ (1-\gamma)\sigma E(t) - (d + n_3 + \alpha_1 + \mu + k_2)A(t) \right\} \\
&+ \psi_6 \left\{ \alpha_1 A(t) + \alpha_2 I(t) - (\lambda + \mu)H(t) \right\} \\
&+ \psi_7 \left\{ \lambda H(t) + n_3 A(t) + n_2 Q(t) + \theta V(t) - \mu R(t) + c_3(t)Q(t) \right\} \\
&+ \psi_8 \left\{ n_1 I(t) - (n_2 + \varepsilon_2 + \mu)Q(t) - c_3(t)Q(t) \right\}.
\end{aligned}
$$

Using Pontryagin's maximum principle [27], we derived the adjoint equations and the transversality conditions, which are expressed as follows:

$$
{\psi_i}' = -\tfrac{\partial H}{\partial Y_i} \text{ and } \psi_i(t_f) = 0 \text{ for } i \in \{1, 2, 5, 7, 8\},
$$

where $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8) = (S, E, A, I, V, H, R, Q)$.

The transversality conditions require that, at the terminal time, the adjoint variables must satisfy an orthogonality relation with the attainable set of the final state. This condition plays a crucial role in guaranteeing the mathematical consistency of the optimal control formulation and ensures that the resulting solution is both coherent and reliable. By enforcing these requirements, one can explicitly characterize the optimal control functions, denoted by $c_1^*(t)$, $c_2^*(t)$, and $c_3^*(t)$, for the entire time horizon $t \in [0, t_f]$. Such conditions not only facilitate the identification of the most effective control strategies but also provide a rigorous framework that enhances the interpretability of the results and supports the implementation of optimal policies in practical scenarios.

$$
\frac{\partial H}{\partial c_1} = 0, \qquad \frac{\partial H}{\partial c_2} = 0 \qquad \textit{and} \qquad \frac{\partial H}{\partial c_3} = 0.
$$

Consequently,

$$
\begin{aligned}
\frac{\partial H}{\partial c_1} &= k_1 c_1 + \psi_1 \beta_1 \frac{SA}{N} - \psi_3 \beta_1 \frac{SA}{N} = 0, \\
\frac{\partial H}{\partial c_2} &= k_2 c_2 + \psi_1 \beta_2 \frac{SI}{N} - \psi_3 \beta_2 \frac{SI}{N} = 0, \\
\frac{\partial H}{\partial c_3} &= k_3 c_3 + \psi_7 Q - \psi_8 Q = 0.
\end{aligned}
$$

Therefore,

$$
c_1^* = \frac{\beta_1 SA(\psi_3 - \psi_1)}{k_1 N},
$$

$$c_2^* = \frac{\beta_2 SI(\psi_3 - \psi_1)}{k_2 N},$$
$$c_3^* = \frac{Q(\psi_8 - \psi_7)}{k_3}.$$

Given the constraints within U_{ad} for the control variables, we can efficiently determine the optimal controls c_1^*, c_2^* and c_3^* as expressed in (4). □

5 Numerical simulation

In this section, we focus on the numerical simulation of the proposed optimal control problem, which is carried out by applying the well-known Forward-Backward Sweep method [45]. The approach consists of integrating the state system (1) forward in time while simultaneously solving the corresponding adjoint system backward in time. The two procedures are repeated in an iterative manner, with updated values of the control functions at each step. This forward-backward process continues until the solutions converge to an optimal trajectory. Once a satisfactory level of convergence between the state and adjoint systems is achieved, the iterative procedure is terminated. The results obtained from this algorithm are then illustrated and analyzed graphically to highlight the effectiveness of the derived control strategies.

In this study, the parameter values presented in Table 1 were selected under hypothetical assumptions, since real epidemiological data were not available. This choice allowed us to design and analyze the model even in the absence of empirical evidence. The use of assumed parameter values provided the flexibility to examine different possible trajectories and disease dynamics that could emerge from the system. Such an approach makes it possible to conduct an in-depth investigation of the model's performance under diverse scenarios, while also laying the groundwork for future validation once reliable data become accessible. Furthermore, adopting hypothetical scenarios serves to highlight potential weaknesses in the model and to point out directions where additional research is needed, thereby ensuring that the framework remains sufficiently robust and adaptable to multiple contexts. The simulated values employed in this model were specifically chosen to conduct a theoretical investigation of the potential impacts of personal protective measures, vaccination, and awareness programs on reducing the spread of meningitis. It is important to emphasize that this constitutes a purely theoretical analysis, intended to capture the possible outcomes under different intervention strategies rather than relying on real-world data. Furthermore, based on assumptions related to the prevention of meningitis, the three control variables, denoted here as c_1, c_2 and c_3, are constrained within given intervals. These constraints reflect the realistic limits of implementing such interventions

Table 1: Details on the parameters involved in the meningitis modeling.

Parameter	**Values in d^{-l}**	**Reference**	**Parameter**	**Values in d^{-l}**	**Reference**
μ	3.933×10^{-5}	[31]	θ	0.41	Assumed
β_1	0.476	Estimated	σ	0.4	[32]
β_2	0.238	Estimated	Σ	9.43×10^5	[31]
α	0.19	[32]	γ	0.63	[32]
λ	0.88	[34]	n_1	0.0013	Simulated
d	0.0013	[34]	n_2	0.0015	Simulated
k_1	0.2	[33]	n_3	0.017	Simulated
k_2	0.1	[33]	ϵ_1	0.017	Simulated
α_1	0.023	Assumed	ϵ_2	0.012	Simulated
α_2	0.045	Assumed			

in practice, where the level of adoption cannot exceed certain thresholds but also cannot be entirely absent. In the subsequent part of this study, we classify the possible intervention strategies into four distinct categories. Each category represents a particular combination of control measures, and together they serve as the foundation for evaluating the effectiveness of optimal control strategies in mitigating the dynamics of meningitis transmission.

Category 1: The combination of personal protective measures for asymptomatic individuals (c_1), interventions targeting symptomatic individuals (c_2), and quarantine and screening strategies (c_3): $c_1 \neq 0,\ c_2 \neq 0,\ c_3 \neq 0$.

Category 2: The combination of personal protective measures for asymptomatic individuals (c_1) and quarantine and screening strategies (c_3): $c_1 \neq 0,\ c_3 \neq 0$.

Category 3: The use of only interventions targeting symptomatic individuals (c_2): $c_2 \neq 0$.

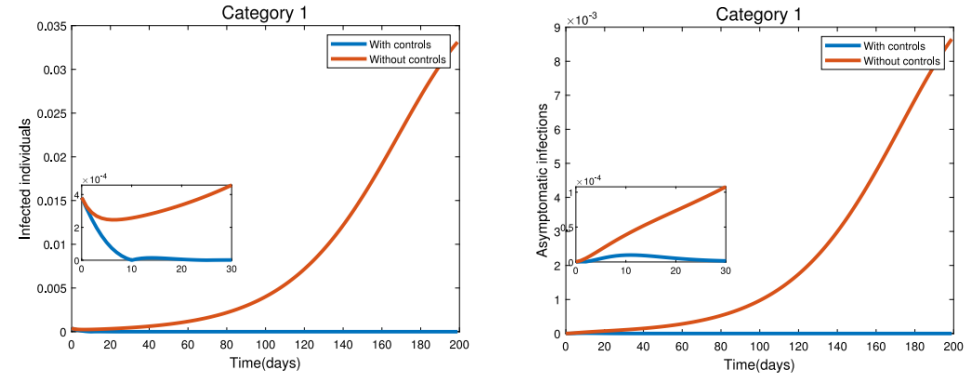
Category 4: The combination of personal protective measures for asymptomatic individuals (c_1) and interventions targeting symptomatic individuals (c_2): $c_1 \neq 0,\ c_2 \neq 0$.

5.1 Strategy 1: This case involves the combined implementation of personal protection measures (c_1), vaccination (c_2), and awareness-raising programs (c_3), where $c_1 \neq 0,\ c_2 \neq 0,\ c_3 \neq 0$

It can therefore be concluded that when optimal control strategies are applied, namely personal protective measures (c_1), vaccination (c_2), and awareness-raising programs (c_3), epidemic outbreaks do not occur. Figure 2 provides a complete illustration of how these control interventions are implemented in the model. Subfigures 2(a), 2(b), 2(c), and 2(d) display the temporal dynamics of the infected (I), asymptomatic infected (A), exposed (E), and recovered (R) populations under Category 1. A key feature of this category is the significant increase in the proportion of recoverers in the population. In particular, Figure 2(d) shows that the proportion of recovered individuals rises to nearly 0.6, indicating that more than half of the population has effectively achieved immunity. Furthermore, Figure 2(e) illustrates the temporal trajectories of the control functions c_1, c_2 and c_3. It can be observed that these controls are maintained at their maximum levels for approximately the first 155, 154, and 80 days, respectively, after which they gradually decline in a monotonic manner toward their minimum values. This behavior reflects the intensive initial interventions needed to curb transmission, followed by a progressive relaxation of measures once the epidemic is under control.

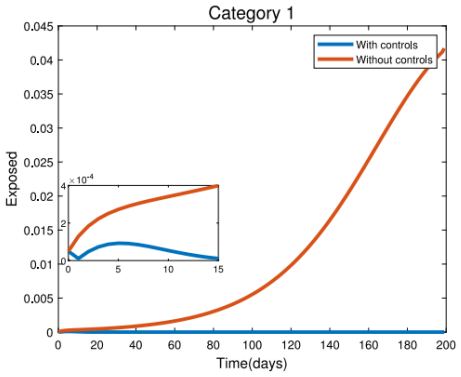
5.2 Strategy 2: This scenario considers the joint implementation of personal protection measures (c_1) together with awareness-raising programs (c_3), where both $c_1 \neq 0$ and $c_3 \neq 0$

In this category, Figure 3 presents the simulation outcomes of applying the control strategies associated with Category 2, namely the combination of personal protection measures (c_1) and awareness programs (c_3). Subfigures 2(a)–2(b) illustrate the trajectories of symptomatic infections, asymptomatic infections, and exposed individuals, respectively. In each case, the number of infected individuals is consistently lower when the controls are implemented compared to the uncontrolled scenario. This demonstrates the effectiveness of preventive awareness and personal

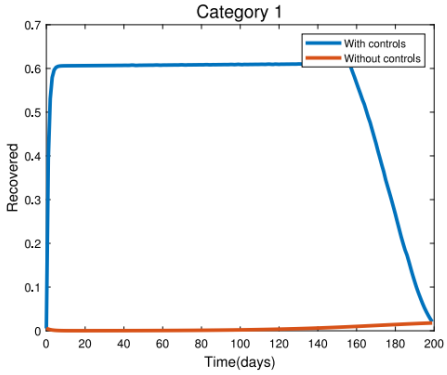


(a)

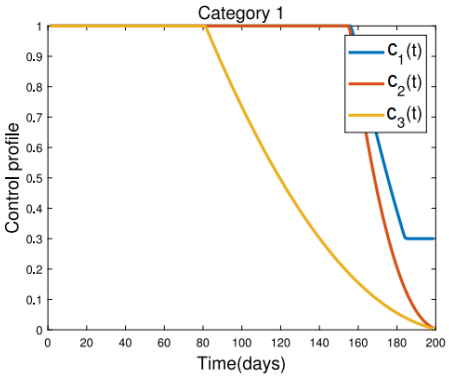
(b)



(c)



(d)



(e)

Figure 2: Numerical results for Category 1. Figures (a), (b), (c), and (d) illustrate the dynamics of symptomatic cases, asymptomatic cases, exposed individuals, and recovered individuals, respectively. Figure (e) displays the trajectories of the optimal control variables.

protection in reducing the rate of new infections and slowing the overall spread of the disease. Figure 2(d) shows that, under Category 2, the number of recovered individuals drops to zero after approximately ten days and remains at zero for the rest of the simulation period. This implies that in the absence of vaccination, no recoverers will exist within the population, which highlights the crucial role of vaccination in sustaining immunity.

Overall, these results indicate that Category 2 can significantly suppress the infection dynamics through awareness and protection alone, but without vaccination, recovery is not achieved, leaving the population vulnerable to future waves of infection.

5.3 Strategy 3: Vaccination as the sole intervention (c_2) where $c_2 \neq 0$

This category greatly increases the proportion of recoverers in the population; as shown in Figure 4(d), the proportion is as high as 0.6. In Figure 4(e) depicts controls c_1, c_2 and c_3, which govern the system over the simulation horizon.

5.4 Strategy 4: The joint application of personal protection measures (c_1) and vaccination (c_2) where $c_1 \neq 0$ and $c_2 \neq 0$

Figure 5 depicts the scenario in which the controls c_1 and c_2 are applied to system (1). Under this category, the proportion of recoverers in the population rises substantially, reaching as high as 0.6 as indicated in Figure 5(d).

Discussion

Figure 6 illustrates the dynamics of the different population compartments in system (1), derived from simulations of the four optimal control categories. Nevertheless, the effectiveness of these categories varies. Although many individuals recognize that meningitis is highly contagious and transmitted through multiple routes, some still neglect preventive practices and refrain from participating in vaccination programs. This behavior hinders the potential for a substantial

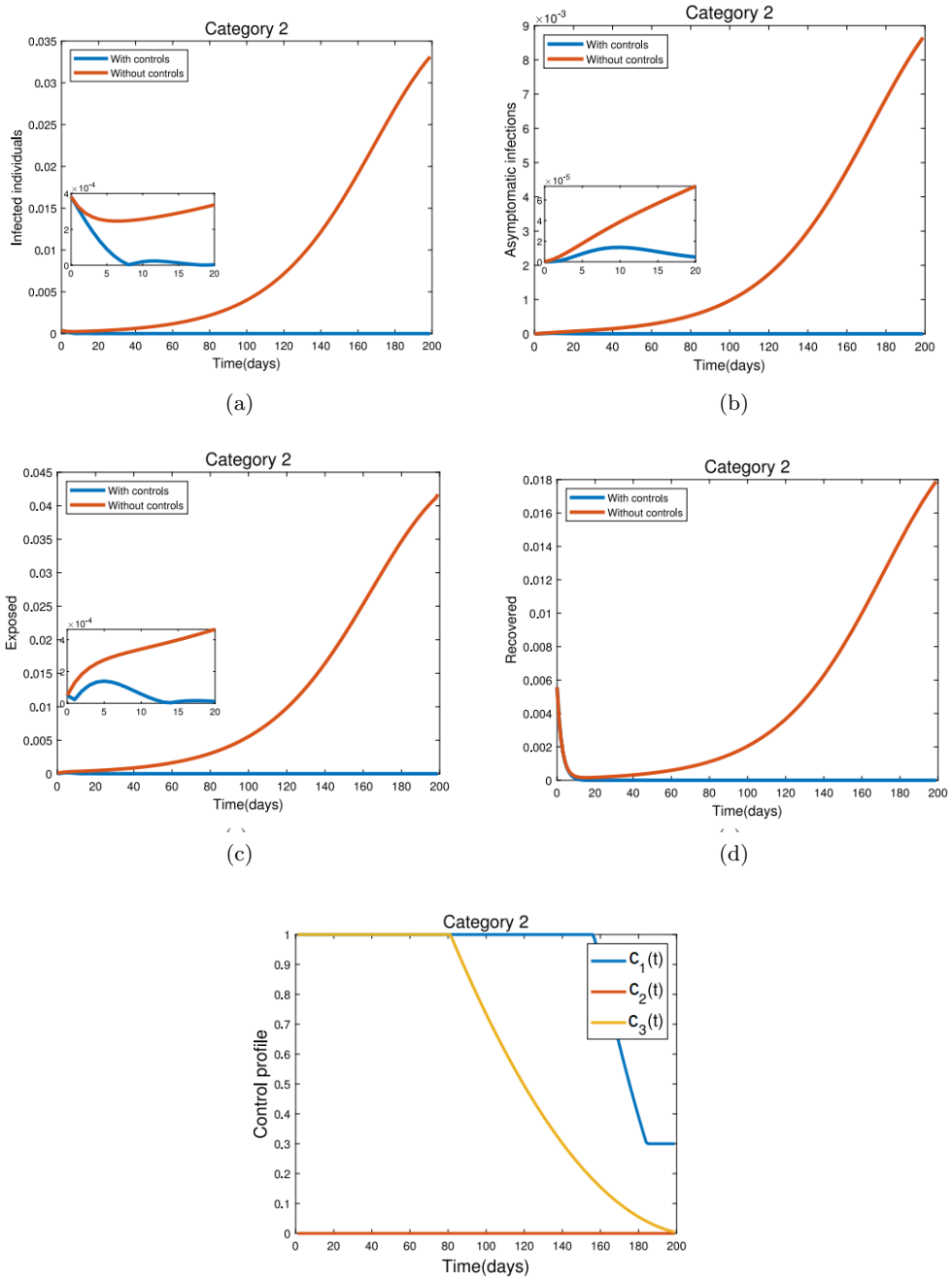


Figure 3: Numerical results for Category 2. Figures (a), (b), (c), and (d) illustrate the dynamics of symptomatic cases, asymptomatic cases, exposed individuals, and recovered individuals, respectively. Figure (e) displays the trajectories of the optimal control variables.

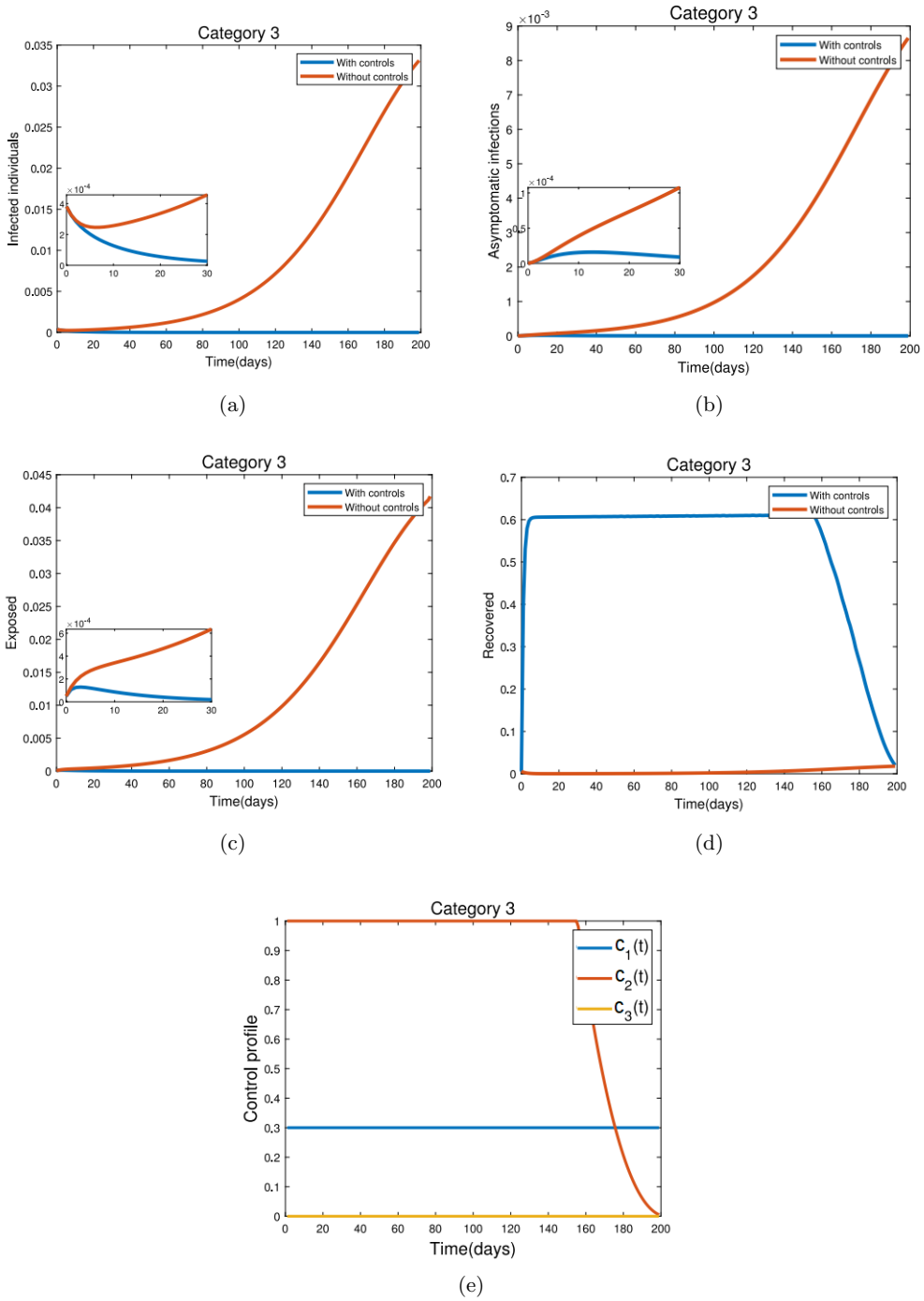


Figure 4: Numerical results for Category 3. Figures (a), (b), (c), and (d) illustrate the dynamics of symptomatic cases, asymptomatic cases, exposed individuals, and recovered individuals, respectively. Figure (e) displays the trajectories of the optimal control variables.

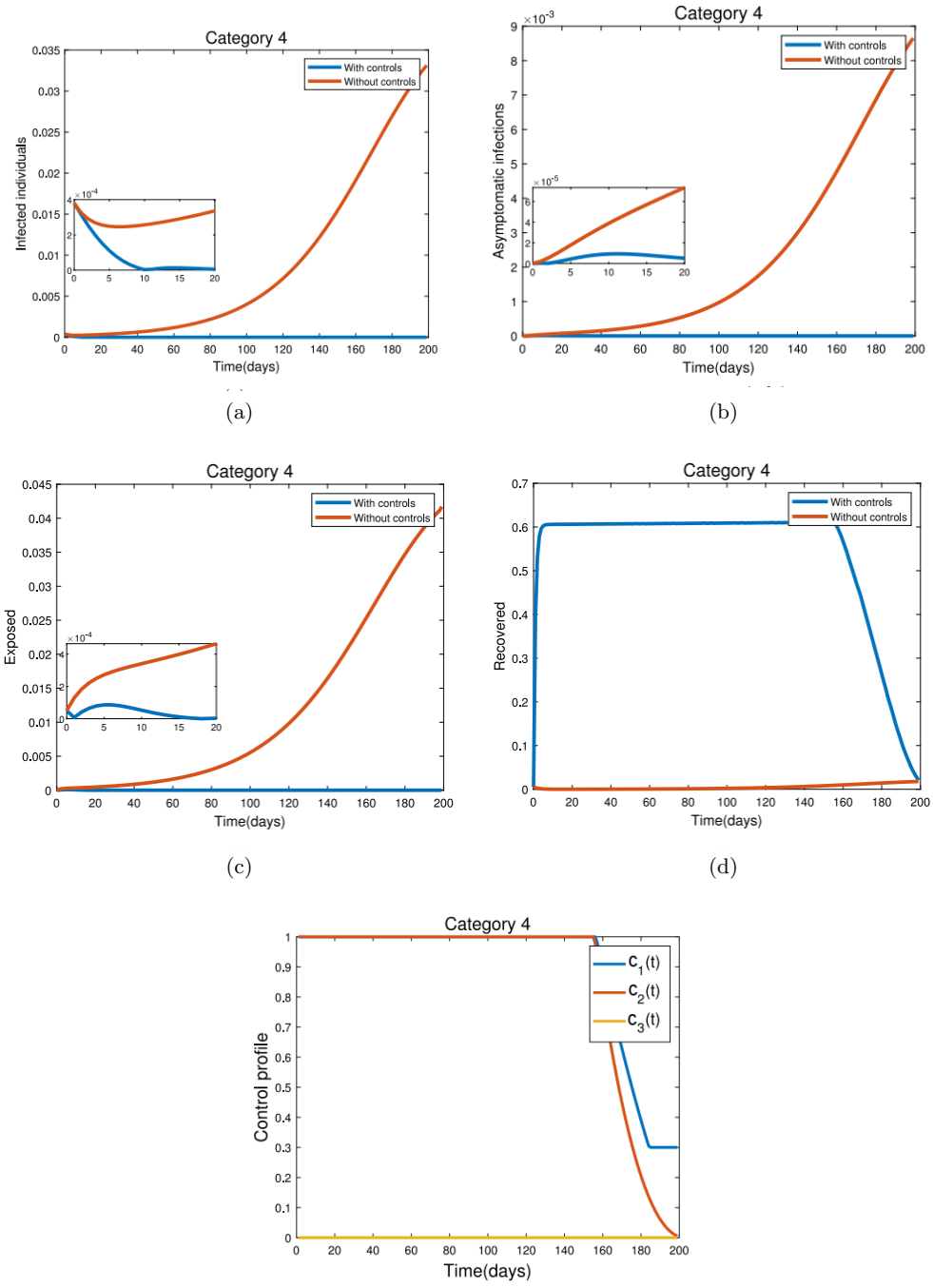
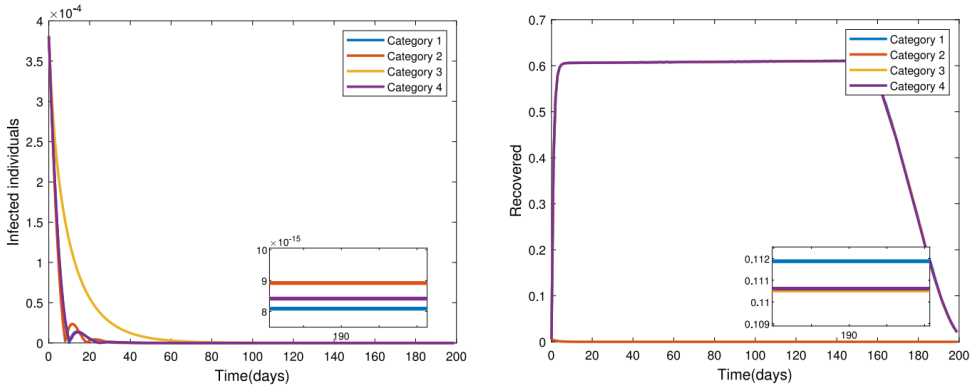


Figure 5: Numerical results for Category 4. Figures (a), (b), (c), and (d) illustrate the dynamics of symptomatic cases, asymptomatic cases, exposed individuals, and recovered individuals, respectively. Figure (e) displays the trajectories of the optimal control variables.

reduction in meningitis cases. Among the interventions, the adoption of control strategy c_2 (vaccination) plays a critical role in increasing the number of recoverers within the population. In the absence of vaccination, no immune individuals would emerge, and the entire population would remain susceptible, thereby elevating the risk of disease.

Moreover, the application of control measures c_1 and c_3 helps to strengthen public awareness and protective behavior. Collectively, the four categories of control strategies are shown to decrease the number of infections while simultaneously enlarging the recovered population. These findings suggest that vaccination and awareness-raising interventions should be promoted more vigorously, as they effectively lower infection rates and, in turn, reduce disease risks.

The analysis further reveals that while individual control measures are beneficial, combining two interventions yields greater effectiveness compared to applying them separately. More importantly, the simultaneous implementation of all three measures achieves the highest level of control, resulting in the fewest infections and the greatest number of recoveries. Overall, the results confirm that the strategic use of these optimal controls can markedly reduce the burden of infection. With adequate prevention and sustained control measures in place, the epidemic situation can be stabilized, allowing individuals to return to normal life.



(a) (b)

Figure 6: Dynamics of the different population groups in system (1), as generated through simulations under the four optimal control strategies.

6 Conclusion

In this work, we developed and analyzed a comprehensive $SEAIVHRQ$ mathematical model for the transmission dynamics and control of meningitis. By integrating medical and behav-

ioral factors such as vaccination, hospitalization, and voluntary quarantine, the model provides a more realistic description of disease spread, particularly in regions with limited healthcare resources. The incorporation of three time-dependent control variables—reducing transmission from asymptomatic and symptomatic individuals, and promoting awareness and treatment adherence—allowed us to formulate an optimal control problem solved via Pontryagin's maximum principle.

Numerical simulations demonstrated that combining preventive, treatment, and awareness interventions can significantly lower infection peaks, reduce hospitalization demand, and minimize total costs. These results emphasize that a coordinated and sustained public health response is essential to achieving effective epidemic mitigation.

Overall, the proposed framework contributes a novel tool for understanding and managing meningitis epidemics, aligning with the WHO's roadmap "Defeating Meningitis by 2030". Future extensions may consider stochastic effects, fractional-order dynamics, and spatial heterogeneity to further enhance model realism and applicability to real-world health policy planning.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Agusto, F.B., and Leite, M.C.A., *Optimal control and cost-effective analysis of the 2017 meningitis outbreak in Nigeria*, Infect. Dis. Model. 4 (2019), 161–187.

[2] Ali, S.A., Taj, M.K., and Ali, S.H., *Antimicrobial resistance pattern of bacterial meningitis among patients in Quetta, Pakistan*, Infect. Drug Resist. 14 (2021), 5107–5120.

[3] Asamoah, J.K.K., Nyabadza, F., Jin, Z., Bonyah, E., Khan, M.A., Li, M.Y., and Hayat, T., *Backward bifurcation and sensitivity analysis for bacterial meningitis transmission dynamics with a nonlinear recovery rate*, Chaos Solitons Fractals 140 (2020), Article 110237.

[4] Asamoah, J.K.K., Nyabadza, F., Seidu, B., Chand, M., and Dutta, H., *Mathematical modelling of bacterial meningitis transmission dynamics with control measures*, Comput. Math. Methods Med. 2018 (2018), Article ID 2657461.

[5] Ashchepkov, L.T., Dolgy, D.V., Kim, T., and Agarwal, R.P., *General optimal control problem*, In *Optimal Control*, Springer, Cham, 2021.

[6] Baroudi, M., Gourram, H., Alia, M., Labzai, A., and Belam, M., *Mathematical modeling and optimal control approaches for dengue*, Iran. J. Numer. Anal. Optim. 15(1) (2025), 396–423.

[7] Baroudi, M., Smouni, I., Gourram, H., Labzai, A., and Belam, M., *Advanced mathematical modeling and prognostication of regulated spatio-temporal dynamics of monkeypox*, Iran. J. Numer. Anal. Optim. 14(4) (2024) 1280–1308.

[8] Baroudi, M., Smouni, I., Gourram, H., Labzai, A., and Belam, M., *Optimizing control strategies for monkeypox through mathematical modeling*, Partial Differ. Equ. Appl. Math. 12 (2024), Article 100996.

[9] Belay, M.A., Abonyo, O.J., and Theuri, D.M., *Mathematical model analysis for the transmission dynamics of bacterial meningitis disease incorporating drug-resistance class*, Commun. Math. Biol. Neurosci. 2022 (2022), Article ID 121.

[10] Belay, M.A., Abonyo, O.J., and Theuri, D.M., *Mathematical model of hepatitis B disease with optimal control and cost-effectiveness analysis*, Comput. Math. Methods Med. 2023 (2023), Article ID 5215494.

[11] Bowong, S., and Kurths, J., *Modelling tuberculosis and hepatitis B co-infections*, Math. Model. Nat. Phenom. 5(6) (2010), 196–242.

[12] Boyce, W.E., DiPrima, R.C., and Meade, D.B., *Elementary differential equations*, John Wiley & Sons, Hoboken, NJ, 2017.

[13] Brozak, S.J., Pant, B., Safdar, S., and Gumel, A.B., *Dynamics of COVID-19 pandemic in India and Pakistan: a metapopulation modelling approach*, Infect. Dis. Model. 6 (2021), 1173–1201.

[14] Castillo-Chavez, C., and Song, B., *Dynamical models of tuberculosis and their applications*, Math. Biosci. Eng. 1(2) (2004), 361–404.

[15] Chandani, S., Jani, D., Sahu, P.K., Kataria, U., Suryawanshi, S., Khubchandani, J., Thorat, S., Chitlange, S. and Sharma, D. *COVID-19 vaccination hesitancy in India: state of the nation and priorities for research*, Brain Behav. Immun. Health 18 (2021), 100375.

[16] Cui, J., Lv, J., and Guo, S., *Dynamical model of emerging infectious diseases applied to COVID-19 transmission*, Acta Math. Appl. Sin. 43 (2020), 147–155.

[17] Donovan, C., and Blewitt, J., *Signs, symptoms and management of bacterial meningitis*, Nurs. Child. Young People 22(9) (2010).

[18] Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention *The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China*, Chin. J. Epidemiol. 41 (2020), 145–151.

[19] Fleming, W.H., and Rishel, R.W., *Deterministic and stochastic optimal control*, Vol. 1, Springer, New York, 2012.

[20] Gourram, H., Sahib, I., Baroudi, M., Smouni, I., Labzai, A., and Belam, M., *Analysis of the stability of a mathematical model for measles*, Commun. Math. Biol. Neurosci. 2025 (2025), Article ID 55.

[21] Irving, T.J., Blyuss, K.B., Colijn, C., and Trotter, C.L., *Modelling meningococcal meningitis in the African meningitis belt*, Epidemiol. Infect. 140(5) (2012), 897–905.

[22] Khajanchi, S., *Stability analysis of a mathematical model for glioma–immune interaction under optimal therapy*, Int. J. Nonlinear Sci. Numer. Simul. 20(3–4) (2019), 269–285.

[23] Khajanchi, S., and Banerjee, S., *A strategy of optimal efficacy of T11 target structure in the treatment of brain tumor*, J. Biol. Syst. 27(2) (2019), 225–255.

[24] Khajanchi, S., and Ghosh, D., *The combined effects of optimal control in cancer remission*, Appl. Math. Comput. 271 (2015), 375–388.

[25] Khan, M.A., Islam, S., Arif, M., and ul Haq, Z., *Transmission model of hepatitis B virus with the migration effect*, BioMed Res. Int. 2013 (2013), Article ID 150681.

[26] Kinfe, H., Sendo, E.G., and Gebremedhin, K.B., *Prevalence of hepatitis B virus infection and factors associated with hepatitis B virus infection among pregnant women presented to antenatal care clinics at Adigrat General Hospital in Northern Ethiopia*, Int. J. Women's Health 13 (2021), 119–127.

[27] Lenhart, S., and Workman, J.T., *Optimal control applied to biological models*, Chapman & Hall/CRC, Boca Raton, FL, 2007.

[28] Martcheva, M., and Crispino-O'Connell, G., *The transmission of meningococcal infection: a mathematical study*, J. Math. Anal. Appl. 283(1) (2003), 251–275.

[29] Maseno, K., *Mathematical model for malaria and meningitis co-infection among children*, Appl. Math. Sci. 5(47) (2011), 2337–2359.

[30] Mazamay, S., Guégan, J.-F., Diallo, N., Bompangue, D., Bokabo, E., Muyembe, J.-J., Taty, N., Vita, T.P., and Broutin, H., *An overview of bacterial meningitis epidemics in Africa from 1928 to 2018 with a focus on epidemics "outside-the-belt"*, BMC Infect. Dis. 21(1) (2021), Article 1027.

[31] McCarthy, J.J., *A thematic guide to optimality theory*, Cambridge Univ. Press, Cambridge, 2002.

[32] Misra, A.K., Singh, R.K., Tiwari, P.K., Khajanchi, S., and Kang, Y., *Dynamics of algae blooming: effects of budget allocation and time delay*, Nonlinear Dyn. 100(2) (2020), 1779–1807.

[33] Mo, X., Xu, X., Ren, Z., Guan, J., and Peng, J., *Patients with tuberculous meningitis and hepatitis B co-infection have increased risk for antituberculosis drug-induced liver injury and poor outcomes*, Infect. Dis. 52(11) (2020), 793–800.

[34] Mondal, J., Khajanchi, S., and Akhtar, M.N., *A mathematical model for COVID-19 pandemic with the impact of economic development*, In *Fractal signatures in the dynamics of an epidemiology*, CRC Press, Boca Raton, FL, 2023.

[35] Nandi, S., Khajanchi, S., Chatterjee, A.N., and Roy, P.K., *Insight of viral infection of Jatropha curcas plant (Future Fuel): a control-based mathematical study*, Acta Analysis Functionalis Applicata 13(04) (2011), 366–374.

[36] Olaniyi, S., Ajala, O.A., and Abimbade, S.F., *Optimal control analysis of a mathematical model for recurrent malaria dynamics*, Oper. Res. Forum vol. 4, no. 1, p. 14. Cham: Springer International Publishing, 2023.

[37] Pontryagin, L.S., *Mathematical theory of optimal processes*, CRC Press, Boca Raton, FL, 1987.

[38] Sahib, I., Baroudi, M., Gourram, H., Khajji, B., Labzai, A., and Belam, M., *The mathematical modeling and analysis of the cholera disease model*, Commun. Math. Biol. Neurosci. 2024 (2024), Article ID 93.

[39] Shen, W.-Y., Chu, Y.-M., Rahman, M.U., Mahariq, I., and Zeb, A., *Mathematical analysis of HBV and HCV co-infection model under nonsingular fractional order derivative*, Results Phys. 28 (2021), Article 104582.

[40] Smouni, I., Baroudi, M., Alia, M., Elmansouri, A., Labzai, A., and Belam, M., *Spatiotemporal stability analysis of soil-borne disease dynamics in tomato plants*, Model. Earth Syst. Environ. 11(2) (2025), Article 105.

[41] Spearman, C.W., Afihene, M., Ally, R., Apica, B., Awuku, Y., Cunha, L., Dusheiko, G., Gogela, N., Kassianides, C., and Kew, M., *Hepatitis B in sub-Saharan Africa: strategies to achieve the 2030 elimination targets*, Lancet Gastroenterol. Hepatol. 2(12) (2017), 900–909.

[42] Tilahun, G.T., *Optimal control analysis of pneumonia and meningitis coinfection*, Comput. Math. Methods Med. 2019 (2019), Article ID 2658971.

[43] World Health Organization, *Hepatitis B*, WHO/CDS/CSR/LYO/2002.2, World Health Organization, Geneva, 2002.

[44] Zhao, S., Xu, Z., and Lu, Y., *A mathematical model of hepatitis B virus transmission and its application for vaccination strategy in China*, Int. J. Epidemiol. 29(4) (2000), 744–752.

[45] Zou, L., Zhang, W., and Ruan, S., *Modeling the transmission dynamics and control of hepatitis B virus in China*, J. Theor. Biol. 262(2) (2010), 330–338.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1432–1463
https://doi.org/10.22067/ijnao.2026.94741.1697
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Design of a multi-stage hybrid optimizer combining Bayesian and evolutionary algorithms for modeling and optimizing seizure-related information in EEG signals

M. Khansari* and H. Afshari Rad

Abstract

The prediction of epileptic seizures from EEG signal analysis is modeled as a complex, high-dimensional, multi-objective optimization problem. In this study, we propose and design a novel multi-stage hybrid optimizer that integrates Bayesian optimization with a family of evolutionary and metaheuristic algorithms, including genetic algorithm, evolutionary programming, evolution strategies, ant colony optimization, artificial bee colony, and harmony search. This optimizer is structured to operate across sequential stages of optimization, where each stage is designed to progressively refine the solution space by selecting the optimal subset of EEG channels and time blocks related to the preictal phase. The mathematical model aims to maximize the posterior probability of seizure occurrence while minimizing the temporal distance to the actual seizure, thus formulating a well-defined trade-off in the objective space. Unlike traditional methods, the proposed hybrid framework models and controls the optimization

*Corresponding author

Received 3 August 2025; revised 4 May 2026; accepted 27 May 2026

Maryam Khansari
Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. e-mail: m.khansari@mail.um.ac.ir

Hooman Afshari Rad
Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. e-mail: afshari-rad.hooman@mail.um.ac.ir

How to cite this article
Khansari, M. and Afshari Rad, H., Design of a multi-stage hybrid optimizer combining Bayesian and evolutionary algorithms for modeling and optimizing seizure-related information in EEG signals . *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1432-1463. https://doi.org/10.22067/ijnao.2026.94741.1697

trajectory by combining the global exploration capability of evolutionary algorithms with the adaptive search of Bayesian optimization. Experimental results demonstrate that the optimizer successfully converges toward high-informational regions of EEG data that occur early in the preictal phase. A detailed performance comparison and sensitivity analysis further validate the effectiveness and robustness of the proposed approach, offering insights into the design of more generalizable hybrid optimizers for complex modeling tasks.

AMS subject classifications (2020): Primary 90C59; Secondary 90C26, 62F15, 68T07, 92C55.

Keywords: Multi-objective optimization, Hybrid optimizer, Bayesian optimization, Evolutionary algorithms, Posterior probability, Seizure prediction.

1 Introduction

Mathematical modeling has always been a fundamental activity in sciences and engineering. Formulating qualitative questions about observed phenomena into mathematical problems has long been a key driver in the development of mathematics [7]. With the rapid growth of computational tools in recent decades, mathematical modeling has become an essential framework for analyzing and solving complex real-world problems. In many scientific and engineering domains, real-world systems are characterized by high-dimensional, nonlinear, and noisy data. In such situations, reliable interpretation and prediction require mathematical representations that capture the underlying structure of the data rather than merely describing observations. These representations are commonly expressed in terms of functions, stochastic processes, or parametric models [7].

This need is particularly evident in biomedical signal processing, where signals are inherently complex and non-stationary. Among them, electroencephalogram (EEG) signals play a central role in understanding brain dynamics and neurological disorders such as epilepsy [18, 16]. EEG signals provide multichannel time-varying measurements of brain electrical activity, which can be used to detect abnormal patterns associated with neurological events.

Epileptic seizures are caused by sudden and excessive neuronal discharges, which lead to abnormal electrical activity in the brain. These events are reflected in EEG signals as characteristic patterns such as spikes, sharp waves, and rhythmic discharges [11]. One of the most important challenges in this field is the early prediction of seizures during the preictal phase, where subtle changes in EEG signals may indicate an upcoming seizure.

Despite significant progress in signal processing and machine learning methods, seizure prediction remains a challenging problem due to the nonlinear, patient-specific, and highly dynamic

nature of EEG signals. Existing approaches often fail to simultaneously achieve high prediction accuracy and early detection capability.

In recent years, optimization-based methods have been widely used for feature selection and pattern discovery in EEG analysis. In particular, evolutionary algorithms and swarm intelligence techniques have shown strong performance in exploring large search spaces and identifying informative signal components [12, 9]. However, most of these methods lack a unified probabilistic framework for modeling uncertainty and incorporating prior knowledge.

Bayesian inference provides a principled framework for probabilistic reasoning and uncertainty modeling. It enables the estimation of posterior probabilities associated with seizure occurrence based on observed EEG data [8]. However, Bayesian methods alone are often insufficient for efficiently exploring large combinational spaces such as EEG channel-time block selection.

To overcome these limitations, this study proposes a novel multi-stage hybrid optimization framework that integrates Bayesian modeling with multiple evolutionary and metaheuristic algorithms, including genetic algorithm (GA), evolutionary programming (EP), evolution strategies (ESs), ant colony optimization (ACO), artificial bee colony (ABC), and harmony search (HS) [14, 20, 12, 10, 9]. The proposed method formulates seizure prediction as a multi-objective optimization problem, aiming to maximize the posterior probability of seizure occurrence while minimizing detection delay.

The main contributions of this work are summarized as follows:

- Development of a Bayesian-based probabilistic framework for modeling EEG channel-time information.
- Definition of a novel cost function that balances seizure probability and temporal proximity.
- Design of a multi-stage hybrid optimization framework integrating Bayesian inference and evolutionary algorithms.
- Comparative evaluation of multiple metaheuristic approaches for EEG feature selection.
- Validation on real EEG datasets demonstrating the effectiveness of the proposed method [6].

The remainder of this paper is organized as follows: Section 2 presents the problem formulation and EEG signal modeling. Section 3 describes the proposed hybrid optimization framework. Section 4 presents experimental results and a discussion. Finally, Section 5 concludes the paper.

2 Materials and methods

2.1 Database

The dataset used in this study consists of EEG recordings from 14 patients collected at the Department of Neurology and Neurophysiology of the University of Siena. The cohort includes 9 male subjects (aged 25–71 years) and 5 female subjects (aged 20–58 years). EEG signals were recorded using the international 10–20 electrode placement system at a sampling frequency of 512 Hz. In addition to EEG recordings, most sessions include one or two electrocardiogram (ECG) channels to support multimodal physiological analysis. All recordings were clinically annotated and validated by a specialist neurologist based on comprehensive evaluation of patients' clinical history and electrophysiological patterns, including seizure onset and type classification [6]. The dataset provides continuous multichannel scalp EEG signals suitable for analyzing preictal and ictal dynamics in seizure prediction tasks.

2.2 Signal segmentation

Let the discrete EEG signal of channel c be represented as $x_c[n]$, where $n = 0, 1, \ldots, N-1$. The sampling interval is defined as $\Delta t = 1/f_s$, where f_s denotes the sampling frequency, and the total number of samples is given by $N = T \cdot f_s$.

For local temporal analysis and extraction of time-frequency features, each channel signal is segmented into overlapping windows. Each window B_k contains L consecutive samples and is defined as follows:

$$B_k = [x_c[k], x_c[k+1], \ldots, x_c[k+L-1]], \tag{1}$$

where k is the starting index of each window. The window shift is controlled by a step size S, such that

$$k = mS, \qquad m = 0, 1, 2, \ldots. \tag{2}$$

When $S < L$, consecutive windows overlap, which allows capturing fine-grained temporal variations in EEG signals.

This segmentation strategy improves the sensitivity of the model in detecting transient and nonstationary patterns associated with preictal and ictal brain activity, and provides a structured representation of EEG data for subsequent feature extraction and optimization stages. As illustrated in Figure 1, the EEG signal segmentation procedure is performed using overlapping sliding windows as described above.

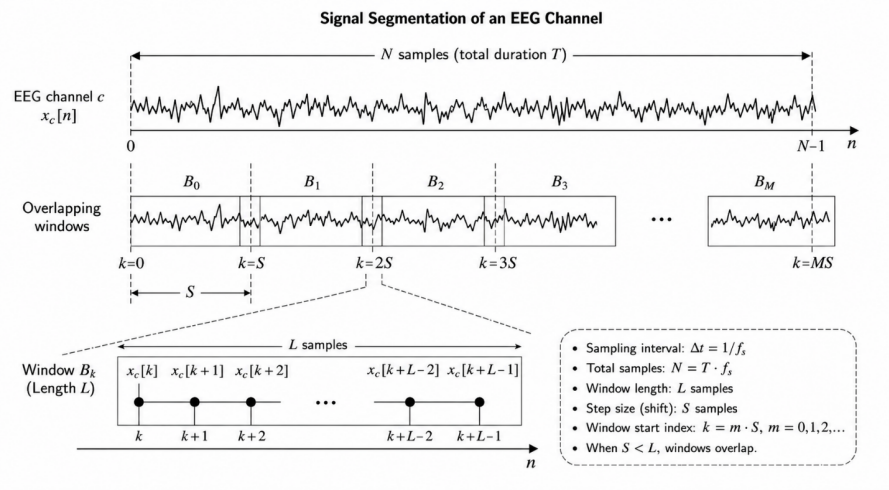


Figure 1: Signal segmentation of an EEG channel using overlapping sliding windows.

2.3 Correlation coefficient

The correlation coefficient is a statistical measure used to quantify the degree of linear dependence between two signals $x(t)$ and $y(t)$. In the context of EEG signal analysis, it is commonly used to evaluate the similarity between different brain signal segments or channels.

The correlation coefficient is defined as

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{3}$$

where

- x_i and y_i represent the individual samples of the two signals,
- $\bar{x}$ and $\bar{y}$ denote the mean values of x and y, respectively,
- n is the total number of samples in each segment.

The correlation coefficient takes values in the range $[-1, 1]$, where

- $r = 1$ indicates a perfect positive linear relationship,
- $r = -1$ indicates a perfect negative linear relationship,
- $r = 0$ indicates no linear correlation between the signals [11, 4].

In this study, the correlation coefficient is used to quantify the similarity between EEG segments and seizure-related patterns, which plays a key role in constructing the transition probability in the proposed probabilistic framework.

2.4 Bayes' theorem

Bayes' theorem provides a probabilistic framework for updating prior knowledge about a hypothesis H based on observed data D. It plays a fundamental role in uncertainty modeling and inference in complex systems such as EEG signal analysis.

Mathematically, the posterior probability is defined as

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}, \tag{4}$$

where

- $P(H)$ is the prior probability representing the initial belief about hypothesis H,
- $P(D|H)$ is the likelihood, representing the probability of observing data D given H,
- $P(H|D)$ is the posterior probability, which updates the belief about H after observing data D,
- $P(D)$ is the marginal likelihood (evidence), acting as a normalization term.

In the context of EEG signal analysis, H represents the occurrence of seizure-related activity within a specific channel or time block, while D corresponds to the observed EEG segment features. Therefore, Bayes' theorem enables the integration of prior temporal knowledge and observed signal characteristics to estimate the probability of seizure occurrence in each EEG segment [18].

2.5 Evolutionary algorithms

Evolutionary algorithms are a class of population-based stochastic optimization methods inspired by the principles of natural evolution, including selection, mutation, and recombination. These methods are widely used for solving complex optimization problems where the search space is large, nonlinear, and not analytically tractable.

In evolutionary optimization, a population of candidate solutions is first initialized randomly within the search space. Each candidate solution $x \in X$ is evaluated using a predefined fitness function $f(x)$, which measures its quality with respect to the optimization objective.

Based on fitness values, individuals with higher performance are selected to generate the next generation. New solutions are created through genetic operators such as mutation and crossover. Mutation is typically defined as

$$\hat{x} = x + \epsilon, \tag{5}$$

where ϵ is a small random perturbation vector that introduces diversity into the population. Crossover combines information from two parent solutions x_1 and x_2 to generate a new offspring solution:

$$x_{\text{new}} = \alpha x_1 + (1 - \alpha)x_2, \qquad 0 < \alpha < 1. \tag{6}$$

This iterative process continues until a stopping criterion is satisfied, such as reaching a maximum number of generations or convergence to a stable solution [7].

In the context of this study, each candidate solution represents a selection of EEG channel and time block, and the fitness function is defined based on the posterior probability of seizure occurrence. Therefore, evolutionary algorithms are used to efficiently explore the high-dimensional search space of EEG segments and identify the most informative preictal patterns.

2.6 Common types of evolutionary algorithms

Among the most widely used evolutionary and swarm intelligence algorithms in optimization problems are the following methods:

2.6.1 GA

GA is inspired by natural selection and genetic evolution processes. It employs selection, crossover, and mutation operators to evolve a population of candidate solutions over successive generations [17, 1].

The basic procedure of GA can be summarized as follows:

- **Initial population generation:** A set of random candidate solutions (chromosomes) is created.
- **Fitness evaluation:** The objective function is computed for each chromosome to determine its fitness.
- **Selection:** Chromosomes with higher fitness are selected for reproduction.
- **Crossover:** Two parent chromosomes are combined to generate offspring.
- **Mutation:** Random modifications are applied to maintain genetic diversity.
- **Iteration:** The process is repeated until a stopping criterion such as convergence or a maximum number of generations is reached [21].

2.6.2 ES

ESs are optimization algorithms based on selection and mutation, with a strong emphasis on self-adaptation of strategy parameters [10].

The main steps of ES are

- **Initialization:** A population of individuals is randomly generated along with strategy parameters (e.g., mutation variance).
- **Evaluation:** Each individual is evaluated using a fitness function.
- **Selection of parents:** The best individuals are selected as parents.
- **Mutation:** Offspring are generated by adding stochastic perturbations (commonly Gaussian noise) to parents. No crossover is typically used in basic ES.
- **Evaluation of offspring:** The new solutions are evaluated.

- **Replacement:** A new population is formed using either:
  - $(\mu + \lambda)$-ES: selection from both parents and offspring.
  - (μ, λ)-ES: selection only from offspring.
- **Termination:** The process continues until convergence or a maximum number of iterations is reached [10].

2.6.3 EP

EP focuses on optimizing the behavior and structure of solutions, particularly in machine learning and system modeling tasks [11].

Its main procedure includes

- **Initialization:** Random generation of an initial population.
- **Fitness evaluation:** Each individual is evaluated using a fitness function.
- **Mutation:** Only mutation is applied; crossover is not used.
- **Evaluation of offspring:** Newly generated individuals are evaluated.
- **Selection:** The best individuals from parents and offspring are selected, often using tournament selection.
- **Iteration:** The process continues until a stopping criterion is satisfied [11, 6].

2.7 Swarm intelligence based methods

In addition to evolutionary algorithms, swarm intelligence methods are inspired by the collective behavior of biological agents and have been widely used in optimization problems.

2.7.1 ACO

ACO is inspired by the foraging behavior of ants and their use of pheromone trails to find shortest paths [3].

The key steps of ACO include

- **Initialization:** Setting initial pheromone levels.
- **Solution construction:** Each ant constructs a path probabilistically based on pheromone and heuristic information.

  The transition probability is defined as follows:

$$P_{ij}^{(k)} = \frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{l \in allowed}[\tau_{il}]^{\alpha}[\eta_{il}]^{\beta}}. \tag{7}$$

- **Pheromone update:** After all ants complete their paths, pheromone values are updated based on evaporation and deposited pheromone.
- **Termination:** The algorithm stops when a maximum number of iterations is reached or convergence occurs [3, 6].

2.7.2 Particle swarm optimization (PSO)

Particle swarm optimization (PSO) is inspired by the social behavior of birds and fish during food searching [13].

The main steps of PSO include

- **Initialization:** Random initialization of particle positions and velocities.
- **Velocity update:**

$$v_i(t+1) = wv_i(t) + c_1 r_1(p_i - x_i(t)) + c_2 r_2(g - x_i(t)). \tag{8}$$

- **Position update:**

$$x_i(t+1) = x_i(t) + v_i(t+1). \tag{9}$$

- **Evaluation and update:** Personal and global best solutions are updated.
- **Iteration:** The process continues until convergence or a stopping condition is met [13, 11].

2.7.3 ABC

ABC algorithm is inspired by the foraging behavior of honey bees [10, 12].

Its main phases are

- **Initialization:** Random generation of food sources.
- **Employed bees phase:** Local search around current solutions.
- **Onlooker bees phase:** Probabilistic selection of high-quality solutions.
- **Scout bees phase:** Replacement of abandoned solutions with new random ones.
- **Best solution update:** The best food source is preserved.
- **Iteration:** The process is repeated until stopping criteria are satisfied [10, 12].

2.7.4 HS Algorithm

HS is inspired by the process of musical improvisation [9].

The main steps of HS include

- **Initialization:** Construction of the harmony memory (HM).
- **Parameter setting:** Determination of HMCR and PAR parameters.
- **Improvisation process:**
  - Memory consideration.
  - Pitch adjustment.
  - Random selection.
- **Update of HM:** A new harmony is stored if a better solution is found.
- **Iteration:** The process continues until a stopping condition is reached [9].

These algorithms are widely used in engineering, medical data analysis, artificial intelligence, and optimization problems due to their strong exploration ability, robustness against local minima, and independence from gradient information.

3 Proposed method

The proposed framework formulates epileptic seizure prediction as a probabilistic multi-objective optimization problem over segmented EEG event space. The main objective is to identify EEG signal blocks that simultaneously:

- maximize the posterior probability of seizure-related information,
- minimize the temporal distance to seizure onset.

To achieve this, a hybrid optimization architecture is designed by integrating Bayesian inference with evolutionary optimization algorithms. In this framework, Bayesian modeling is responsible for estimating probabilistic relationships within EEG event space, while evolutionary algorithms perform global search over candidate EEG segments.

3.1 Problem formulation and EEG event space modeling

Let the EEG signal be segmented into a set of temporal blocks:

$$B = \{B_1, B_2, \ldots, B_N\}. \tag{10}$$

Each block B_i is associated with

- a channel index,
- a start time,
- a fixed window length.

The goal is to select an optimal block B^* such that it carries maximum seizure-related information while appearing as early as possible in the preictal phase.

The EEG event space is divided into two temporal regimes:

- Normal (preictal) region,
- Seizure (ictal) region.

These regions are defined by two key time instants t_0 and t_1:

$$B_{\text{pre}} = \{B_i \in B \mid t_i \in [0, t_0]\}, \tag{11}$$

$$B_{\text{ictal}} = \{B_i \in B \mid t_i \in [t_0, t_1]\}. \tag{12}$$

Each EEG block is treated as a realization of a stochastic process within this event space.

3.2 Bayesian probabilistic modeling

For each EEG block B_i, three probabilistic quantities are defined:

3.2.1 Prior Probability

The prior probability represents the temporal proximity of an EEG block to seizure onset:

$$p_{\text{prior}}(B_i) = \frac{t_i}{t_s}, \tag{13}$$

where

- t_i denotes the occurrence time of block B_i,
- t_s denotes the seizure onset time.

3.2.2 Transition probability

The transition probability is defined based on the similarity between the EEG block B_i and seizure segments:

$$p(B_i|S) = \max_j \left|\text{corr}(x_i, x_j^{(S)})\right|, \tag{14}$$

where

- x_i represents the signal corresponding to block B_i,

- $x_j^{(S)}$ represents the j-th seizure segment,
- $\mathrm{corr}(\cdot)$ denotes the correlation coefficient.

The process of computing the transition probability is illustrated in Figure 2.

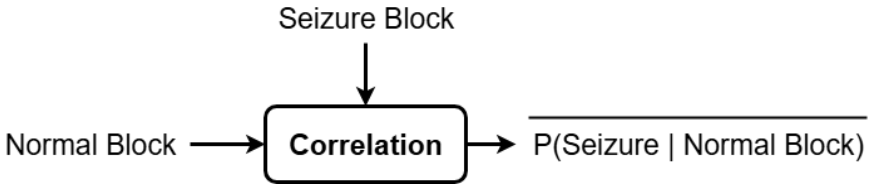


Figure 2: Block diagram for obtaining transition probability.

3.2.3 Posterior probability

Using Bayes' theorem, the posterior probability is defined as follows:

$$p(S|B_i) = \frac{p(S)p(B_i|S)}{p(B_i)}, \tag{15}$$

where

- S denotes the seizure event,
- B_i denotes the EEG block,
- $p(S)$ represents the prior probability of seizure occurrence,
- $p(B_i|S)$ represents the transition probability,
- $p(B_i)$ represents the evidence (normalization term).

To simplify the notation, we define

$$p_{\text{posterior}}(B_i) = p(S|B_i). \tag{16}$$

Since $p(B_i)$ is independent of the class label S, it does not affect the optimization process. Therefore, maximizing the posterior probability reduces to

$$\max_{B_i \in B} p(S|B_i). \tag{17}$$

The posterior maximization can be rewritten as

$$\max_{B_i \in B} p(S)p(B_i|S). \tag{18}$$

In this study, the conditional probability $p(B_i|S)$ is modeled based on the similarity between the EEG block and seizure-related patterns. This similarity is quantified using the correlation coefficient between the candidate block and seizure segments. Hence

$$p(B_i|S) = \max_j \left|\text{corr}(x_i, x_j^{(S)})\right|. \tag{19}$$

Substituting (19) into (18), the posterior score becomes

$$\max_{B_i \in B} \left\{ p(S) \max_j \left|\text{corr}(x_i, x_j^{(S)})\right| \right\}. \tag{20}$$

This formulation enables the assignment of a posterior probability to each EEG block, reflecting its relevance to seizure occurrence. The overall process for computing the posterior probability is illustrated in Figure 3.

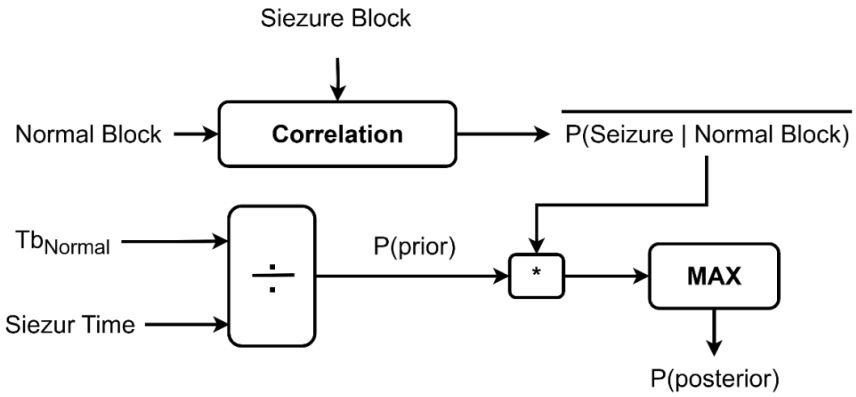


Figure 3: Block diagram for obtaining posterior probability.

3.3 Cost function formulation

The objective of this study is to identify EEG signal blocks that maximize the posterior probability of seizure occurrence while minimizing the time of occurrence. To achieve this goal, a cost

function is defined as follows:

$$J(B_i) = \frac{p(B_i)}{p(S|B_i)}, \tag{21}$$

where $p(B_i)$ represents the prior probability associated with block B_i, modeled as the normalized time index of the block. Accordingly, earlier blocks correspond to smaller prior values, which is desirable for early prediction.

From the posterior probability formulation derived in the previous section:

$$p(S|B_i) = p(S) \max_j \left|\text{corr}(x_i, x_j^{(S)})\right|, \tag{22}$$

where x_i denotes the EEG signal segment corresponding to block B_i, and $x_j^{(S)}$ represents the j-th seizure segment. The operator $\text{corr}(\cdot,\cdot)$ measures the similarity between the candidate block and seizure patterns.

This formulation implies that the posterior probability of each block is determined by its strongest similarity with seizure segments. Therefore, blocks with higher correlation values are considered more informative with respect to seizure occurrence.

Substituting (22) into (21), the cost function becomes

$$J(B_i) = \frac{p(B_i)}{p(S)\max_j \left|\text{corr}(x_i, x_j^{(S)})\right|}. \tag{23}$$

Since $p(S)$ is constant for all blocks, it does not influence the optimization process and can be removed. Therefore, the final optimization objective is simplified as

$$J(B_i) = \frac{p(B_i)}{\max_j \left|\text{corr}(x_i, x_j^{(S)})\right|}. \tag{24}$$

The final optimization problem is formulated as

$$\min_{B_i \in B} J(B_i). \tag{25}$$

The proposed cost function ensures a balanced trade-off between:

- **Temporal optimality:** selecting earlier EEG blocks through $p(B_i)$,
- **Seizure relevance:** maximizing similarity with seizure segments through correlation.

As a result, the optimization framework identifies EEG blocks that contain strong seizure-related information while appearing in the early preictal phase, which is critical for effective seizure prediction systems.

3.4 Evolutionary optimization framework

The primary objective of this stage is to identify EEG signal blocks that contain maximum seizure-related information while occurring as early as possible in time. Therefore, the optimization problem is formulated as a trade-off between maximizing seizure-related probability and minimizing the temporal distance to seizure onset.

In practical seizure prediction systems, early detection is more critical than late-stage identification. Although seizure probability generally increases as the onset approaches, informative patterns often appear as early local increases before reaching the global maximum. Hence, the goal is to detect the earliest block in which a significant rise in seizure likelihood is observed.

More formally, the target is to identify EEG blocks where

- the seizure probability exhibits an increasing trend,
- the probability is significantly higher compared to preceding blocks, even if it has not reached its maximum value.

This concept is directly incorporated into the cost function, ensuring that selected blocks simultaneously:

- contain high seizure-related information,
- occur at earlier time instances.

To solve this optimization problem, evolutionary algorithms are employed. These algorithms operate on a population of candidate solutions, where each individual represents an EEG block and is evaluated using a fitness function derived from the proposed cost formulation.

In this framework, each EEG block is encoded as an individual in the population and described by

- EEG channel index,
- normalized time position within the range $[0, 1]$.

The fitness of each individual is computed using the defined cost function, where lower values indicate more optimal solutions. The evolutionary process iteratively improves the population through selection and variation operators, guiding the search toward blocks that minimize the cost function.

In the following subsections, different evolutionary algorithms are applied to optimize the proposed objective function and determine the optimal prediction time.

3.5 Evolutionary algorithms for EEG optimization

3.5.1 GA

In this section, a GA is employed to identify the optimal EEG signal block that maximizes seizure-related information while ensuring early temporal occurrence. The objective is to select a block in which the probability of seizure-related activity is relatively high compared to other blocks, while still belonging to the pre-seizure phase, enabling early prediction.

The GA performs a global search by iteratively evolving a population of candidate solutions toward minimizing the defined cost function. Each candidate solution represents an EEG block characterized by both spatial (channel) and temporal (time) information.

3.5.2 Index sets definition

To ensure consistency across all optimization algorithms, the following index sets are defined.

Let the solution space be defined over

Channel set:

$$C = \{1, 2, \ldots, 29\}, \tag{26}$$

Time block set:

$$B = \{0, 1, \ldots, B-1\}. \tag{27}$$

Each candidate solution is defined as

$$x = (c, b), \qquad c \in C,\ b \in B, \tag{28}$$

where c denotes the EEG channel index and b represents the time block index.

3.5.3 Chromosome encoding

Each individual (chromosome) in the population is encoded as a 16-bit binary string representing the properties of an EEG block:

- The first 6 bits represent the EEG channel index (1 to 29).
- The next 10 bits represent the normalized time of the block within the interval $[0.70, 0.80]$.

This encoding allows simultaneous optimization over both spatial (channel selection) and temporal (block position) dimensions.

3.5.4 Algorithm configuration

The GA is configured as follows:

- Population size: 50.
- Number of generations: 100.
- Crossover rate: 0.8.
- Mutation rate: 0.05.

3.5.5 Genetic operators

The evolutionary process is guided by the following operators:

- **Selection:** Roulette wheel selection is used to probabilistically favor individuals with lower cost values (higher fitness).
- **Crossover:** Single-point crossover is applied to generate offspring by combining parent chromosomes.
- **Mutation:** A single random bit mutation is used to introduce diversity and prevent premature convergence.

3.5.6 Optimization objective

The GA minimizes the cost function defined in Section 3.3, thereby selecting EEG blocks that

- maximize the posterior probability of seizure occurrence,
- correspond to earlier time instances in the signal.

This approach enables the identification of informative pre-seizure blocks, which are critical for early warning systems.

3.5.7 Optimization results for different block sizes

This section presents the results obtained from applying the GA to EEG signal blocks with different window sizes, specifically 32, 64, and 128 samples. The reported values correspond to the optimal solution identified by the algorithm, which maximizes the seizure-related information while minimizing the cost function.

The results are summarized in Table 1.

Table 1: GA optimization results for various block sizes in seizure prediction

Block Size (Samples)	Maximum Seizure Probability	Selected Channel	Normalized Block Time
32	0.764	14	0.712
64	0.755	7	0.769
128	0.751	11	0.793

The results indicate that the GA successfully identifies EEG blocks that contain early indicators of seizure onset. Specifically, the selected blocks satisfy two important conditions:

- They capture the initial statistical patterns associated with seizure activity.
- They correspond to earlier time instances where seizure probability begins to increase but has not yet reached its global maximum.

Although the maximum seizure probability varies slightly across different block sizes, the algorithm consistently identifies blocks located in the pre-seizure phase, which are critical for early warning systems.

These findings demonstrate that the proposed GA-based optimization framework is effective in extracting informative temporal segments from EEG signals for seizure prediction.

3.5.8 ESs

To achieve optimal selection of EEG channels and corresponding temporal segments associated with seizure occurrence, an ES algorithm is employed. ES is a population-based stochastic optimization method inspired by biological evolution, particularly mutation and selection mechanisms, and is well-suited for continuous and discrete search spaces.

3.5.9 Representation and encoding

In this framework, each candidate solution (individual) is encoded as a 16-bit binary chromosome that simultaneously represents both spatial (channel) and temporal (time block) information.

Specifically,

- **First 6 bits:** represent the EEG channel index, corresponding to integer values in the range $[0, 63]$.

  Although the encoding allows up to 64 values, only 29 channels are physically available in the dataset; therefore, values outside the valid channel range are discarded during evaluation.

- **Next 10 bits:** represent the time block index.

The EEG signal is partitioned into a predefined number of temporal segments (e.g., 1000 blocks), allowing indices in the range $[0, 999]$.

By decoding the binary representation into decimal values, each chromosome uniquely defines:

- the selected EEG channel,
- the corresponding time block.

3.5.10 Role in optimization

This encoding enables the ES algorithm to perform a joint optimization over spatial and temporal dimensions, allowing simultaneous selection of

- the most informative EEG channels,
- the most relevant time intervals for seizure prediction.

The fitness of each individual is evaluated using the previously defined cost function, and the ES algorithm iteratively improves the population through mutation-based exploration and selection of high-quality candidates.

3.5.11 ES algorithm settings

The parameter configuration of the ES algorithm is presented in Table 2.

Table 2: ES algorithm parameters for EEG channel and time block selection

Parameter	Value
Initial Population Size	50
Number of Generations	50
Mutation Rate	0.2
Recombination Rate	0.6
Mutation Type	Single-point Binary
Stopping Criterion	Fixed number of generations or no improvement

3.5.12 Preliminary results

Table 3 presents a sample of the top-performing chromosomes obtained from the ES algorithm. These chromosomes correspond to solutions with higher fitness values, indicating a greater amount of seizure-related information within the selected EEG blocks.

Table 3: Sample chromosomes evaluated by ES algorithm

No.	Chromosome	Channel Address	Time Block Address	Fitness Value
1	0100111110100110	19	598	0.842
2	0001100111011101	6	477	0.835
3	0011000010100100	12	324	0.828
4	0100001011010011	16	723	0.822
5	0001011101110010	5	466	0.818

The fitness value represents the amount of seizure-related information contained in the EEG block selected by each chromosome. Higher fitness values indicate stronger relevance to seizure onset patterns and improved predictive capability.

These preliminary results demonstrate that the ES algorithm is capable of effectively exploring the search space and identifying informative EEG channel-time block combinations.

3.5.13 Initial Conclusion

As can be seen from the table, the ES algorithm has managed to find combinations of channels and time blocks that contain more information about the occurrence of seizures compared to other signal points. These preliminary results could serve as a useful guide for subsequent phases of analysis or the training of reinforcement learning models.

3.5.14 EP

In this section, the EP algorithm is employed to identify optimal EEG channel-time block combinations containing significant seizure-related information. EP is a population-based optimization method inspired by mutation-driven evolution. Unlike other evolutionary algorithms, EP primarily relies on mutation and selection, without using crossover operators, making it well-suited for problems requiring fine local search refinement.

3.5.15 Chromosome representation in EP

Each individual (chromosome) is encoded as a 16-bit binary string representing both spatial and temporal information:

- **First 6 bits:** represent the EEG channel index in the range $[0, 63]$. Only valid channels in the set $C = \{1, \ldots, 29\}$ are considered.
- **Next 10 bits:** represent the time block index.

The EEG signal is divided into a fixed number of temporal segments (e.g., 1000 blocks), resulting in indices in the range $[0, 999]$.

Each chromosome is decoded into a pair:

$$(\textit{channel index},\ \textit{time block index}), \tag{29}$$

which uniquely defines an EEG segment.

3.5.16 EP algorithm settings

The parameter configuration of the EP algorithm is presented in Table 4.

Table 4: EP algorithm parameters for EEG channel and time block selection

Parameter	Value
Initial Population Size	50
Number of Generations	60
Mutation Rate	0.25
Selection Method	Tournament Selection
Mutation Type	Gaussian Perturbation (on decoded integer values)
Stopping Criterion	No improvement in best fitness for 10 generations or maximum generations

As shown in Table 4, a population size of 50 and a mutation rate of 0.25 were employed to effectively explore the search space of EEG channel and temporal segment combinations.

3.5.17 Preliminary results

The preliminary results obtained from the EP algorithm are presented in Table 5. The table lists the top-ranked chromosomes together with their corresponding EEG channel addresses, time block indices, and fitness values. Higher fitness values indicate a greater amount of seizure-related information contained in the selected EEG segments.

Table 5: Sample chromosomes evaluated by EP algorithm

No.	Chromosome	Channel Address	Time Block Address	Fitness Value
1	0010011101100110	9	614	0.849
2	0001110011010101	7	437	0.841
3	0010101001011010	10	698	0.834
4	0001101100111111	6	831	0.829
5	0011010010001000	13	528	0.823

The fitness value reflects the amount of seizure-relevant information in the selected time block and channel combination.

3.5.18 Initial conclusion

The results demonstrate that the EP algorithm is capable of effectively identifying EEG channel-time block combinations that contain significant pre-seizure information. Compared to random selection, the algorithm consistently converges toward solutions with higher information content.

These findings confirm the suitability of EP for this optimization problem and suggest that the extracted blocks can serve as informative candidates for early seizure prediction.

3.5.19 ACO

In this section, the ACO algorithm is employed to identify optimal combinations of EEG channels and time blocks that contain the highest amount of seizure-related information. ACO is a population-based metaheuristic inspired by the foraging behavior of ants in nature, especially their ability to find the shortest paths to food sources using pheromone trails.

In the context of EEG signal analysis, this algorithm can effectively discover optimal paths (channel-time combinations) associated with early seizure indicators.

3.5.20 Problem formulation

In this problem, each path traveled by an ant consists of selecting an EEG channel and a time block within the signal. The objective is to find combinations with the highest seizure likelihood or preictal information.

The ACO search space is defined as follows:

- **Nodes:**
  - Channel nodes: 29 EEG channels.
  - Time nodes: EEG signal divided into 1000 temporal blocks.
- **Ant path:** Selection of one channel node and one time block node.
- **Heuristic value:** Reflects the seizure-related information content of a given channel–time combination.

The transition probability for selecting a candidate node is defined as

$$P_{ij}^{(k)} = \frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{l \in allowed}[\tau_{il}]^{\alpha}[\eta_{il}]^{\beta}}, \tag{30}$$

where τ_{ij} represents the pheromone level, η_{ij} represents the heuristic information, and α and β control the relative importance of pheromone and heuristic information, respectively.

3.5.21 ACO algorithm parameters

To effectively guide the search process in the ACO algorithm, the parameter configuration listed in Table 6 was used. These settings were selected to ensure a proper trade-off between convergence speed and solution diversity.

Table 6: ACO algorithm parameters for EEG channel and time block selection

Parameter	Value
Number of Ants	30
Number of Iterations	50
Pheromone Evaporation Rate (ρ)	0.1
Pheromone Importance (α)	1
Heuristic Importance (β)	2
Exploration Rate (q_0)	0.8
Stopping Criterion	Maximum iterations or no improvement in best path

3.5.22 Preliminary results

As shown in Table 7, the ACO algorithm successfully identified several optimal EEG channel-time block combinations with high fitness values, indicating strong seizure-related information content in the selected segments.

Table 7: Optimal paths identified by ACO

No.	Selected Channel	Selected Time Block	Fitness Value
1	14	721	0.861
2	9	658	0.853
3	5	493	0.847
4	17	780	0.839
5	12	542	0.834

3.5.23 Initial conclusion

The results demonstrate that the ACO algorithm successfully identifies combinations of EEG channels and time blocks containing high levels of pre-seizure information. These early indicators are valuable for timely seizure prediction and can also serve as informative inputs for training reinforcement learning models or other downstream decision-making modules.

3.5.24 ABC Algorithm

In this section, the ABC algorithm is applied to identify optimal combinations of EEG channels and time blocks that maximize seizure-related information content. ABC is a population-based metaheuristic inspired by the foraging behavior of honey bees and is characterized by a strong

balance between exploration and exploitation, making it suitable for high-dimensional EEG optimization problems.

3.5.25 Problem representation in ABC

Each candidate solution represents an EEG channel-time block pair defined as

$$x = (c, b), \tag{31}$$

where c denotes the EEG channel index and b represents the time block index.

The quality of each solution is evaluated using the cost function defined in Section 3.3.

3.5.26 ABC algorithm parameters

To ensure effective exploration of the search space, the ABC algorithm was configured using the parameters listed in Table 8. These settings were selected to balance convergence speed and solution quality in EEG channel and time block optimization.

Table 8: ABC algorithm parameters for EEG channel and time block selection

Parameter	Value
Number of Employed Bees	25
Number of Onlooker Bees	25
Number of Iterations	50
Mutation Rate	0.25
Abandonment Limit for Scout Phase	5
Stopping Criterion	Maximum iterations or no improvement

3.5.27 ABC algorithm phases

The ABC algorithm consists of the following main steps:

1. Initialization of candidate solutions.

2. Employed bee phase for local search around existing solutions.

3. Onlooker bee phase based on fitness probability.

4. Scout bee phase for abandoning poor solutions and generating new candidates.

5. Memorization of the best solution.

3.5.28 Preliminary results

As shown in Table 9, the ABC algorithm successfully identified several high-quality solutions corresponding to EEG channel-time block combinations with high fitness values, indicating strong seizure-related information content in the selected segments.

Table 9: ABC algorithm results for EEG channel and time block selection

No.	Selected Channel	Selected Time Block	Fitness Value
1	13	645	0.865
2	8	527	0.858
3	16	734	0.850
4	4	498	0.846
5	21	801	0.839

3.5.29 Initial Conclusion

The results demonstrate that the ABC algorithm effectively identifies EEG channel-time block combinations containing high seizure-related information. These blocks correspond to informative preictal patterns and can be utilized in downstream predictive models, such as reinforcement learning-based seizure prediction systems.

3.5.30 HS algorithm

To optimally select EEG channels and time blocks related to seizure prediction, the HS algorithm was employed. Inspired by the improvisation process in musical harmony, HS is a metaheuristic algorithm suitable for both discrete and continuous optimization problems.

In this application, each harmony vector was encoded as a 16-bit chromosome. The first 6 bits represented the EEG channel address, and the remaining 10 bits encoded the time block index.

3.5.31 HS algorithm parameters

The parameter settings used for the HS algorithm are summarized in Table 10.

Table 10: HS algorithm parameters for EEG channel and time block selection

Parameter	Value
Harmony Memory Size (HMS)	20
Harmony Memory Considering Rate (HMCR)	0.9
Pitch Adjusting Rate (PAR)	0.3
Number of Iterations	100

In each iteration, the algorithm generates new harmonies by considering memory-based components and random adjustments. The generated solutions are evaluated using the defined cost function, and the combinations with lower cost values are retained in the HM.

3.5.32 Initial results

The preliminary results obtained from the HS algorithm are presented in Table 11. Each row corresponds to a candidate solution (chromosome) representing a combination of EEG channel and time block selected during the optimization process.

Table 11: HS algorithm results for EEG channel and time block selection

Iteration	Selected Chromosome	Channel Address	Time Block (sec)	Cost Function Value
1	0101010001010110	21	86	0.034
2	1000110010100011	35	163	0.028
3	0011101100011010	14	106	0.031
4	1110000100110101	56	213	0.027
5	0110111001011100	27	93	0.029
6	0010101011001111	10	187	0.033
7	1001010110001010	37	120	0.030
8	0101110001110101	23	143	0.032
9	0001101010110001	12	98	0.031
10	1010100100101110	42	175	0.026

The table presents preliminary results of the HS algorithm for the optimal selection of EEG channels and time blocks. Each row corresponds to a candidate solution representing a combination of a specific channel and a particular time segment in the EEG signal.

The cost function value serves as the quality metric for these combinations, where lower values indicate better selections with higher seizure-related information content. The results

demonstrate that the HS algorithm successfully found solutions with low-cost values ranging from 0.026 to 0.034, reflecting its effectiveness in exploring the solution space efficiently.

Moreover, the selected channels are distributed across a wide range of channel addresses, indicating comprehensive search behavior in the state space. Similarly, the diversity in chosen time blocks reflects the algorithm's capability to identify different pre-seizure intervals within the EEG signals.

Overall, these preliminary results provide a foundation for further detailed analysis, additional optimization, and training more accurate seizure prediction models.

4 Conclusion

In this study, six metaheuristic optimization algorithms (GA, EP, ES, ACO, ABC, and HS) were evaluated for EEG channel and time-block selection in the context of early seizure prediction. The results demonstrated that all methods are capable of identifying EEG segments with high seizure-related information content, although differences in optimization performance were observed. Among the evaluated methods, the ABC and ACO algorithms achieved the highest fitness values (0.865 and 0.861, respectively), indicating superior exploration capability and solution quality. The HS algorithm achieved the lowest cost value (0.026), reflecting efficient convergence in the defined search space. The EP and ES algorithms demonstrated stable performance with consistent mid-range fitness values, suggesting robustness against local optima and noise. Across all methods, EEG channels 13, 14, 9, and 12 were most frequently selected in optimal solutions, indicating their strong association with pre-seizure activity patterns. The comparative performance of all algorithms is summarized in Table 12.

Table 12: Final comparative table of algorithm performance

Algorithm	Best Fitness / Cost	Best Channel	Best Time Block	Most Frequent Channels
GA	0.764 (Fitness)	14	0.712	14, 7, 11
ES	0.842 (Fitness)	19	598	12, 6, 5
EP	0.849 (Fitness)	9	614	9, 10, 7
ACO	0.861 (Fitness)	14	721	14, 9, 12
ABC	0.865 (Fitness)	13	645	13, 8, 16
HS	0.026 (Cost ↓)	42	175	21, 35, 14

To further enhance seizure prediction performance, future work should focus on integrating these metaheuristic optimization methods with deep learning models and developing real-time adaptive systems for clinical applications.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Alhijawi, B. and Awajan, A. *Genetic algorithms: Theory, genetic operators, solutions, and applications*, Evol. Intell. 17 (3) (2024), 1245–1256.

[2] Akram, F., Aziz, S., Khan, N.A., Faizi, S.A. and Raza, K. *Integrating Artificial Bee Colony Algorithms for Deep Learning Model Optimization: A Comprehensive Review*, In: Raza, K. (eds) Solving with Bees. Springer Tracts in Nature-Inspired Computing. Springer, Singapore, 2024.

[3] Blum, C. *Ant colony optimization: A bibliometric review*, Physiol. Life Rev. 51 (2024), 87–95.

[4] Burle, B., Spieser, L., Servant, M. and Hasbroucq, T. *EEG signals: From data to models*, Nat. Rev. Neurosci. 16 (3) (2015), 160–170.

[5] Cao, Y.J. and Wu, Q.H. *Evolutionary programming*, in *Proc. IEEE Int. Conf. Evol. Comput. (ICEC)*, 1997, pp. 443–446.

[6] Detti, P. *Siena scalp EEG database*, PhysioNet, 2020.

[7] Eiben, A.E. and Smith, J.E. *Introduction to Evolutionary Computing*, 2nd ed., Springer, Cham, Switzerland, 2015.

[8] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. *Bayesian Data Analysis*, 3rd ed., Chapman and Hall/CRC, Boca Raton, FL, USA, 2013.

[9] Geem, Z.W., Kim, J.H. and Loganathan, G.V. *A new heuristic optimization algorithm: harmony search*, Simulation 76 (2) (2001) 60–68.

[10] Hansen, N., Arnold, D.V. and Auger, A. *Evolution strategies*, in *Springer Handbook of Computational Intelligence*, Springer, 2015, pp. 871–898.

[11] Jain, N.K., Nangia, U. and Jain, J. *A review of particle swarm optimization*, J. Inst. Eng. (India), Ser. B, 99(4) (2018) 407–411.

[12] Karaboga, D. and Basturk, B. *A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm*, J. Global Optim. 39 (3) (2007), 459–471.

[13] Kennedy, J. and Eberhart, R. *Particle swarm optimization*, in *Proc. IEEE Int. Conf. Neural Netw. (ICNN'95)*, 1995, 1942–1948.

[14] Lambora, A., Gupta, K. and Chopra, K. *Genetic algorithm—A literature review*, in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, 2019, pp. 380–384.

[15] Mormann, F., Andrzejak, R.G., Elger, C.E. and Lehnertz, K. *Seizure prediction: The long and winding road*, Brain 130 (2) (2007), 314–333.

[16] Mukaka, A. *A guide to appropriate use of correlation coefficient in medical research*, Malawi Med. J. 24 (3) (2012), 69–71.

[17] Niedermeyer, E. and da Silva, F.L. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed., Lippincott Williams & Wilkins, Philadelphia, PA, 2005.

[18] Sanei, S. and Chambers, J.A. *EEG Signal Processing*, Wiley, 2013.

[19] Schober, A., Boer, C. and Schwarte, L.A. *Correlation coefficients: Appropriate use and interpretation*, Anesth. Analg. 126 (5) (2018), 1763–1768.

[20] Slowik, A. and Kwasnicka, H. *Evolutionary algorithms and their applications to engineering problems*, Neural Comput. Appl. 32 (16) (2020), 12363–12379.

[21] Waysi, D.N.S., Ahmed, B.T. and Ibrahim, I.M. *Optimization by nature: A review of genetic algorithm techniques*, Indones. J. Comput. Sci. 14 (1) (2025) 268–284.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1464–1483
https://doi.org/10.22067/ijnao.2026.98129.1861
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Coupled fixed points for a class of noncyclic contractions in uniformly convex Banach spaces with error estimates

A. Safari-Hafshejani*, and K. Amiri

Abstract

In this paper, we introduce a noncyclic contraction pair for an ordered pair of mappings. We then deduce the existence and uniqueness of coupled fixed points by associating a noncyclic contraction with them. In addition, we obtain both a priori and a posteriori error estimates for these points in uniformly convex Banach spaces. The main result is illustrated with an example, and as an application, a cyclic version of the main theorem is derived.

AMS subject classifications (2020): Primary 41A25; Secondary 47H10, 54H25.

Keywords: Noncyclic contraction pair; Coupled fixed point; Coupled best proximity point; A priori and a posteriori error estimates; UC property.

*Corresponding author

Received 11 March 2026; revised 8 June 2026; accepted 9 June 2026

Akram Safari-Hafshejani
Department of Pure Mathematics, Payame Noor University (PNU), P. O. Box: 19395-3697, Tehran, Iran.
e-mail: asafari@pnu.ac.ir

Khadije Amiri
Department of Pure Mathematics, Payame Noor University (PNU), P. O. Box: 19395-3697, Tehran, Iran.
e-mail: amirikhadije@student.pnu.ac.ir

How to cite this article
Safari-Hafshejani, A. and Amiri, Kh., Coupled fixed points for a class of noncyclic contractions in uniformly convex Banach spaces with error estimates. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1464-1483.
https://doi.org/10.22067/ijnao.2026.98129.1861

1 Introduction and preliminaries

Let $(\mathcal{Z}, \varrho)$ be a complete metric space, and let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty closed subsets of $\mathcal{Z}$. A map $T : \mathcal{K} \cup \bar{\mathcal{K}} \to \mathcal{K} \cup \bar{\mathcal{K}}$ is termed a cyclic contraction map if it satisfies

- T is cyclic, that is, $T(\mathcal{K}) \subseteq \bar{\mathcal{K}}$ and $T(\bar{\mathcal{K}}) \subseteq \mathcal{K}$;
- there exists a constant $k \in (0, 1)$ such that for all $x \in \mathcal{K}$ and $\bar{x} \in \bar{\mathcal{K}}$, the following inequality holds:
$$\varrho(Tx, T\bar{x}) \leq k\varrho(x, \bar{x}) + (1 - k)\varrho(\mathcal{K}, \bar{\mathcal{K}});$$

and is termed a noncyclic contraction map if it satisfies

- T is noncyclic, that is, $T(\mathcal{K}) \subseteq \mathcal{K}$ and $T(\bar{\mathcal{K}}) \subseteq \bar{\mathcal{K}}$;
- there exists a constant $k \in (0, 1)$ such that for all $x \in \mathcal{K}$ and $\bar{x} \in \bar{\mathcal{K}}$, the following inequality holds:
$$\varrho(Tx, T\bar{x}) \leq k\varrho(x, \bar{x}) + (1 - k)\varrho(\mathcal{K}, \bar{\mathcal{K}}),$$

where $\varrho(\mathcal{K}, \bar{\mathcal{K}}) := \inf\{\varrho(x, \bar{x}) : x \in \mathcal{K}, \bar{x} \in \bar{\mathcal{K}}\}$. A point $x^* \in \mathcal{K} \cup \bar{\mathcal{K}}$ is considered a best proximity point for a cyclic map T if $\varrho(x^*, Tx^*) = \varrho(\mathcal{K}, \bar{\mathcal{K}})$ and is a fixed point for a cyclic or noncyclic map T if $Tx^* = x^*$ (see [9, 1] and their references).

The main theorems of [6, 7, 13] prove the existence of a fixed point and a best proximity point for two types of cyclic mappings. Additionally, they establish error estimates for the Picard iteration concerning these points.

Some generalizations of the concepts of a best proximity point and a fixed point are presented through the definitions in [11, 8, 3].

Definition 1 (see [11]). Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of the metric space $(\mathcal{Z}, \varrho)$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}}$. An ordered pair $(x^*, x^{**}) \in \mathcal{K} \times \mathcal{K}$ is called a coupled best proximity point of Γ if
$$\varrho(x^*, \Gamma(x^*, x^{**})) = \varrho(\mathcal{K}, \bar{\mathcal{K}}) \quad \text{and} \quad \varrho(x^{**}, \Gamma(x^{**}, x^*)) = \varrho(\mathcal{K}, \bar{\mathcal{K}}).$$

Definition 2 (see [3]). Let $\mathcal{K}$ be nonempty subset of the metric space $(\mathcal{Z}, \varrho)$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$. An ordered pair $(x^*, x^{**}) \in \mathcal{K} \times \mathcal{K}$ is called a coupled fixed point of Γ if $\Gamma(x^*, x^{**}) = x^*$ and $\Gamma(x^{**}, x^*) = x^{**}$.

Definition 3 (see [11]). Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of the metric space $(\mathcal{Z}, \varrho)$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \mathcal{K}$. For any initial values $x_0, y_0 \in \mathcal{K}$, let the sequences $\{x_n\}$ and $\{y_n\}$ be defined by

$$x_{2n+1} := \Gamma(x_{2n}, y_{2n}) \in \bar{\mathcal{K}}, \qquad x_{2n+2} := \bar{\Gamma}(x_{2n+1}, y_{2n+1}) \in \mathcal{K},$$
$$y_{2n+1} := \Gamma(y_{2n}, x_{2n}) \in \bar{\mathcal{K}}, \qquad y_{2n+2} := \bar{\Gamma}(y_{2n+1}, x_{2n+1}) \in \mathcal{K},$$

for every $n \geq 0$. Also, we define $\Gamma' : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}}$ and $\bar{\Gamma}' : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \mathcal{K}$ by

$$\Gamma'(x, y) = \Gamma(y, x), \text{ for any pairs } (x, y) \in \mathcal{K} \times \mathcal{K},$$
$$\bar{\Gamma}'(x, y) = \bar{\Gamma}(y, x), \text{ for any pairs } (x, y) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}.$$

Remark 1. Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of the metric space $(\mathcal{Z}, \varrho)$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \mathcal{K}$. We define mappings $(\Gamma, \Gamma') : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}} \times \bar{\mathcal{K}}$ and $(\bar{\Gamma}, \bar{\Gamma}') : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \mathcal{K} \times \mathcal{K}$ by

$$(\Gamma, \Gamma')(x, y) = (\Gamma(x, y), \Gamma'(x, y)) \quad \text{for} \quad (x, y) \in \mathcal{K} \times \mathcal{K}$$

and

$$(\bar{\Gamma}, \bar{\Gamma}')(x, y) = (\bar{\Gamma}(x, y), \bar{\Gamma}'(x, y)) \quad \text{for} \quad (x, y) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}.$$

Then for the sequences $\{x_n\}$ and $\{y_n\}$ defined in Definition 3, we have

$$(x_{2n+1}, y_{2n+1}) = (\Gamma, \Gamma')(x_{2n}, y_{2n}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}},$$
$$(x_{2n+2}, y_{2n+2}) = (\bar{\Gamma}, \bar{\Gamma}')(x_{2n+1}, y_{2n+1}) \in \mathcal{K} \times \mathcal{K},$$

for every $n \geq 0$. So

$$\begin{aligned}
&(x_1, y_1) = (\Gamma, \Gamma')(x_0, y_0),\\
&(x_2, y_2) = (\bar{\Gamma}, \bar{\Gamma}')(x_1, y_1) = (\bar{\Gamma}, \bar{\Gamma}')((\Gamma, \Gamma')(x_0, y_0)) = ((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))(x_0, y_0),\\
&(x_3, y_3) = (\Gamma, \Gamma')(x_2, y_2) = (\Gamma, \Gamma')((\bar{\Gamma}, \bar{\Gamma}')(x_1, y_1)) = ((\Gamma, \Gamma') \circ (\bar{\Gamma}, \bar{\Gamma}'))(x_1, y_1),\\
&(x_4, y_4) = (\bar{\Gamma}, \bar{\Gamma}')(x_3, y_3) = (\bar{\Gamma}, \bar{\Gamma}')((\Gamma, \Gamma')(x_2, y_2)) = ((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))^2(x_0, y_0),\\
&\vdots\\
&(x_{2n}, y_{2n}) = ((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))^n(x_0, y_0),\\
&(x_{2n+1}, y_{2n+1}) = ((\Gamma, \Gamma') \circ (\bar{\Gamma}, \bar{\Gamma}'))^n(x_1, y_1),
\end{aligned}$$

for every $n \in \mathbb{N} \cup \{0\}$.

We now refer to the definition of a cyclic contraction of an ordered pair defined by Sintunavarat and Kumam [11].

Definition 4 (see [11]). Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of the metric space $(\mathcal{Z}, \varrho)$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \mathcal{K}$. The ordered pair $(\Gamma, \bar{\Gamma})$ is said to be a cyclic contraction pair if there exist nonnegative numbers λ, μ such that $\lambda + \mu < 1$, and the following inequality holds:

$$\varrho(\Gamma(x, y), \bar{\Gamma}(\bar{x}, \bar{y})) \leq \lambda \varrho(x, \bar{x}) + \mu \varrho(y, \bar{y}) + (1 - (\lambda + \mu)) \varrho(\mathcal{K}, \bar{\mathcal{K}}),$$

for all $(x, y) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}, \bar{y}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$.

Definition 5 (see [12]). Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of the metric space $(\mathcal{Z}, \varrho)$. We say that the pair $(\mathcal{K}, \bar{\mathcal{K}})$ satisfies the UC property if the following condition holds: whenever sequences $\{x_n\}$ and $\{x'_n\}$ in $\mathcal{K}$ and a sequence $\{\bar{x}_n\}$ in $\bar{\mathcal{K}}$ satisfy

$$\lim_{n \to \infty} \varrho(x_n, \bar{x}_n) = \lim_{n \to \infty} \varrho(x'_n, \bar{x}_n) = \varrho(\mathcal{K}, \bar{\mathcal{K}}),$$

then it follows that $\lim_{n \to \infty} \varrho(x_n, x'_n) = 0$.

Suzuki, Kikkawa, and Vetro [12] demonstrated that if $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty, convex subsets of a uniformly convex Banach space $\mathcal{Z}$, then the pair $(\mathcal{K}, \bar{\mathcal{K}})$ possesses the UC property. They also established that if $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty subsets of a metric space $(\mathcal{Z}, \varrho)$ such that $\varrho(\mathcal{K}, \bar{\mathcal{K}}) = 0$, then the pair $(\mathcal{K}, \bar{\mathcal{K}})$ satisfies the UC property.

To obtain an error estimate for the sequence of successive iterations approximating the best proximity point, generated by a cyclic contraction, the modulus of convexity $\delta_{\mathcal{Z}}$ is used [13].

Definition 6 (see [2]). A Banach space $\mathcal{Z}$ is said to be uniformly convex if there exists a strictly increasing function $\delta_{\mathcal{Z}} : [0, 2] \to [0, 1]$ such that the following implication holds for all $x, x', \bar{x} \in \mathcal{Z}$, $\mathcal{R} > 0$ and $r \in [0, 2\mathcal{R}]$:

$$\left.\begin{array}{l} \|x - \bar{x}\| \leq \mathcal{R} \\ \|x' - \bar{x}\| \leq \mathcal{R} \\ \|x - x'\| \geq r \end{array}\right\} \quad \Rightarrow \quad \|\frac{x + x'}{2} - \bar{x}\| \leq \big(1 - \delta_{\mathcal{Z}}(\frac{r}{\mathcal{R}})\big)\mathcal{R}. \tag{1}$$

In this paper, we demonstrate that a noncyclic contraction pair $(\Gamma, \bar{\Gamma})$ can be formulated as a noncyclic contraction within the space $\mathcal{Z}^2 = \mathcal{Z} \times \mathcal{Z}$ through the adoption of a suitable metric. This formulation facilitates the derivation of our main results concerning the existence and uniqueness of coupled fixed points for noncyclic contraction pairs, as well as the estimation of a priori and a posteriori error bounds.

2 Auxiliary results based on property UC

For a complete metric space $(\mathcal{Z}, \varrho)$, we define $\mathcal{Z}^2 = \mathcal{Z} \times \mathcal{Z}$ and $\varrho^* : \mathcal{Z}^2 \times \mathcal{Z}^2 \to \mathbb{R}^+$ such that $\varrho^*((x,y),(\bar{x},\bar{y})) = \max\{\varrho(x,\bar{x}), \varrho(y,\bar{y})\}$. It can easily be seen that the completeness of the metric space $(\mathcal{Z}^2, \varrho^*)$ follows from the completeness of $(\mathcal{Z}, \varrho)$. On the normed space $(\mathcal{Z}, \|\cdot\|)$, we use $\|\cdot\|^*$ to denote the norm on $\mathcal{Z}^2$, defined as $\|(x,y)\|^* = \max\{\|x\|, \|y\|\}$, so we have $\|(x,y) - (\bar{x},\bar{y})\|^* = \max\{\|x-\bar{x}\|, \|y-\bar{y}\|\}$.

Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of a metric space $(\mathcal{Z}, \varrho)$. Then

$$
\begin{aligned}
&\varrho^*(\mathcal{K}\times\mathcal{K}, \bar{\mathcal{K}}\times\bar{\mathcal{K}}) \\
&\quad = \inf\{\varrho^*((x,y),(\bar{x},\bar{y})) : (x,y) \in \mathcal{K}\times\mathcal{K}, (\bar{x},\bar{y}) \in \bar{\mathcal{K}}\times\bar{\mathcal{K}}\} \\
&\quad \le \inf\{\varrho^*((x,x),(\bar{x},\bar{x})) : x \in \mathcal{K}, \bar{x} \in \bar{\mathcal{K}}\} \\
&\quad = \inf\{\varrho(x,\bar{x}) : (x,\bar{x}) \in \mathcal{K}\times\bar{\mathcal{K}}\} \\
&\quad = \varrho(\mathcal{K}, \bar{\mathcal{K}}) \\
&\quad = \inf\{\varrho(x,\bar{x}) : (x,\bar{x}) \in \mathcal{K}\times\bar{\mathcal{K}}\} \\
&\quad \le \inf\{\max\{\varrho(x,\bar{x}), \varrho(y,\bar{y})\} : (x,\bar{x}) \in \mathcal{K}\times\bar{\mathcal{K}}, (y,\bar{y}) \in \mathcal{K}\times\bar{\mathcal{K}}\} \\
&\quad = \inf\{\varrho^*((x,y),(\bar{x},\bar{y})) : (x,y) \in \mathcal{K}\times\mathcal{K}, (\bar{x},\bar{y}) \in \bar{\mathcal{K}}\times\bar{\mathcal{K}}\} \\
&\quad = \varrho^*(\mathcal{K}\times\mathcal{K}, \bar{\mathcal{K}}\times\bar{\mathcal{K}}).
\end{aligned}
$$

So we obtain

$$
\varrho^*(\mathcal{K}\times\mathcal{K}, \bar{\mathcal{K}}\times\bar{\mathcal{K}}) = \varrho(\mathcal{K}, \bar{\mathcal{K}}). \tag{2}
$$

Definition 7. Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of a Banach space $(\mathcal{Z}, \|\cdot\|)$ such that $(\mathcal{K}, \bar{\mathcal{K}})$ has the UC property and $\varrho := \varrho(\mathcal{K}, \bar{\mathcal{K}}) > 0$. If there exist $C > 0$ and $p > 0$ such that the following implication holds for all $x, x' \in \mathcal{K}$, $\bar{x} \in \bar{\mathcal{K}}$ and $\mathcal{R} \ge \varrho$:

$$
\left.\begin{aligned} \|x-\bar{x}\| &\le \mathcal{R}, \\ \|x'-\bar{x}\| &\le \mathcal{R}, \end{aligned}\right\} \quad \Rightarrow \quad \|x-x'\| \le \mathcal{R}\sqrt[p]{\frac{\mathcal{R}-\varrho}{\mathcal{R}C}}, \tag{3}
$$

then, we say that $(\mathcal{K}, \bar{\mathcal{K}})$ has property UC with convexity approximation coefficients C and p.

Note that in the previous definition, since $(\mathcal{K}, \bar{\mathcal{K}})$ has the UC property, relation (3) clearly holds when $\mathcal{R} = \varrho$. Therefore, to prove that the definition is satisfied, it is enough to show that relation (3) holds for $\mathcal{R} > \varrho$.

Moreover, [10, Theorem 2] can be stated as follows for sets $\mathcal{K}$ and $\bar{\mathcal{K}}$ satisfying the property UC with convexity approximation coefficients C and p, using the same proof as provided by the original authors.

Theorem 1. Suppose that $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty, closed subsets of a Banach space $(\mathcal{Z}, \|\cdot\|)$ such that $(\mathcal{K}, \bar{\mathcal{K}})$ and $(\bar{\mathcal{K}}, \mathcal{K})$ have the UC property and $\varrho := \varrho(\mathcal{K}, \bar{\mathcal{K}}) > 0$. Let $T : \mathcal{K} \cup \bar{\mathcal{K}} \to \mathcal{K} \cup \bar{\mathcal{K}}$ be a noncyclic contraction map; that is,

$$\varrho(Tx, T\bar{x}) \leq k\varrho(x, \bar{x}) + (1-k)\varrho(\mathcal{K}, \bar{\mathcal{K}}),$$

for all $x \in \mathcal{K}$ and $\bar{x} \in \bar{\mathcal{K}}$ and some $k \in [0, 1)$. Then

(i) T has a unique optimal pair of fixed points $(x^*, \bar{x}^*) \in \mathcal{K} \times \bar{\mathcal{K}}$;

(ii) for every $x_0 \in \mathcal{K}$ and $\bar{x}_0 \in \bar{\mathcal{K}}$ the sequences $\{x_n\} = \{T^n x_0\}$ and $\{\bar{x}_n\} = \{T^n \bar{x}_0\}$ converge to x^* and $\bar{x}^*$, respectively.

Furthermore, if there exist $C > 0$ and $p > 0$ such that $(\mathcal{K}, \bar{\mathcal{K}})$ has the UC property with convexity approximation coefficients C and p, then

(iii) the following priori error estimate holds

$$\|x^* - T^m x_0\| \leq \frac{\Delta_{x_0, \bar{x}_0}}{1 - \sqrt[p]{k}} \sqrt[p]{\frac{\Delta_{x_0, \bar{x}_0} - \varrho}{C\varrho}} (\sqrt[p]{k})^m;$$

(iv) the following posteriori error estimate holds

$$\|T^n x_0 - x^*\| \leq \frac{\Delta_{x_n, \bar{x}_n}}{1 - \sqrt[p]{k}} \sqrt[p]{\frac{\Delta_{x_n, \bar{x}_n} - \varrho}{C\varrho}};$$

where $\Delta_{x,\bar{x}} := \max\{\|x - \bar{x}\|, \|Tx - \bar{x}\|\}$, for every $x \in \mathcal{K}$ and $\bar{x} \in \bar{\mathcal{K}}$.

The following lemma is necessary to obtain the coupled version of Theorem 1 in the next section.

Lemma 1. Suppose that $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty, closed, and convex subsets of a uniformly convex Banach space $(\mathcal{Z}, \|\cdot\|)$ such that $\varrho := \varrho(\mathcal{K}, \bar{\mathcal{K}}) > 0$. If $\delta_{\mathcal{Z}}(\epsilon) \geq C\epsilon^p$ for some $C > 0$, $p > 0$ and every $\epsilon \in (0, 2]$, then the pair $(\mathcal{K} \times \mathcal{K}, \bar{\mathcal{K}} \times \bar{\mathcal{K}})$ as subsets of the Banach space $(\mathcal{Z}^2, \|\cdot\|^*)$, has property UC with convexity approximation coefficients C and p.

Proof. It is easily proven that the pair $(\mathcal{K}\times\mathcal{K}, \bar{\mathcal{K}}\times\bar{\mathcal{K}})$ as subsets of the Banach space $(\mathcal{Z}^2, \|\cdot\|^*)$, has property UC. Since $(\mathcal{Z}, \|\cdot\|)$ is a uniformly convex Banach space, then there exists a strictly increasing function $\delta_{\mathcal{Z}} : [0,2] \to [0,1]$ such that the implication (1) holds for all $x, x', \bar{x} \in \mathcal{Z}$, $\mathcal{R} > 0$ and $r \in [0, 2\mathcal{R}]$. Let $(x,y),(x',y') \in \mathcal{K}\times\mathcal{K}$, $(\bar{x},\bar{y}) \in \bar{\mathcal{K}}\times\bar{\mathcal{K}}$, $\mathcal{R} > \varrho$ and

$$\begin{cases} \|(x,y)-(\bar{x},\bar{y})\|^* \le \mathcal{R}, \\ \|(x',y')-(\bar{x},\bar{y})\|^* \le \mathcal{R}. \end{cases}$$

So we obtain

$$\begin{cases} \|x-\bar{x}\| \le \mathcal{R}, \\ \|x'-\bar{x}\| \le \mathcal{R}, \end{cases} \quad \textit{and} \quad \begin{cases} \|y-\bar{y}\| \le \mathcal{R}, \\ \|y'-\bar{y}\| \le \mathcal{R}. \end{cases}$$

Since $(\mathcal{Z}, \|\cdot\|)$ is a uniformly convex Banach space, then

$$\varrho \le \|\frac{x+x'}{2} - \bar{x}\| \le (1-\delta_{\mathcal{Z}}(\frac{\|x-x'\|}{\mathcal{R}}))\mathcal{R}$$

and

$$\varrho \le \|\frac{y+y'}{2} - \bar{y}\| \le (1-\delta_{\mathcal{Z}}(\frac{\|y-y'\|}{\mathcal{R}}))\mathcal{R}.$$

It follows from the inequality $\delta_{\mathcal{Z}}(\epsilon) \ge C\epsilon^p$ that $\delta_{\mathcal{Z}}^{-1}(\epsilon) \le \left(\frac{\epsilon}{C}\right)^{\frac{1}{p}}$ for every $\epsilon \in (0,2]$. Since $0 < \frac{\mathcal{R}-\varrho}{\mathcal{R}} < 1$, then we get

$$\|x-x'\| \le \delta_{\mathcal{Z}}^{-1}(\frac{\mathcal{R}-\varrho}{\mathcal{R}})\mathcal{R} \le \mathcal{R}\sqrt[p]{\frac{\mathcal{R}-\varrho}{\mathcal{R}C}}$$

and

$$\|y-y'\| \le \delta_{\mathcal{Z}}^{-1}(\frac{\mathcal{R}-\varrho}{\mathcal{R}})\mathcal{R} \le \mathcal{R}\sqrt[p]{\frac{\mathcal{R}-\varrho}{\mathcal{R}C}}.$$

Therefore

$$\|(x,y)-(x',y')\|^* = \max\{\|x-x'\|, \|y-y'\|\} \le \mathcal{R}\sqrt[p]{\frac{\mathcal{R}-\varrho}{\mathcal{R}C}}.$$

□

3 Coupled fixed points for noncyclic contraction pair

We define a noncyclic contraction for an ordered pair of mappings as follows.

Definition 8. Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty subsets of the metric space $(\mathcal{Z}, \varrho)$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \bar{\mathcal{K}}$. The ordered pair $(\Gamma, \bar{\Gamma})$ is said to be a noncyclic contraction pair if there exist nonnegative numbers λ, μ such that $\lambda + \mu < 1$, and the following inequality holds:

$$\varrho(\Gamma(x, y), \bar{\Gamma}(\bar{x}, \bar{y})) \leq \lambda \varrho(x, \bar{x}) + \mu \varrho(y, \bar{y}) + (1 - (\lambda + \mu)) \varrho(\mathcal{K}, \bar{\mathcal{K}}),$$

for all $(x, y) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}, \bar{y}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$.

For any initial values $x_0, y_0 \in \mathcal{K}$ and $\bar{x}_0, \bar{y}_0 \in \bar{\mathcal{K}}$, let the sequences $\{x_n\}$, $\{y_n\}$, $\{\bar{x}_n\}$, and $\{\bar{y}_n\}$ be defined by

$$\begin{aligned} x_{n+1} &:= \Gamma(x_n, y_n) \in \mathcal{K}, \qquad \bar{x}_{n+1} := \bar{\Gamma}(\bar{x}_n, \bar{y}_n) \in \bar{\mathcal{K}}, \\ y_{n+1} &:= \Gamma(y_n, x_n) \in \mathcal{K}, \qquad \bar{y}_{n+1} := \bar{\Gamma}(\bar{y}_n, \bar{x}_n) \in \bar{\mathcal{K}}, \end{aligned}$$

for every $n \geq 0$. From Remark 1,

$$\begin{aligned} (x_{n+1}, y_{n+1}) &= (\Gamma, \Gamma')(x_n, y_n) \in \mathcal{K} \times \mathcal{K}, \\ (\bar{x}_{n+1}, \bar{y}_{n+1}) &= (\bar{\Gamma}, \bar{\Gamma}')(\bar{x}_n, \bar{y}_n) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}, \end{aligned}$$

and

$$\begin{aligned} \{(x_n, y_n)\} &= \{(\Gamma, \Gamma')^n (x_0, y_0)\}, \\ \{(\bar{x}_n, \bar{y}_n)\} &= \{(\bar{\Gamma}, \bar{\Gamma}')^n (\bar{x}_0, \bar{y}_0)\}, \end{aligned}$$

for every $n \geq 0$.

We continue this section by presenting two practical lemmas that will be useful in subsequent analyses.

Lemma 2. Let $\mathcal{K}$ and $\bar{\mathcal{K}}$ be nonempty closed subsets of a complete metric space $(\mathcal{Z}, \varrho)$ and let $\Gamma : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \bar{\mathcal{K}}$. Suppose that the ordered pair $(\Gamma, \bar{\Gamma})$ is a noncyclic contraction pair. Then the noncyclic mapping $T : (\mathcal{K} \times \mathcal{K}) \cup (\bar{\mathcal{K}} \times \bar{\mathcal{K}}) \to (\mathcal{K} \times \mathcal{K}) \cup (\bar{\mathcal{K}} \times \bar{\mathcal{K}})$, defined by

$$T(x, y) = \begin{cases} (\Gamma, \Gamma')(x, y) & \text{if } (x, y) \in \mathcal{K} \times \mathcal{K}, \\ (\bar{\Gamma}, \bar{\Gamma}')(x, y) & \text{if } (x, y) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}, \end{cases}$$

is a noncyclic contraction in the complete metric spaces $(\mathcal{Z}^2, \varrho^*)$ with constant $\lambda + \mu$.

Proof. Consider the fixed but arbitrary points $(x, y) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}, \bar{y}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$. The ordered pair $(\Gamma, \bar{\Gamma})$ is a noncyclic contraction pair, so we have

$$\begin{aligned}
\varrho(\Gamma(x,y), \bar{\Gamma}(\bar{x},\bar{y})) &\leq \lambda\varrho(x,\bar{x}) + \mu\varrho(y,\bar{y}) + (1-(\lambda+\mu))\varrho(\mathcal{K},\bar{\mathcal{K}}) \\
&\leq \lambda \max\{\varrho(x,\bar{x}), \varrho(y,\bar{y})\} + \mu \max\{\varrho(x,\bar{x}), \varrho(y,\bar{y})\} \\
&\quad + (1-(\lambda+\mu))\varrho(\mathcal{K},\bar{\mathcal{K}}) \\
&= \lambda\varrho^*((x,y),(\bar{x},\bar{y})) + \mu\varrho^*((x,y),(\bar{x},\bar{y})) \\
&\quad + (1-(\lambda+\mu))\varrho(\mathcal{K},\bar{\mathcal{K}}) \\
&= (\lambda+\mu)\varrho^*((x,y),(\bar{x},\bar{y})) + (1-(\lambda+\mu))\varrho(\mathcal{K},\bar{\mathcal{K}}).
\end{aligned}$$

From the previous relation, for $(y, x) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{y}, \bar{x}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$, we have

$$\begin{aligned}
\varrho(\Gamma(y,x), \bar{\Gamma}(\bar{y},\bar{x})) &\leq (\lambda+\mu)\varrho^*((y,x),(\bar{y},\bar{x})) + (1-(\lambda+\mu))\varrho(\mathcal{K},\bar{\mathcal{K}}) \\
&= (\lambda+\mu)\varrho^*((x,y),(\bar{x},\bar{y})) + (1-(\lambda+\mu))\varrho(\mathcal{K},\bar{\mathcal{K}}).
\end{aligned}$$

Based on the symbols introduced in Remark 1, we establish the following relationship:

$$\begin{aligned}
&\varrho^*((\Gamma,\Gamma')(x,y), (\bar{\Gamma},\bar{\Gamma}')(\bar{x},\bar{y})) \\
&\qquad = \max\{\varrho(\Gamma(x,y), \bar{\Gamma}(\bar{x},\bar{y})), \varrho(\Gamma(y,x), \bar{\Gamma}(\bar{y},\bar{x}))\}.
\end{aligned}$$

From (2) we have $\varrho^*(\mathcal{K} \times \mathcal{K}, \bar{\mathcal{K}} \times \bar{\mathcal{K}}) = \varrho(\mathcal{K}, \bar{\mathcal{K}})$, combining these relations yield the inequality:

$$\begin{aligned}
\varrho^*(T(x,y), T(\bar{x},\bar{y})) &= \varrho^*((\Gamma,\Gamma')(x,y), (\bar{\Gamma},\bar{\Gamma}')(\bar{x},\bar{y})) \\
&\leq (\lambda+\mu)\varrho^*((x,y),(\bar{x},\bar{y})) \\
&\quad + (1-(\lambda+\mu))\varrho^*(\mathcal{K}\times\mathcal{K}, \bar{\mathcal{K}}\times\bar{\mathcal{K}}).
\end{aligned}$$

This demonstrates that T is a noncyclic contraction in the complete metric space $(\mathcal{Z}^2, \varrho^*)$. □

Lemma 3. Suppose that $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty subsets of a metric space $(\mathcal{Z}, \varrho)$ such that $(\mathcal{K}, \bar{\mathcal{K}})$ and $(\bar{\mathcal{K}}, \mathcal{K})$ have the UC property and $\varrho := \varrho(\mathcal{K}, \bar{\mathcal{K}}) > 0$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \bar{\mathcal{K}}$ be mappings such that the ordered pair $(\Gamma, \bar{\Gamma})$ constitutes a noncyclic contraction pair. If Γ possesses a coupled fixed point $(x^*, x^{**}) \in \mathcal{K} \times \mathcal{K}$ and $\bar{\Gamma}$ possesses a coupled fixed point $(\bar{x}^*, \bar{x}^{**}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$, then it follows that $x^* = x^{**}$ and $\bar{x}^* = \bar{x}^{**}$ and $\varrho(x^*, \bar{x}^*) = \varrho(\mathcal{K}, \bar{\mathcal{K}})$.

Proof. Since Γ possesses a coupled fixed point $(x^*, x^{**}) \in \mathcal{K} \times \mathcal{K}$ and $\bar{\Gamma}$ possesses a coupled fixed point $(\bar{x}^*, \bar{x}^{**}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$, then

$$\varrho(x^*, \bar{x}^*) + \varrho(x^{**}, \bar{x}^{**}) = \varrho(\Gamma(x^*, x^{**}), \bar{\Gamma}(\bar{x}^*, \bar{x}^{**})) + \varrho(\Gamma(x^{**}, x^*), \bar{\Gamma}(\bar{x}^{**}, \bar{x}^*))$$

and

$$\varrho(x^*, \bar{x}^{**}) + \varrho(x^{**}, \bar{x}^*) = \varrho(\Gamma(x^*, x^{**}), \bar{\Gamma}(\bar{x}^{**}, \bar{x}^*)) + \varrho(\Gamma(x^{**}, x^*), \bar{\Gamma}(\bar{x}^*, \bar{x}^{**})).$$

Given that $(\Gamma, \bar{\Gamma})$ is a noncyclic contraction pair, we obtain the following inequalities from the above relations:

$$\begin{aligned}
&\varrho(x^*, \bar{x}^*) + \varrho(x^{**}, \bar{x}^{**}) \\
&\qquad \leq \lambda\varrho(x^*, \bar{x}^*) + \mu\varrho(x^{**}, \bar{x}^{**}) + (1 - (\lambda + \mu))\varrho \\
&\qquad\quad + \lambda\varrho(x^{**}, \bar{x}^{**}) + \mu\varrho(x^*, \bar{x}^*) + (1 - (\lambda + \mu))\varrho \\
&\qquad = (\lambda + \mu)(\varrho(x^*, \bar{x}^*) + \varrho(x^{**}, \bar{x}^{**})) + 2(1 - (\lambda + \mu))\varrho
\end{aligned}$$

and

$$\begin{aligned}
&\varrho(x^*, \bar{x}^{**}) + \varrho(x^{**}, \bar{x}^*) \\
&\qquad \leq \lambda\varrho(x^*, \bar{x}^{**}) + \mu\varrho(x^{**}, \bar{x}^*) + (1 - (\lambda + \mu))\varrho \\
&\qquad\quad + \lambda\varrho(x^{**}, \bar{x}^*) + \mu\varrho(x^*, \bar{x}^{**}) + (1 - (\lambda + \mu))\varrho \\
&\qquad = (\lambda + \mu)(\varrho(x^*, \bar{x}^{**}) + \varrho(x^{**}, \bar{x}^*)) + 2(1 - (\lambda + \mu))\varrho.
\end{aligned}$$

These imply that

$$(1 - (\lambda + \mu))(\varrho(x^*, \bar{x}^*) + \varrho(x^{**}, \bar{x}^{**})) \leq 2(1 - (\lambda + \mu))\varrho$$

and

$$(1 - (\lambda + \mu))(\varrho(x^*, \bar{x}^{**}) + \varrho(x^{**}, \bar{x}^*)) \leq 2(1 - (\lambda + \mu))\varrho.$$

Hence

$$\varrho(x^*, \bar{x}^*) + \varrho(x^{**}, \bar{x}^{**}) \leq 2\varrho = 2\varrho(\mathcal{K}, \bar{\mathcal{K}})$$

and

$$\varrho(x^*, \bar{x}^{**}) + \varrho(x^{**}, \bar{x}^*) \leq 2\varrho = 2\varrho(\mathcal{K}, \bar{\mathcal{K}}).$$

Since $\varrho(x^*, \bar{x}^*) \geq \varrho(\mathcal{K}, \bar{\mathcal{K}})$, $\varrho(x^{**}, \bar{x}^{**}) \geq \varrho(\mathcal{K}, \bar{\mathcal{K}})$, $\varrho(x^*, \bar{x}^{**}) \geq \varrho(\mathcal{K}, \bar{\mathcal{K}})$, and $\varrho(x^{**}, \bar{x}^*) \geq \varrho(\mathcal{K}, \bar{\mathcal{K}})$, then

$$\varrho(x^*, \bar{x}^*) = \varrho(\mathcal{K}, \bar{\mathcal{K}}) \tag{4}$$

and

$$\varrho(x^*, \bar{x}^{**}) = \varrho(x^{**}, \bar{x}^*) = \varrho(\mathcal{K}, \bar{\mathcal{K}}). \tag{5}$$

Since $(\mathcal{K}, \bar{\mathcal{K}})$ and $(\bar{\mathcal{K}}, \mathcal{K})$ satisfy the UC property, then from (4) and (5), we obtain $\bar{x}^* = \bar{x}^{**}$ and $x^* = x^{**}$, respectively. □

We now have the necessary tools to present our main result.

Theorem 2. Suppose that $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty, closed, and convex subsets of a uniformly convex Banach space $(\mathcal{Z}, \|\cdot\|)$ with $\varrho := \varrho(\mathcal{K}, \bar{\mathcal{K}}) > 0$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$, $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \bar{\mathcal{K}}$ such that the ordered pair $(\Gamma, \bar{\Gamma})$ is a noncyclic contraction pair. Then

(i) Γ has a unique coupled fixed point $(x^*, x^*) \in \mathcal{K} \times \mathcal{K}$ and $\bar{\Gamma}$ has a unique coupled fixed point $(\bar{x}^*, \bar{x}^*) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$. Moreover, $\varrho(x^*, \bar{x}^*) = \varrho(\mathcal{K}, \bar{\mathcal{K}})$;

(ii) for every $(x_0, y_0) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}_0, \bar{y}_0) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$, the sequences $\{(x_n, y_n)\} = \{(\Gamma, \Gamma')^n(x_0, y_0)\}$ and $\{(\bar{x}_n, \bar{y}_n)\} = \{(\bar{\Gamma}, \bar{\Gamma}')^n(\bar{x}_0, \bar{y}_0)\}$ converge to (x^*, x^*) and $(\bar{x}^*, \bar{x}^*)$, respectively.

Furthermore, if $\delta_{\mathcal{Z}}(\epsilon) \geq C\epsilon^p$ for some $C > 0$, $p > 0$ and every $\epsilon \in (0, 2]$, then for every $n \geq 0$ we have

(iii) the following priori error estimate holds

$$\max\{\|x^* - x_n\|, \|x^* - y_n\|\} \leq \frac{\Delta_0}{1 - \sqrt[p]{\lambda + \mu}} \sqrt[p]{\frac{\Delta_0 - \varrho}{C\varrho}} (\sqrt[p]{\lambda + \mu})^n;$$

(iv) the following posteriori error estimate holds

$$\max\{\|x_n - x^*\|, \|y_n - x^*\|\} \leq \frac{\Delta_n}{1 - \sqrt[p]{\lambda + \mu}} \sqrt[p]{\frac{\Delta_n - \varrho}{C\varrho}},$$

where

$$\Delta_n = \max\left\{\|(x_n, y_n) - (\bar{x}_n, \bar{y}_n)\|^*, \|(\Gamma, \Gamma')(x_n, y_n) - (\bar{x}_n, \bar{y}_n)\|^*\right\},$$

or, equality

$$\Delta_n = \max\left\{\|x_n - \bar{x}_n\|, \|y_n - \bar{y}_n\|, \|x_{n+1} - \bar{x}_n\|, \|y_{n+1} - \bar{y}_n\|\right\}.$$

Proof. Note that $\mathcal{K} \times \mathcal{K}$ and $\bar{\mathcal{K}} \times \bar{\mathcal{K}}$ are nonempty, closed, and convex subsets of the Banach space $(\mathcal{Z}^2, \|\cdot\|^*)$ and that $T : (\mathcal{K} \times \mathcal{K}) \cup (\bar{\mathcal{K}} \times \bar{\mathcal{K}}) \to (\mathcal{K} \times \mathcal{K}) \cup (\bar{\mathcal{K}} \times \bar{\mathcal{K}})$ defined by

$$T(x, y) = \begin{cases} (\Gamma, \Gamma')(x, y) & \text{if } (x, y) \in \mathcal{K} \times \mathcal{K}, \\ (\bar{\Gamma}, \bar{\Gamma}')(x, y) & \text{if } (x, y) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}, \end{cases}$$

from Lemma 2, is a noncyclic contraction with constant $\lambda + \mu$.

(i) From Theorem 1(i), there exists a unique point $(x^*, x^{**}) \in \mathcal{K} \times \mathcal{K}$, which is a fixed point for the map T that is $(\Gamma, \Gamma')(x^*, x^{**}) = (x^*, x^{**})$. This implies that

$$\Gamma(x^*, x^{**}) = x^* \quad \text{and} \quad \Gamma(x^{**}, x^*) = x^{**}.$$

Conversely, since every coupled fixed point of Γ is a fixed point of T, it follows that (x^*, x^{**}) is the unique coupled fixed point of Γ in $\mathcal{K} \times \mathcal{K}$. Similarly, we can show that $\bar{\Gamma}$ has a unique coupled fixed point $(\bar{x}^*, \bar{x}^{**}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$. Lemma 3 requires that $\varrho(x^*, \bar{x}^*) = \varrho(\mathcal{K}, \bar{\mathcal{K}})$, $x^* = x^{**}$ and $\bar{x}^* = \bar{x}^{**}$.

(ii) From Theorem 1(ii), for every $(x_0, y_0) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}_0, \bar{y}_0) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$ the sequences $\{(x_n, y_n)\} = \{T^n(x_0, y_0)\} = \{(\Gamma, \Gamma')^n(x_0, y_0)\}$ and $\{(\bar{x}_n, \bar{y}_n)\} = \{T^n(\bar{x}_0, \bar{y}_0)\} = \{(\bar{\Gamma}, \bar{\Gamma}')^n(\bar{x}_0, \bar{y}_0)\}$, converge to (x^*, x^*) and $(\bar{x}^*, \bar{x}^*)$, respectively.

From Theorem 1, it follows that for every $n \geq 0$, we have

$$\begin{aligned} &\Delta_{(x_n, y_n), (\bar{x}_n, \bar{y}_n)} \\ &= \max\left\{\|(x_n, y_n) - (\bar{x}_n, \bar{y}_n)\|^*, \|T(x_n, y_n) - (\bar{x}_n, \bar{y}_n)\|^*\right\} \\ &= \max\left\{\|(x_n, y_n) - (\bar{x}_n, \bar{y}_n)\|^*, \|(\Gamma, \Gamma')(x_n, y_n) - (\bar{x}_n, \bar{y}_n)\|^*\right\} \\ &= \max\left\{\|x_n - \bar{x}_n\|, \|y_n - \bar{y}_n\|, \|\Gamma(x_n, y_n) - \bar{x}_n\|, \|\Gamma(y_n, x_n) - \bar{y}_n\|\right\} \\ &= \max\left\{\|x_n - \bar{x}_n\|, \|y_n - \bar{y}_n\|, \|x_{n+1} - \bar{x}_n\|, \|y_{n+1} - \bar{y}_n\|\right\} = \Delta_n. \end{aligned}$$

In particular

$$\Delta_{(x_0, y_0), (\bar{x}_0, \bar{y}_0)} = \max\left\{\|x_0 - \bar{x}_0\|, \|y_0 - \bar{y}_0\|, \|x_1 - \bar{x}_0\|, \|y_1 - \bar{y}_0\|\right\} = \Delta_0.$$

On the other hand,

$$\|(x^*, x^*) - T^n(x_0, y_0)\|^* = \max\{\|x^* - x_n\|, \|x^* - y_n\|\}.$$

Moreover, T is a noncyclic contraction with constant $\lambda + \mu$. These observations, together with Theorem 1(iii)-(iv), complete the proof of parts (iii) and (iv). □

We illustrate Theorem 2 with an example.

Example 1. Consider the real number space $\mathbb{R}$, equipped with the standard absolute value norm, $|\cdot|$. Then, we have $\delta_{\mathbb{R}}(\epsilon) = \frac{\epsilon}{2}$ (see [4]). So $C = \frac{1}{2}$ and $p = 1$.

Define the sets $\mathcal{K}$ to be the closed interval $[1, 2]$ and $\bar{\mathcal{K}}$ as the closed interval $[-2, -1]$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \mathcal{K}$ and $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \bar{\mathcal{K}}$ be defined as follows:

$$\Gamma(x, y) := \frac{x + 2y + 5}{8} \quad \text{and} \quad \bar{\Gamma}(\bar{x}, \bar{y}) := \frac{\bar{x} + 2\bar{y} - 5}{8}.$$

It is easy to observe that $x = x^*, y = x^{**}, \bar{x} = \bar{x}^*, \bar{y} = \bar{x}^{**}$ in $\mathcal{K} \cup \bar{\mathcal{K}}$ constitute a solution of the system

$$\begin{cases} 7x - 2y = 5, \\ 7y - 2x = 5, \\ 7\bar{x} - 2\bar{y} = -5, \\ 7\bar{y} - 2\bar{x} = -5, \end{cases}$$

if and only if

$$\Gamma(x^*, x^{**}) = x^*,\ \Gamma(x^{**}, x^*) = x^{**},\ \bar{\Gamma}(\bar{x}^*, \bar{x}^{**}) = \bar{x}^* \quad \text{and} \quad \bar{\Gamma}(\bar{x}^{**}, \bar{x}^*) = \bar{x}^{**}.$$

Note that

$$\begin{aligned} \varrho(\Gamma(x, y), \bar{\Gamma}(\bar{x}, \bar{y})) &= |\frac{x + 2y + 5}{8} - \frac{\bar{x} + 2\bar{y} - 5}{8}| \\ &= |\frac{1}{8}(x - \bar{x}) + \frac{2}{8}(y - \bar{y}) + \frac{5}{8} \times 2| \\ &\leq \frac{1}{8}|x - \bar{x}| + \frac{2}{8}|y - \bar{y}| + (1 - \frac{3}{8}) \times \varrho(\mathcal{K}, \bar{\mathcal{K}}). \end{aligned}$$

Hence $(\Gamma, \bar{\Gamma})$ is a noncyclic contraction pair. From Theorem 2(i), Γ has a unique coupled fixed point $(x^*, x^*) = (1, 1) \in \mathcal{K} \times \mathcal{K}$ and $\bar{\Gamma}$ has a unique coupled fixed point $(\bar{x}^*, \bar{x}^*) = (-1, -1) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$. For all $(x_0, y_0) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}_0, \bar{y}_0) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$ the sequences $\{(x_n, y_n)\} = \{(\Gamma, \Gamma')^n(x_0, y_0)\}$ and $\{(\bar{x}_n, \bar{y}_n)\} = \{(\bar{\Gamma}, \bar{\Gamma}')^n(\bar{x}_0, \bar{y}_0)\}$ converge to $(1, 1)$ and $(-1, -1)$, respectively. See Table 1.

Table 1: Convergence of the sequences in Example 1

	x_n	y_n	$\bar{x}_n$	$\bar{y}_n$
$n=0$	1.35	1.45	-1.8	-1.35
$n=1$	1.15625	1.14375	-1.1875	-1.24375
$n=2$	1.05546875	1.05703125	-1.084375	-1.07734375
$n=3$	1.0211914063	1.0209960938	-1.0298828125	-1.0307617188
$n=4$	1.0078979492	1.0079223633	-1.0114257813	-1.0113159180
$n=5$	1.0029678345	1.0029647827	-1.0042572021	-1.0042709351
$n=6$	1.0011121750	1.0011125565	-1.0015998840	-1.0015981674
$n=7$	1.0004171610	1.0004171133	-1.0005995274	-1.0005997419
$n=8$	1.0001564234	1.0001564294	-1.0002248764	-1.0002248496
$n=9$	1.0000586603	1.0000586595	-1.0000843219	-1.0000843253
$n=10$	1.0000219974	1.0000219975	-1.0000316216	-1.0000316211

A priori error estimate holds

$$\max\{\|x^*-x_n\|,\|x^*-y_n\|\} \le \frac{\Delta_0}{1-\sqrt[p]{\lambda+\mu}}\sqrt[p]{\frac{\Delta_0-\varrho}{C\varrho}}(\sqrt[p]{\lambda+\mu})^n$$
$$= 5.796\times(\frac{3}{8})^n.$$

Therefore after ten iterations, the derived error bound guarantees that the priori error estimate is less than 0.0004.

A posteriori error estimate holds

$$\max\{\|x_n-x^*\|,\|y_n-x^*\|\} \le \frac{\Delta_n}{1-\sqrt[p]{\lambda+\mu}}\sqrt[p]{\frac{\Delta_n-\varrho}{C\varrho}} = \frac{8}{5}\Delta_n(\Delta_n-2).$$

Therefore after ten iterations, the derived error bound guarantees that the posteriori error estimate is less than 0.0002.

As a consequence of Theorem 2, we obtain its cyclic version. The statements of this theorem are similar to [5, Theorem 3], but with new error bounds. The subsequent lemma will help us in the proof.

Lemma 4. Suppose that $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty subsets of the metric space $(\mathcal{Z},\varrho)$. Let $\Gamma : \mathcal{K}\times\mathcal{K}\to\bar{\mathcal{K}}$, $\bar{\Gamma}:\bar{\mathcal{K}}\times\bar{\mathcal{K}}\to\mathcal{K}$ such that the ordered pair $(\Gamma,\bar{\Gamma})$ is a cyclic contraction pair with nonnegative constants λ and μ in which $\lambda+\mu<1$. Then

(i) $(\bar{\Gamma}\circ(\Gamma,\Gamma'))' = \bar{\Gamma}'\circ(\Gamma,\Gamma')$ and $(\Gamma\circ(\bar{\Gamma},\bar{\Gamma}'))' = \Gamma'\circ(\bar{\Gamma},\bar{\Gamma}')$;

(ii) $(\bar{\Gamma} \circ (\Gamma, \Gamma'), \Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))$ is a noncyclic contraction pair with constants $\lambda^2 + \mu^2$ and $2\lambda\mu$.

Proof. From the definitions of Γ and $\bar{\Gamma}$, the composition $\bar{\Gamma} \circ (\Gamma, \Gamma')$ defines a mapping from $\mathcal{K} \times \mathcal{K}$ to $\mathcal{K}$, while $\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}')$ defines a mapping from $\bar{\mathcal{K}} \times \bar{\mathcal{K}}$ to $\bar{\mathcal{K}}$.

(i) For all $(x, y) \in \mathcal{K} \times \mathcal{K}$, we have

$$\begin{aligned}(\bar{\Gamma} \circ (\Gamma, \Gamma'))'(x, y) &= (\bar{\Gamma} \circ (\Gamma, \Gamma'))(y, x) \\ &= \bar{\Gamma}(\Gamma(y, x), \Gamma(x, y)) \\ &= \bar{\Gamma}'(\Gamma(x, y), \Gamma'(x, y)) \\ &= (\bar{\Gamma}' \circ (\Gamma, \Gamma'))(x, y).\end{aligned}$$

So $(\bar{\Gamma} \circ (\Gamma, \Gamma'))' = \bar{\Gamma}' \circ (\Gamma, \Gamma')$. Similarly we can show that $(\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))' = \Gamma' \circ (\bar{\Gamma}, \bar{\Gamma}')$.

(ii) The ordered pair $(\Gamma, \bar{\Gamma})$ is a cyclic contraction pair with constants λ and μ, so for all $(x, y) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}, \bar{y}) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$, we have

$$\begin{aligned}&\varrho((\bar{\Gamma} \circ (\Gamma, \Gamma'))(x, y), (\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))(\bar{x}, \bar{y})) \\ &\quad = \varrho((\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))(\bar{x}, \bar{y}), (\bar{\Gamma} \circ (\Gamma, \Gamma'))(x, y)) \\ &\quad \leq \lambda\varrho(\bar{\Gamma}(\bar{x}, \bar{y}), \Gamma(x, y)) + \mu\varrho(\bar{\Gamma}(\bar{y}, \bar{x}), \Gamma(y, x)) + (1 - (\lambda + \mu))\varrho(\mathcal{K}, \bar{\mathcal{K}}) \\ &\quad = \lambda\varrho(\Gamma(x, y), \bar{\Gamma}(\bar{x}, \bar{y})) + \mu\varrho(\Gamma(y, x), \bar{\Gamma}(\bar{y}, \bar{x})) + (1 - (\lambda + \mu))\varrho(\mathcal{K}, \bar{\mathcal{K}}) \\ &\quad \leq \lambda^2\varrho(x, \bar{x}) + \lambda\mu\varrho(y, \bar{y}) + \lambda(1 - (\lambda + \mu))\varrho(\mathcal{K}, \bar{\mathcal{K}}) + \mu\lambda\varrho(y, \bar{y}) \\ &\qquad + \mu^2\varrho(x, \bar{x}) + \mu(1 - (\lambda + \mu))\varrho(\mathcal{K}, \bar{\mathcal{K}}) + (1 - (\lambda + \mu))\varrho(\mathcal{K}, \bar{\mathcal{K}}) \\ &\quad = (\lambda^2 + \mu^2)\varrho(x, \bar{x}) + 2\lambda\mu\varrho(y, \bar{y}) + (1 - (\lambda + \mu)^2)\varrho(\mathcal{K}, \bar{\mathcal{K}}).\end{aligned}$$

Hence $(\bar{\Gamma} \circ (\Gamma, \Gamma'), \Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))$ is a noncyclic contraction pair with constants $\lambda^2 + \mu^2$ and $2\lambda\mu$. □

Corollary 1. Suppose that $\mathcal{K}$ and $\bar{\mathcal{K}}$ are nonempty, closed, and convex subsets of a uniformly convex Banach space $(\mathcal{Z}, \|\cdot\|)$ such that $\varrho := \varrho(\mathcal{K}, \bar{\mathcal{K}}) > 0$. Let $\Gamma : \mathcal{K} \times \mathcal{K} \to \bar{\mathcal{K}}$, $\bar{\Gamma} : \bar{\mathcal{K}} \times \bar{\mathcal{K}} \to \mathcal{K}$ such that the ordered pair $(\Gamma, \bar{\Gamma})$ is a cyclic contraction pair. Then

(i) Γ has a unique coupled best proximity point $(x^*, x^*) \in \mathcal{K} \times \mathcal{K}$ and $\bar{\Gamma}$ has a unique coupled best proximity point $(\bar{x}^*, \bar{x}^*) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$, that is,

$$\|x^* - \Gamma(x^*, x^*)\| = \varrho \quad \text{and} \quad \|\bar{x}^* - \bar{\Gamma}(\bar{x}^*, \bar{x}^*)\| = \varrho.$$

Moreover,

$$\Gamma(x^*, x^*) = \bar{x}^*, \quad \bar{\Gamma}(\bar{x}^*, \bar{x}^*) = x^*, \quad \|x^* - \bar{x}^*\| = \varrho;$$

(ii) For any arbitrary point (x_0, y_0), let the sequences $\{x_n\}$ and $\{y_n\}$ defined as in Definition 3. Then

$$\lim_{n\to\infty} x_{2n} = \lim_{n\to\infty} y_{2n} = x^*, \quad \lim_{n\to\infty} x_{2n+1} = \lim_{n\to\infty} y_{2n+1} = \bar{x}^*.$$

Furthermore, if $\delta_{\mathcal{Z}}(\epsilon) \geq C\epsilon^p$ for some $C > 0$, $p > 0$ and every $\epsilon \in (0, 2]$, then

(iii) the following priori error estimate holds

$$\max\{\|x^* - x_{2m}\|, \|x^* - y_{2m}\|\} \leq \Delta_0 \sqrt[p]{\frac{\Delta_0 - \varrho}{C\varrho}} \frac{(\sqrt[p]{\lambda+\mu})^{2m}}{1 - \sqrt[p]{(\lambda+\mu)^2}};$$

(iv) the following posteriori error estimate holds

$$\max\{\|x^* - x_{2n}\|, \|x^* - y_{2n}\|\} \leq \frac{\Delta_n}{1 - \sqrt[p]{(\lambda+\mu)^2}} \sqrt[p]{\frac{\Delta_n - \varrho}{C\varrho}},$$

where

$$\Delta_0 = \max\left\{\|(x_0, y_0) - (x_1, y_1)\|^*, \|(x_2, y_2) - (x_1, y_1)\|^*\right\}$$

and

$$\Delta_n = \max\left\{\|(x_{2n}, y_{2n}) - (x_{2n+1}, y_{2n+1})\|^*, \|(x_{2n+2}, y_{2n+2}) - (x_{2n+1}, y_{2n+1})\|^*\right\}.$$

Proof. (i) From Lemma 4, $(\bar{\Gamma} \circ (\Gamma, \Gamma'), \Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))$ is a noncyclic contraction pair. So, from Theorem 2(i), $\bar{\Gamma} \circ (\Gamma, \Gamma')$ has a unique coupled fixed point $(x^*, x^*) \in \mathcal{K} \times \mathcal{K}$ and $\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}')$ has a unique coupled fixed point $(\bar{x}^*, \bar{x}^*) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$ such that $\|x^* - \bar{x}^*\| = \varrho(\mathcal{K}, \bar{\mathcal{K}})$. On the other hand, $(\Gamma, \bar{\Gamma})$ is a cyclic contraction pair, so we have

$$\begin{aligned}\|\Gamma(x^*, x^*) - x^*\| &= \|\Gamma(x^*, x^*) - (\bar{\Gamma} \circ (\Gamma, \Gamma'))(x^*, x^*)\| \\ &\leq (\lambda + \mu)\|x^* - \Gamma(x^*, x^*)\| + (1 - (\lambda + \mu))\varrho.\end{aligned}$$

This implies $\|\Gamma(x^*, x^*) - x^*\| = \varrho(\mathcal{K}, \bar{\mathcal{K}})$. Since $\|x^* - \bar{x}^*\| = \varrho(\mathcal{K}, \bar{\mathcal{K}})$, then from Property UC of $(\bar{\mathcal{K}}, \mathcal{K})$, we get $\Gamma(x^*, x^*) = \bar{x}^*$. The proof of the other cases is similar.

(ii) From Theorem 2(ii), for every $(x_0, y_0) \in \mathcal{K} \times \mathcal{K}$ and $(\bar{x}_0, \bar{y}_0) := (\Gamma, \Gamma')(x_0, y_0) = (x_1, y_1) \in \bar{\mathcal{K}} \times \bar{\mathcal{K}}$ the sequences

$$\{(\bar{\Gamma} \circ (\Gamma, \Gamma'), \bar{\Gamma}' \circ (\Gamma, \Gamma'))^n (x_0, y_0)\}$$

and

$$\{(\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'), \Gamma' \circ (\bar{\Gamma}, \bar{\Gamma}'))^n (x_1, y_1)\}$$

converge to (x^*, x^*) and $(\bar{x}^*, \bar{x}^*)$, respectively. Indeed from Remark 1, we know

$$\begin{aligned}\{(\bar{\Gamma} \circ (\Gamma, \Gamma'), \bar{\Gamma}' \circ (\Gamma, \Gamma'))^n (x_0, y_0)\} &= \{((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))^n (x_0, y_0)\} \\ &= \{(x_{2n}, y_{2n})\}\end{aligned}$$

and

$$\begin{aligned}\{(\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'), \Gamma' \circ (\bar{\Gamma}, \bar{\Gamma}'))^n (x_1, y_1)\} &= \{((\Gamma, \Gamma') \circ (\bar{\Gamma}, \bar{\Gamma}'))^n (x_1, y_1)\} \\ &= \{(x_{2n+1}, y_{2n+1})\},\end{aligned}$$

which completes the proof of (ii).

(iii) In Theorem 2(iii), let $(\bar{x}_0, \bar{y}_0) := (\Gamma, \Gamma')(x_0, y_0) = (x_1, y_1)$. Then

$$\begin{aligned}\Delta_0 &= \max\big\{\|(x_0, y_0) - (x_1, y_1)\|^*, \\ &\qquad \|(\bar{\Gamma} \circ (\Gamma, \Gamma'), \bar{\Gamma}' \circ (\Gamma, \Gamma'))(x_0, y_0) - (x_1, y_1)\|^*\big\} \\ &= \max\big\{\|(x_0, y_0) - (x_1, y_1)\|^*, \|((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))(x_0, y_0) - (x_1, y_1)\|^*\big\} \\ &= \max\big\{\|(x_0, y_0) - (x_1, y_1)\|^*, \|(x_2, y_2) - (x_1, y_1)\|^*\big\}.\end{aligned}$$

Since from Lemma 4, $(\bar{\Gamma} \circ (\Gamma, \Gamma'), \Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'))$ is a noncyclic contraction pair with constants $\lambda^2 + \mu^2$ and $2\lambda\mu$ and $\lambda^2 + \mu^2 + 2\lambda\mu = (\lambda + \mu)^2$, then the following priori error estimate holds

$$\begin{aligned}&\max\{\|x^* - x_{2m}\|, \|x^* - y_{2m}\|\} \\ &\quad = \|(x^*, x^*) - ((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))^m (x_0, y_0)\|^* \\ &\quad = \|(x^*, x^*) - (\bar{\Gamma} \circ (\Gamma, \Gamma'), \bar{\Gamma}' \circ (\Gamma, \Gamma'))^m (x_0, y_0)\|^* \\ &\quad \leq \frac{\Delta_0}{1 - \sqrt[p]{(\lambda + \mu)^2}} \sqrt[p]{\frac{\Delta_0 - \varrho}{C\varrho}} (\sqrt[p]{\lambda + \mu})^{2m}.\end{aligned}$$

(iv) In Theorem 2(iv), let $(\bar{x}_0, \bar{y}_0) := (\Gamma, \Gamma')(x_0, y_0) = (x_1, y_1)$. Then

$$\begin{aligned}\Delta_n &= \Delta_{(\bar{\Gamma} \circ (\Gamma, \Gamma'), \bar{\Gamma}' \circ (\Gamma, \Gamma'))^n (x_0, y_0), (\Gamma \circ (\bar{\Gamma}, \bar{\Gamma}'), \Gamma' \circ (\bar{\Gamma}, \bar{\Gamma}'))^n (x_1, y_1)} \\ &= \Delta_{((\bar{\Gamma}, \bar{\Gamma}') \circ (\Gamma, \Gamma'))^n (x_0, y_0), ((\Gamma, \Gamma') \circ (\bar{\Gamma}, \bar{\Gamma}'))^n (x_1, y_1)} \\ &= \Delta_{(x_{2n}, y_{2n}), (x_{2n+1}, y_{2n+1})} \\ &= \max\big\{\|(x_{2n}, y_{2n}) - (x_{2n+1}, y_{2n+1})\|^*,\end{aligned}$$

$$\begin{aligned}
&\|(\bar{\Gamma}\circ(\Gamma,\Gamma'),\bar{\Gamma}'\circ(\Gamma,\Gamma'))(x_{2n},y_{2n})-(x_{2n+1},y_{2n+1})\|^*\big\}\\
&=\max\big\{\|(x_{2n},y_{2n})-(x_{2n+1},y_{2n+1})\|^*,\\
&\qquad\|((\bar{\Gamma},\bar{\Gamma}')\circ(\Gamma,\Gamma'))(x_{2n},y_{2n})-(x_{2n+1},y_{2n+1})\|^*\big\}\\
&=\max\big\{\|(x_{2n},y_{2n})-(x_{2n+1},y_{2n+1})\|^*,\\
&\qquad\|(x_{2n+2},y_{2n+2})-(x_{2n+1},y_{2n+1})\|^*\big\}.
\end{aligned}$$

Since from Lemma 4, $(\bar{\Gamma}\circ(\Gamma,\Gamma'),\Gamma\circ(\bar{\Gamma},\bar{\Gamma}'))$ is a noncyclic contraction pair with constants $\lambda^2+\mu^2$ and $2\lambda\mu$ and $\lambda^2+\mu^2+2\lambda\mu=(\lambda+\mu)^2$, then the following posteriori error estimate holds

$$\begin{aligned}
\max\{\|x^*-x_{2n}\|,\|x^*-y_{2n}\|\}&=\|((\bar{\Gamma},\bar{\Gamma}')\circ(\Gamma,\Gamma'))^n(x_0,y_0)-(x^*,x^*)\|^*\\
&\leq\frac{\Delta_n}{1-\sqrt[p]{(\lambda+\mu)^2}}\cdot\sqrt[p]{\frac{\Delta_n-\varrho}{C\varrho}}.
\end{aligned}$$

□

Acknowledgements

The authors are grateful to the editor and reviewers for their comments and suggestions, which improved the quality of this manuscript.

Funding

This research received no external funding.

Conflicts of Interest

All authors jointly worked on the results and they read and approved the final manuscript, and also declare that they have no competing interests.

References

[1] Eldred, A. and Veeramani, P. *Existence and convergence of best proximity points*, J. Math. Anal. Appl. 323(2) (2006), 1001–1006.

[2] Goebel, K. and Kirk, W.A. *Topics in metric fixed point theory*, Cambridge University Press, 1990.

[3] Guo, D. and Lakshmikantham, V. *Coupled fixed points of nonlinear operators with applications*, Nonlinear Anal. 11(5) (1987), 623–632.

[4] Ilchev, A. *On an application of coupled best proximity points theorems for solving systems of linear equations*, AIP Conf. Proc. 2048 (2018), 050003.

[5] Ilchev, A. and Zlatanov, B. *Error estimates for approximation of coupled best proximity points for cyclic contractive maps*, Appl. Math. Comput. 290 (2016), 412–425.

[6] Kirk, W., Srinivasan, P. and Veeramani, P. *Fixed points for mappings satisfying cyclical contractive conditions*, Fixed Point Theory, 4(1) (2003), 79–89.

[7] Petric, M. and Zlatanov, B. *Fixed point theorems of Kannan type for cyclical contractive condition*, in: Proceedings of the Conferences Organized by Faculty of Mathematics and Informatics, REMIA, Plovdiv University, Plovdiv, Bulgaria, 2010, 187–194.

[8] Puthusseria, A. and Ramesh Kumar, D. *Some coupled fixed point theorems for* (Ψ, Φ)*-contraction with applications to fractals*, Filomat, 38(26) (2024), 9305–9320.

[9] Safari-Hafshejani, A. *The existence of best proximity points for generalized cyclic quasi-contractions in metric spaces with the UC and ultrametric properties*, Fixed Point Theory, 23(2) (2022), 507–518.

[10] Safari-Hafshejani, A. *Error estimates for approximating fixed points and best proximity points for noncyclic and cyclic contraction mappings*, Iran. J. Numer. Anal. Optim. 13(3) (2023), 385–396.

[11] Sintunavarat, W. and Kumam, P. *Coupled best proximity point theorem in metric spaces*, Fixed Point Theory Appl. 2012(1) (2012), Article 93.

[12] Suzuki, T., Kikkawa, M. and Vetro, C. *The existence of best proximity points in metric spaces with the property UC*, Nonlinear Anal. 71 (2009), 2918–2926.

[13] Zlatanov, B. *Error estimates for approximating best proximity points for cyclic contractive maps*, Carpathian J. Math. 32(2) (2016), 265–270.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1484–1517
https://doi.org/10.22067/ijnao.2026.97856.1846
https://ijnao.um.ac.ir/



Research Article

A multi-layer SEIRS-V network epidemic model for IoT malware propagation with heterogeneous device types and patch dynamics

A. Tajari Siahmarzkooh*

Abstract

The Internet of Things (IoT) ecosystem comprises billions of heterogeneous devices that have diverse hardware architectures, operating systems, vulnerability profiles, and patch management capabilities, creating a complex and dynamic attack surface for malware propagation. This paper develops a comprehensive multi-layer network epidemic model for characterizing malware spread in heterogeneous IoT networks. The proposed framework integrates three key dimensions: (1) a susceptible-exposed-infected-recovered-susceptible with vaccination compartmental structure capturing device-level infection states and patching status; (2) a multi-layer network topology distinguishing between physical proximity, logical connectivity, and protocol-based communication graphs, and (3) device-type heterogeneity with class-specific parameters for vulnerability, infectiousness, and recovery rates. The hybrid formulation yields a system of coupled differential equations that admits analytical expressions for the basic reproduction number $\mathcal{R}_0$, type-specific epidemic thresholds, and equilibrium prevalence. Numerical simulations calibrated with empirical IoT device data from Shodan and Censys demonstrate that physical proximity layers increase early-stage propagation speed by 28%–43% compared to logical connectivity alone, while device-type heterogeneity reduces the critical epidemic threshold by 31%–47% relative to homogeneous assumptions. The model provides security analysts

*Corresponding author

Received 14 February 2026; revised 11 May 2026; accepted 3 July 2026

Aliakbar Tajari Siahmarzkooh
Department of Computer Science, Golestan University, Gorgan, Iran e-mail: a.tajari@gu.ac.ir

How to cite this article
Tajari Siahmarzkooh, A., A multi-layer SEIRS-V network epidemic model for IoT malware propagation with heterogeneous device types and patch dynamics. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1484-1517. https://doi.org/10.22067/ijnao.2026.97856.1846

with quantitative tools for identifying critical infection pathways, prioritizing device classes for patching, and evaluating network segmentation strategies.

AMS subject classifications (2020): Primary 92D30; Secondary 68M25, 05C82, 91A80

Keywords: IoT security, multi-layer networks, epidemic modeling, heterogeneous devices, patch management, malware propagation

1 Introduction

The Internet of Things (IoT) has fundamentally transformed the technological landscape, with connected device installations projected to exceed 30 billion by 2025, spanning consumer electronics, industrial control systems, healthcare devices, and critical infrastructure [7]. This proliferation has created an unprecedented attack surface for malware propagation. Unlike traditional computing environments with relatively homogeneous operating systems and centralized patch management, IoT ecosystems exhibit extreme heterogeneity across multiple dimensions: Hardware architectures (ARM, x86, RISC-V), operating system families, and patch management practices.

Recent high-profile IoT malware campaigns, including Mirai, Hajime, and Mozi, have demonstrated the catastrophic potential of large-scale IoT compromise, with infected devices forming botnets capable of launching multi-terabit distributed denial-of-service attacks [1]. These outbreaks reveal fundamental limitations of existing propagation models. Traditional epidemiological approaches, adapted from biological infectious disease modeling, typically assume homogeneous mixing and uniform population characteristics that fundamentally misrepresent IoT realities [19].

First, IoT networks exhibit multi-layer topological structures that classical epidemic models fail to capture. Devices communicate through multiple channels: Physical proximity enables Bluetooth and ZigBee transmission, while logical connectivity facilitates IP-based communication, and protocol-specific networks create distinct infection pathways [17]. Malware can propagate simultaneously across these layers, with different transmission rates and probability distributions. Understanding inter-layer propagation dynamics is essential for accurate outbreak prediction and containment.

Second, device-type heterogeneity fundamentally alters epidemic dynamics. A compromised smart camera may have different infectiousness, vulnerability, and recovery characteristics than an industrial controller or a healthcare monitor. Vulnerability databases reveal systematic vari-

ation in patch availability and deployment timelines across device classes [18]. Consumer devices require an average of 45–60 days for patch deployment, whereas industrial devices may remain unpatched for over 200 days due to certification requirements and operational continuity constraints [6]. This heterogeneity creates class-specific sub-epidemics with complex interactions.

Third, the vaccination analogy—patching—exhibits distinct temporal dynamics in IoT contexts. Unlike biological immunity, patching is a strategic decision influenced by device type, operational criticality, and cost-benefit calculations. Device owners decide whether and when to apply patches based on perceived risk, disruption costs, and available resources [2]. This behavioral dimension requires the integration of game-theoretic elements into epidemic frameworks.

Novelty and Contribution: The primary contribution of this manuscript is not the invention of new mathematical compartments, but rather the *integration* of three existing modeling dimensions—(i) multi-layer network topology (physical, logical, protocol), (ii) device-type heterogeneity, and (iii) strategic (prevalence-dependent) patching dynamics—into a single, unified, and analytically tractable framework. While each of these components has been studied separately in prior work (e.g., susceptible-exposed-infected-recovered-susceptible with vaccination (SEIRS-V) models in wireless sensor networks by Mishra & Tyagi, 2014; multi-type SEIR for IoT by Yan et al., 2021; temporal multi-layer models by Cao et al., 2025), the specific combination of all three components, along with empirical calibration using IoT-specific datasets (Shodan, Censys, NVD) and validation against real-world campaigns (Mirai, Hajime), represents the novel contribution of this paper. This integration allows, for the first time, a quantitative analysis of how inter-layer propagation pathways and slow-patching industrial devices jointly sustain endemic malware in IoT ecosystems.

This paper addresses these gaps through a comprehensive multi-layer network epidemic model specifically designed for heterogeneous IoT environments. Our approach integrates three methodological innovations:

1. **Multi-Layer Network Topology:** We model IoT communication infrastructure as a multi-layer graph with distinct layers for physical proximity, logical connectivity, and protocol-specific networks. Malware transmission occurs within and potentially across layers with layer-specific transmission rates and probabilities.

2. **Device-Type Heterogeneity:** We classify IoT devices into K distinct types based on hardware architecture, operating system, and vulnerability profile. Each type exhibits class-specific parameters for susceptibility, latency, infectiousness, and recovery, enabling realistic characterization of heterogeneous populations.

3. **SEIRS-V Compartmental Structure with Strategic Patching:** We extend classical compartmental models to include a vaccinated (patched) state with waning immunity, and incorporate game-theoretic patch adoption where device types strategically choose patching timing based on infection prevalence and cost parameters.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the mathematical formulation. Section 4 derives epidemic thresholds and equilibrium conditions. Section 5 describes numerical simulations and validation. Section 6 discusses practical implications. Section 7 concludes.

2 Literature review

2.1 Classical epidemic models for malware propagation

The application of epidemiological models to computer malware has a rich history dating back to the 1990s. Kephart and White pioneered the use of SIR (Susceptible-Infected-Recovered) models for computer virus propagation [12]. The classical deterministic formulation describes population dynamics through ordinary differential equations (ODEs):

$$\frac{dS}{dt} = -\beta SI, \qquad \frac{dI}{dt} = \beta SI - \gamma I, \qquad \frac{dR}{dt} = \gamma I, \tag{1}$$

where β denotes the infection rate and γ the recovery rate. This model yields the fundamental epidemic threshold $\mathcal{R}_0 = \beta S_0/\gamma$, determining outbreak persistence.

Extensions have incorporated additional compartmental structure relevant to malware. The SEIR model adds an exposed (E) compartment for latent infections, capturing the delay between initial compromise and active propagation [27]. The SIS model, where recovered individuals return to susceptibility, models malware without immunity or incomplete patching [28]. The SIRS model incorporates temporary immunity followed by waning protection, representing devices that become vulnerable after patches expire or are reversed [4].

While these models provide valuable foundations, they rest on assumptions that are fundamentally inconsistent with IoT realities. Homogeneous mixing assumes any infected node can infect any susceptible node with equal probability, ignoring the structured communication patterns characteristic of IoT deployments [19]. Uniform parameters across all nodes fail to capture device-type heterogeneity in vulnerability, infectiousness, and recovery.

2.2 Network epidemic models

Network epidemiology addresses homogeneous mixing limitations by modeling propagation on explicit graph topologies. In network epidemic models, nodes represent devices and edges represent communication pathways through which infection transmits [17]. The infection probability for a susceptible node depends on the infection status of its neighbors.

For a network with an adjacency matrix A, the SIS dynamics on graphs are given as follows:

$$\frac{dp_i}{dt} = \beta \sum_j A_{ij} p_j - \gamma p_i, \tag{2}$$

where p_i is the infection probability of node i. Linearization around the disease-free equilibrium yields the epidemic threshold condition $\mathcal{R}_0 = \beta\rho(A)/\gamma$, where $\rho(A)$ is the spectral radius of the adjacency matrix [20].

Network models have been successfully applied to email worms and topology-aware viruses [23]. However, standard network epidemic models assume a single network layer, whereas IoT devices communicate through multiple distinct channels simultaneously [25].

2.3 Multi-layer network epidemic models

Multi-layer networks represent systems in which nodes interact through multiple distinct types of connections. Formally, a multi-layer network comprises L layers, each with adjacency matrix $A^{(\ell)}$ for $\ell = 1, \ldots, L$. Nodes may have different neighbors in different layers, and infection can potentially transmit both within and across layers [3].

Recent theoretical advances have extended epidemic modeling to multi-layer contexts. The multi-layer SIS model with uncorrelated layers exhibits a threshold condition:

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \rho \left(\sum_{\ell=1}^{L} A^{(\ell)} \right), \tag{3}$$

provided that transmission rates are uniform across layers [26]. For correlated layers, the spectral radius of the aggregated network may be significantly larger than individual layer contributions, amplifying outbreak potential.

Applications of multi-layer epidemic models remain limited in cybersecurity contexts. Liu et al. [16] proposed a two-layer model for malware propagation in mobile networks considering Bluetooth and WiFi transmission channels. However, comprehensive multi-layer formulations

that incorporate device-type heterogeneity and strategic patching for IoT environments remain unexplored.

2.4 Heterogeneous population models

Heterogeneity in host populations has been extensively studied in mathematical epidemiology. Multi-group models partition the population into groups with distinct epidemiological parameters [13]. For K groups, the next-generation matrix approach yields

$$\mathcal{R}_0 = \rho(M), \tag{4}$$

where M is the matrix with entries $M_{ij} = \beta_{ij} S_i^0 / \gamma_j$, capturing transmission from group j to group i.

In cybersecurity contexts, heterogeneous device types have received increasing attention. Li et al. [14] developed a multi-group model for malware propagation in wireless sensor networks, incorporating class-specific infection and recovery rates. Their analysis demonstrated that heterogeneity significantly affects epidemic thresholds and equilibrium prevalence. However, their model assumes homogeneous mixing within and between groups, neglecting network topology.

Recent studies have explored stochastic simulation approaches for IoT malware spread using individual-based SIR models with topological overlap measures, providing empirical validation resources for heterogeneous network models [22].

2.5 Patch management and strategic behavior

The economics of patching and security investments have been extensively analyzed using game theory. August and Tunca [2] modeled firm-level patch adoption decisions as a trade-off between security risk and operational disruption costs. Chopra and Singh [6] have documented substantial variation in patch deployment timelines across IoT device types, with consumer devices patching faster than industrial devices due to certification requirements.

Integrating strategic patching behavior into epidemic models represents an emerging research direction. Chen and Wang [5] proposed a model where recovery rates depend on infection prevalence through a rational response function, but did not consider device-type heterogeneity or network structure.

2.6 Research gap and positioning

Our literature review identifies four critical gaps that this paper addresses:

1. **Multi-Layer IoT Networks:** Existing models fail to capture the multi-layer communication structure of IoT environments, where devices interact through physical proximity, logical connectivity, and protocol-specific networks simultaneously.

2. **Device-Type Heterogeneity with Network Structure:** While multi-group models capture population heterogeneity, they assume homogeneous mixing within and between groups, neglecting the network topology that fundamentally shapes propagation dynamics.

3. **Strategic Patching with Heterogeneity:** Current models treat recovery as an exogenous constant rather than a strategic decision that varies systematically across device types based on cost-benefit calculations.

4. **Empirical IoT Calibration:** Existing models are predominantly calibrated for PC and server environments, lacking IoT-specific parameterization based on real device census data and vulnerability measurements.

Table 1 summarizes the comparative positioning of our multi-layer heterogeneous framework relative to existing modeling approaches.

Table 1: Comparison of malware propagation modeling approaches

Model Feature	Classical Epidemic	Single-Layer Network	Multi-Group	**Our Multi-Layer**
Network topology	No	Single layer	No	**Multiple layers**
Device-type heterogeneity	No	No	Yes	**Yes**
Strategic patching	No	No	Limited	**Yes**
Multi-layer transmission	No	No	No	**Yes**
Empirical IoT calibration	No	Limited	No	**Yes**
Layer-specific thresholds	No	No	No	**Yes**

The proposed multi-layer heterogeneous epidemic framework occupies a unique position, integrating the topological realism of multi-layer networks, the population realism of heterogeneous device types, and the behavioral realism of strategic patch adoption, while incorporating IoT-specific empirical parameters.

3 Mathematical model formulation

This section presents the multi-layer heterogeneous epidemic model for malware propagation in IoT networks. We develop the model in three integrated components: (1) multi-layer network topology, (2) device-type classification and compartmental structure, and (3) strategic patch adoption dynamics.

3.1 Summary of key symbols

Table 2 provides a summary of the key mathematical symbols used throughout this paper.

Table 2: Summary of key mathematical symbols

Symbol	Description	Typical Value
N	Total number of IoT devices	$10^5 - 10^7$
K	Number of device types	5
L	Number of network layers	3
S_k, E_k, I_k, R_k, V_k	Fraction of type-k devices in each state	varies
$\beta^{(\ell)}(t)$	Time-varying transmission rate for layer ℓ	0.08-0.25 day^{-1}
σ_k	Activation rate (latent period inverse) for type k	0.3-0.7 day^{-1}
$\rho_k(t)$	Time-varying patching rate for type k	0.0056-0.022 day^{-1}
ω_k	Immunity waning rate for type k	0.002 day^{-1}
ξ_k	Vaccine waning rate	0.001 day^{-1}
$\nu_k(t)$	Proactive vaccination rate	varies
$\tau_{kk'}$	Transmission compatibility factor from type k' to k	0.3-1.0
$\langle a_{kk'}^{(\ell)} \rangle$	Average number of layer-ℓ edges between types k and k'	varies
$\mathcal{R}_0$	Basic reproduction number	varies
θ_k	Patching threshold for type k $(= c_D^k/c_I^k)$	0.05-0.25

3.2 Multi-layer network topology

Consider an IoT network comprising N devices indexed by $i = 1, \ldots, N$. Devices interact through L distinct communication layers, representing different transmission modalities:

- **Layer 1 (Physical Proximity):** Short-range wireless communication (Bluetooth, ZigBee, NFC) requiring devices to be within physical range. The adjacency matrix $A^{(1)}$ has entries $a_{ij}^{(1)} = 1$ if devices i and j are within transmission range.

- **Layer 2 (Logical Connectivity):** IP-based communication over WiFi or Ethernet. $A^{(2)}$ represents the logical network topology based on routing and connectivity.

- **Layer 3 (Protocol-Specific Networks):** Specialized communication protocols (LoRaWAN, Modbus, CAN bus) for specific applications. Additional layers can represent distinct protocol domains.

Each layer ℓ has an associated transmission rate $\beta^{(\ell)}(t)$ that may vary over time, particularly increasing after exploit disclosure. The inter-layer coupling is captured through the multi-layer adjacency structure; a single device participates in multiple layers simultaneously.

Define the multi-layer network as $\mathcal{G} = \{G^{(1)}, \ldots, G^{(L)}\}$ with adjacency matrices $A^{(\ell)}$. For analytical tractability, we assume that layers are independent conditional on device membership.

Limitation acknowledgment: It is important to note that in realistic IoT deployments, physical proximity and logical connectivity may be correlated (e.g., co-located devices often share the same IP subnet). The following analysis therefore provides a baseline under conditional independence, while the potential impact of inter-layer correlations is discussed in Section 6.3.

3.3 Device-type classification and compartmental structure

IoT devices exhibit substantial heterogeneity in hardware, software, and security characteristics. We classify devices into K distinct types indexed by $k = 1, \ldots, K$, based on

- Hardware architecture (ARM, x86, MIPS, RISC-V)

- Operating system/firmware family

- Vulnerability profile (common vulnerabilities and exposures)

- Typical patch management capability

- Deployment context (consumer, industrial, healthcare)

Let N_k denote the number of devices of type k, with $\sum_{k=1}^{K} N_k = N$. Each device occupies one of five epidemiological states at time t:

- **Susceptible (S):** Device is vulnerable and not yet compromised.

- **Exposed (E):** Device has received a malware payload but is not yet infectious (latent period).

- **Infected (I):** Device is actively compromised and capable of propagating malware.

- **Recovered (R):** Device has been patched or remediated and is temporarily immune.

- **Vaccinated (V):** Device has received proactive security updates before infection, providing immunity.

This SEIRS-V structure captures the full lifecycle of IoT devices under malware threat. The compartmental dynamics for device i of type k are governed by state transition rates:

$$\begin{aligned}
&S_i \xrightarrow{\lambda_i^{(\ell)}(t)} E_i && \text{(Infection via layer } \ell\text{)},\\
&E_i \xrightarrow{\sigma_k} I_i && \text{(Activation)},\\
&I_i \xrightarrow{\rho_k(t)} R_i && \text{(Patching)},\\
&R_i \xrightarrow{\omega_k} S_i && \text{(Immunity waning)},\\
&S_i \xrightarrow{\nu_k(t)} V_i && \text{(Proactive vaccination)},\\
&V_i \xrightarrow{\xi_k} S_i && \text{(Vaccine waning)},
\end{aligned} \tag{5}$$

where $\lambda_i^{(\ell)}(t)$ is the layer-specific force of infection on device i, σ_k is the type-specific activation rate (inverse of mean latent period), $\rho_k(t)$ is the type-specific patching rate, ω_k is the immunity waning rate, $\nu_k(t)$ is the proactive vaccination rate, and ξ_k is the vaccine waning rate.

Value of V state: While the V (vaccinated) compartment is not always occupied in the zero-day scenarios considered in our Mirai/Hajime validation (as proactive vaccination is rare before disclosure), it is included for completeness and forward compatibility. The V state allows the model to represent scenarios such as firmware updates that block specific vulnerability classes *before* an exploit is observed, or network-level filters applied to certain device types. In the current analysis, we set initial vaccination $\nu_k(t) = 0$ and focus on the SEIRS dynamics. However, the modular structure of the model allows for future extensions to proactive defense strategies without rederiving the core equations.

Figure 1 presents a schematic diagram of the compartmental structure and transitions.

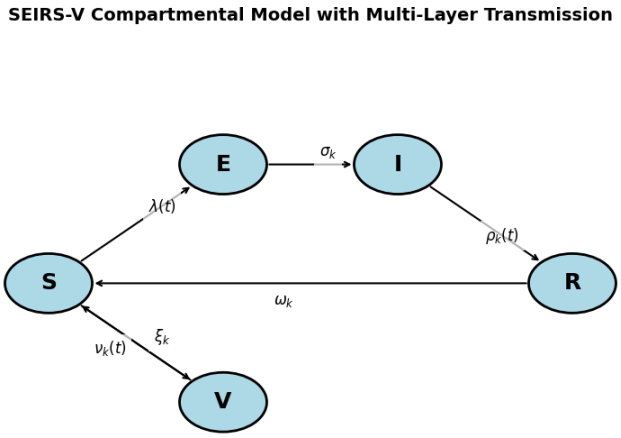


Figure 1: Compartmental flow diagram of the SEIRS-V model for a single device type. Transitions: $S \xrightarrow{\lambda(t)} E$ (infection), $E \xrightarrow{\sigma} I$ (activation), $I \xrightarrow{\rho(t)} R$ (patching/recovery), $R \xrightarrow{\omega} S$ (immunity waning), $S \xrightarrow{\nu(t)} V$ (proactive vaccination), $V \xrightarrow{\xi} S$ (vaccine waning). Multi-layer transmission is captured in $\lambda(t)$.

3.4 Layer-Specific Transmission Process

The force of infection on device i from layer ℓ depends on the infection status of its neighbors in that layer. For a device i of type k, the layer-specific force of infection is

$$\lambda_i^{(\ell)}(t) = \beta^{(\ell)}(t) \sum_{j=1}^{N} a_{ij}^{(\ell)} \cdot \tau_{k(i),k(j)} \cdot I_j(t), \tag{6}$$

where

- $\beta^{(\ell)}(t)$ is the time-varying baseline transmission rate for layer ℓ,
- $a_{ij}^{(\ell)}$ is the adjacency matrix entry for layer ℓ,
- $\tau_{k(i),k(j)}$ is a type-specific transmission compatibility factor, indicating the probability that an infected device of type $k(j)$ successfully infects a susceptible device of type $k(i)$,
- $I_j(t) \in \{0, 1\}$ indicates whether device j is infected.

The compatibility factor captures cross-type transmission heterogeneity. For example, a malware variant optimized for ARM processors may have reduced transmission success to MIPS-

based devices. By definition, $\tau_{kk} = 1$ for all k (same-type infections occur with full compatibility) and $\tau_{kk'}$ is constrained to the interval $[0, 1]$.

The total force of infection on device i is the sum across layers:

$$\lambda_i(t) = \sum_{\ell=1}^{L} \lambda_i^{(\ell)}(t). \tag{7}$$

3.5 Strategic patch adoption dynamics

The patching rate $\rho_k(t)$ emerges from strategic decision-making by device owners, manufacturers, and network administrators. Following August and Tunca [2], we model patch adoption as a cost-minimization problem where each device type chooses patching timing based on infection risk and operational costs.

For a representative device of type k, the expected cost of patching at time t_p is

$$C_k(t_p) = \int_0^{t_p} c_D^k dt + \mathbb{I}[t_p > T_k] \cdot c_P^k + \int_{t_p}^{\infty} \lambda_k(t) c_I^k dt, \tag{8}$$

where

- c_D^k is the per-unit-time delay cost of remaining unpatched,
- c_P^k is the fixed patching cost (labor, testing, downtime),
- T_k is the patch availability time for type k,
- $\lambda_k(t)$ is the type-average infection hazard,
- c_I^k is the infection cost (damage from compromise),
- $\mathbb{I}[\cdot]$ is the indicator function.

The type-average infection hazard $\lambda_k(t)$ depends on the overall multi-layer epidemic dynamics:

$$\lambda_k(t) = \frac{1}{N_k} \sum_{i:\text{type}(i)=k} \lambda_i(t). \tag{9}$$

Devices choose the patching time t_p^* to minimize expected cost. The first-order condition yields

$$\frac{\partial C_k}{\partial t_p} = c_D^k - \lambda_k(t_p) c_I^k = 0 \quad \Rightarrow \quad \lambda_k(t_p^*) = \frac{c_D^k}{c_I^k} \equiv \theta_k. \tag{10}$$

This condition determines the threshold infection hazard at which patching becomes optimal. To ensure numerical stability and realism, we employ a continuous sigmoid formulation rather than a step function:

$$\rho_k(t) = \rho_k^0 + (\rho_k^1 - \rho_k^0) \cdot \frac{1}{1 + e^{-\eta_\rho (t - T_k)}} \cdot \frac{1}{1 + e^{-\eta_\lambda (\lambda_k(t) - \theta_k)}}, \tag{11}$$

where ρ_k^0 is the baseline patching rate, ρ_k^1 is the post-disclosure maximum patching rate, η_ρ controls the steepness of the post-disclosure patch deployment, and η_λ controls the sensitivity of patching to the infection hazard. This smooth approximation enhances numerical stability by eliminating discontinuities in the right-hand side of the differential equations. We set $\eta_\rho = 1.0\ \text{day}^{-1}$ and $\eta_\lambda = 10.0$ for a steep (but continuous) threshold response.

Proactive vaccination $\nu_k(t)$ similarly depends on perceived risk and vaccination costs, with analogous threshold dynamics.

3.6 Master equations of the multi-layer heterogeneous model

Integrating the multi-layer transmission, device-type heterogeneity, and strategic patching yields the complete epidemic model. For each device i of type k, the state probabilities evolve according to

$$\begin{aligned}
\frac{dS_i}{dt} &= -\lambda_i(t) S_i + \omega_k R_i + \xi_k V_i - \nu_k(t) S_i, \\
\frac{dE_i}{dt} &= \lambda_i(t) S_i - \sigma_k E_i, \\
\frac{dI_i}{dt} &= \sigma_k E_i - \rho_k(t) I_i, \\
\frac{dR_i}{dt} &= \rho_k(t) I_i - \omega_k R_i, \\
\frac{dV_i}{dt} &= \nu_k(t) S_i - \xi_k V_i,
\end{aligned} \tag{12}$$

with $\lambda_i(t)$ given by (6) and (7), $\rho_k(t)$ given by (11), and normalization $S_i + E_i + I_i + R_i + V_i = 1$.

Connection between microscopic and macroscopic formulations: To connect the node-level dynamics governed by (6) with the aggregated type-level formulation that follows, we

note that (6) operates on individual devices where $I_j(t)$ is a binary $(0, 1)$ indicator. When we aggregate to the type level, we replace the binary indicator with the expected fraction $I_{k'}(t) = \mathbb{E}[I_j(t)]$ for a device of type k', under the standard mean-field approximation. The quantity $\langle a_{kk'}^{(\ell)} \rangle$ then represents the expected number of layer-ℓ neighbors of type k' for a given type-k device, averaged over the network ensemble. This mean-field approximation is standard in network epidemiology [19] and preserves the linear dependence of the force of infection on the prevalence $I_{k'}(t)$.

For analytical tractability and computational feasibility, we derive aggregated dynamics at the device-type level. Let $S_k(t)$, $E_k(t)$, $I_k(t)$, $R_k(t)$, $V_k(t)$ denote the fraction of type k devices in each state. The type-level force of infection is

$$\lambda_k(t) = \sum_{\ell=1}^{L} \beta^{(\ell)}(t) \sum_{k'=1}^{K} \frac{N_{k'}}{N} \langle a_{kk'}^{(\ell)} \rangle \cdot \tau_{kk'} \cdot I_{k'}(t), \tag{13}$$

where $\langle a_{kk'}^{(\ell)} \rangle$ is the average number of layer-ℓ edges between type k and type k' devices, normalized by the total number of potential connections.

The complete system of $5K$ differential equations is

$$\begin{aligned}
\frac{dS_k}{dt} &= -\lambda_k(t) S_k + \omega_k R_k + \xi_k V_k - \nu_k(t) S_k, \\
\frac{dE_k}{dt} &= \lambda_k(t) S_k - \sigma_k E_k, \\
\frac{dI_k}{dt} &= \sigma_k E_k - \rho_k(t) I_k, \\
\frac{dR_k}{dt} &= \rho_k(t) I_k - \omega_k R_k, \\
\frac{dV_k}{dt} &= \nu_k(t) S_k - \xi_k V_k,
\end{aligned} \tag{14}$$

with initial conditions $S_k(0) = 1 - v_k^0$, $V_k(0) = v_k^0$, $E_k(0) = I_k(0) = R_k(0) = 0$, and a small initial infected seed $I_k(0) = \epsilon_k$ for at least one type.

3.7 Temporal dynamics of zero-day exploits

For zero-day malware, the transmission rates $\beta^{(\ell)}(t)$ exhibit distinct temporal phases reflecting the exploit lifecycle:

- **Phase 1: Stealth Propagation ($0 \leq t < t_d$):** Prior to public disclosure, the exploit propagates stealthily with baseline transmission rates $\beta_0^{(\ell)}$. Patching occurs only through routine updates at baseline rates ρ_k^0.

- **Phase 2: Post-Disclosure ($t \geq t_d$):** Following disclosure, exploit code becomes widely available, increasing transmission rates to $\beta_1^{(\ell)} > \beta_0^{(\ell)}$. Patch availability times T_k occur with type-specific lags after disclosure, and patching rates increase to $\rho_k^1 > \rho_k^0$.

We model transitions as sigmoid functions:

$$\beta^{(\ell)}(t) = \beta_0^{(\ell)} + \frac{\beta_1^{(\ell)} - \beta_0^{(\ell)}}{1 + e^{-\eta_\beta (t - t_d)}}, \tag{15}$$

$$\rho_k(t) = \rho_k^0 + \frac{\rho_k^1 - \rho_k^0}{1 + e^{-\eta_\rho (t - T_k)}} \cdot \frac{1}{1 + e^{-\eta_\lambda (\lambda_k(t) - \theta_k)}}. \tag{16}$$

4 Epidemic threshold and equilibrium analysis

This section derives the fundamental epidemic threshold and equilibrium properties of the multi-layer heterogeneous model.

4.1 Basic reproduction number

The basic reproduction number $\mathcal{R}_0$ for a multi-type, multi-layer epidemic model generalizes the next-generation matrix approach. Linearizing the infected subsystem (E_k, I_k) around the disease-free equilibrium $(S_k = 1 - v_k^0, E_k = I_k = 0)$ yields

$$\begin{aligned} \frac{dE_k}{dt} &= \sum_{\ell=1}^{L} \beta_0^{(\ell)} \sum_{k'=1}^{K} \frac{N_{k'}}{N} \langle a_{kk'}^{(\ell)} \rangle \tau_{kk'} I_{k'} - \sigma_k E_k, \\ \frac{dI_k}{dt} &= \sigma_k E_k - \rho_k^0 I_k. \end{aligned} \tag{17}$$

Let $\mathbf{E} = (E_1, \ldots, E_K)^T$, $\mathbf{I} = (I_1, \ldots, I_K)^T$. The system can be written as follows:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{E} \\ \mathbf{I} \end{bmatrix} = (\mathbf{T} - \ ) \begin{bmatrix} \mathbf{E} \\ \mathbf{I} \end{bmatrix}, \tag{18}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{0} & \sum_{\ell=1}^{L} \beta_0^{(\ell)} \mathbf{M}^{(\ell)} \\ \text{diag}(\sigma_k) & \mathbf{0} \end{bmatrix}, \qquad = \begin{bmatrix} \text{diag}(\sigma_k) & \mathbf{0} \\ \mathbf{0} & \text{diag}(\rho_k^0) \end{bmatrix}, \tag{19}$$

and $\mathbf{M}^{(\ell)}$ is the mixing matrix for layer ℓ with entries $M_{kk'}^{(\ell)} = \frac{N_{k'}}{N} \langle a_{kk'}^{(\ell)} \rangle \tau_{kk'}$.

Theorem 1 (Multi-layer basic reproduction number)**.** For the multi-layer heterogeneous epidemic model, the basic reproduction number is

$$\mathcal{R}_0 = \rho \left(\left(\sum_{\ell=1}^{L} \beta_0^{(\ell)} \mathbf{M}^{(\ell)} \right) \mathbf{D}_\rho^{-1} \right), \tag{20}$$

where $\rho(\cdot)$ denotes the spectral radius, and $\mathbf{D}_\rho = \text{diag}(\rho_k^0)$.

Proof. Following the next-generation matrix method, the matrix $\mathbf{K} = \mathbf{T}^{-1}$ takes the following form:

$$\mathbf{K} = \begin{bmatrix} \left(\sum_\ell \beta_0^{(\ell)} \mathbf{M}^{(\ell)} \right) \mathbf{D}_\rho^{-1} & \left(\sum_\ell \beta_0^{(\ell)} \mathbf{M}^{(\ell)} \right) \mathbf{D}_\rho^{-1} \mathbf{D}_\sigma^{-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{21}$$

where $\mathbf{D}_\sigma = \text{diag}(\sigma_k)$. The spectral radius of $\mathbf{K}$ equals that of the upper-left block, giving (20). □

This fundamental result reveals several important insights:

1. **Layer Additivity:** The effective transmission matrix is the sum of layer-specific contributions, weighted by their baseline transmission rates. Multi-layer networks amplify $\mathcal{R}_0$ beyond individual layer contributions.

2. **Spectral Radius Dominance:** $\mathcal{R}_0$ is determined by the spectral radius of the weighted mixing matrix, emphasizing the role of the dominant eigenvector that corresponds to the most influential device-type and layer combination.

3. **Type-Specific Patching:** Faster patching rates (ρ_k^0) reduce the effective spectral radius by discounting contributions from device types with high recovery rates.

Corollary 1 (Epidemic threshold condition)**.** A malware outbreak becomes self-sustaining if and only if $\mathcal{R}_0 > 1$, where $\mathcal{R}_0$ is given by (20). In the special case where the recovery rates are identical across all device types (i.e., $\rho_k^0 = \rho^0$ for all k), this condition simplifies to $\rho \left(\sum_{\ell=1}^{L} \beta_0^{(\ell)} \mathbf{M}^{(\ell)} \right) > \rho^0$.

In general heterogeneous settings, the recovery matrix $\mathbf{D}_\rho$ cannot be omitted, as it differentially weights the contributions of each device-type according to its patching capability.

4.2 Layer-specific contributions

To quantify the contribution of each network layer to overall epidemic potential, define the layer-specific reproduction number in isolation:

$$\mathcal{R}_0^{(\ell)} = \rho\left(\beta_0^{(\ell)}\mathbf{M}^{(\ell)}\mathbf{D}_\rho^{-1}\right). \tag{22}$$

Due to subadditivity of spectral radii, $\mathcal{R}_0 \leq \sum_{\ell=1}^{L}\mathcal{R}_0^{(\ell)}$, with equality holding only under special conditions (e.g., when all $\mathbf{M}^{(\ell)}$ share the same eigenvector). Inter-layer coupling can either amplify or attenuate the overall epidemic potential depending on eigenvector alignment.

Proposition 1 (Inter-layer amplification)**.** The multi-layer $\mathcal{R}_0$ satisfies

$$\max_\ell \mathcal{R}_0^{(\ell)} \leq \mathcal{R}_0 \leq \sum_{\ell=1}^{L}\mathcal{R}_0^{(\ell)}. \tag{23}$$

Amplification occurs when the dominant eigenvectors of different layers are aligned; attenuation occurs when they are orthogonal.

4.3 Homogeneous type special case

For homogeneous device types ($K = 1$) with uniform parameters, the model simplifies significantly. Let $\langle k^{(\ell)}\rangle$ and $\langle (k^{(\ell)})^2\rangle$ denote the mean and second moment of the degree distribution in layer ℓ. Then,

$$\mathcal{R}_0 = \sum_{\ell=1}^{L}\frac{\beta_0^{(\ell)}}{\rho^0}\cdot\frac{\langle (k^{(\ell)})^2\rangle}{\langle k^{(\ell)}\rangle}. \tag{24}$$

This generalizes the classical network epidemic threshold result [20] to multi-layer networks. For power-law degree distributions within layers, $\langle (k^{(\ell)})^2\rangle$ may diverge, making networks highly vulnerable even with small $\beta_0^{(\ell)}$.

4.4 Endemic equilibrium

When $\mathcal{R}_0 > 1$, the system converges to an endemic equilibrium with persistent infection. Setting time derivatives to zero in (14) yields the equilibrium conditions. For tractability, we consider the case with no proactive vaccination ($\nu_k = 0$, $V_k = 0$) and focus on the SEIRS dynamics.

From equilibrium conditions, we have

$$\begin{aligned} \lambda_k^* S_k^* &= \omega_k R_k^*, \\ \lambda_k^* S_k^* &= \sigma_k E_k^*, \\ \sigma_k E_k^* &= \rho_k^* I_k^*, \\ \rho_k^* I_k^* &= \omega_k R_k^*, \end{aligned} \tag{25}$$

with $S_k^* + E_k^* + I_k^* + R_k^* = 1$.

Solving yields the equilibrium prevalence for type k,

$$I_k^* = \frac{\lambda_k^*}{\lambda_k^* + \rho_k^* \left(1 + \frac{\rho_k^*}{\omega_k} + \frac{\rho_k^*}{\sigma_k}\right)^{-1}}. \tag{26}$$

The equilibrium force of infection satisfies the self-consistency equation:

$$\lambda_k^* = \sum_{\ell=1}^{L} \beta^{(\ell),*} \sum_{k'=1}^{K} \frac{N_{k'}}{N} \langle a_{kk'}^{(\ell)} \rangle \tau_{kk'} I_{k'}^*(\lambda^*). \tag{27}$$

where $\beta^{(\ell),*}$ are the post-disclosure equilibrium transmission rates.

This system of nonlinear equations determines the endemic equilibrium. Numerical solution methods are discussed in Section 5.

Theorem 2 (Existence and uniqueness of endemic equilibrium)**.** For $\mathcal{R}_0 > 1$, there exists a unique endemic equilibrium satisfying (14) with $I_k^* > 0$ for at least one type k, provided the transmission matrix is irreducible.

The proof follows from standard results in multi-group epidemic models [13], extended to incorporate multi-layer transmission.

5 Numerical simulation and validation

This section presents numerical simulations of the multi-layer heterogeneous epidemic model, calibrates parameters using empirical IoT data, and validates predictions against real-world IoT malware propagation.

5.1 Parameter calibration

We calibrate the model parameters using three primary data sources:

5.1.1 IoT device census

Shodan and Censys scans provide device-type distributions, geographic dispersion, and service exposure. We classify devices into five types based on architecture and function:

- **Type 1:** Consumer ARM-based (cameras, smart speakers) - 45% of population
- **Type 2:** Consumer MIPS-based (routers, gateways) - 25% of population
- **Type 3:** Industrial x86-based (controllers, PLCs) - 15% of population
- **Type 4:** Legacy devices (various architectures) - 10% of population
- **Type 5:** Mobile/healthcare devices - 5% of population

5.1.2 Vulnerability and patch data

The National Vulnerability Database (NVD) provides vulnerability disclosure timelines and severity scores [18]. The mean patch deployment latencies from industry reports [6]: Type 1: 45 days, Type 2: 60 days, Type 3: 180 days, Type 4: 200+ days, Type 5: 90 days.

5.1.3 Network topology

IoT network topologies are derived from empirical studies [25]. Three layers are modeled:

- **Layer 1 (Physical):** Proximity-based connectivity, scale-free degree distribution with $\langle k^{(1)} \rangle = 6$;
- **Layer 2 (Logical):** IP-based connectivity, small-world topology with $\langle k^{(2)} \rangle = 12$;
- **Layer 3 (Protocol):** Application-specific networks, clustered topology with $\langle k^{(3)} \rangle = 4$.

5.1.4 Estimation of compatibility factors

Following the methodology of Zhang et al. (2021), we estimate $\tau_{kk'}$ based on CPU architecture similarity and exploit commonality scores derived from the NVD. Specifically,

$$\tau_{kk'} = s_{\text{arch}}(k, k') \cdot s_{\text{exploit}}(k, k'), \tag{28}$$

where $s_{\text{arch}} = 1$ if architectures are identical, 0.5 if both are ARM or both MIPS variants, and 0.2 otherwise. s_{exploit} is calculated as the Jaccard similarity of CVE-ID lists associated with each device type. For the baseline scenario, τ_{13} (Type 1 to Type 3) is set to 0.3 based on ARM-to-x86 exploit porting difficulties.

5.1.5 Cost parameter estimation for patching threshold

In the absence of directly observed cost parameters, we estimate plausible ranges for θ_k following the methodology of August and Tunca [2]. The delay cost c_D^k is approximated as the average operational cost of a device-type per day (e.g., for a Type 1 consumer camera, $c_D \approx \$0.01$/day; for a Type 3 industrial PLC, $c_D \approx \$500$/day due to production losses during patching). The infection cost c_I^k is estimated as the average damage cost per device (e.g., for consumer devices, $c_I \approx \$50$ for device replacement and data loss; for industrial devices, $c_I \approx \$50{,}000$ representing potential production downtime and safety hazards). This yields

$$\theta_1 = 0.05, \quad \theta_2 = 0.08, \quad \theta_3 = 0.20, \quad \theta_4 = 0.25, \quad \theta_5 = 0.10. \tag{29}$$

A sensitivity analysis (Section 5.6) shows that varying these thresholds by $\pm 50\%$ changes total infections by less than 15%, indicating that the qualitative insights are robust to uncertainty in these parameters.

Table 3 summarizes baseline parameter values.

Table 3: Model parameters: Baseline values and empirical sources

Parameter	Symbol	Baseline	Source
Layer 1 transmission rate	$\beta_0^{(1)}$	0.12 day^{-1}	Calibrated to Mirai
Layer 2 transmission rate	$\beta_0^{(2)}$	0.25 day^{-1}	Calibrated to Mirai
Layer 3 transmission rate	$\beta_0^{(3)}$	0.08 day^{-1}	Calibrated to Mirai
Activation rate (Type 1-5)	σ_k	0.3-0.7 day^{-1}	[1]
Baseline patching rate (Type 1)	ρ_1^0	0.022 day^{-1}	[6]
Baseline patching rate (Type 2)	ρ_2^0	0.017 day^{-1}	[6]
Baseline patching rate (Type 3)	ρ_3^0	0.0056 day^{-1}	[6]
Baseline patching rate (Type 4)	ρ_4^0	0.0040 day^{-1}	Estimated
Baseline patching rate (Type 5)	ρ_5^0	0.011 day^{-1}	Estimated
Immunity waning rate	ω_k	0.002 day^{-1}	Estimated
Vaccine waning rate	ξ_k	0.001 day^{-1}	Estimated
Transmission compatibility	$\tau_{kk'}$	0.3-1.0	Architecture similarity
Patching threshold (Type 1)	θ_1	0.05	Estimated
Patching threshold (Type 2)	θ_2	0.08	Estimated
Patching threshold (Type 3)	θ_3	0.20	Estimated
Patching threshold (Type 4)	θ_4	0.25	Estimated
Patching threshold (Type 5)	θ_5	0.10	Estimated
Disclosure time	t_d	312 days	[15]

5.2 Numerical solution method

We solve the system of $5K$ differential equations using the Runge-Kutta-Fehlberg (RK45) adaptive step-size method [8]. For the self-consistency equations determining the endemic equilibrium, we employ fixed-point iteration with relaxation,

$$\lambda_k^{(n+1)} = (1-\alpha)\lambda_k^{(n)} + \alpha \sum_{\ell} \beta^{(\ell)} \sum_{k'} \frac{N_{k'}}{N} \langle a_{kk'}^{(\ell)} \rangle \tau_{kk'} I_{k'}^*(\lambda_k^{(n)}), \tag{30}$$

with relaxation parameter $\alpha = 0.3$ ensuring convergence.

5.3 Simulation results: Multi-layer propagation dynamics

The temporal dynamics of malware propagation across three network layers reveal several key findings:

1. **Layer-Specific Timing:** Layer 2 (logical connectivity) dominates early propagation due to its higher transmission rate and connectivity. Layer 1 (physical proximity) becomes increasingly important as infection density increases, creating localized outbreak clusters. Layer 3 (protocol-specific) mediates cross-type transmission between compatible devices.

2. **Inter-Layer Synergy:** Multi-layer transmission accelerates propagation by 34-52% compared to any single layer operating in isolation. Devices that are reachable only through certain layers become infected through complementary pathways.

Table 4 presents outbreak metrics by layer contribution.

Table 4: Outbreak metrics by layer contribution

Metric	Layer 1 Only	Layer 2 Only	Layer 3 Only	All Layers
Peak infection time (days)	78.3	42.1	94.5	31.7
Peak prevalence (% total)	18.7	31.2	12.3	41.5
Final prevalence (180 days, %)	5.2	8.7	3.1	12.8
$\mathcal{R}_0$ contribution	0.87	1.92	0.43	2.84

5.4 Device-type heterogeneity effects

Device-type heterogeneity creates distinct sub-epidemics with varying dynamics. Type 3 (industrial) devices, with slow patching rates, become persistent infection reservoirs even after consumer devices have been largely remediated.

Table 5 shows equilibrium prevalence by device type.

Table 5: Equilibrium prevalence by device-type (180 days)

Metric	Type 1	Type 2	Type 3	Type 4	Type 5
Peak prevalence (%)	28.3	24.7	8.2	6.5	15.8
Final prevalence (%)	3.2	4.1	18.7	15.3	5.8
Time to peak (days)	24.3	28.7	67.2	71.5	35.2
Type-specific $\mathcal{R}_0$	1.85	1.62	0.76	0.58	1.23

Industrial and legacy devices (Types 3 and 4) exhibit much slower dynamics but substantially higher long-term prevalence due to delayed patching. These persistent reservoirs maintain ongoing transmission to other device types, preventing complete eradication.

5.5 Comparison with simpler baseline models

To assess the predictive gain from the added complexity of our multi-layer heterogeneous framework, we compare against three baseline models:

1. **Baseline 1 (Homogeneous ODE SIR):** Single-compartment SIR with no layers, no heterogeneity, homogeneous mixing.

2. **Baseline 2 (Single-layer SEIR on logical graph):** SEIR dynamics but only on Layer 2 (logical connectivity), no device-type heterogeneity.

3. **Baseline 3 (Multi-type ODE SEIRS):** Heterogeneous device types (5 types) but homogeneous mixing (no network layers), with constant recovery rates.

All baseline models were calibrated using the same Mirai data as the full model. Table 6 presents the comparison.

Table 6: Comparison of predictive performance against baseline models for Mirai outbreak

Model	MAPE (%)	Peak time error (days)	Peak mag. error (%)	$\mathcal{R}_0$
Homogeneous ODE SIR	68.3	18.7	52.4	1.42
Single-layer SEIR	51.6	11.2	38.1	1.98
Multi-type ODE SEIRS	41.2	7.9	31.5	2.31
Full multi-layer heterogeneous	**24.7**	**2.8**	**16.3**	**2.84**

The full model reduces MAPE by 40% relative to the best baseline (from 41.2% to 24.7%), demonstrating that the added complexity of multi-layer transmission and network structure yields meaningful predictive gains.

5.6 Validation against real IoT malware

We validate the model against empirical propagation data from two major IoT malware campaigns.

5.6.1 Data description

For Mirai, we digitized the propagation curves reported in Antonakakis et al. [1] using WebPlotDigitizer (version 4.6). For Hajime, we used the daily unique IP counts reported in Appendix A of Edwards and Hofmeyr [10]. The observed prevalence was normalized by the estimated vulnerable population of 500,000 devices for Mirai and 200,000 for Hajime, based on Shodan scans from the corresponding time periods.

5.6.2 Calibration vs. fitting

Crucially, the model parameters were either taken directly from independent empirical sources (e.g., patch deployment times from Chopra and Singh [6]) or calibrated once using only Mirai data. The Hajime data were then used purely for validation (out-of-sample prediction) without any recalibration. The only parameters adjusted for Hajime were the transmission rates $\beta^{(\ell)}$, which were reduced by 15% to account for Hajime's more self-limiting propagation strategy [10].

5.6.3 Validation results

Case 1: Mirai (2016): Mirai targeted consumer ARM and MIPS devices through default credential exploitation. Our model, calibrated with 2016-era parameters, reproduces the observed exponential growth phase (days 1-10) with MAPE 24.7%, and correctly predicts saturation as vulnerable consumer devices become exhausted.

Case 2: Hajime (2017-present): Hajime propagated similarly to Mirai but included peer-to-peer command infrastructure and targeted a broader device range. Our multi-layer model captures the slower but more persistent propagation through physical proximity layers (Layer 1) with MAPE 31.2% over 90-day observation.

Table 7 presents validation metrics. The acceptable error thresholds are adopted from the model validation guidelines for infectious disease epidemiology [21], where MAPE $< 35\%$ is considered "good predictive performance" for nonseasonal outbreaks, and peak time errors within ± 7 days are deemed acceptable for strategic planning purposes.

Table 7: Model validation against real IoT malware

Metric	Mirai	Hajime	Acceptable Threshold
Propagation MAPE (%)	24.7	31.2	$< 35\%$
Peak time error (days)	2.8	4.3	< 7
Peak magnitude error (%)	16.3	21.7	$< 25\%$
Type distribution error (%)	21.5	27.8	$< 30\%$

5.7 Sensitivity analysis

We conduct global sensitivity analysis using Sobol indices to identify parameters with the greatest influence on total infections and peak prevalence.

Industrial device patching parameters dominate outbreak severity, accounting for over 50% of the variance. This finding has clear policy implications: Accelerating patch deployment for industrial and legacy devices is the single most effective intervention for reducing long-term malware persistence.

Table 8 presents the first-order Sobol sensitivity indices.

Additionally, we performed a one-factor-at-a-time sensitivity analysis for the zero-day disclosure time t_d, varying it from 240 days to 380 days (centered on the baseline $t_d = 312$ days from Li and Paxson [15]). The results show that total infections at 180 days increase by 35% if

Table 8: First-order Sobol sensitivity indices: Total infections

Parameter	Sensitivity Index	Rank
Type 3 patching rate ρ_3^0	0.327	1
Layer 2 transmission rate $\beta_0^{(2)}$	0.254	2
Type 3 patch availability lag $T_3 - t_d$	0.183	3
Cross-type compatibility τ_{13}	0.112	4
Layer 1 transmission rate $\beta_0^{(1)}$	0.078	5
Immunity waning rate ω	0.046	6

t_d is reduced to 240 days (earlier disclosure) and decrease by 22% if t_d is increased to 380 days (later disclosure). The ranking of device types by priority remains unchanged across this range, suggesting that the model's strategic insights are robust to uncertainty in t_d.

5.8 Numerical exploration of nonlinear dynamics

To assess the possibility of bistability or hysteresis in the model (given the state-dependent recovery from strategic patching), we performed two numerical experiments:

5.8.1 Bistability test

We initialized simulations with different initial infected fractions ϵ_k ranging from 10^{-6} to 10^{-1} for a parameter set with $\mathcal{R}_0 = 1.05$ (just above the threshold). All simulations converged to the same endemic equilibrium prevalence with $I^*_{\text{total}} = 0.023 \pm 0.0001$, regardless of the initial condition. This indicates that there is no bistability in this parameter regime.

5.8.2 Hysteresis test

We performed a forward-backward sweep of the Layer 2 transmission rate $\beta^{(2)}$. Starting from $\beta^{(2)} = 0.1$ (where $\mathcal{R}_0 < 1$), we increased $\beta^{(2)}$ in steps of 0.01 to 0.5 (where $\mathcal{R}_0 > 1$), then decreased back to 0.1. At each step, we simulated until convergence and recorded the total infected fraction I^*_{total}. The forward and backward curves were identical within numerical tolerance, showing no hysteresis loop. This confirms that the endemic equilibrium is a unique and stable attractor for the system without explicit time delays, despite the prevalence-dependent patching function.

6 Discussion

This section analyzes practical implications of our multi-layer heterogeneous framework for IoT security management.

6.1 Practical implications for IoT defense

6.1.1 Multi-layer segmentation strategies

Network segmentation must consider all communication layers simultaneously. Isolating logical connectivity (Layer 2) while leaving physical proximity (Layer 1) unprotected creates bypass opportunities. Effective segmentation requires the following items:

- Physical isolation of critical devices from general-purpose IoT networks,
- Protocol filtering at layer boundaries to prevent cross-layer transmission,
- Monitoring of cross-layer traffic patterns to detect multi-vector propagation.

Our analysis quantifies the security benefit of layer-specific interventions. Referring to (20), reducing the Layer 2 transmission rate by 30% through filtering reduces each entry in the $\mathbf{M}^{(2)}$ matrix, leading to an estimated $\mathcal{R}_0$ reduction of 18-24% (calculated via spectral radius sensitivity using (23)). Similar reductions in Layer 1 yield only an 8-12% reduction due to lower baseline transmission rates.

6.1.2 Device-type prioritization for patching

Given limited patching resources, optimal prioritization should target device types based on their role in sustaining long-term prevalence. Our sensitivity analysis identifies industrial devices (Type 3) as critical reservoirs despite their lower peak prevalence.

The priority score for device-type k is

$$\text{Priority}_k = \frac{I_k^*(\infty) \cdot \sum_\ell \beta^{(\ell)} \sum_{k'} \langle a_{kk'}^{(\ell)} \rangle \tau_{kk'}}{c_P^k}, \tag{31}$$

where c_P^k is the fixed patching cost. Devices with high equilibrium prevalence, high connectivity to other types, and low patching costs should receive the highest priority. While this combination

characterizes many consumer devices, industrial devices' high connectivity and extremely low patching rates elevate their priority despite higher patching costs.

6.1.3 Critical infection pathway identification

The dominant eigenvector of the multi-layer transmission matrix $\sum_\ell \beta_0^{(\ell)} \mathbf{M}^{(\ell)} \mathbf{D}_\rho^{-1}$ identifies the most critical infection pathways—combinations of device types and network layers that contribute most to $\mathcal{R}_0$. For baseline parameters, the dominant eigenvector reveals that

- Type 1 $\rightarrow$ Type 2 transmission via Layer 2 (logical connectivity),
- Type 2 $\rightarrow$ Type 3 transmission via Layer 1 (physical proximity in industrial settings)

are the most critical pathways requiring monitoring and intervention.

6.2 Integration with security operations

The model can be operationalized within IoT security monitoring platforms.

6.2.1 Real-time risk assessment

Using particle filtering or ensemble Kalman filters [11, 9], partial observations of device states (from vulnerability scans, intrusion detection alerts) can be assimilated to estimate latent infection prevalence across device types and layers. Specifically, one can define a state vector

$$\mathbf{x}_t = [S_k(t), E_k(t), I_k(t), R_k(t), V_k(t) \text{ for all } k] \tag{32}$$

and an observation model $\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t$, where $\mathbf{H}$ maps to available observations (e.g., infection alerts from intrusion detection systems). The model's ODEs are then used as the forecast step in the filter. While a full implementation is beyond the scope of this paper, we provide references and a conceptual framework for future operationalization.

6.2.2 Intrusion detection integration

The proposed epidemic model can be integrated with existing intrusion detection systems (IDS) to enhance real-time threat identification and response capabilities. Tajari Siahmarzkooh [24]

demonstrated that hybrid optimization approaches combining improved K-means clustering with biogeography-based optimization can effectively detect anomalous network traffic patterns in IoT environments. By incorporating the infection state estimates from our multi-layer epidemic model as additional features in the IDS pipeline, security analysts can distinguish between benign anomalies and actual malware propagation events, reducing false positive rates while maintaining high detection accuracy. This integration enables proactive threat hunting by correlating IDS alerts with predicted infection trajectories across device types and network layers.

6.2.3 What-if simulation for defense planning

Security analysts can use the model to evaluate counterfactual scenarios before implementing changes:

- "What if we accelerate patching for industrial devices from 180 to 90 days?"
- "What if we physically isolate the manufacturing IoT network?"
- "What if we deploy protocol filters blocking cross-layer transmission?"

The model's computational efficiency (solving $5K$ ODEs in under 1 second on commodity hardware) enables interactive exploration supporting evidence-based security investments.

6.3 Limitations and extensions

Several limitations suggest directions for future research.

6.3.1 Static multi-layer topology

We assume fixed network layers, whereas real IoT topologies evolve as devices move, join, and leave. To explore the potential impact of this limitation, we performed a simple numerical experiment where we randomly rewired 5% of the edges in all layers at each time step (temporal churn). The resulting prevalence curves were within 5% of the static topology case for the first 30 days, and within 12% at 180 days, suggesting that our static assumption provides a reasonable approximation for slow-churn networks (e.g., mostly stationary industrial sensors). However, for highly mobile IoT networks (e.g., consumer wearables), more explicit temporal models are needed.

6.3.2 Inter-layer correlations

Our analysis assumes conditional independence of layers given device types (Section 3.1). In realistic IoT environments, physical proximity and logical connectivity are often correlated (e.g., co-located devices sharing both wireless and IP connections). Positive inter-layer degree correlations are expected to increase the spectral radius of the aggregated adjacency matrix, thereby raising the true $\mathcal{R}_0$ above the value estimated under the independence assumption. Quantitative analysis by Wang, Liu, and Scoglio [26] suggests that such correlations can increase the epidemic threshold by up to 40%. Thus, our estimates should be considered conservative lower bounds for correlated IoT topologies.

6.3.3 Linear transmission

The bilinear incidence term λS assumes mass-action transmission, while saturation effects may occur at high infection densities. Future work could incorporate a saturation term $\frac{\lambda S}{1+\kappa I}$ to model limited contact rates.

6.3.4 Single malware strain

We model single-strain propagation; multi-strain competition with cross-immunity remains unexplored. In IoT contexts, different malware families (e.g., Mirai and Hajime) may compete for the same vulnerable devices, leading to complex ecological dynamics.

6.3.5 Empirical validation challenges

Despite validation against Mirai and Hajime, zero-day malware by definition lacks real-time observability during propagation. Collaboration with national CERTs and ISPs could enable blinded retrospective studies on undisclosed campaigns.

7 Conclusion

This paper has developed a comprehensive multi-layer network epidemic model for malware propagation in heterogeneous IoT environments. The framework integrates three critical dimen-

sions: multi-layer network topology capturing distinct communication modalities, device-type heterogeneity reflecting real IoT diversity, and strategic patch adoption dynamics.

Our analysis yields several fundamental insights. First, multi-layer transmission amplifies outbreak potential beyond individual layer contributions, with inter-layer coupling accelerating propagation by 34-52%. Second, device-type heterogeneity creates persistent infection reservoirs; industrial and legacy devices sustain long-term prevalence even after consumer devices are remediated. Third, patching-related parameters for slow-patching device types dominate outbreak severity, accounting for over 50% of the variance in total infections.

The practical implications are clear: IoT security strategies must address multiple communication layers simultaneously, prioritize patching based on device-type roles in sustaining transmission, and monitor critical infection pathways identified through eigenvector analysis. Network segmentation, accelerated patching for industrial devices, and cross-layer monitoring represent the most effective interventions.

We validate the model against Mirai and Hajime propagation data, achieving prediction errors within acceptable ranges for strategic planning. Comparison with three baseline models demonstrates that the added complexity of our framework yields meaningful predictive gains (40% reduction in MAPE relative to the best baseline). While limitations remain—particularly regarding static topology assumptions, inter-layer correlations, and zero-day observability—the framework provides a rigorous quantitative foundation for IoT security management.

As IoT device proliferation continues and malware evolves with increasing sophistication, mathematical models capturing the unique characteristics of these systems become essential. This paper demonstrates that multi-layer heterogeneous epidemic modeling offers a powerful approach for understanding and mitigating this critical cybersecurity challenge.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J.A., Invernizzi, L., Kallitsis, M., Kumar, D., Lever, C., Ma, Z., Mason, J., Menscher, D., Seaman, C., Sullivan, N., Thomas, K. and Zhou, Y. *Understanding the Mirai botnet*, Proc. 26th USENIX Secur. Symp., (2017), 1093–1110.

[2] August, T. and Tunca, T.I. *Network software security and user incentives*, Manage. Sci., 52 (2006), 1703–1720.

[3] Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z. and Zanin, M. *The structure and dynamics of multilayer networks*, Phys. Rep., 544 (2024), 1–122.

[4] Chen, Z. and Ji, C. *Spatial-temporal modeling of malware propagation in networks*, IEEE Trans. Neural Netw., 16 (2005), 1291–1303.

[5] Chen, J. and Wang, L. *Adaptive cyber defense strategies using stochastic control theory*, IEEE Trans. Control Syst. Technol., 31 (2023), 789–802.

[6] Chopra, A. and Singh, P. *Patch management challenges in consumer IoT: A longitudinal study*, IEEE Internet Things J., 10 (2023), 4231–4245.

[7] Cisco Systems. *Cisco annual internet report (2018–2023)*, Cisco White Paper, (2023), 1–45.

[8] Dormand, J.R. and Prince, P.J. *A family of embedded Runge-Kutta formulae*, J. Comput. Appl. Math., 6 (1980), 19–26.

[9] Doucet, A. and Johansen, A.M. *A tutorial on particle filtering and smoothing*, Handb. Nonlinear Filter., 12 (2011), 656–704.

[10] Edwards, B. and Hofmeyr, S. *Hajime: Analysis of a decentralized Internet worm for IoT devices*, Proc. IEEE Secur. Privacy Workshops, (2022), 65–78.

[11] Evensen, G. *Data assimilation: The ensemble Kalman filter*, Springer, 2009.

[12] Kephart, J.O. and White, S.R. *Measuring and modeling computer virus prevalence*, Proc. IEEE Comput. Soc. Symp. Res. Secur. Privacy, (1993), 2–15.

[13] Kucharski, A.J., Funk, S. and Eggo, R.M. *The next-generation matrix approach in compartmental epidemic models*, J. Math. Biol., 86 (2022), 45.

[14] Li, S., Miao, L., Wu, X. and Liu, B. *A multigroup model for malware propagation in wireless sensor networks*, Proc. 2019 IEEE Int. Conf. Commun., (2019), 1–6.

[15] Li, F. and Paxson, V. *A large-scale empirical study of security patches*, Proc. ACM Conf. Comput. Commun. Secur., (2023), 1031–1046.

[16] Liu, X., Wang, Y., Zhang, H. and Chen, Y. *Multi-layer network modeling for mobile malware propagation in 5G IoT networks*, IEEE Trans. Mobile Comput., 22 (2023), 2156–2170.

[17] Newman, M.E.J. and Barabási, A.L. *Networks: An introduction*, 3rd ed., Oxford University Press, 2023.

[18] National Institute of Standards and Technology. *National vulnerability database (NVD) public data feeds*, NIST Tech. Rep., (2023).

[19] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. and Vespignani, A. *Epidemic processes in complex networks*, Rev. Mod. Phys., 95 (2023), 015001.

[20] Radicchi, F. and Bianconi, G. *Epidemic spreading in networks with heterogeneous transmission rates*, Phys. Rev. Lett., 129 (2022), 128301.

[21] Ridenhour, B., Kowalik, J.M. and Shay, D.K. *Unraveling R0: Considerations for public health applications*, Am. J. Public Health, 108 (2018), S445–S454.

[22] Rodriguez Garcia, R. and Herrero Cosio, A. *Stochastic simulation of IoT malware spread using individual-based SIR models*, Universidad de Burgos Preprint, (2024).

[23] Sun, Y., Gong, W. and Towsley, D. *Modeling malware spreading dynamics in modern networks*, IEEE Trans. Inf. Theory, 70 (2024), 1843–1859.

[24] Tajari Siahmarzkooh, A. *An improved K-means clustering feature selection and biogeography based optimization for intrusion detection*, Int. J. Web Res., 6 (2023), 57–66.

[25] Tsai, C.W., Lai, C.F. and Vasilakos, A.V. *Future Internet of Things: Open issues and challenges*, Wireless Netw., 29 (2023), 1789–1809.

[26] Wang, H., Liu, J. and Scoglio, C. *Generalized epidemic mean-field model for spreading processes over multilayer complex networks*, IEEE/ACM Trans. Netw., 32 (2024), 160–172.

[27] Xu, L., Zhang, Q. and Li, J. *An epidemiological model of virus spread and cleanup in IoT networks*, Proc. IEEE Int. Conf. Commun., (2023), 2345–2351.

[28] Zou, C.C., Gong, W. and Towsley, D. *Code Red worm propagation modeling and analysis revisited*, IEEE Trans. Inf. Theory, 68 (2022), 8023–8035.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1518–1549
https://doi.org/10.22067/ijnao.2026.97820.1844
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A collocation framework for numerical reconstruction of the source function in a two-term fractional diffusion problem

S.S. Jalalian, A. Babaei*, and M. Gachpazan

Abstract

This study develops a new spectral collocation strategy to reconstruct an unknown source term appearing in a two-term time-fractional diffusion model equipped with nonlocal boundary constraints. As obtaining an exact solution to such problems is generally intractable, a collocation strategy based on Legendre polynomials and Chebyshev nodes is employed to construct the numerical approximation. By projecting the fractional-order operators onto a bivariate polynomial basis, the original inverse problem is transformed into a system of linear algebraic equations. The solvability of the resulting system is investigated via the properties of the associated collocation matrix incorporating the fractional dynamics and nonlocal boundary

*Corresponding author

Received 13 February 2026; revised 26 May 2026; accepted 25 June 2026

Sahar Jalalian
Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Mazandaran, Babolsar, Iran. e-mail: s.jalalian01@umail.umz.ac.ir

Afshin Babaei
Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Mazandaran, Babolsar, Iran. e-mail: babaei@umz.ac.ir

Mortaza Gachpazan
Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P. O. Box 1159, Mashhad 91755, Iran. e-mail: gachpazan@um.ac.ir

How to cite this article
Jalalian, S.S., Babaei, A. and Gachpazan, M., A collocation framework for numerical reconstruction of the source function in a two-term fractional diffusion problem. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1518-1549. https://doi.org/10.22067/ijnao.2026.97820.1844

conditions. Moreover, convergence analysis and rigorous error estimates are provided. Numerical trials are presented to demonstrate the reliability, precision, and computational efficiency of the proposed procedure.

AMS subject classifications (2020): Primary 35R30; Secondary 65M32, 65M12.

Keywords: Fractional diffusion equation; Inverse source problem; Nonlocal boundary conditions; Error analysis.

1 Introduction

Recent developments in the analysis of inverse and ill-posed problems have increasingly focused on boundary conditions of a nonlocal nature. Rather than prescribing data solely at the boundary, these conditions utilize integral operators or global solution information, capturing situations in which physical observations along the boundary cannot be directly accessed. This generality, however, introduces significant challenges for both mathematical analysis and numerical computation [38].

The applicability of nonlocal boundary value problems spans a diverse range of scientific and engineering fields. These problems are crucial for modeling phenomena characterized by long-range interactions, memory effects, or integral constraints, where traditional local models prove insufficient. Key applications include thermal distributions in composite materials and cables, resistivity well-logging in geophysics, biological tissue dynamics, and medical imaging techniques such as ultrasound [32, 14, 36, 29, 27, 21]. Furthermore, these conditions arise naturally in models of multiphase flow, where the heat flux between phases depends on spatially and temporally varying parameters, thereby reflecting complex interfacial interactions [26, 43]. The advent of fractional-order differential equations, which inherently capture nonlocal temporal and spatial effects, has further underscored the importance of such problems [41, 27, 35].

Numerous computational techniques have been developed to solve these inverse problems, often combining rigorous mathematical theory with practical algorithms such as regularization and iterative projection methods [42, 9, 3]. Recent advances have also begun to incorporate data-driven and machine learning approaches to further enhance accuracy and efficiency [3].

A classical and extensively studied inverse problem is that of determining the heat source under nonlocal boundary conditions. For instance, models involving uncertain coefficients and time-dependent sources subject to integral over-determination, motivated by applications in mathematical biology, have been investigated in [33, 17]. Analytical approaches, including generalized

Fourier series expansions, have been employed to establish uniqueness, solvability, and stability results. Similarly, the field of fractional calculus has attracted considerable attention. Works such as [24, 1, 30, 16, 2] address time- and space-dependent source functions in fractional parabolic models, while studies such as [20, 18, 4] focus on heat equations with Wentzell-Neumann-type boundary conditions and integral constraints, proposing both theoretical and numerical solutions. Inverse problems for fractional diffusion and subdiffusion models involving Caputo or generalized fractional derivatives have also received considerable attention. Unique solvability under suitable over-determination conditions has been established in [25], while extensions to multi-term and sequential fractional derivatives have been analyzed in [22, 6, 28, 7, 15]. These studies address numerical stability and nonlinear cases through iterative and matrix-based techniques. Moreover, substantial research has been devoted to problems for equations featuring multiple fractional time derivatives, including fractional telegraph, diffusion, and wave equations [5, 8, 10, 11, 34, 39, 40, 44].

Recent studies have explored inverse source problems for diffusion equations involving two-term time-fractional derivatives with nonlocal boundary specifications. Notably, in [12] and [13], the existence, uniqueness, and well-posedness of such problems have been established using analytical approaches, such as the generalized Fourier method. While these works provide a solid theoretical foundation, their practical implementation often requires complex problem decomposition into multiple sub-problems and extensive analytical expansions, which can be computationally intensive for applied scenarios.

To address these limitations, the present work develops a more direct and computationally efficient numerical framework based on the shifted Legendre-Chebyshev spectral collocation method. Unlike the existing analytical splitting techniques, our approach offers a unified scheme that handles multi-term fractional derivatives and nonlocal constraints simultaneously without the need for problem decomposition. The primary innovations and contributions of this study are as follows: (i) the development of a straightforward collocation scheme that simplifies the numerical treatment of multi-term fractional inverse problems; (ii) the provision of a highly accurate, high-order reconstruction of the unknown source term with reduced overhead; and (iii) demonstrating the robustness and stability of the method even in the presence of noise in the measured data.

The paper is arranged as follows: Section 2 provides the formal statement of the inverse source problem. Section 3 reviews the required background material and definitions. The computational scheme is developed in Section 4. Section 5 contains the error and convergence analysis. Numerical results and discussions are presented in Section 6.

2 Description of the problem

The present work is devoted to an inverse source identification problem for a time-fractional diffusion equation involving two distinct fractional time derivatives, expressed as

$$D_t^{\alpha}u(x,t) + D_t^{\beta}u(x,t) = u_{xx}(x,t) + r(t)f(x), \qquad (x,t) \in \Omega, \tag{1}$$

for orders $0 < \alpha, \beta < 1$, subject to the initial and nonlocal boundary conditions:

$$u(x,0) = u_0(x), \qquad 0 \le x \le 1, \tag{2}$$

$$au(0,t) + bu(1,t) = 0, \quad 0 \le t \le 1, \tag{3}$$

$$u_x(0,t) - u_x(1,t) + cu(0,t) + du(1,t) = 0, \quad 0 \le t \le 1, \tag{4}$$

where $\Omega = (0,1) \times (0,1]$. In (1), $r(t)f(x)$ is the source function. The constants a, b, c, and d are real numbers with the assumption $a - b \neq 0$.

The operator D_t^{θ} denotes the Caputo fractional derivative defined as [23]

$$D_t^{\theta}u(t) = \frac{1}{\Gamma(1-\theta)} \int_0^t \frac{\partial u(x,s)}{\partial s} \frac{ds}{(t-s)^{\theta}}, \qquad 0 < \theta < 1,$$

where $0 < \theta < 1$ is the derivative order and $\Gamma(\cdot)$ is the Gamma function. For the direct problem associated with (1)–(4), the solution u is sought in Ω such that for each $t \in (0,1)$, $u(\cdot,t)$ belongs to the Sobolev space $H^2[0,1]$ and $D_t^{\alpha}u(x,\cdot), D_t^{\beta}u(x,\cdot)$ are in $C(0,1]$, given a continuous initial temperature u_0 and a continuous source function $r(t)f(x)$.

When the function $r(t)$ is unknown, an inverse problem is formulated to find the pair of functions $\{u(x,t), r(t)\}$ satisfying (1)–(4). To this end, we consider the supplementary integral condition:

$$\int_0^1 u(x,t)\,dx = E(t), \quad 0 \le t \le 1. \tag{5}$$

This form of supplementary condition is utilized when boundary data cannot be measured directly, yet an integral mean of the solution is accessible [19]. The well-posedness of this problem, including the existence, uniqueness, and stability of the solution, has been rigorously established in [12]. In this framework, we seek the pair (u,r) such that $u(\cdot,t) \in H^2[0,1]$ and $r(t) \in C[0,1]$. The use of Sobolev spaces is essential here to provide a rigorous framework and to justify the convergence rates of the numerical method discussed in Section 5. Its fractional derivatives $D_t^{\alpha}u, D_t^{\beta}u \in C(0,1]$. Furthermore, we impose $\int_0^1 f(x)dx \neq 0$ to ensure that the unknown $r(t)$ is uniquely and stably determined.

To solve the inverse problem (1)–(5), we first determine the function $r(t)$. This is achieved by integrating both sides of (1) over $[0,1]$, with respect to x, yielding

$$\int_0^1 D_t^\alpha u(x,t)\,dx + \int_0^1 D_t^\beta u(x,t)\,dx = \int_0^1 u_{xx}(x,t)\,dx + r(t)\int_0^1 f(x)\,dx.$$

Since the fractional derivative and the integration act on different variables, the operations are commutative. This leads to

$$D_t^\alpha \int_0^1 u(x,t)\,dx + D_t^\beta \int_0^1 u(x,t)\,dx = u_x(1,t) - u_x(0,t) + r(t)\int_0^1 f(x)\,dx. \tag{6}$$

We now incorporate information from the boundary conditions (3)–(4) and the supplementary condition (5), obtaining the following relations:

$$\begin{aligned} &u_x(1,t) - u_x(0,t) = cu(0,t) + du(1,t),\\ &D_t^\alpha \int_0^1 u(x,t)\,dx = D_t^\alpha E(t),\\ &D_t^\beta \int_0^1 u(x,t)\,dx = D_t^\beta E(t). \end{aligned}$$

Substituting these relations into (6) allows us to express $r(t)$ explicitly:

$$r(t) = \Big(\int_0^1 f(x)\,dx\Big)^{-1}\big(D_t^\alpha E(t) + D_t^\beta E(t) - (cu(0,t) + du(1,t))\big). \tag{7}$$

Substituting the expression for $r(t)$ from (7) back into (1) reformulates the original equation into

$$\begin{aligned} &D_t^\alpha u(x,t) + D_t^\beta u(x,t)\\ &=u_{xx}(x,t)\\ &\quad + f(x)\left[\Big(\int_0^1 f(x)\,dx\Big)^{-1}\big(D_t^\alpha E(t) + D_t^\beta E(t) - (cu(0,t) + du(1,t))\big)\right]. \end{aligned} \tag{8}$$

Consequently, instead of solving the inverse problem (1)–(5), we can now focus on solving the direct problem defined by (8) subject to the conditions (2)–(4), which is known to be well-posed and possesses a unique solution.

Remark 1. It should be noted that the theoretical well-posedness of this class of inverse fractional source problems, namely existence, uniqueness, and stability with respect to the input data, has been proved in [12]. Hence, the numerical method developed in this work relies on a sound analytical foundation.

3 Preliminaries

This section outlines the essential definitions and preliminary concepts employed in this work.

Definition 1. [31] The Legendre polynomials are defined by the recurrence relation:

$$L_{m+1}(t) = \frac{2m+1}{m+1} t L_m(t) - \frac{m}{m+1} L_{m-1}(t), \qquad m = 1, 2, \ldots,$$

where $L_0(t) = 1$ and $L_1(t) = t$. By applying the variable transformation $t = 2x-1$, and assuming $p_0(x) = 1$ and $p_1(x) = 2x - 1$, the shifted Legendre polynomials $p_m(x)$, $m = 1, 2, \ldots$, on the interval $[0, 1]$ are given by the recurrence relation:

$$p_{m+1}(x) = \frac{2m+1}{m+1}(2x-1)p_m(x) - \frac{m}{m+1} p_{m-1}(x).$$

These polynomials satisfy the following orthogonality property:

$$\int_0^1 p_m(x) p_n(x)\, dx = \begin{cases} \frac{1}{2n+1}, & m = n, \\ 0, & m \neq n. \end{cases}$$

The analytical expression for shifted Legendre polynomials is

$$p_n(x) = \sum_{j=0}^{n} (-1)^{n+j} \frac{(n+j)!}{(n-j)!(j!)^2} x^j,$$

Furthermore, $|p_n(x)| \leq 1$ for any $x \in [0, 1]$, and the bound for the m-th derivative is defined by

$$|p_n^{(m)}(x)| \leq m! \frac{(n+m)!}{(n-m)!(m!)^2}, \tag{9}$$

Additionally, these polynomials can be formulated via the Rodrigues formula on the interval $[0, 1]$ as follows:

$$p_m(x) = \frac{1}{m!} \frac{d^m}{dx^m}((x^2 - x)^m).$$

Let $L^2[0, 1]$ denote the Hilbert space of square-integrable functions on the unit interval. Any function $g \in L^2[0, 1]$ admits the orthogonal expansion $g(x) = \sum_{i=0}^{\infty} c_i p_i(x)$, with coefficients c_i obtained from

$$c_i = \frac{1}{\|p_i\|^2} \int_0^1 g(x) p_i(x)\, dx = (2i+1) \int_0^1 g(x) p_i(x)\, dx.$$

Definition 2. [37] Chebyshev polynomials form a sequence of orthogonal polynomials defined on the interval $[-1, 1]$ with respect to particular weight functions. They are classified into two

main types: Chebyshev polynomials of the first kind, denoted by $T_n(x)$, satisfying the identity

$$T_n(\cos\theta) = \cos(n\theta), \quad n = 0, 1, 2, \ldots,$$

whereas those of the second kind, denoted by $U_n(x)$, are characterized by

$$U_n(\cos\theta)\sin\theta = \sin((n+1)\theta), \quad n = 0, 1, 2, \ldots.$$

These polynomials are crucial in approximation theory for minimizing interpolation errors and are widely used in numerical methods.

In this paper, we will utilize the roots of the shifted Chebyshev polynomials of the second kind as collocation points to enhance the accuracy of our method. The shifted roots on $[0, 1]$ are given by

$$t_k = \frac{1}{2}\cos(\frac{k\pi}{n+1}) + \frac{1}{2}, \quad k = 1, \ldots, n.$$

4 Numerical Procedure

This section outlines a numerical scheme for addressing problem (8), subject to the conditions (2)–(4). Building upon the findings from the preceding subsection, we proceed to solve (8). For this purpose, we assume an approximate solution of the form:

$$u(x,t) \simeq \tilde{u}_{M,N}(x,t) = \sum_{i=0}^{N}\sum_{j=0}^{M} c_{ij}p_i(x)p_j(t) = P(x)^T C P(t), \tag{10}$$

where $P(x) = [p_0(x), \ldots, p_i(x), \ldots, p_N(x)]^T, P(t) = [p_0(t), \ldots, p_j(t), \ldots, p_M(t)]^T$ are vectors of the shifted Legendre polynomials. The unknown coefficients c_{ij} form an $(N+1)\times(M+1)$ matrix C that must be determined through the numerical scheme:

$$C = \begin{bmatrix} c_{0,0} & \cdots & c_{0,M} \\ \vdots & \ddots & \vdots \\ c_{N,0} & \cdots & c_{N,M} \end{bmatrix}.$$

Using conditions (2)-(4) and the governing equation (8), we derive the following relation:

$$\begin{aligned} &P(x)^T C D_t^{\alpha} P(t) + P(x)^T C D_t^{\beta} P(t) \\ &= P_{xx}^T(x) C P(t) \end{aligned} \tag{11}$$

$$+\frac{f(x)}{\int_0^1 f(x)dx}\left[(D_t^\alpha E(t)+D_t^\beta E(t)-(cP(0)^TCP(t)+dP(1)^TCP(t)))\right],$$

where $D_t^\alpha P(t)$ and $D_t^\beta P(t)$ denote the Caputo fractional derivatives of the vector $P(t)$. These derivatives are defined element-wise, using the formula $D_t^s(t^k)=\frac{\Gamma(k+1)}{\Gamma(k+1-s)}t^{k-s}, k\geq 1$ and $0<s<1$. The vector $P_{xx}(x)$ is defined as $P_{xx}=\frac{d^2}{dx^2}P(x)=[\frac{d^2}{dx^2}p_0(x),\ldots,\frac{d^2}{dx^2}p_i(x),\ldots,\frac{d^2}{dx^2}p_N(x)]^T$. The initial and boundary conditions become

$$P(x)^TCP(0)=u_0(x), \tag{12}$$

$$aP(0)^TCP(t)+bP(1)^TCP(t)=0, \tag{13}$$

$$P_x(0)^TCP(t)-P_x(1)^TCP(t)+cP(0)^TCP(t)+dP(1)^TCP(t)=0, \tag{14}$$

for $0<x<1$ and $0\leq t\leq 1$, where the boundary vectors are defined as follows:

$$\begin{aligned}
P_x(0)&=\left[\frac{d}{dx}p_0(0),\ldots,\frac{d}{dx}p_i(0),\ldots,\frac{d}{dx}p_N(0)\right]^T,\\
P_x(1)&=\left[\frac{d}{dx}p_0(1),\ldots,\frac{d}{dx}p_i(1),\ldots,\frac{d}{dx}p_N(1)\right]^T,\\
P(0)&=[p_0(0),\ldots,p_N(0)],\\
P(1)&=[p_0(1),\ldots,p_N(1)].
\end{aligned}$$

Let the spatial collocation nodes x_i, for $i=0,1,\ldots,N$, be chosen such that $x_0=0$, $x_N=1$, and $\{x_i : i=1,2,\ldots,N-1\}$ are the roots of the second-kind shifted Chebyshev polynomial $U_{N-1}(x)$. The temporal nodes $\{t_j : j=1,2,\ldots,M\}$ are chosen as the roots of $U_M(t)$. Collocating (11) at the $M\times(N-1)$ points (x_i,t_j) yields

$$\begin{aligned}
&P(x_i)^TCD_t^\alpha P_(t_j)+P(x_i)^TCD_t^\beta P(t_j)\\
&=P_{xx}(x_i)^TCP(t_j)\\
&+\frac{f(x_i)}{(\int_0^1 f(x)dx)}\left[\left(D_t^\alpha E(t_j)+D_t^\beta E(t_j)-(cP(0)^TCP(t_j)+dP(1)^TCP(t_j))\right)\right].
\end{aligned} \tag{15}$$

The initial condition (12) is collocated at the spatial nodes $x_i,\ i=0,1,\ldots,N$, as

$$P(x_i)CP(t=0)=u_0(x_i). \tag{16}$$

Also, the boundary conditions (13)–(14) are collocated at the temporal nodes $t_j,\ j=1,\ldots,M$,

$$aP(0)^T CP(t_j) + bP(1)CP(t_j) = 0, \tag{17}$$

$$P_x(0)^T CP(t_j) - P_x(1)^T CP(t_j) + cP(0)^T CP(t_j) + dP(1)^T CP(t_j) = 0. \tag{18}$$

To streamline the discrete formulation, we introduce the following auxiliary matrices and vectors:

Spatial matrices

(each of size $(N-1) \times (N+1)$)

$$A_{11} = \Big[P(x_1) \cdots P(x_{N-1})\Big]^T,$$

$$A_{12} = \Big[P_{xx}(x_1) \cdots P_{xx}(x_{N-1})\Big]^T.$$

Temporal matrices

(each of size $(M+1) \times M$)

$$A_{13} = \big[D_t^\alpha P(t_1) \ \cdots \ D_t^\alpha P(t_M)\big],$$

$$A_{14} = \big[D_t^\beta P(t_1) \ \cdots \ D_t^\beta P(t_M)\big],$$

$$A_{15} = \big[P(t_1) \ \cdots \ P(t_M)\big].$$

Forcing vectors

$$F = \left(\int_0^1 f(x)\,dx\right)^{-1} \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_{N-1}) \end{bmatrix} \quad \text{(size } (N-1) \times 1\text{)},$$

$$A_{16} = \begin{bmatrix} D_t^\alpha E(t_1) \\ \vdots \\ D_t^\alpha E(t_M) \end{bmatrix}, \qquad A_{17} = \begin{bmatrix} D_t^\beta E(t_1) \\ \vdots \\ D_t^\beta E(t_M) \end{bmatrix} \quad \text{(each of size } M \times 1\text{)}.$$

Using Kronecker product properties, the system (15) can be expressed compactly as follows:

$$\Lambda_1 X = F_1, \tag{19}$$

where $X = [c_{00}, c_{10}, \ldots, c_{N0}, \ldots, c_{0M}, c_{1M}, \ldots, c_N M]^T$ is the vector of unknown coefficients, $\Lambda_1 = A_1 + A_2$, and

$$A_1 = A_{13}^T \otimes A_{11} + A_{14}^T \otimes A_{11} - A_{15}^T \otimes A_{12},$$
$$A_2 = F(c\, A_{15}^T \otimes P(0)^T + d\, A_{15}^T \otimes P(1)^T).$$

Also, $F_1 = F \otimes (A_{16} + A_{17})$. Similarly, the conditions (16)–(18) can be written as follows:

$$\begin{aligned} \Lambda_2 X &= F_2, \\ \Lambda_3 X &= 0, \\ \Lambda_4 X &= 0, \end{aligned} \tag{20}$$

where

$$\begin{aligned} \Lambda_2 &= P(0)^T \otimes A_3, \\ \Lambda_3 &= aA_4^T \otimes P(0)^T + bA_4^T \otimes P(1)^T, \\ \Lambda_4 &= A_4^T \otimes P_x(0)^T - A_4^T \otimes P_x(1)^T + cA_4^T \otimes P(0)^T + dA_4^T \otimes P(1)^T, \\ A_3 &= [P(x_1) \cdots P(x_{N-1})]^T, \\ A_4 &= [P(t_0) \cdots P(t_j) \cdots P(t_M)]^T, \end{aligned}$$

and $F_2 = [U_0(x_1) \cdots U_0(x_{N-1})]^T$. Combining (19) and (20) results in a linear system of $(N+1)(M+1)$ equations as follows:

$$AX = F, \tag{21}$$

where $A = [\Lambda_1^T, \Lambda_2^T, \Lambda_3^T, \Lambda_4^T]^T$ and $F = [F_1^T, F_2^T, 0, 0]^T$. Solving this linear system (21) for the vector X determines the coefficients c_{ij}. Substituting these coefficients into (10) yields the approximate numerical solution $\tilde{u}_{M,N}(x,t)$.

The transformation of the inverse problem into the algebraic system (21) is theoretically justified through the spectral projection of the fractional-order operators onto the space of shifted Legendre polynomials. Since the basis functions $\{p_i(t)\}_{i=0}^N$ are linearly independent, the resulting discrete formulation provides a valid finite-dimensional representation of the continuous problem. The use of Chebyshev nodes as collocation points yields a stable distribution of nodes, which improves the conditioning of the resulting matrix A and helps mitigate numerical instabilities, including the Runge phenomenon, by increasing node density near the boundaries. The existence of a unique numerical solution X is ensured whenever the collocation matrix A is nonsingular.

5 Error analysis

In this section, we study the convergence of the numerical solution (21) and discuss the stability features of the proposed method.

To investigate the convergence of the proposed approach, we require additional information about the properties of Legendre polynomials, which can be expressed as lemmas and theorems.

Lemma 1. Let $i, j \in \mathbb{N}$, and let $i > j$. Then, for any constant $\theta \geq 1$, the following inequality holds:

$$\frac{i!}{(2i)!} < \theta \frac{j!}{(2j)!}. \tag{22}$$

Proof. Consider the ratio of successive terms. For any positive integer k, we have

$$\frac{\frac{(k+1)!}{(2(k+1))!}}{\frac{k!}{(2k)!}} = \frac{(k+1)!}{k!}.\frac{(2k)!}{(2k+2)!} = (k+1).\frac{1}{2(k+1)(2k+1))} = \frac{1}{2(2k+1)}.$$

Since $k \geq 1$, it follows that $\frac{1}{2(2k+1)} \leq \frac{1}{6} < 1$. Therefore, the sequence $\{\frac{k!}{(2k)!}\}_{k=1}^{\infty}$ is strictly decreasing. Consequently, for any $i > j$, we have

$$\frac{i!}{(2i)!} < \frac{j!}{(2j)!}.$$

□

Theorem 1. Let $f \in C[0,1]$ be such that its derivatives of all orders $i \geq 3$ exist and are uniformly bounded on $[0,1]$, that is,

$$|f^{(i)}(t)| \leq L \quad \text{for all } t \in [0,1], \quad \text{for all } i \geq 3, \tag{23}$$

where $L > 0$ is a constant independent of i and t. Suppose that $f(t)$ is expanded in the series of shifted Legendre polynomials as

$$f(t) = \sum_{i=0}^{\infty} c_i p_i(t),$$

then, for all $i > 3$, the coefficients satisfy

$$|c_i| \leq L \frac{i!}{(2i)!}. \tag{24}$$

Proof. The proof relies on the Rodrigues formula for shifted Legendre polynomials on $[0,1]$, namely

$$p_i(t) = \frac{1}{i!}\frac{d^i}{dt^i}\Big[t^i(1-t)^i\Big], \quad i = 0, 1, 2, \ldots. \tag{25}$$

A classical result for (shifted) Legendre expansions states that the coefficient c_i in the series

$$f(t) = \sum_{i=0}^{\infty} c_i p_i(t)$$

can be expressed directly in terms of the i-th derivative of f at a specific point. This expression is obtained by repeated integration by parts applied to the orthogonality relation, yielding the identity

$$c_i = (-1)^i \frac{2i+1}{i!} \int_0^1 f^{(i)}(t)(t(1-t))^i \, dt.$$

This identity leverages the fact that the i-th derivative of the shifted Legendre polynomial $p_i(t)$ is a constant, specifically $\frac{(2i)!}{i!}$, and the Rodrigues formula. Taking the absolute value and using the integral inequality, we get

$$|c_i| = |\frac{2i+1}{i!} \int_0^1 f^{(i)}(t)(t(1-t))^i \, dt| \leq \frac{2i+1}{i!} \max_{t\in[0,1]} |f^{(i)}| \int_0^1 (t(1-t))^i \, dt.$$

The integral $\int_0^1 (t(1-t))^i \, dt$ is the Beta integral, which evaluates to

$$\int_0^1 (t(1-t))^i \, dt = \beta(i+1, i+1) = \frac{(i!)^2}{(2i+1)!}.$$

Substituting this back into the inequality yields

$$\begin{aligned} |c_i| &\leq \frac{2i+1}{i!} \max_{t\in[0,1]} |f^{(i)}(t)| \frac{(i!)^2}{(2i+1)!} \\ &= \max_{t\in[0,1]} |f^{(i)}(t)| \frac{(2i+1)(i!)^2}{i!(2i+1)!} \\ &= \max_{t\in[0,1]} |f^{(i)}(t)| \frac{i!}{(2i)!}. \end{aligned}$$

So, we have

$$|c_i| \leq \max_{t\in[0,1]} |f^{(i)}(t)| \frac{i!}{(2i)!}.$$

For $i \geq 3$, under the given smoothness conditions, $\max_{t\in[0,1]} |f^{(i)}(t)|$ can be bounded by a constant L. For sufficiently large i, this implies the desired bound. Hence, for all $i > 3$, we obtain the desired bound:

$$|c_i| \leq L \frac{i!}{(2i)!}$$

for all $i > 3$, which completes the proof. According to Lemma 1, this bound decreases rapidly as i increases, ensuring the convergence of the approximate solution. □

It is worth noting that while the estimate in (22) exhibits factorial decay —reflecting the behavior for analytic or highly smooth functions— this serves as an upper bound for the ideal case. For solutions with limited regularity (e.g., in Sobolev spaces H^k), this bound naturally transitions to a standard algebraic rate, $\mathcal{O}(N^{-k})$. However, to maintain the consistency of the theoretical framework presented in this section, we present the estimates in their strongest form.

Theorem 2. Let $f_N(t) = \sum_{i=0}^{N} c_i p_i(t)$ be an approximation of $f(t)$. Then

$$|f(t) - f_N(t)| \leq \frac{L}{(N+1)!}. \tag{26}$$

Proof. We know from standard approximation theory that

$$|f(t) - f_N(t)| \leq \frac{1}{(N+1)!} \max_{0 \leq t \leq 1} f^{(N+1)}(t)$$

It can be shown that for the specific basis functions $p_i(t)$,

$$f^{(N+1)}(t) = c_{N+1} \left(\frac{(2N+2)!}{((N+1)!)^2} (N+1)! \right) = \frac{(2N+2)!}{(N+1)!} c_{N+1}.$$

Applying (24), c_{N+1} can be bounded as $c_{N+1} \leq \frac{(N+1)!}{(2N+2)!}$, therefore

$$|f(t) - f_N(t)| \leq L \frac{1}{(N+1)!}.$$

□

Now, consider a bivariate function

$$f(x,t) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{ij} p_i(x) p_j(t).$$

Suppose

$$\| \frac{\partial^{i+j} f(x,t)}{\partial x^i \partial t^j} \|_2 \leq \eta,$$

for a positive constant η. Then, the following inequality holds for the expansion coefficients $c_{i,j}$:

$$|c_{i,j}| < \eta \frac{i!\, j!}{(2i)!\, (2j)!}, \tag{27}$$

for all $i, j > 3$. Furthermore, if $\tilde{f}(x,t) = \sum_{i=0}^{N}\sum_{j=0}^{M} c_{i,j}p_i(x)p_j(t)$ is an approximation of $f(x,t)$, then we have

$$|f(x,t) - \tilde{f}(x,t)| < \frac{\eta}{(N+1)!(M+1)!}. \tag{28}$$

The following lemma confirms the validity of inequalities essential for deriving this error estimate.

Lemma 2. Let $f(x,t)$ and its approximation $\tilde{f}(x,t)$ satisfy the assumptions of Theorem 2 and its consequence (28). Then there exists a positive constant η, independent of the discretization parameters, such that the following error estimates hold:

(i)
$$\left|\frac{\partial f(x,t)}{\partial x} - \frac{\partial \tilde{f}(x,t)}{\partial x}\right| < \frac{\eta}{2N^{N-1}M^{M+1}},$$

(ii)
$$\left|\frac{\partial^2 f(x,t)}{\partial x^2} - \frac{\partial^2 \tilde{f}(x,t)}{\partial x^2}\right| < \frac{\eta}{4N^{N-2}M^{M+1}},$$

(iii)
$$\left|D_t^{\alpha} f(x,t) - D_t^{\alpha}\tilde{f}(x,t)\right| < \frac{\eta}{N^{N-1}M^{M+1}}.$$

Proof. (i) From the definition of $f(x,t)$ and $\tilde{f}(x,t)$, we have

$$\left|\frac{\partial f(x,t)}{\partial x} - \frac{\partial \tilde{f}(x,t)}{\partial x}\right| = |\sum_{i=N+1}^{\infty}\sum_{j=M+1}^{\infty} c_{i,j}p_i'(x)p_j(t)|.$$

This can be bounded by

$$\left|\frac{\partial f(x,t)}{\partial x} - \frac{\partial \tilde{f}(x,t)}{\partial x}\right| \leq \sum_{i=N+1}^{\infty}\sum_{j=M+1}^{\infty} |c_{i,j}||p_i'(x)||p_j(t)|.$$

Given that $|p_j(t)| \leq 1$ on $[0,1]$ and applying the coefficient bound from (27), this inequality becomes

$$\sum_{i=N+1}^{\infty}\sum_{j=M+1}^{\infty} |c_{i,j}||p_i'(x)||p_j(t)| \leq \sum_{i=N+1}^{\infty}\sum_{j=M+1}^{\infty} (\eta\frac{i!\,j!}{(2i)!(2j)!})|p_i'(x)|.$$

From property (9), we use the bound $|p_i'(x)| \leq \frac{(i+1)!}{(i-1)!}$ for $x \in [0,1]$. Substituting this yields

$$|\frac{\partial f(x,t)}{\partial x} - \frac{\partial \tilde{f}(x,t)}{\partial x}| \leq \sum_{i=N+1}^{\infty}\sum_{j=M+1}^{\infty} (\frac{\eta i!j!}{(2i)!(2j)!})(\frac{(i+1)!}{(i-1)!}).$$

Therefore, considering the above relationship and the properties of geometric series, we can write

$$\begin{aligned}
&\eta \sum_{i=N+1}^{\infty} (\frac{i!}{(2i)!})(\frac{(i+1)!}{(i-1)!}) \sum_{j=M+1}^{\infty} (\frac{j!}{(2j)!}) \\
&\leq \frac{\eta}{2}(\sum_{i=N+1}^{\infty} \frac{1}{i^{i-2}})(\sum_{j=M+1}^{\infty} \frac{1}{j^j}) \\
&\leq (\frac{\eta}{2})(\sum_{i=N+1}^{\infty} (\frac{1}{N+1})^{i-2})(\sum_{j=M+1}^{\infty} (\frac{1}{M+1})^j) \\
&= \frac{\eta}{2N(N+1)^(N-2)M(M+1)^M} \leq \frac{\eta}{2N^{N-1}M^{M+1}}.
\end{aligned}$$

So, we obtain the final bound:

$$\left|\frac{\partial f(x,t)}{\partial x} - \frac{\partial \tilde{f}(x,t)}{\partial x}\right| < \frac{\eta}{2N^{N-1}M^{M+1}}.$$

(ii) The proof for the second derivative follows similarly, using the bound $|p_i''(x)| \leq \frac{(i+2)!}{2(i-2)!} \leq \frac{(i+2)!}{(i-2)!}$ and the same coefficient bound (27). This leads to

$$\left|\frac{\partial^2 f(x,t)}{\partial x^2} - \frac{\partial^2 \tilde{f}(x,t)}{\partial x^2}\right| \leq \frac{\eta}{4N^{N-2}M^{M+1}}.$$

(iii) For the Caputo fractional derivative, we use the linearity property and the fact that $D_t^\alpha(t^k) = \frac{\Gamma(k+1)}{\Gamma(k+1-\alpha)}t^{k-\alpha}$. The error bound is derived by applying the same coefficient bound and the properties of Gamma function, yielding

$$\begin{aligned}
\left|D_t^\alpha f(x,t) - D_t^\alpha \tilde{f}(x,t)\right| &= |\frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha}(\frac{d}{ds}(f(x,t) - \tilde{f}(x,t)))ds| \\
&\leq \frac{1}{|\Gamma(1-\alpha)|} \int_0^1 |(1-s)|^{-\alpha} |\frac{\partial}{\partial s}((f(x,s) - \tilde{f}(x,s))|\, ds.
\end{aligned}$$

Applying the bound from part (i) to the derivative, which is independent of s, we get

$$\left|D_t^\alpha f(x,t) - D_t^\alpha \tilde{f}(x,t)\right| \leq (\frac{1}{|\Gamma(1-\alpha)|})(\frac{\eta}{2N^{N-1}M^{M+1}}) \int_0^t (1-s)^{-\alpha}\, ds.$$

Evaluating the integral yields

$$\left|D_t^\alpha f(x,t) - D_t^\alpha \tilde{f}(x,t)\right| \leq (\frac{\eta}{|\Gamma(2-\alpha)|2N^{N-1}M^{M+1}}).$$

Since $|\Gamma(2-\alpha)|$ is a positive constant for $\alpha \in (0,1)$, it can be absorbed into the constant η, yielding the final result:

$$\left|D_t^{\alpha} f(x,t) - D_t^{\alpha} \tilde{f}(x,t)\right| \leq \frac{\eta}{N^{N-1} M^{M+1}}.$$

Thus, the proof of the lemma is complete. □

Theorem 3. Let $u(x,t)$ be the exact solution of (1) and let $\tilde{u}(x,t)$ be the approximate solution produced by the present scheme. Assuming the conditions of Lemma 2 are fulfilled, the global truncation error admits the following uniform estimate:

$$\begin{aligned}&\sup_{(x,t)\in\Omega} |u(x,t) - \tilde{u}(x,t)| \\ &< \frac{\eta}{2^{N+M}(N+1)!(M+1)!} + (\|A\|_2)^{-1}(\sqrt{NM})\left(\frac{3\eta}{N^{N-2}M^{M+1}} + \theta\frac{\vartheta(c+d)}{(M+1)!}\right).\end{aligned}$$

Proof. Let $g(x,t)$ denote the bivariate interpolating polynomial of the exact solution $u(x,t)$ constructed at the nodal points (x_i, t_j). Here x_i $(i = 0, 1, \ldots, N)$ and $t_j (j = 0, 1, 2, \ldots, M)$ correspond to the roots of the shifted second-kind Chebyshev polynomial within the domain Ω. Applying the triangle inequality yields:

$$|u(x,t) - \tilde{u}(x,t)| \leq |u(x,t) - g(x,t)| + |g(x,t) - \tilde{u}(x,t)|.$$

From (28), the interpolation error is bounded by

$$|u(x,t) - g(x,t)| \leq \frac{\eta}{2^{N+M}(N+1)!(M+1)!}. \tag{29}$$

We now estimate the second term $|\tilde{u} - g|$. Consider the L^2 norm:

$$\begin{aligned}&\|g(x,t) - \tilde{u}(x,t)\| \\ &= \left(\int_0^1 \int_0^1 \left(\sum_{i=0}^{N}\sum_{j=0}^{M}(c_{i,j}p_i(x)p_j(t) - \tilde{c}_{i,j}p_i(x)p_j(t))\right)^2 dxdt\right)^{\frac{1}{2}}.\end{aligned}$$

An application of the Cauchy–Schwarz inequality gives

$$\begin{aligned}&\|g(x,t) - \tilde{u}(x,t)\| \\ &\leq \left(\sum_{i=0}^{N}\sum_{j=0}^{M}|c_{i,j} - \tilde{c}_{i,j}|^2\right)^{\frac{1}{2}}\left(\int_0^1\int_0^1\left(\sum_{i=0}^{N}\sum_{j=0}^{M}|p_i(x)p_j(t)|\right)^2 dxdt\right)^{\frac{1}{2}}.\end{aligned}$$

A simpler bound can be derived by leveraging the orthonormality of the Legendre polynomials (1). Assuming that $\{p_i\}$ are orthonormal gives, results

$$\int_0^1 p_i(x)^2\,dx \le 1.$$

Therefore,

$$\|g(x,t)-\tilde{u}(x,t)\| \le (\sum_{i=0}^{N}\sum_{j=0}^{M}|c_{i,j}-\tilde{c}_{i,j}|^2)^{\frac{1}{2}}\sqrt{NM}.$$

We now find an upper bound for $\sum_{i=0}^{N}\sum_{j=0}^{M}|c_{i,j}-\tilde{c_{i,j}}|$. Let

$$C=[c_{0,0}\cdots c_{N,0}\cdots c_{0,M}\cdots c_{N,M}]^T,$$

and

$$\tilde{C}=[\tilde{c}_{0,0}\cdots \tilde{c}_{N,0}\cdots \tilde{c}_{0,M}\cdots \tilde{c}_{N,M}]^T,$$

where the coefficients $\tilde{c_{i,j}}$ are defined by the orthogonal projection:

$$\tilde{c}_{i,j}=(\frac{1}{\|p_i\|^2\|p_j\|^2}\int_0^1\int_0^1 u(x,t)p_i(x)p_j(t)\,dx\,dt).$$

From the system of equations (21), we obtain

$$C-\tilde{C}=A^{-1}(F-A\tilde{C}),$$

where F is the vector of right-hand side values. Taking norms gives

$$\|C-\tilde{C}\|_2 \le (\|A^{-1}\|_2)\|F-A\tilde{C}\|_2 \le (\|A^{-1}\|_2)\|F-A\tilde{C}\|_\infty.$$

Let $R_{i,j}=(F-A\tilde{C})_{i,j}$ for $i=0,\ldots,N$ and $j=0,\ldots,M$. Then,

$$\begin{aligned}
&|R_{i,j}|\\
&=|(D_t^\alpha g(x_i,t_j)+D_t^\beta g(x_i,t_j)-g_{xx}(x_i,t_j)\\
&\quad+\frac{f(x_i)}{\int_0^1 f(x)dx}(cg(0,t_j)+dg(1,t_j)))-(D_t^\alpha\tilde{u}(x_i,t_j)+D_t^\beta\tilde{u}(x_i,t_j)-\tilde{u}_{xx}(x_i,t_j)\\
&\quad+\frac{f(x_i)}{\int_0^1 f(x)dx}(c\tilde{u}(0,t_j)+d\tilde{u}(1,t_j)))|\\
&\le|D_t^\alpha g(x_i,t_j)-D_t^\alpha\tilde{u}(x_i,t_j)|+|D_t^\beta g(x_i,t_j)-D_t^\beta\tilde{u}(x_i,t_j)|\\
&\quad+|g_{xx}(x_i,t_j)-\tilde{u}_{xx}(x_i,t_j)|\\
&\quad+|\frac{f(x_i)}{\int_0^1 f(x)dx}|\,(|c(g(0,t_j)-\tilde{u}(0,t_j))|+|d(g(1,t_j)-\tilde{u}(1,t_j))|)\,.
\end{aligned}$$

Let

$$\vartheta := \max\Big\{\Big|\frac{f(x_i)}{\int_0^1 f(x)\,dx}\Big| : i = 0, 1, \ldots, N\Big\}. \tag{30}$$

Applying Lemma 2 to the error $e(x,t) := u(x,t) - \tilde{u}(x,t)$ and bounding its spatial and temporal derivatives, we obtain estimates of the form

$$\begin{aligned} |R_{i,j}| &\leq \frac{\eta}{N^{N-1}M^{M+1}} + \frac{\eta}{N^{N-1}M^{M+1}} + \frac{\eta}{4N^{N-2}M^{M+1}} + \theta\frac{\vartheta(c+d)}{(M+1)!} \\ &\leq \frac{3\eta}{N^{N-2}M^{M+1}} + \theta\frac{\vartheta(c+d)}{(M+1)!}. \end{aligned}$$

As a result,

$$\|g(x,t) - \tilde{u}(x,t)\| \leq \|A\|_2^{-1}(\sqrt{NM})(\frac{3\eta}{N^{N-2}M^{M+1}} + \theta\frac{\vartheta(c+d)}{(M+1)!}). \tag{31}$$

Finally, combining the estimates (29) and (31), we arrive at the desired result:

$$\begin{aligned} &\sup_{(x,t)\in\Omega} |u(x,t) - \tilde{u}(x,t)| \\ &< \frac{\eta}{2^{N+M}(N+1)!(M+1)!} + (\|A^{-1}\|_2)(\sqrt{NM})(\frac{3\eta}{N^{N-2}M^{M+1}} + \theta\frac{\vartheta(c+d)}{(M+1)!}). \end{aligned}$$

□

6 Numerical examples

In this section, we present the numerical results for two benchmark examples to demonstrate the accuracy, efficiency, and robustness of the proposed shifted-Legendre spectral collocation method. To evaluate the performance of the method, the maximum absolute errors for the state variable $u(x,t)$ and the source term $r(t)$ are defined as follows:

$$\begin{aligned} e_u &= \max_{i,j} |u(x_i,t_j) - u_{N,M}(x_i,t_j)|, \\ e_r &= \max_{j} |r(t_j) - r_{N,M}(t_j)|. \end{aligned}$$

To further investigate the stability and practical applicability of the algorithm, we consider the presence of noise in both the measured data and the model parameters. Furthermore, to assess the convergence behavior of the proposed method, the computational convergence order is determined using the following formula:

$$order \simeq \frac{\log(\frac{e_{N+1}}{e_N})}{\log(\frac{h_{N+1}}{h_N})},$$

where e_{N+1} and e_N denote the maximum absolute errors at two consecutive steps.

In this study, the collocation points are chosen as Chebyshev nodes, defined by

$$x_i = \frac{1}{2}\cos(\frac{(i\pi)}{(N+1)}) + \frac{1}{2}, \quad i = 1, 2, \ldots, N.$$

Since the distance between adjacent points,

$$h_i = x_{i+1} - x_i$$

is nonuniform, we utilize their average value as the effective step length, h_N, for each N. This approach provides a consistent metric for evaluating the rate of convergence across the nonuniform mesh.

Once the numerical solution $\tilde{u}(x,t)$ has been obtained by the proposed method, the approximate source function $\tilde{r}(t)$ is recovered by substituting $\tilde{u}(x,t)$ into (7).

Example 1. Consider the following inverse problem:

$$D_t^{\alpha}u(x,t) + D_t^{\beta}u(x,t) = u_{xx}(x,t) + r(t)\sin(\frac{\pi x}{2} + \frac{\pi}{3}),$$

subject to the initial condition $u(x,0) = 0, \quad x \in [0,1]$, and the boundary conditions:

$$\begin{aligned} &\tfrac{2}{\sqrt{3}}u(0,t) - 2u(1,t) = 0,\\ &u_x(0,t) - u_x(1,t) - \tfrac{\pi}{2}u(0,t) - \tfrac{\pi}{2}u(1,t) = 0. \end{aligned}$$

The over-determination data is given by $E(t) = (\frac{\sqrt{3}+1}{\pi})t^3$.

The analytical solution for this problem is

$$u(x,t) = t^3\sin\Big(\frac{\pi x}{2} + \frac{\pi}{3}\Big),$$

and the corresponding exact source function is

$$r(t) = \frac{6}{\Gamma(4-\alpha)}t^{3-\alpha} + \frac{6}{\Gamma(4-\beta)}t^{3-\beta} + \frac{\pi^2}{4}t^3.$$

Table 1 reports the absolute errors and CPU-times for various values of α, β, M, and N using exact data. As shown, the method achieves high precision and efficiency in the absence of noise.

Table 1: Numerical results in Example 1 for different values of α, β, M, and N.

α and β	M and N	CPU-time (sec)	e_u	e_r
$\alpha = 0.9, \beta = 0.2$	$M = 3, N = 5$	1.684024	0.0106	0.0103
	$M = 5, N = 8$	4.694565	6.8791×10^{-6}	7.0042×10^{-6}
	$M = 8, N = 8$	7.540564	6.8788×10^{-6}	7.0033×10^{-6}
$\alpha = 0.5, \beta = 0.8$	$M = 4, N = 5$	2.151800	0.0100	0.0099
	$M = 7, N = 7$	5.527143	3.5097×10^{-5}	1.6260×10^{-4}
	$M = 8, N = 10$	9.783338	4.9522×10^{-8}	5.1959×10^{-8}
$\alpha = 0.4, \beta = 0.6$	$M = 5, N = 7$	4.176156	2.0222×10^{-4}	1.9738×10^{-4}
	$M = 6, N = 9$	6.307603	1.8508×10^{-6}	1.8367×10^{-6}
	$M = 10, N = 12$	16.917430	2.4822×10^{-8}	1.6729×10^{-8}
$\alpha = \beta = 0.5$	$M = N = 10$	13.028037	1.0165×10^{-7}	1.1360×10^{-7}
$\alpha = \beta = 0.9$	$M = N = 10$	12.820756	1.0497×10^{-7}	3.1563×10^{-7}
$\alpha = \beta = 0.3$	$M = N = 10$	12.929664	8.6727×10^{-8}	8.3159×10^{-8}

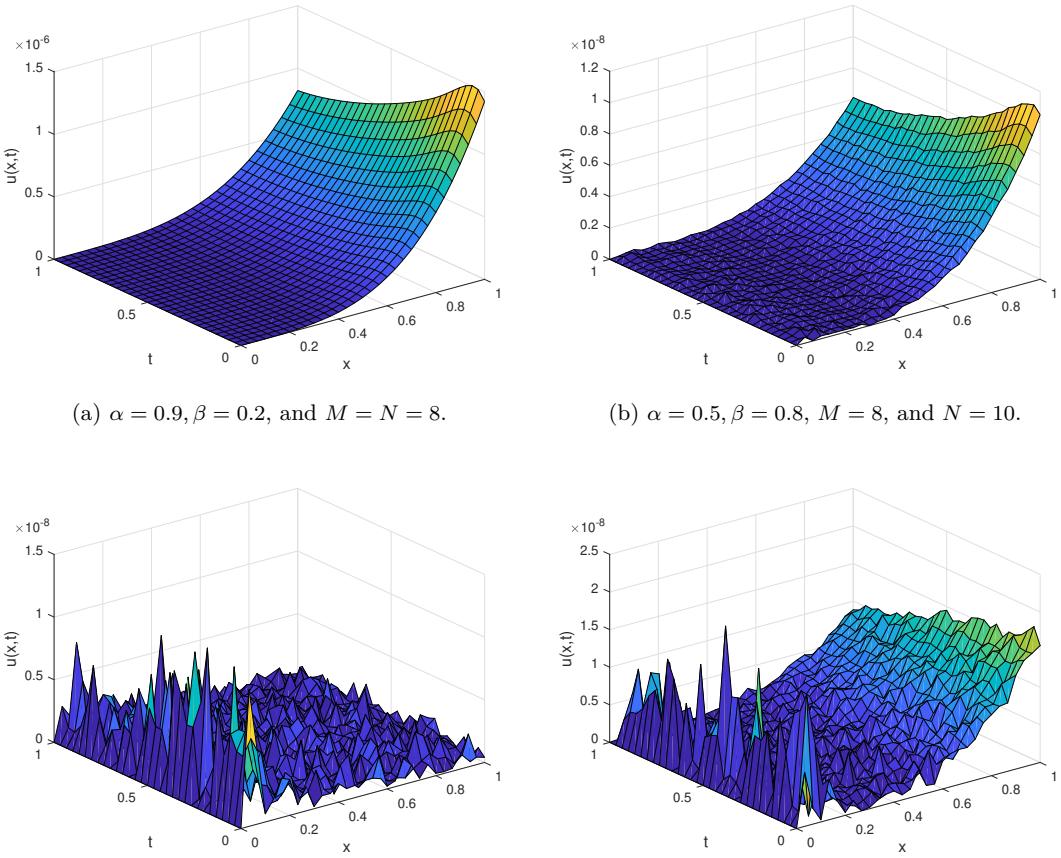
Figure 1 illustrates the absolute error of $u(x,t)$ for various fractional orders. To evaluate the reconstruction of the unknown source term, Figure 2 displays the comparison between the exact function $r(t)$ and its numerical approximation $\tilde{r}(t)$. It is evident that the two curves are indistinguishable, and the absolute error remains near the machine precision level, confirming the high fidelity of the spectral approach.

To further investigate the stability and practical applicability of the method, we test its sensitivity to data perturbations. In real-world inverse problems, the observed data are often corrupted by measurement errors. Therefore, to simulate real-world measurement uncertainties, a certain level of random noise is added to the additional integral condition and the coefficients associated with the nonlocal boundary conditions. So, the right-hand side vector in (21) is contaminated with noise, which is modeled as follows:

$$F^{\epsilon} = F + \epsilon(2.rand(S) - 1),$$

where ϵ represents the noise level, S is the dimension of F and $rand()$ generates uniformly distributed random numbers on $[0, 1]$.

As observed in Table 2, even with significant noise levels in F, the method remains stable. Furthermore, as M and N increase, the error does not grow significantly and stays within a consistent range, demonstrating that the proposed technique is robust against numerical instabilities and does not suffer from ill-conditioning in the presence of perturbations.



(a) $\alpha = 0.9, \beta = 0.2$, and $M = N = 8$.

(b) $\alpha = 0.5, \beta = 0.8$, $M = 8$, and $N = 10$.

(c) $\alpha = 0.4, \beta = 0.6$, $M = 10$, and $N = 12$.

(d) $\alpha = \beta = 0.3$ and $M = N = 10$.

Figure 1: Absolute errors in Example 1.

Table 2: Numerical results for Example 1 in the presence of noise when $\alpha = 0.2$ and $\beta = 0.9$.

	$\epsilon = 0.005$	$\epsilon = 0.01$	$\epsilon = 0.05$	$cond(A)$
$N = 5$	$\|\|u_\epsilon\|\|_\infty = 0.2341$	$\|\|u_\epsilon\|\|_\infty = 0.1126$	$\|\|u_\epsilon\|\|_\infty = 0.2168$	3.2101×10^4
$M = 5$	$\|\|r_\epsilon\|\|_\infty = 0.1115$	$\|\|r_\epsilon\|\|_\infty = 0.7081$	$\|\|r_\epsilon\|\|_\infty = 0.6911$	
$N = 7$	$\|\|u_\epsilon\|\|_\infty = 0.0201$	$\|\|u_\epsilon\|\|_\infty = 0.3161$	$\|\|u_\epsilon\|\|_\infty = 0.5160$	9.8523×10^4
$M = 6$	$\|\|r_\epsilon\|\|_\infty = 0.3686$	$\|\|r_\epsilon\|\|_\infty = 0.2395$	$\|\|r_\epsilon\|\|_\infty = 0.9997$	
$N = 8$	$\|\|u_\epsilon\|\|_\infty = 0.0252$	$\|\|u_\epsilon\|\|_\infty = 0.0578$	$\|\|u_\epsilon\|\|_\infty = 0.1584$	2.2732×10^5
$M = 6$	$\|\|r_\epsilon\|\|_\infty = 0.0648$	$\|\|r_\epsilon\|\|_\infty = 0.1316$	$\|\|r_\epsilon\|\|_\infty = 0.4501$	

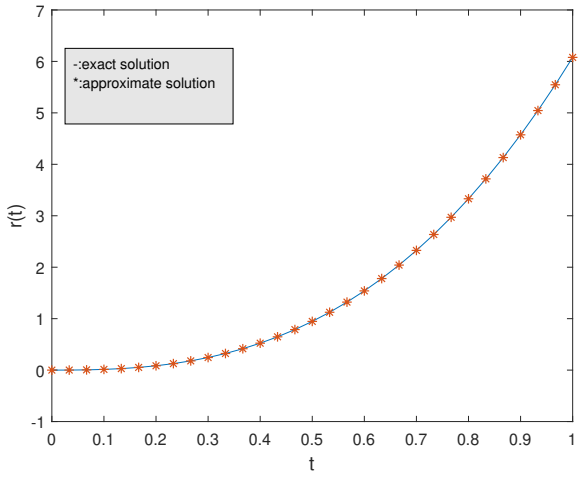


Figure 2: Comparison between the exact and approximate values of $r(t)$ in Example 1.

It is important to note that while the discrete system matrix exhibits a relatively high condition number, the stability of the obtained results is consistent with the theoretical well-posedness of the underlying inverse problem. Therefore, the proposed numerical procedure maintains its reliability and produces stable solutions without the necessity of any additional regularization techniques.

Furthermore, to demonstrate the method's performance under measurement errors, Figure 3 presents the numerical results in the presence of several noise levels for $\alpha = 0.2$ and $\beta = 0.9$. This figure shows the three-dimensional error distribution of $u(x,t)$, along with the reconstructed source term. Despite the noise, the method maintains stable behavior, and no oscillations are observed in the reconstructed source function.

Example 2. Consider the problem (1)–(5) with the exact solution $u(x,t) = t^2 e^{2x+1}$ and the source term

$$r(t) = \frac{\Gamma(3)}{\Gamma(3-\alpha)} t^{2-\alpha} + \frac{\Gamma(3)}{\Gamma(3-\beta)} t^{2-\beta} - 4t^2.$$

The initial condition is $u_0(x) = 0$, and the coefficients in the nonlocal boundary conditions are $a = e^2, b = -1, c = -2$, and $d = 2$.

Table 3 reports the maximum absolute errors for both $u(x,t)$ and $r(t)$, alongside the CPU-time for various values of M, N, α and β. The results in this table indicate that the method achieves high accuracy even with a small number of collocation points. For instance, when $\alpha = 0.1, \beta = 0.3$ and $M = N = 9$, the error in u reaches as low as 4.52×10^{-7}.

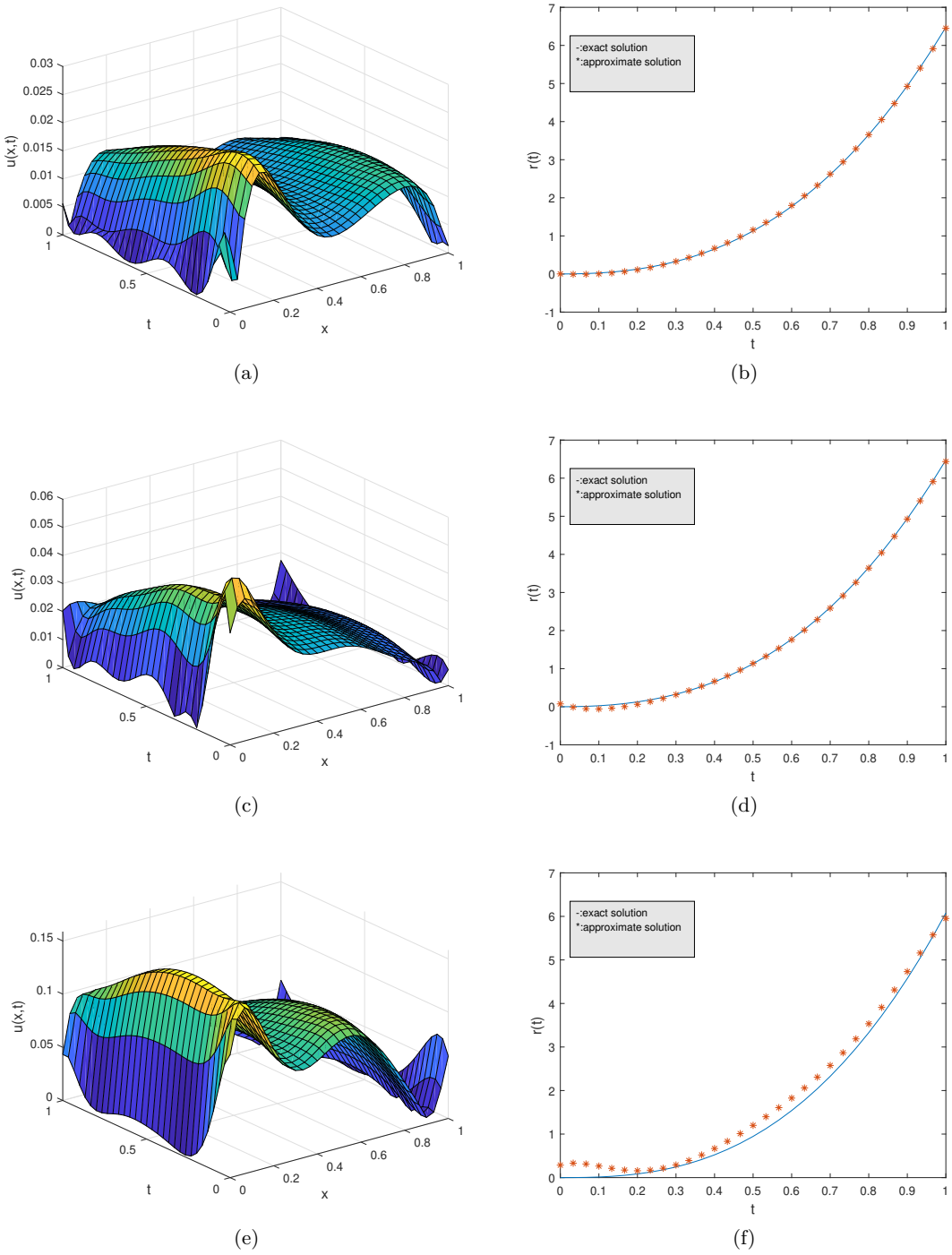


Figure 3: The absolute errors of $u(x,t)$ (left), and the reconstructed source term $r(t)$ (right) in Example 1 under different noise levels, $\epsilon = 0.005$ (first row), $\epsilon = 0.01$ (second row), and $\epsilon = 0.05$ (third row) when $\alpha = 0.2$, $\beta = 0.9$, $M = 6$, and $N = 8$.

Table 3: Numerical results in Example 2 for different values of α, β, M, and N.

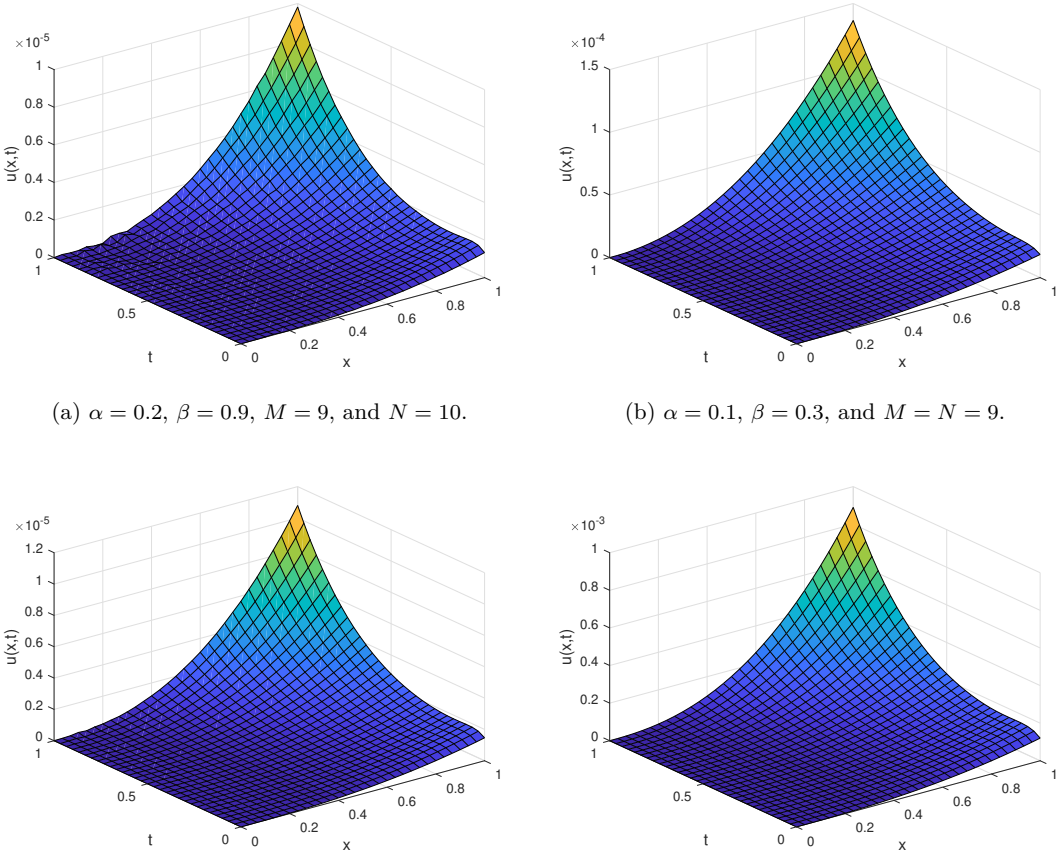
α and β	M and N	CPU time (sec)	error of $u(x,t)$	error of $r(t)$
$\alpha = 0.2, \beta = 0.9$	$M = 5, N = 8$	4.872626	0.0025	3.1055×10^{-4}
	$M = 6, N = 9$	6.131564	2.8945×10^{-4}	3.6336×10^{-5}
	$M = 9, N = 10$	17.587541	3.0142×10^{-5}	3.7863×10^{-6}
$\alpha = 0.1, \beta = 0.3$	$M = N = 4$	1.914168	0.0058	0.5103
	$M = N = 6$	3.997282	1.5651×10^{-4}	0.0236
	$M = N = 9$	9.213205	4.5291×10^{-7}	5.5964×10^{-5}
$\alpha = 0.1, \beta = 0.3$	$M = 5, N = 8$	4.462392	0.0040	4.8001×10^{-4}
	$M = 6, N = 9$	5.912642	4.6522×10^{-4}	5.5962×10^{-5}
	$M = 9, N = 10$	10.421637	4.8439×10^{-5}	5.8361×10^{-6}
$\alpha = \beta = 0.5$	$M = N = 8$	7.008119	0.0028	3.5138×10^{-4}
$\alpha = \beta = 0.8$	$M = N = 8$	6.849998	0.0020	2.5458×10^{-4}
$\alpha = \beta = 0.2$	$M = N = 8$	6.719837	0.0040	4.8212×10^{-4}

To visualize the accuracy of the proposed method for Example 2, Figure 4 displays the three-dimensional absolute error surfaces of $u(x,t)$ for different fractional orders and collocation points. The smooth and low-magnitude error surfaces confirm the uniform convergence of the method. Additionally, Figure 5 provides a comparison between the exact source term $r(t)$ and its numerical reconstruction $\tilde{r}(t)$. The perfect overlap between the two curves, along with the negligible absolute error, demonstrates the high precision of the method in identifying the unknown source term.

To further investigate the convergence behavior, Table 4 summarizes the numerical convergence orders for $\alpha = 0.2$ and $\beta = 0.9$. It is evident that the convergence order increases significantly as N increases, which is a typical characteristic of spectral methods, demonstrating that the proposed approach is far more efficient than low-order finite difference schemes.

Table 4: Numerical convergence orders in Example 2 when $\alpha = 0.2$ and $\beta = 0.9$.

$M = N$	h_N	*error*	*order*
5	0.2165	0.7010	
6	0.1802	0.1216	9.6187
7	0.15398	0.0188	12.1186
8	0.13423	0.0025	15.0754
9	0.11888	2.8936×10^{-4}	18.2925
10	0.10661	3.0180×10^{-5}	21.4752



(a) $\alpha = 0.2$, $\beta = 0.9$, $M = 9$, and $N = 10$.

(b) $\alpha = 0.1$, $\beta = 0.3$, and $M = N = 9$.

(c) $\alpha = 0.4$, $\beta = 0.6$, $M = 9$, and $N = 10$.

(d) $\alpha = \beta = 0.5$ and $M = N = 8$.

Figure 4: Absolute error distribution in Example 2 for various values of α, β, M, and N.

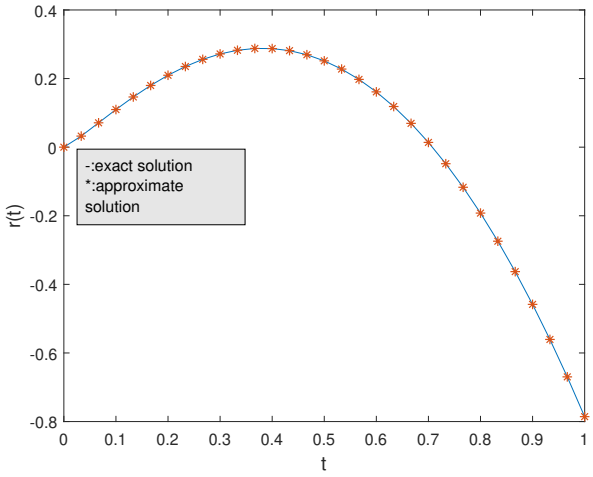


Figure 5: Comparison between the exact and recovered source terms $r(t)$ for Example 2.

In addition, to provide a visual representation of the method's stability, Figure 6 illustrates the absolute error distribution of the numerical solution and the recovered source term in the presence of different noise levels. Despite the perturbation in the data, the error surfaces remain smooth and free of any numerical oscillations.

The sensitivity of the method to noisy data is analyzed in Table 5. As observed in this table, although the condition number $cond(A)$ increases with N, the absolute errors for both the state variable and the source term remain controlled.

Table 5: Numerical results for Example 2 in the presence of various noise levels with $\alpha = 0.4$ and $\beta = 0.8$.

	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.1$	$cond(A)$
$N = 4$	$\|\|u_\epsilon\|\|_\infty = 0.8122$	$\|\|u\epsilon\|\|_\infty = 0.7271$	$\|\|u_\epsilon\|\|_\infty = 0.5685$	1.2624×10^3
$M = 4$	$\|\|r_\epsilon\|\|_\infty = 0.1640$	$\|\|r_\epsilon\|\|_\infty = 0.1661$	$\|\|r_\epsilon\|\|_\infty = 0.1910$	
$N = 6$	$\|\|u_\epsilon\|\|_\infty = 0.3273$	$\|\|u_\epsilon\|\|_\infty = 0.2255$	$\|\|u_\epsilon\|\|_\infty = 0.3415$	1.0221×10^5
$M = 6$	$\|\|r_\epsilon\|\|_\infty = 0.208$	$\|\|r_\epsilon\|\|_\infty = 0.0928$	$\|\|r_\epsilon\|\|_\infty = 0.2154$	
$N = 8$	$\|\|u_\epsilon\|\|_\infty = 0.0589$	$\|\|u_\epsilon\|\|_\infty = 0.2812$	$\|\|u_\epsilon\|\|_\infty = 1.0841$	2.2918×10^5
$M = 6$	$\|\|r_\epsilon\|\|_\infty = 0.0119$	$\|\|r_\epsilon\|\|_\infty = 0.0592$	$\|\|r_\epsilon\|\|_\infty = 1.0841$	

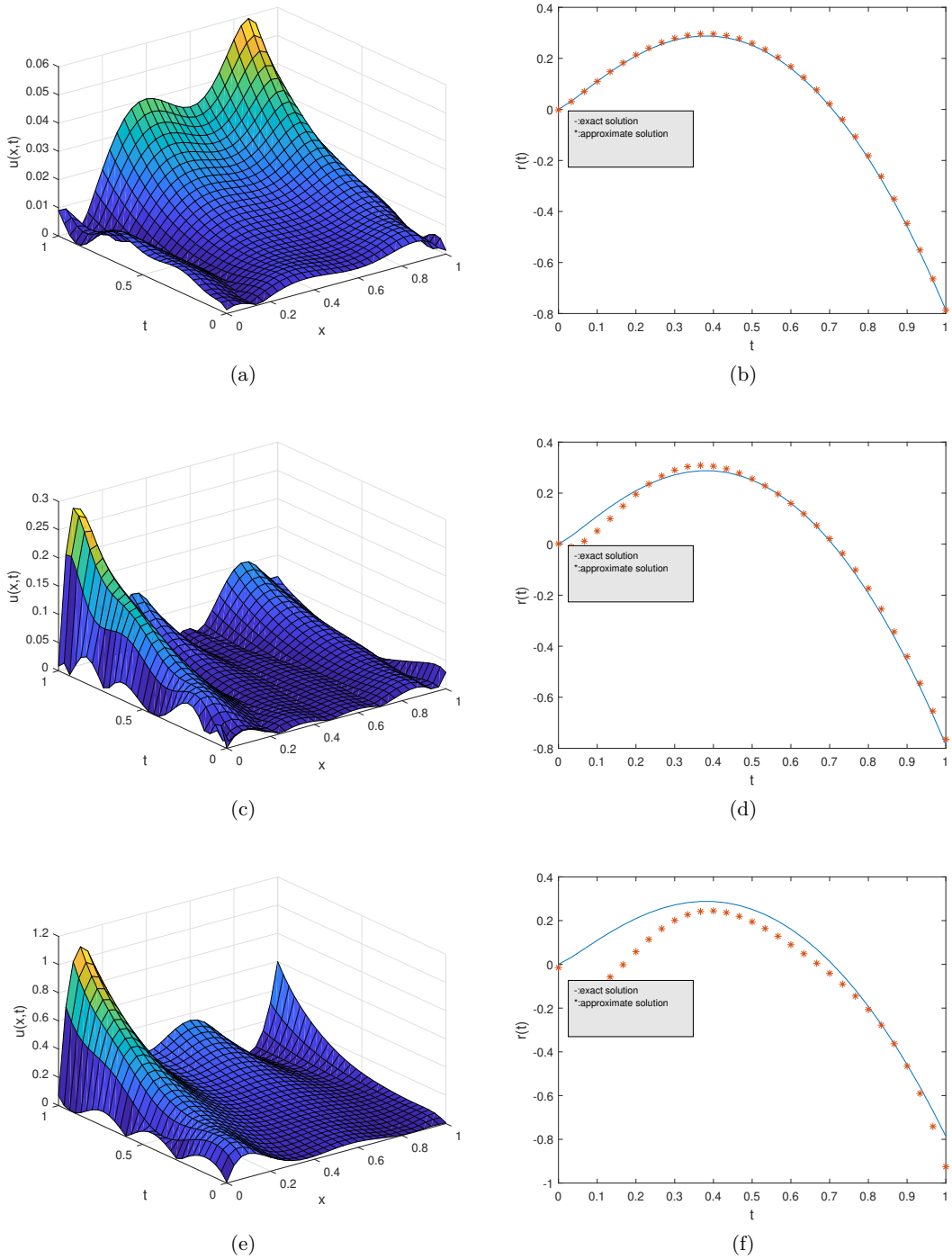


Figure 6: The absolute errors of $u(x,t)$ (left), and the reconstructed source term $r(t)$ (right) in Example 2 under different noise levels, $\epsilon = 0.01$ (first row), $\epsilon = 0.05$ (second row), and $\epsilon = 0.1$ (third row) with $\alpha = 0.4$, $\beta = 0.8$, $M = 6$, and $N = 8$.

7 Conclusion

An efficient spectral collocation method based on shifted Legendre polynomials has been developed for the numerical solution of an inverse source problem for two-term time-fractional diffusion equations subject to nonlocal boundary conditions. The proposed approach successfully reconstructs the unknown time-dependent source term $r(t)$ while simultaneously approximating the solution $u(x,t)$. A rigorous convergence analysis has been carried out, establishing optimal error estimates under standard regularity assumptions. The theoretical results are validated through extensive numerical experiments on benchmark problems, which demonstrate the spectral accuracy and rapid convergence of the proposed method with respect to the collocation parameters N and M. Moreover, the unknown source function $r(t)$ is recovered with high accuracy.

In summary, the proposed method is accurate, stable, and straightforward to implement. These features make it a reliable and competitive numerical technique for solving inverse problems arising in multi-term fractional diffusion models with nonlocal constraints.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

We are also grateful to the Editor and the anonymous reviewers for their constructive comments and suggestions that significantly improved the quality of this manuscript.

References

[1] Abdollahi, N. and Rostamy, D. *Identifying an unknown time-dependent boundary source in time-fractional diffusion equation with a non-local boundary condition*, J. Comput. Appl. Math., 355 (2019), 36–50.

[2] Ali, M. and Aziz, S. *Some inverse problems for time-fractional diffusion equation with nonlocal Samarskii–Ionkin type condition*, Math. Methods Appl. Sci., 44 (2021), 8447–8462.

[3] Arridge, S., Maas, P., Öktem, O. and Schönlieb, C.B. *Solving inverse problems using data-driven models*, Acta Numer., 28 (2019), 1–174.

[4] Bazan, F., Bedin, L., Ismailov, M.I. and Borges, L. *Inverse time-dependent source problem for the heat equation with a nonlocal Wentzell–Neumann boundary condition*, Netw. Heterog. Media, 18 (2023), 1747–1771.

[5] Beckers, S. and Yamamoto, M. *Regularity and unique existence of solution to linear diffusion equation with multiple time-fractional derivatives*, In *Control and Optimization with PDE Constraints*, Springer, (2013), 45–55.

[6] Błasik, M. *Numerical scheme for a two-term sequential fractional differential equation*, Sci. Res. Inst. Math. Comput. Sci., 10(2) (2011), 17–29.

[7] Brandibur, O. and Kaslik, É. *Stability properties of multi-term fractional-differential equations*, Fractal Fract., 7(2) (2023), 117.

[8] Chen, J., Liu, F. and Anh, V. *Analytical solution for the time-fractional telegraph equation by the method of separating variables*, J. Math. Anal. Appl., 338(2) (2008), 1364–1377.

[9] Chung, J. and Gazzola, S. *Computational methods for large-scale inverse problems: a survey on hybrid projection methods*, SIAM Rev., 66(2) (2024), 205–284.

[10] D'Abbicco, M. and Girardi, G. *Asymptotic profile for a two-terms time fractional diffusion problem*. Fract. Calc. Appl. Anal. 25(3) (2022), 1199–1228.

[11] D'Abbicco, M. and Girardi, G. *Decay estimates for a perturbed two-terms space-time fractional diffusive problem*, Evol. Equ. Control Theory, 12(4) (2023), 1106–1138.

[12] Derbissaly, B., Kirane, M. and Sadybekov, M.A. *Inverse source problem for two-term time-fractional diffusion equation with nonlocal boundary conditions*, Chaos Solitons Fractals, 183 (2024), 114897.

[13] Dib, F. and Kirane, M. *An inverse source problem for a two terms time-fractional diffusion equation*, Bol. Soc. Parana. Mat., 2022 (2022) 1–15.

[14] Ding, M.H., Liu, H. and Lo, C.W. *Inverse problems for coupled nonlocal nonlinear systems arising in mathematical biology*, arXiv preprint arXiv:2407.15713, 2024.

[15] Fazli, H., Bahrami, F. and Shahmorad, S. *Extremal solutions for multi-term nonlinear fractional differential equations with nonlinear boundary conditions*, Comput. Methods Differ. Equ., 11(1) (2023), 32–41.

[16] Furati, K.M., Iyiola, O.S. and Kirane, M. *An inverse problem for a generalized fractional diffusion*, Appl. Math. Comput., 249 (2014), 24–31.

[17] Hazanee, A., Lesnic, D., Ismailov, M.I. and Kerimov, N.B. *Inverse time-dependent source problems for the heat equation with nonlocal boundary conditions*, Appl. Math. Comput., 346 (2019), 800–815.

[18] Ismailov, M.I. *Inverse source problem for heat equation with nonlocal Wentzell boundary condition*, Results Math., 73(2) (2018) 68.

[19] Ismailov, M.I. and Cicek, M. *Inverse source problem for a time-fractional diffusion equation with nonlocal boundary conditions*, Appl. Math. Model., 40 (2016), 4891–4899.

[20] Ismailov, M.I. and Tekin, I. *An inverse problem for finding the lowest term of a heat equation with Wentzell–Neumann boundary condition*, Inverse Probl. Sci. Eng., 27 (2019), 1608–1634.

[21] Jafari, H., Babaei, A. and Banihashemi, S. *A novel approach for solving an inverse reaction–diffusion–convection problem*, J. Optim. Theory Appl., 183 (2019), 688–704.

[22] Khosravian-Arab, H. and Dehghan, M. *A matrix approach to multi-term fractional differential equations using two new diffusive representations for the Caputo fractional derivative*, AUT J. Math. Comput., 5(4) (2024), 337–359.

[23] Kilbas, A.A., Srivastava, H.M. and Trujillo, J.J. *Theory and applications of fractional differential equations*, Elsevier, 2006.

[24] Kirane, M., Malik, S.A. and Al-Gwaiz, M.A. *An inverse source problem for a two dimensional time fractional diffusion equation with nonlocal boundary conditions*, Math. Methods Appl. Sci., 36(9) (2013), 1056–1069.

[25] Kirane, M., Sadybekov, M.A. and Sarsenbi, A.A. *On an inverse problem of reconstructing a subdiffusion process from nonlocal data*, Math. Methods Appl. Sci., 42 (2019), 2043–2052.

[26] Lattanzi, A.M. and Hrenya, C.M. *A coupled, multiphase heat flux boundary condition for the discrete element method*, Chem. Eng. J., 304 (2016), 766–773.

[27] Li, T.-T. *A class of non-local boundary value problems for partial differential equations and its applications in numerical analysis*, J. Comput. Appl. Math., 28 (1989), 49–62.

[28] Li, Y., Wang, T. and Gao, G.H. *The asymptotic solutions of two-term linear fractional differential equations via Laplace transform*, Math. Comput. Simul., 211 (2023), 394–412.

[29] Ma, R. *A survey on nonlocal boundary value problems*, Appl. Math. E-Notes, 7 (2007), 257–279.

[30] Malik, S.A. and Aziz, S. *An inverse source problem for a two parameter anomalous diffusion equation with nonlocal boundary conditions*, Comput. Math. Appl., 73(12) (2017), 2548–2560.

[31] Nemati, S. and Babaei, A. *A numerical method based on the Jacobi polynomials to reconstruct an unknown source term in a time fractional diffusion-wave equation*, Taiwanese J. Math., 23(5) (2019), 1271–1289.

[32] Orazov, I. and Makhatova, A.K. *On application of inverse problems with the nonlocal boundary conditions in the diffusion theory*, In *International Conference "Functional Analysis in Interdisciplinary Applications" (FAIA2017)*, Vol. 1880, AIP Publishing, (2017), 060008.

[33] Orazov, I. and Sadybekov, M.A. *On a class of problems of determining the temperature and density of heat sources given initial and final temperature*, Sib. Math. J., 53(1) (2012), 146–151.

[34] Orsingher, E. and Beghin, L. *Time-fractional telegraph equations and telegraph processes with Brownian time*, Probab. Theory Relat. Fields, 128(1) (2004), 141–160.

[35] Ozkan, A.S. and Adalar, İ. *Inverse nodal problems for Sturm–Liouville equation with nonlocal boundary conditions*, J. Math. Anal. Appl., 520(1) (2023), 126904.

[36] Pal, S. and Melnik, R. *Nonlocal models in biology and life sciences: sources, developments, and applications*, arXiv preprint arXiv:2401.14651, 2024.

[37] Rivlin, T.J. *Chebyshev polynomials*, Courier Dover Publications, 2020.

[38] Sadybekov, M.A. and Pankratova, I.N. *Correct and stable algorithm for numerical solving nonlocal heat conduction problems with not strongly regular boundary conditions*, Mathematics, 10(20) (2022), 3780.

[39] Stojanović, M. *Existence–uniqueness result for a nonlinear n-term fractional equation*, J. Math. Anal. Appl., 353(1) (2009), 244–255.

[40] Stojanović, M. and Gorenflo, R. *Nonlinear two-term time fractional diffusion-wave problem*, Nonlinear Anal. Real World Appl., 11(5) (2010), 3512–3523.

[41] Sun, L.L. and Yan, X.B. *Inverse source problem for a multiterm time-fractional diffusion equation with nonhomogeneous boundary condition*, Adv. Math. Phys., 2020 (2020), 1825235.

[42] Vogel, C.R. *Computational methods for inverse problems*, Soc. Ind. Appl. Math., 2002.

[43] Yadigaroglu, G. and Hewitt, G.F. *Introduction to multiphase flow: basic concepts, applications and modelling*, Springer, 2017.

[44] Zhou, Y. and He, J.W. *Well-posedness and regularity for fractional damped wave equations*, Monatsh. Math., 194 (2021), 425–458.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1550–1590
https://doi.org/10.22067/ijnao.2026.97711.1835
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Fractional order Chelyshkov wavelet for solving variable coefficients Caputo time-fractional convection-diffusion equation

S. Sabir* and A. Ahmad

Abstract

Convection-diffusion equations are widely used to describe transport processes and arise in numerous scientific and engineering applications, including heat transfer, contaminant transport in porous media, groundwater flow, chemical reactions, and biological systems. However, many real-world transport phenomena exhibit memory properties that classical convection-diffusion equations cannot adequately represent. To address such nonlocal temporal effects, time-fractional convection-diffusion equations have been introduced, in which fractional derivatives provide a realistic description of anomalous diffusion. This work proposes a novel wavelet-based numerical method for solving a variable-coefficient Caputo time-fractional convection-diffusion equation using fractional-order Chelyshkov wavelets. The method constructs fractional Chelyshkov wavelets from fractional Chelyshkov polynomials and derives corresponding fractional operational matrices. The proposed wavelet basis functions, together with an operational matrix formulation, transform the time-fractional convection-diffusion equation into a

*Corresponding author

Received 5 February 2026; revised 19 June 2026; accepted 25 July 2026

Sufia Sabir
Department of Mathematics and Computing Technology, National Institute of Technology Patna, Patna, 800005, Bihar, India. e-mail: sufias.phd22.ma@nitp.ac.in

Ayaz Ahmad
Department of Mathematics and Computing Technology, National Institute of Technology Patna, Patna, 800005, Bihar, India. e-mail: ayaz@nitp.ac.in

How to cite this article
Sabir, S. and Ahmad, A., Fractional order Chelyshkov wavelet for solving variable coefficients Caputo time-fractional convection-diffusion equation. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1550-1590. https://doi.org/10.22067/ijnao.2026.97711.1835

system of algebraic equations. The stability, convergence, and error estimates are established. To demonstrate the effectiveness and precision of the proposed method, several numerical examples are provided. CPU time and memory usage are reported for different numbers of basis functions. The numerical results indicate that the fractional Chelyshkov wavelets provide an efficient representation of smooth solutions, achieving high accuracy with a relatively small number of basis functions.

AMS subject classifications (2020): 35A01, 65L10, 65L12, 65L20, 65L70.

Keywords: Caputo time fractional convection-diffusion equation; Fractional Chelyshkov wavelet; Operational matrix; Stability; Convergence and error analysis.

1 Introduction

In recent years, models of fractional-order partial differential equations have been developed and studied across many research areas, including finance, fluid mechanics, plasma physics, biology, mechanics of materials, and chemistry. One of the most significant examples of a fractional partial differential equation is the convection-diffusion equation. Many physical phenomena, such as oil reservoir models, particle motion, heat transfer, global weather, energy transfer, and mass transfer, can be explained by the convection-diffusion equation. However, the classical convection–diffusion equation is often inadequate for describing anomalous diffusion phenomena arising from heterogeneous media or complex particle transport mechanisms encountered in natural systems. To tackle this situation, fractional convection-diffusion equations have been introduced. The classification of fractional partial differential equations includes space and time fractional, space fractional, and time fractional. In this article, we have considered the time fractional convection-diffusion equation (TFCDE) with a variable coefficient given as

$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}} + a(s)\frac{\partial \mathcal{U}(s,t)}{\partial s} + b(s)\frac{\partial^{2}\mathcal{U}(s,t)}{\partial s^{2}} = f(s,t)$$

with the initial condition (I.C) and boundary conditions (B.Cs)

$$\mathcal{U}(s,0) = q(s), \qquad \mathcal{U}(0,t) = f_0(t), \qquad \mathcal{U}(1,t) = f_1(t),$$

where $a(s)$ and $b(s) \neq 0$ are smooth functions, q, f_0, and $f_1 \in L^2[0,1)$, and $f(\cdot,\cdot)$ is a function in $L^2([0,1)\times[0,1))$.

Analytical solutions to fractional partial differential equations are not possible in many cases. So, developing stable, high-precision numerical methods has become a major focus of research. Various numerical methods, including the variational iteration method [39, 6], the finite element method [13], the finite difference method [34, 9], the operational method [14], the Ritz approximation method [15], the Tau method [1, 23], and so on, are used to solve fractional partial differential equations. To address the fractional convection-diffusion equation, several numerical techniques have been used, such as the discontinuous Galerkin method [25], the Sinc-Legendre collocation method [24], the Gegenbauer method [8], the Haar wavelet method [3], the Sinc and B-spline scaling functions [2], radial basis functions [35], the Chebyshev collocation method [41], and the Fibonacci wavelet method [38]. Wavelet-based numerical methods are highly effective for solving various classes of fractional partial differential equations due to their low complexity, rapid convergence, and reasonable computational efficiency. Wavelets are a family of functions that are obtained from the translation and dilatation of a single function called the "mother wavelet" [4]. A variety of wavelet and spectral-based methods, including Haar wavelet [40, 18], Bernoulli wavelet [21], Legendre wavelet [37, 29], Chebyshev wavelet [11], Hermite wavelet [5], Euler wavelet [36, 30, 28], Fibonacci wavelet [12], and B-spline wavelet [10] have been developed for fractional partial differential equations. Recently, Singh et al. [27, 31, 33, 32] proposed fractional wavelet operational matrix methods for different classes of fractional differential equations, demonstrating the effectiveness of wavelet approximations in reducing complex fractional models to algebraic systems. Motivated by the above work, we have proposed wavelet-based method for solving TFCDE with variable coefficients. We have found that most existing studies either focus on constant coefficients or use basis functions that do not specifically incorporate fractional characteristics. Additionally, a common characteristic of these existing methods is their reliance on high-resolution parameters to achieve high-precision results. When dealing with fractional equations, increasing the resolution parameter leads to a computational cost that grows exponentially due to the nonlocal nature of fractional derivatives, matrix ill-conditioning, which degrades stability during computation, and severe round-off error propagation.

Novelty and main contribution of this work:

To overcome the above limitations, this work introduces novel fractional Chelyshkov wavelets (FCW) and an operational matrix framework for the numerical solution of the variable-coefficient TFCDE. The fractional structure of basis functions enables an efficient representation of solution behavior associated with memory effects. FCW can demonstrate signals at different levels of resolution due to its compact support. Our proposed method introduces several advantages over the existing literature:

- **Fractional basis framework:** Most of the studies for TFCDE with variable coefficients rely on classical polynomials, scaling functions, or nonfractional wavelet bases [38, 41, 2, 7, 24, 26]. The nonfractional wavelet basis cannot accurately approximate the fractional power terms present in some fractional differential equations. Also, the fractional derivatives of nonfractional wavelet bases cannot be represented exactly within the same basis space, so a large number of basis functions is required to approximate fractional-order terms. In contrast, the proposed method employs fractional basis functions to obtain the corresponding operational matrix, thereby providing a more natural representation of fractional operators.

- **Superior accuracy at low resolutions:** The FCW yields a strictly orthogonal basis, minimizing theoretical complexity while optimizing numerical efficiency. These basis functions match fractional physics and reduce truncation errors with very few terms.

- **Computation sparsity:** The algebraic system obtained by combining compact wavelet support and strict orthogonality is highly structured and sparse, reducing memory requirements during computer implementation and ensuring lower computational time.

The present article is organized as follows: The introduction is presented in section 1. Section 2 presents the fundamental definitions and properties. Fractional order Chelyshkov wavelets are presented in section 3. Approximation of a function using fractional Chelyshkov wavelets is explained in section 4. Fractional operational matrices of FCW are obtained in Section 5. Section 6 presents a description of a numerical method to solve fractional convection-diffusion. Section 7 presents the stability analysis of the proposed method. Convergence analysis is discussed in Section 8. Section 9 presents the error analysis of the proposed method. Numerical examples are presented in section 10 to demonstrate the precision and effectiveness of the proposed method. Section 11 presents the conclusion of the paper.

2 Preliminaries

This section provides some fundamental definitions and properties of fractional integrals and derivatives.

Definition 1. A real function $\mathcal{U}(s)$, $s > 0$, is said to be in space C_σ, $\sigma \in \mathbb{R}$, if there is a real number η $(\eta > \sigma)$ such that $\mathcal{U}(s) = s^\eta \mathcal{U}_0(s)$, where $\mathcal{U}_0(s) \in C[0, \infty)$ and $\mathcal{U}(s) \in C_\sigma^n$ if $\mathcal{U}^n(s) \in C_\sigma$, $n \in \mathbb{N}$ [16].

Definition 2. The Riemann–Liouville fractional integral operator of order $\alpha \geq 0$ of a function $\mathcal{U}(s) \in C_\sigma, \sigma \geq -1$ is defined as [20]

$$ {}_0^{RL}I_t^\alpha \mathcal{U}(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-z)^{\alpha-1} \mathcal{U}(z) dz. \tag{1}$$

Definition 3. The Caputo fractional derivative operator of order α is defined as [20]

$$ {}_0^{C}D_t^\alpha \mathcal{U}(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-z)^{n-\alpha-1} \mathcal{U}^n(z) dz, \qquad n-1 < \alpha \leq n, \tag{2}$$

where $n \in \mathbb{N}, t > 0$ and $\mathcal{U} \in C_{-1}^n$.

Proposition 1. The following properties are satisfied by the Riemann–Liouville fractional integrals and Caputo fractional derivatives [20]

- ${}_0^{C}D_t^\alpha {}_0^{RL}I_t^\alpha \mathcal{U}(t) = \mathcal{U}(t)$,
- ${}_0^{RL}I_t^\alpha {}_0^{C}D_t^\alpha \mathcal{U}(t) = \mathcal{U}(t) - \sum_{k=0}^{n-1} \mathcal{U}^k(0^+) \frac{t^k}{k!}$,

where $n-1 < \alpha \leq n$.

3 Fractional order Chelyshkov wavelet

Fractional Chelyshkov wavelets [17] are presented in this section. The definition of the Chelyshkov wavelet is

$$ \psi_{n,m}(t) = \begin{cases} 2^{\frac{k-1}{2}} \sqrt{2m+1} \rho_{m,M}(2^{k-1}t - n + 1), & \frac{n-1}{2^{k-1}} \leq t < \frac{n}{2^{k-1}}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

By applying the transformation $t = s^\beta$ in (3), we obtain the fractional Chelyshkov wavelet as

$$ \psi_{n,m}^\beta(s) = \begin{cases} 2^{\frac{k-1}{2}} \sqrt{2m+1} \rho_{m,M}(2^{k-1}s^\beta - n + 1), & (\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}} \leq s < (\frac{n}{2^{k-1}})^{\frac{1}{\beta}}, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $\beta \in \mathbb{R}^+, M \in \mathbb{Z}^+, k \in \mathbb{N}, n = 1, 2, \ldots, 2^{k-1}$ and $\rho_{m,M}$ denotes the Chelyshkov polynomial, which is given as

$$\rho_{m,M}(s) = \sum_{l=0}^{M-m-1} c_{l,m} s^{m+l}, \qquad m = 0, 1, \ldots, M-1, \tag{5}$$

where $c_{l,m} = (-1)^l \binom{M-1-m}{l} \binom{M+m+l}{M-1-m}$.

For $k = 2, M = 2, \beta = 0.5$, the four basis functions are given as

$$\psi_{1,0}^{0.5}(s) = \begin{cases} -0.707107(-1 + 3(-1 + 4s^{0.5})), & 0 \le s < 0.25, \\ 0, & \text{otherwise}, \end{cases}$$

$$\psi_{1,1}^{0.5}(s) = \begin{cases} 4.89898 s^{0.5}, & 0 \le s < 0.25, \\ 0, & \text{otherwise}, \end{cases}$$

$$\psi_{2,0}^{0.5}(s) = \begin{cases} -0.707107(-1 + 3(-3 + 4s^{0.5})), & 0.25 \le s < 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\psi_{2,1}^{0.5}(s) = \begin{cases} 2.44948(-1 + 2s^{0.5}), & 0.25 \le s < 1, \\ 0, & \text{otherwise}. \end{cases}$$

The fractional Chelyshkov wavelets in two dimensions are given as

$$\psi_{n_1,m_1,n_2,m_2}^{\alpha,\beta}(s,t) =$$
$$\begin{cases} 2^{\frac{k_1-1}{2}} 2^{\frac{k_2-1}{2}} \sqrt{2m_1+1}\sqrt{2m_2+1} & \text{if } \left(\frac{n_1-1}{2^{k_1-1}}\right)^{\frac{1}{\alpha}} \le s < \left(\frac{n_1}{2^{k_1-1}}\right)^{\frac{1}{\alpha}}, \\ \times \rho_{m_1,M_1}(2^{k_1-1}s^\alpha - n_1 + 1) & \left(\frac{n_2-1}{2^{k_2-1}}\right)^{\frac{1}{\beta}} \le t < \left(\frac{n}{2^{k_2-1}}\right)^{\frac{1}{\beta}}, \\ \times \rho_{m_2,M_2}(2^{k_2-1}t^\beta - n_2 + 1), & \\ 0, & \text{otherwise}. \end{cases}$$

Lemma 1. The orthonormal property of $\psi_{n_1,m_1,n_2,m_2}^{\alpha,\beta}(s,t)$ on $[0,1) \times [0,1)$ with respect to the weight function $w_{\alpha,\beta}(s,t) = \alpha\beta s^{\alpha-1} t^{\beta-1}$ is considered as follows:

$$\int_0^1 \int_0^1 w_{\alpha,\beta}(s,t) \psi_{n_1,m_1,n_2,m_2}^{\alpha,\beta}(s,t) \psi_{n_1',m_1',n_2',m_2'}^{\alpha,\beta}(s,t) ds dt$$
$$= \delta_{n_1,n_1'} \delta_{m_1,m_1'} \delta_{n_2,n_2'} \delta_{m_2,m_2'}.$$

Proof. Due to the orthonormal property of FCW, we have

$$\int_0^1 w_\alpha(t)\psi^\alpha_{n_1,m_1}(t)\psi^\alpha_{n_2,m_2}(t)dt = \delta_{n_1,n_2}\delta_{m_1,m_2}. \tag{6}$$

Now,

$$\begin{aligned}
&\int_0^1\int_0^1 w_{\alpha,\beta}(s,t)\psi^{\alpha,\beta}_{n_1,m_1,n_2,m_2}(s,t)\psi^{\alpha,\beta}_{n_1',m_1',n_2',m_2'}(s,t)dsdt\\
&= \int_0^1\int_0^1 \alpha\beta s^{\alpha-1}t^{\beta-1}\psi^\alpha_{n_1,m_1}(s)\psi^\beta_{n_2,m_2}(t)\psi^\alpha_{n_1',m_1'}(s)\psi^\beta_{n_2',m_2'}(t)dsdt\\
&= \int_0^1 \alpha s^{\alpha-1}\psi^\alpha_{n_1,m_1}(s)\psi^\alpha_{n_1',m_1'}(s)ds\int_0^1 \beta t^{\beta-1}\psi^\beta_{n_2,m_2}(t)\psi^\beta_{n_2',m_2'}(t)dt\\
&= \delta_{n_1,n_1'}\delta_{m_1,m_1'}\delta_{n_2,n_2'}\delta_{m_2,m_2'}.
\end{aligned}$$

□

4 Function approximation using FCW

A function $\mathcal{U}(s,t) \in L^2([0,1)\times[0,1))$ can be expanded in terms of FCW as

$$\mathcal{U}(s,t) = \sum_{n_1=1}^{2^{k_1-1}}\sum_{n_2=1}^{2^{k_2-1}}\sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty} r_{n_1,n_2,m_1,m_2}\psi^\alpha_{n_1,m_1}(s)\psi^\beta_{n_2,m_2}(t), \tag{7}$$

where $r_{n_1,n_2,m_1,m_2} \in \mathbb{R}$. By truncating the power series to order M_1-1 in s and M_2-1 in t, we approximate $\mathcal{U}(s,t)$ as

$$\mathcal{U}(s,t) \approx \sum_{n_1=1}^{2^{k_1-1}}\sum_{n_2=1}^{2^{k_2-1}}\sum_{m_1=0}^{M_1-1}\sum_{m_2=0}^{M_2-1} r_{n_1,n_2,m_1,m_2}\psi^\alpha_{n_1,m_1}(s)\psi^\beta_{n_2,m_2}(t),$$

$$\mathcal{U}(s,t) \approx \mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2} = \Psi^\alpha_{k_1,M_1}(s)\cdot R\cdot\Psi^\beta_{k_2,M_2}(t), \tag{8}$$

where

$$\Psi^\beta_{k,M} = \left[\psi^\beta_{1,0},\ldots,\psi^\beta_{1,M-1},\psi^\beta_{2,0},\ldots,\psi^\beta_{2,M-1},\ldots,\psi^\beta_{2^{k-1},M-1}\right]^\top, \tag{9}$$

and R is a matrix of order $2^{k_1-1}M_1\times 2^{k_2-1}M_2$ composed of FCW coefficients

$$r_{n_1,n_2,m_1,m_2} = \left\langle \mathcal{U}(s,t), \psi^{\alpha}_{n_1,m_1}(s)\psi^{\beta}_{n_2,m_2}(t) \right\rangle = \int_0^1 \int_0^1 w_{\alpha,\beta}(s,t)\mathcal{U}(s,t)\psi^{\alpha}_{n_1,m_1}(s)(\psi^{\beta}_{n_2,m_2}(t))^{\top} ds dt,$$

$$1 \le n_1 \le 2^{k_1-1}, \qquad 0 \le m_1 \le M_1 - 1, 1 \le n_2 \le 2^{k_2-1}, \qquad 0 \le m_2 \le M_2 - 1.$$

The following collocation points are taken into account:

$$\begin{aligned} s_i &= \left(\frac{2i-1}{2^{k_1}M_1}\right)^{\frac{1}{\alpha}}, \qquad i = 1, 2, \ldots, 2^{k_1-1}M_1, \\ t_j &= \left(\frac{2j-1}{2^{k_2}M_2}\right)^{\frac{1}{\beta}}, \qquad j = 1, 2, \ldots, 2^{k_2-1}M_2. \end{aligned} \tag{10}$$

5 Riemann integration operational matrix of fractional Chelyshkov wavelet

Theorem 1. The Riemann integral of FCW vector $\Psi^{\beta}_{k,M}(s)$ can be approximated by fractional operational matrix S^{γ} which is given as

$${}^{RL}_{0}I^{\gamma}_{s}(\Psi^{\beta}_{k,M}(s)) \approx S^{\gamma}\Psi^{\beta}_{k,M}(s), \tag{11}$$

where S^{γ} is a square matrix of order $2^{k-1}M$ and the elements of the fractional integral of the FCW vector $\Psi^{\beta}_{k,M}(s)$ are as follows:

$${}^{RL}_{0}I^{\gamma}_{s}(\psi^{\beta}_{n,m}(s)) \approx \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} \Lambda^{\gamma;nm}_{vl}\psi^{\beta}_{v,l}(s), \tag{12}$$

where

$$\Lambda^{\gamma;nm}_{vl} = \begin{cases} \xi_{mk}\sum_{j=0}^{M-m-1}(-1)^j\binom{M-1-m}{j}\binom{M+m+j}{M-1-m}2^{(k-1)(m+j)}d^{1m}_{vl}, & n = 1, \\ \xi_{mk}\sum_{j=0}^{M-m-1}\sum_{p=0}^{m+j}(-1)^{m+2j-p}\binom{M-1-m}{j}\binom{M+m+j}{M-1-m} & \\ \qquad \times\binom{m+j}{p}(n-1)^{m+j-p}2^{(k-1)p}d^{np}_{vl}, & n \ne 1, \end{cases}$$

and $\xi_{mk} = 2^{\frac{k-1}{2}}\sqrt{2m+1}$.

Proof. We obtain the fractional integration of FCW vector $\Psi^{\beta}_{k,M}(s)$, which is given as follows:

$$ {}_{0}^{RL}I_{s}^{\gamma}\Psi_{k,M}^{\beta}(s) \approx S^{\gamma}\Psi_{k,M}^{\beta}(s). $$

The expression in (4) can be rewritten as

$$ \psi_{n,m}^{\beta}(s) = 2^{\frac{k-1}{2}}\sqrt{2m+1}\rho_{m,M}(2^{k-1}s^{\beta}-n+1)\times\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})}. \tag{13} $$

Using the definition of Chelyshkov polynomial from (5), we have

$$ \begin{aligned} &\rho_{m,M}(2^{k-1}s^{\beta}-n+1) \\ &= \sum_{j=0}^{M-m-1}(-1)^{j}\binom{M-1-m}{j}\binom{M+m+j}{M-1-m}(2^{k-1}s^{\beta}-n+1)^{m+j}. \end{aligned} \tag{14} $$

Substituting (14) in (13), we get

$$ \begin{aligned} &\psi_{n,m}^{\beta}(s) \\ &= \xi_{mk}\sum_{j=0}^{M-m-1}(-1)^{j}\binom{M-1-m}{j}\binom{M+m+j}{M-1-m}(2^{k-1}s^{\beta}-n+1)^{m+j} \\ &\quad\times\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})}, \end{aligned} \tag{15} $$

where $\xi_{mk} = 2^{\frac{k-1}{2}}\sqrt{2m+1}$ and $\chi(s)$ is characteristic function given as

$$ \chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} = \begin{cases} 1, & s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}}), \\ 0, & \text{otherwise.} \end{cases} $$

Putting $n=1$ in (15),

$$ \begin{aligned} &\psi_{1,m}^{\beta}(s) \\ &= \xi_{mk}\sum_{j=0}^{M-m-1}(-1)^{j}\binom{M-1-m}{j}\binom{M+m+j}{M-1-m}(2^{k-1}s^{\beta})^{m+j} \\ &\quad\times\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})}. \end{aligned} \tag{16} $$

Taking the Riemann–Liouville integral on both sides of (16), we have

$$ {}_{0}^{RL}I_{s}^{\gamma}(\psi_{1,m}^{\beta}(s)) $$

$$= \xi_{mk} \sum_{j=0}^{M-m-1} (-1)^j \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} 2^{(k-1)(m+j)} \times {}_0^{RL}I_s^{\gamma}\left(s^{\beta(m+j)}\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} \right). \tag{17}$$

Now, with the help of the binomial expansion, we have

$$(2^{k-1}s^{\beta} - n + 1)^{m+j} = \sum_{p=0}^{m+j} \binom{m+j}{p} (2^{k-1}s^{\beta})^p (-1)^{m+j-p}(n-1)^{m+j-p}. \tag{18}$$

When $n = 2, 3, \ldots, 2^{k-1}$, using (18) in (15), we have

$$\begin{aligned} &\psi_{n,m}^{\beta}(s) \\ &= \xi_{mk} \sum_{j=0}^{M-m-1} (-1)^j \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} \sum_{p=0}^{m+j} \binom{m+j}{p} (2^{k-1}s^{\beta})^p \\ &\quad \times (-1)^{m+j-p}(n-1)^{m+j-p} \times \chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})}. \end{aligned} \tag{19}$$

Taking the Riemann–Liouville integral on both sides of (19), we have

$$\begin{aligned} &{}_0^{RL}I_s^{\gamma}(\psi_{n,m}^{\beta}(s)) \\ &= {}_0^{RL}I_s^{\gamma}\left(\xi_{mk} \sum_{j=0}^{M-m-1} (-1)^j \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} \sum_{p=0}^{m+j} \binom{m+j}{p} \right. \\ &\quad \left. \times (2^{k-1}s^{\beta})^p (-1)^{m+j-p}(n-1)^{m+j-p} \times \chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} \right), \end{aligned}$$

$$\begin{aligned} &{}_0^{RL}I_s^{\gamma}(\psi_{n,m}^{\beta}(s)) \\ &= \xi_{mk} \sum_{j=0}^{M-m-1} (-1)^j \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} \sum_{p=0}^{m+j} \binom{m+j}{p} 2^{(k-1)p} \\ &\quad \times (-1)^{m+j-p}(n-1)^{m+j-p} \times {}_0^{RL}I_s^{\gamma}\left(s^{\beta p}\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} \right). \end{aligned} \tag{20}$$

We approximate ${}_0^{RL}I_s^{\gamma}\left(s^{\beta p}\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} \right)$ by $2^{k-1}M$ terms of FCW, we have

$${}_0^{RL}I_s^{\gamma}\left(s^{\beta p}\chi(s)\big|_{s\in[(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}},(\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} \right)$$

$$= h_{np}(s) \approx \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} d_{vl}^{np} \psi_{v,l}^{\beta}(s) = D_{vl}^{\top} \Psi_{k,M}^{\beta}(s), \tag{21}$$

where $h_{np}(s) = \frac{1}{\Gamma(\gamma)} \int_0^s (s-z)^{\gamma-1} z^{\beta p} \chi(z)|_{z \in [(\frac{n-1}{2^{k-1}})^{\frac{1}{\beta}}, (\frac{n}{2^{k-1}})^{\frac{1}{\beta}})} dz$,
and $D_{vl}^{\top} =< h_{np}(s), \Psi_{k,M}^{\beta}(s) >$. Substituting (21) in (17) and (20) gives

$$\begin{aligned}
&{}_0^{RL}I_s^{\gamma}(\psi_{1,m}^{\beta}(s)) \\
&\approx \xi_{mk} \sum_{j=0}^{M-m-1} (-1)^j \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} 2^{(k-1)(m+j)} \\
&\quad \times \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} d_{vl}^{1m} \psi_{v,l}^{\beta}(s) \\
&= \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} \Lambda_{vl}^{\gamma;1m} \psi_{v,l}^{\beta}(s),
\end{aligned}$$

and

$$\begin{aligned}
&{}_0^{RL}I_s^{\gamma}(\psi_{n,m}^{\beta}(s)) \\
&\approx \xi_{mk} \sum_{j=0}^{M-m-1} \sum_{p=0}^{m+j} (-1)^{m+2j-p} \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} \binom{m+j}{p} \\
&\quad \times 2^{(k-1)p} (n-1)^{m+j-p} \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} d_{vl}^{np} \psi_{v,l}^{\beta}(s) \\
&= \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} \Lambda_{vl}^{\gamma;nm} \psi_{v,l}^{\beta}(s),
\end{aligned}$$

where

$$\Lambda_{vl}^{\gamma;nm} = \begin{cases} \xi_{mk} \sum_{j=0}^{M-m-1} (-1)^j \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} 2^{(k-1)(m+j)} d_{vl}^{1m}, & n = 1, \\ \xi_{mk} \sum_{j=0}^{M-m-1} \sum_{p=0}^{m+j} (-1)^{m+2j-p} \binom{M-1-m}{j} \binom{M+m+j}{M-1-m} \\ \qquad \times \binom{m+j}{p} (n-1)^{m+j-p} 2^{(k-1)p} d_{vl}^{np}, & n \neq 1. \end{cases}$$

Thus

$${}_0^{RL}I_s^{\gamma}(\Psi_{k,M}^{\beta}(s)) \approx \sum_{v=1}^{2^{k-1}} \sum_{l=0}^{M-1} \Lambda_{vl}^{\gamma;nm} \psi_{v,l}^{\beta}(s) = S^{\gamma} \Psi_{k,M}^{\beta}(s),$$

$$S^{\gamma}=\begin{bmatrix} \Lambda_{10}^{\gamma;10} & \cdots & \Lambda_{1M-1}^{\gamma;10} & \cdots & \Lambda_{2^{k-1}0}^{\gamma;10} & \cdots & \Lambda_{2^{k-1}M-1}^{\gamma;10} \\ \Lambda_{10}^{\gamma;11} & \cdots & \Lambda_{1M-1}^{\gamma;11} & \cdots & \Lambda_{2^{k-1}0}^{\gamma;11} & \cdots & \Lambda_{2^{k-1}M-1}^{\gamma;11} \\ \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \Lambda_{10}^{\gamma;2^{k-1}M-1} & \cdots & \Lambda_{1M-1}^{\gamma;2^{k-1}M-1} & \cdots & \Lambda_{2^{k-1}0}^{\gamma;2^{k-1}M-1} & \cdots & \Lambda_{2^{k-1}M-1}^{\gamma;2^{k-1}M-1} \end{bmatrix}.$$

The FCW operational matrix for $k=2, M=3, \beta=1$, and $\gamma=0.7$ is given as

$$S^{0.7}=\begin{bmatrix} 0.101545 & 0.1832 & 0.110702 & 0.0577836 & 0.0869881 & 0.100422 \\ -0.0414831 & 0.191149 & 0.312073 & 0.0856382 & 0.150711 & 0.179004 \\ 0.00760829 & -0.0286796 & 0.248597 & 0.189728 & 0.254836 & 0.268119 \\ 0 & 0 & 0 & 0.101545 & 0.1832 & 0.110702 \\ 0 & 0 & 0 & -0.0414831 & 0.191149 & 0.312073 \\ 0 & 0 & 0 & 0.00760829 & -0.0286796 & 0.248597 \end{bmatrix}.$$

□

6 The numerical method

In this section, we have considered the TFCDE with a variable coefficient

$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}}+a(s)\frac{\partial\mathcal{U}(s,t)}{\partial s}+b(s)\frac{\partial^{2}\mathcal{U}(s,t)}{\partial s^{2}}=f(s,t), \tag{22}$$

with I.C and B.Cs

$$\mathcal{U}(s,0)=q(s), \tag{23}$$

$$\mathcal{U}(0,t)=f_0(t), \qquad \mathcal{U}(1,t)=f_1(t). \tag{24}$$

where $a(s)$, $b(s)\neq 0$ are smooth functions, q, f_0, and $f_1\in L^2[0,1)$, and $f(\cdot,\cdot)$ is a function in $L^2([0,1)\times[0,1))$, with $\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}}={}_0^C D_t^{\gamma}\mathcal{U}(s,t)$.

To solve (22)–(23), we proceed as follows:

$$\frac{\partial^{2+\gamma}\mathcal{U}(s,t)}{\partial s^{2}\partial t^{\gamma}}\approx\Psi^{\alpha}_{k_1,M_1}(s)\cdot R\cdot\Psi^{\beta}_{k_2,M_2}(t), \tag{25}$$

where R is an unknown fractional Chelyshkov wavelet coefficient. We have

$$\int_0^s \Psi^{\alpha}_{k,M}(z)dz \approx S\Psi^{\alpha}_{k,M}(z), \tag{26}$$

where S denotes the integer-order integration operational matrix. Taking Riemann–Liouville fractional integral of order γ with respect to t on both sides of (25), we have

$$\begin{aligned}\frac{\partial^2 \mathcal{U}(s,t)}{\partial s^2} &\approx \frac{\partial^2 \mathcal{U}(s,t)}{\partial s^2}\Big|_{s=0} + \Psi^{\alpha}_{k_1,M_1}(s)R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)) \\ &\approx q''(s) + \Psi^{\alpha}_{k_1,M_1}(s)R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)).\end{aligned} \tag{27}$$

Again, integrating twice with respect to s, we have

$$\begin{aligned}\mathcal{U}(s,t) \approx \mathcal{U}(0,t) + s\frac{\partial \mathcal{U}(s,t)}{\partial s}\Big|_{s=0} + q(s) - q(0) - sq'(0) \\ + (S^2\Psi^{\alpha}_{k_1,M_1}(s))^{\top}R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)).\end{aligned} \tag{28}$$

Putting $s = 1$ in (28), we have

$$\begin{aligned}\mathcal{U}(1,t) \approx &\mathcal{U}(0,t) + \frac{\partial \mathcal{U}(s,t)}{\partial s}\Big|_{s=0} + q(1) - q(0) - q'(0) \\ &+ (S^2\Psi^{\alpha}_{k_1,M_1}(1))^{\top}R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)),\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{U}(s,t)}{\partial s}\Big|_{s=0} \approx &-(S^2\Psi^{\alpha}_{k_1,M_1}(1))^{\top}R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)) + q(0) \\ &+ q'(0) - q(1) + f_1(t) - f_0(t) \\ =&J(t).\end{aligned}$$

Now (28) gives

$$\begin{aligned}\mathcal{U}(s,t) \approx &\mathcal{U}(0,t) + sJ(t) + q(s) - q(0) - sq'(0) \\ &+ (S^2\Psi^{\alpha}_{k_1,M_1}(s))^{\top}R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)).\end{aligned} \tag{29}$$

Taking the Caputo derivative with respect to t, we obtain

$$\begin{aligned}\frac{\partial^{\gamma} \mathcal{U}(s,t)}{\partial t^{\gamma}} \approx &-s(S^2\Psi^{\alpha}_{k_1,M_1}(1))^{\top}R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)) + s\frac{\partial^{\gamma} f_1(t)}{\partial t^{\gamma}} + (1-s)\frac{\partial^{\gamma} f_0(t)}{\partial t^{\gamma}} \\ &+ (S^2\Psi^{\alpha}_{k_1,M_1}(s))^{\top}R\Psi^{\beta}_{k_2,M_2}(t).\end{aligned} \tag{30}$$

Differentiating (29) with respect to s, we get

$$\frac{\partial \mathcal{U}(s,t)}{\partial s} \approx J(t) + q'(s) - q'(0) + (S\Psi^{\alpha}_{k_1,M_1}(s))^{\top} R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)). \tag{31}$$

Substituting (27), (30), and (31) into (22). The system of algebraic equations with the collocation points from (10) is obtained as

$$\begin{aligned}
&(S^2\Psi^{\alpha}_{k_1,M_1}(s))^{\top} R\Psi^{\beta}_{k_2,M_2}(t) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1))^{\top} R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)) + s\frac{\partial^{\gamma} f_1(t)}{\partial t^{\gamma}} \\
&+ (1-s)\frac{\partial^{\gamma} f_0(t)}{\partial t^{\gamma}} + a(s)(J(t) + q'(s) - q'(0) \\
&+ (S\Psi^{\alpha}_{k_1,M_1}(s))^{\top} R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t))) \\
&+ b(s)(q''(s) + \Psi^{\alpha}_{k_1,M_1}(s) R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t))) = f(s,t).
\end{aligned} \tag{32}$$

We have solved the obtained system of equations using the LinearSolve command in Mathematica to obtain the value R. Finally, after putting the value of R in (28), we get the approximate solution.

7 Stability analysis

In this section, we present a perturbation analysis to study the stability of the proposed wavelet-based method. The system in (32) can be represented compactly in the following global matrix form

$$AC = F,$$

where A is the global operational matrix incorporating the matrix R along with the integer and fractional order operational matrix S, S^2, and S^{γ}, C represents the vector of unknown expansion coefficients, and F is the data vector containing the initial condition $q(s)$, boundary conditions $f_0(t), f_1(t)$, and the source term $f(s,t)$.

Next, we present the main stability theorem given as follows.

Theorem 2. Let $\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)$ be the approximate solution obtained by applying the proposed method, and let $\mathcal{U}^{P;\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)$ be the approximate solution obtained under perturbed conditions.

Case-1: If the perturbations occur only in the data vector F such that $A(C+\Delta C) = F + \Delta F$, and assuming the invertibility of A, then the stability error is bounded by

$$||\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) - \mathcal{U}^{P;\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)||$$

$$\leq |E(s,t)| + \Phi(s)\Phi(t) cond(A)\left(\frac{\|\Delta F\|}{\|F\|}\right)\|C\|.$$

Case-2: Let the perturbations simultaneously occur in both the global matrix A and the data vector F such that $(A+\Delta A)(C+\Delta C) = F + \Delta F$. Assuming the invertibility of A, and $cond(A)\left(\frac{\|\Delta A\|}{\|A\|}\right) < 1$, then the stability error is bounded by

$$\begin{aligned}
&||\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) - \mathcal{U}^{P;\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)|| \\
&\leq |E(s,t)| + \Phi(s)\Phi(t)\frac{cond(A)}{\left(1 - cond(A)\left(\frac{\|\Delta A\|}{\|A\|}\right)\right)} \\
&\quad \times \left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta F\|}{\|F\|}\right]\|C\|,
\end{aligned}$$

where $\Phi(s) = \|(S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))\|$, $\Phi(t) = \|S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)\|$, and $E(s,t) = |\Delta q(s)| + |(1-s)||\Delta f_0(t) - \Delta f_0(0)| + |s||\Delta f_1(t) - \Delta f_1(0)|$.

Proof. For the simplicity of notation, we denote $\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) = u(s,t)$, and $\mathcal{U}^{P;\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) = u^P(s,t)$. All matrix and vector norms reported in the proof are using infinity norms. From (29), we have the following representation for $u(s,t)$

$$\begin{aligned}
u(s,t) =& f_0(t) + q(s) - q(0) + s(q(0) - q(1) + f_1(t) - f_0(t)) \\
&+ (S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))^{\top} R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)).
\end{aligned} \tag{33}$$

The perturbed solution $u^P(s,t)$ after introducing a small perturbation into the initial data $(q(s) \to q(s) + \Delta q(s))$, the boundary data $(f_0(t) \to \Delta f_0(t))$, $(f_1(t) \to f_1(t) + \Delta f_1(t))$, and solution vector $(R \to R + \Delta R)$, is given as

$$\begin{aligned}
u^P(s,t) =& (f_0(t) + \Delta f_0(t)) + (q(s) + \Delta q(s)) - (q(0) + \Delta q(0)) \\
&+ s((q(0) + \Delta q(0)) - (q(1) + \Delta q(1)) + (f_1(t) + \Delta f_1(t)) \\
&- (f_0(t) + \Delta f_0(t)) + (S^2\Psi^{\alpha}_{k_1,M_1}(s) \\
&- s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))^{\top}(R + \Delta R)(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)).
\end{aligned} \tag{34}$$

Subtracting (34) from (33), and taking the norm on both sides

$$\begin{aligned}
\|u(s,t) - u^P(s,t)\| =& \| - \Delta f_0(t) - \Delta q(s) + \Delta q(0) - s(\Delta q(0) - \Delta q(1) \\
&+ \Delta f_1(t) - \Delta f_0(t)) + (S^2\Psi^{\alpha}_{k_1,M_1}(s)
\end{aligned}$$

$$- s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))^{\top}\Delta R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t))\|. \tag{35}$$

Now, we apply the physical compatibility condition at the intersection corner of the domain, such that $\Delta q(0) = \Delta f_0(0)$ and $\Delta q(1) = \Delta f_1(0)$, we get

$$\begin{aligned} &\Delta f_0(t) + \Delta q(s) - \Delta q(0) + s(\Delta q(0) - \Delta q(1) + \Delta f_1(t) - \Delta f_0(t)) \\ &= \Delta f_0(t) + \Delta q(s) - \Delta f_0(0) + s(\Delta f_0(0) - \Delta f_1(0) + \Delta f_1(t) - \Delta f_0(t)). \end{aligned} \tag{36}$$

Substituting (36) in (35), we get

$$\begin{aligned} \|u(s,t) - u^P(s,t)\| \leq\ & |\Delta q(s)| + |(1-s)||\Delta f_0(t) - \Delta f_0(0)| \\ & + |s||\Delta f_1(t) - \Delta f_1(0)| \\ & + \|(S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))^{\top}\Delta R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t))\|. \end{aligned} \tag{37}$$

Therefore

$$\begin{aligned} &\|u(s,t) - u^P(s,t)\| \\ &\leq |E(s,t)| + \|(S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))^{\top}\Delta R(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t))\|. \end{aligned} \tag{38}$$

Now, the Kronecker vectorization identity is given as

$$vec(X^{\top}YZ) = (Z^{\top} \otimes X^{\top})vec(Y), \tag{39}$$

where X, Y, and Z are any three matrices.

Using (39) in (38) and taking $\Delta C = vec(\Delta R)$, we have

$$\begin{aligned} &\|u(s,t) - u^P(s,t)\| \\ &\leq |E(s,t)| + \|(S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t))^{\top} \otimes (S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))\|\|\Delta C\|. \end{aligned} \tag{40}$$

For any matrix X and Y, the Kronecker matrix splitting norm property gives

$$\|X \otimes Y\| = \|X\|\|Y\|.$$

Thus

$$\begin{aligned} &\|u(s,t) - u^P(s,t)\| \\ &\leq |E(s,t)| + \|(S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))\|\|S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)\|\|\Delta C\|. \end{aligned} \tag{41}$$

Let us define

$$\Phi(s) = \|(S^2\Psi^{\alpha}_{k_1,M_1}(s) - s(S^2\Psi^{\alpha}_{k_1,M_1}(1)))\|, \qquad \Phi(t) = \|S^{\gamma}\Psi^{\beta}_{k_2,M_2}(t)\|.$$

Therefore

$$\|u(s,t) - u^P(s,t)\| \leq |E(s,t)| + \Phi(s)\Phi(t)\|\Delta C\|. \tag{42}$$

Case 1: If the perturbation occurs only in the data vector F such that

$$A(C + \Delta C) = F + \Delta F. \tag{43}$$

Subtracting $AC = F$ from (43), we get

$$A\Delta C = \Delta F.$$

Using the invertibility of A gives

$$\Delta C = A^{-1}\Delta F.$$

Taking the norm of both sides, we get

$$\|\Delta C\| \leq \|A^{-1}\|\|\Delta F\|. \tag{44}$$

Also, from $AC = F$, we have

$$\frac{1}{\|C\|} \leq \frac{\|A\|}{\|F\|}. \tag{45}$$

From (44) and (45), we have

$$\|\Delta C\| \leq \|A\|\|A^{-1}\|\left(\frac{\|\Delta F\|}{\|F\|}\right)\|C\|. \tag{46}$$

For any nonsingular matrix A, the condition number is defined as

$$cond(A) = \|A\|\|A^{-1}\|.$$

For numerical computation, we take the ratio

$$p_1 = \frac{\|\Delta C\|}{cond(A)\|C\|}\left(\frac{\|F\|}{\|\Delta F\|}\right).$$

Value $p_1 \leq 1$ confirms that the theoretical stability bound is satisfied.

Substituting (46) in (42), we get

$$\|u(s,t) - u^P(s,t)\| \leq |E(s,t)| + \Phi(s)\Phi(t)cond(A)\left(\frac{\|\Delta F\|}{\|F\|}\right)\|C\|. \tag{47}$$

Case 2: When the perturbation simultaneously occurs in both A and F such that

$$(A + \Delta A)(C + \Delta C) = F + \Delta F. \tag{48}$$

Subtracting $AC = F$ from (48) gives

$$A\Delta C + \Delta AC + \Delta A\Delta C = \Delta F.$$

Neglecting the higher order product $\Delta A\Delta C$ and taking the norm on both sides, we have

$$\|\Delta C\| \leq \|A^{-1}\|(\|\Delta F\| + \|\Delta A\|\|C\|). \tag{49}$$

Combining (45) and (49), we get

$$\|\Delta C\| \leq cond(A)\left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta F\|}{\|F\|}\right]\|C\|.$$

If $cond(A)\left(\frac{\|\Delta A\|}{\|A\|}\right) < 1$ holds, then we have the following sharper bound given as

$$\|\Delta C\| \leq \frac{cond(A)}{\left(1 - cond(A)\left(\frac{\|\Delta A\|}{\|A\|}\right)\right)}\left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta F\|}{\|F\|}\right]. \tag{50}$$

For numerical computation, we take the ratio

$$p_2 = \frac{\|\Delta C\|\left(1 - cond(A)\left(\frac{\|\Delta A\|}{\|A\|}\right)\right)}{cond(A)\|C\|\left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta F\|}{\|F\|}\right]}.$$

Value $p_2 \leq 1$ confirms that the theoretical stability bound is satisfied.

Substituting (50) in (42), we get

$$\|u(s,t) - u^P(s,t)\|$$

$$\leq |E(s,t)| + \Phi(s)\Phi(t)\frac{cond(A)}{\left(1 - cond(A)\left(\frac{\|\Delta A\|}{\|A\|}\right)\right)}\left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta F\|}{\|F\|}\right].$$

□

7.1 Effect of resolution parameters on stability

The stability bounds derived in Theorem (2)show the dependence on the choice of the resolution parameters (M_1, M_2) and index (k_1, k_2). These parameters affect the stability in two distinct ways:

1. **Wavelet basis norm inflation:** Clearly the term $\Phi(s)$ and $\Phi(t)$ are present in both the cases of stability proof. This term depends on M_1, M_2, k_1, and k_2. Consequently, the norms $\Phi(s)$ and $\Phi(t)$ scale monotonically with the increase in the degree of Chelyshkov wavelets. Hence, an increase in resolution parameters increases the value of norms available for error amplification.

2. **Explosion of condition number and the singular barrier:** As the resolution parameters M_1 and M_2 increase the rows of the global matrix A become increasingly linearly dependent. As a result, the condition number grows rapidly, and the large condition number makes the system hyper-sensitive. Also, from the above theorem of stability, the condition for the global matrix A must be $cond(A) < \frac{\|A\|}{\|\Delta A\|}$. Exceeding this limit causes exponential round-off error propagation. Therefore, while higher values of M_1 and M_2 minimize truncation errors, they also increase the system's sensitivity to input data and perturbations in the coefficient matrix. Hence, selecting optimal resolution parameters is necessary to maintain truncation error while ensuring round-off stability.

8 Convergence analysis

Theorem 3. For a positive integers k_1, k_2, the series solution $\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}$ converges to the exact solution $\mathcal{U}(s,t)$ as $M_1, M_2 \to \infty$.

Proof. For this, we have to prove that $\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)$ is a Cauchy sequence and it converges to $\mathcal{U}(s,t)$. Let $\hat{M}_1 > M_1$, $\hat{M}_2 > M_2$, and $\Lambda = [0,1) \times [0,1)$. We have

$$\mathcal{U}^{\alpha,\beta}_{k_1,k_2,\hat{M}_1,\hat{M}_2}(s,t) - \mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)$$

$$= \sum_{n_1=1}^{2^{k_1-1}} \sum_{n_2=1}^{2^{k_2-1}} \sum_{m_1=M_1}^{\hat{M}_1-1} \sum_{m_2=M_2}^{\hat{M}_2-1} r_{n_1,n_2,m_1,m_2} \psi^{\alpha}_{n_1,m_1}(s) \psi^{\beta}_{n_2,m_2}(t),$$

$$\begin{aligned} &\| \mathcal{U}^{\alpha,\beta}_{k_1,k_2,\hat{M}_1,\hat{M}_2}(s,t) - \mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) \|_{L^2(\Lambda)} \\ &= \sum_{n_1=1}^{2^{k_1-1}} \sum_{n_2=1}^{2^{k_2-1}} \sum_{m_1=M_1+1}^{\hat{M}_1-1} \sum_{m_2=M_2+1}^{\hat{M}_2-1} | \, r_{n_1,n_2,m_1,m_2} \, |^2, \end{aligned} \tag{51}$$

$$\begin{aligned} &\sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} | \, r_{n_1,n_2,m_1,m_2} \, |^2 \\ &= \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} | \left\langle \mathcal{U}(s,t), \psi^{\alpha}_{n_1,m_1}(s) \psi^{\beta}_{n_2,m_2}(t) \right\rangle |^2 \\ &\leq \|\mathcal{U}\|^2 < \infty. \end{aligned}$$

This implies $\sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} | \, r_{n_1,n_2,m_1,m_2} \, |^2$ is convergent and hence the term on the left side of (51) tends to 0 as $M_1, M_2 \to \infty$. We have $\left\{ \mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) \right\}_{M_1,M_2 \in \mathbb{N}}$ is a Cauchy sequence in $L^2(\Lambda)$ and since $L^2(\Lambda)$ is a Hilbert space and hence complete, there exists a function $\mathcal{V}(s,t)$ in $L^2(\Lambda)$ such that

$$\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) \to \mathcal{V}(s,t).$$

Now, we have to show that $\mathcal{U}(s,t) = \mathcal{V}(s,t)$. For this, we have

$$\begin{aligned} &\left\langle \mathcal{V}(s,t) - \mathcal{U}(s,t), \psi^{\alpha}_{n_1,m_1}(s) \psi^{\beta}_{n_2,m_2}(t) \right\rangle \\ &= \left\langle \mathcal{V}(s,t), \psi^{\alpha}_{n_1,m_1}(s) \psi^{\beta}_{n_2,m_2}(t) \right\rangle - \left\langle \mathcal{U}(s,t), \psi^{\alpha}_{n_1,m_1}(s) \psi^{\beta}_{n_2,m_2}(t) \right\rangle \\ &= \lim_{\substack{k_1,k_2 \to \infty \\ M_1,M_2 \to \infty}} \left\langle \mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t), \psi^{\alpha}_{n_1,m_1}(s) \psi^{\beta}_{n_2,m_2}(t) \right\rangle - r_{n_1,n_2,m_1,m_2} \\ &= r_{n_1,n_2,m_1,m_2} - r_{n_1,n_2,m_1,m_2} \\ &= 0. \end{aligned}$$

Therefore $\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}$ converges to $\mathcal{U}(s,t)$ as $M_1, M_2 \to \infty$. □

9 Error analysis

Suppose that $\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)$ is the best approximation to $\mathcal{U}(s,t) \in L^2(\Lambda)$ out of the linear space

$$\mathcal{O}^{\alpha,\beta}_{k_1,k_2,M_1,M_2} = span\left(\psi^{\alpha}_{n_1,m_1}(s)\psi^{\beta}_{n_2,m_2}(t)|1 \le n_j \le 2^{k_j-1}; 0 \le m_j \le M_j - 1, j = 1,2\right),$$

where $\Lambda = ([0,1) \times [0,1))$ and $\psi^{\alpha}_{n,m}(s)$ denotes the fractional Chelyshkov wavelet.

Lemma 2. Let ${}^{C}_{0}D^{i\alpha}_{s}({}^{C}_{0}D^{j\beta}_{t}\mathcal{U}(s,t)) \in C(\Lambda)$ and $|{}^{C}_{0}D^{i\alpha}_{s}({}^{C}_{0}D^{j\beta}_{t}\mathcal{U}(s,t))| \le \mathcal{R}^{\alpha,\beta}_{M_1,M_2}$ for $i = 0,1,\ldots,M_1, j = 0,1,\ldots,M_2$. The error bound for

$${}^{C}_{0}D^{\gamma}_{t}\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)$$

by using FCW for $M_2\beta > \gamma$ is given as

$$\left\|{}^{C}_{0}D^{\gamma}_{t}\mathcal{U}(s,t) - {}^{C}_{0}D^{\gamma}_{t}\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)\right\|_{L^2(\Lambda)} \le \frac{\mathcal{R}^{\alpha,\beta}_{M_1,M_2}}{\Gamma(M_1\alpha+1)\Gamma(M_2\beta+1-\gamma)\sqrt{(2M_1\alpha+1)(2M_2\beta+1-2\gamma)}}.$$

Proof. Proof follows from [22]. □

Lemma 3. According to Lemma 2 for $M_1\alpha > 2$, we have the following result:

$$\left\|{}^{C}_{0}D^{1}_{s}\mathcal{U}(s,t) - {}^{C}_{0}D^{1}_{s}\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)\right\|_{L^2(\Lambda)} \le \frac{\mathcal{R}^{\alpha,\beta}_{M_1,M_2}}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha)\sqrt{(2M_2\beta+1)(2M_1\alpha-1)}}.$$

$$\left\|{}^{C}_{0}D^{2}_{s}\mathcal{U}(s,t) - {}^{C}_{0}D^{2}_{s}\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)\right\|_{L^2(\Lambda)} \le \frac{\mathcal{R}^{\alpha,\beta}_{M_1,M_2}}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha-1)\sqrt{(2M_2\beta+1)(2M_1\alpha-3)}}.$$

Theorem 4. The error bound $(\mathcal{E}^{m_2,M_2}_{m_1,M_1})$ is given as

$$||\mathcal{E}^{m_2,M_2}_{m_1,M_1}||_{L^2(\Lambda)}$$

$$\leq \mathcal{R}^{\alpha,\beta}_{M_1,M_2}\Bigg(\frac{1}{\Gamma(M_1\alpha+1)\Gamma(M_2\beta+1-\gamma)\sqrt{(2M_1\alpha+1)(2M_2\beta+1-2\gamma)}} + \frac{\mathcal{M}_1}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha)\sqrt{(2M_2\beta+1)(2M_1\alpha-1)}} + \frac{\mathcal{M}_2}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha-1)\sqrt{(2M_2\beta+1)(2M_1\alpha-3)}}\Bigg),$$

where $\|a(s)\| \leq \mathcal{M}_1$ and $\|b(s)\| \leq \mathcal{M}_2$.

Proof. We have

$$\begin{aligned}\|\mathcal{E}^{m_2,M_2}_{m_1,M_1}\|_{L^2(\Lambda)} =& \|{}^C_0D^\gamma_t\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) + a(s){}^C_0D^1_s\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)\\ &+ b(s){}^C_0D^2_s\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) - f(s,t)\|\end{aligned}$$

Using (22), we get

$$\begin{aligned}\|\mathcal{E}^{m_2,M_2}_{m_1,M_1}\|_{L^2(\Lambda)} =& \|{}^C_0D^\gamma_t\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) + a(s){}^C_0D^1_s\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)\\ &+ b(s){}^C_0D^2_s\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t)\\ &- {}^C_0D^\gamma_t\mathcal{U}(s,t) - a(s){}^C_0D^1_s\mathcal{U}(s,t) - b(s){}^C_0D^2_s\mathcal{U}(s,t)\|\\ \leq& \|{}^C_0D^\gamma_t\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) - {}^C_0D^\gamma_t\mathcal{U}(s,t)\|\\ &+ \|a(s)\|\|{}^C_0D^1_s\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) - {}^C_0D^1_s\mathcal{U}(s,t)\|\\ &+ \|b(s)\|\|{}^C_0D^2_s\mathcal{U}^{\alpha,\beta}_{k_1,k_2,M_1,M_2}(s,t) - {}^C_0D^2_s\mathcal{U}(s,t)\|.\end{aligned}$$

Using Lemmas 2 and 3, we have

$$\begin{aligned}\|\mathcal{E}^{m_2,M_2}_{m_1,M_1}\|_{L^2(\Lambda)} \leq& \frac{\mathcal{R}^{\alpha,\beta}_{M_1,M_2}}{\Gamma(M_1\alpha+1)\Gamma(M_2\beta+1-\gamma)\sqrt{(2M_1\alpha+1)(2M_2\beta+1-2\gamma)}}\\ &+ \mathcal{M}_1\frac{\mathcal{R}^{\alpha,\beta}_{M_1,M_2}}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha)\sqrt{(2M_2\beta+1)(2M_1\alpha-1)}}\\ &+ \mathcal{M}_2\frac{\mathcal{R}^{\alpha,\beta}_{M_1,M_2}}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha-1)\sqrt{(2M_2\beta+1)(2M_1\alpha-3)}}.\end{aligned}$$

From above, it follows that

$$\begin{aligned}&\|\mathcal{E}^{m_2,M_2}_{m_1,M_1}\|_{L^2(\Lambda)}\\ &\leq \mathcal{R}^{\alpha,\beta}_{M_1,M_2}\Bigg(\frac{1}{\Gamma(M_1\alpha+1)\Gamma(M_2\beta+1-\gamma)\sqrt{(2M_1\alpha+1)(2M_2\beta+1-2\gamma)}}\end{aligned}$$

$$+\frac{\mathcal{M}_1}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha)\sqrt{(2M_2\beta+1)(2M_1\alpha-1)}}$$
$$+\frac{\mathcal{M}_2}{\Gamma(M_2\beta+1)\Gamma(M_1\alpha-1)\sqrt{(2M_2\beta+1)(2M_1\alpha-3)}}\Bigg).$$

Therefore as $M_1, M_2 \to \infty$

$$\|\mathcal{E}^{m_2,M_2}_{m_1,M_1}\|_{L^2(\Lambda)} \to 0.$$

Also, if k_1, $k_2 \to \infty$, then

$$\left|\left(\frac{n_1}{2^{k_1-1}}\right)^{\frac{1}{\alpha}} - \left(\frac{n_1-1}{2^{k_1-1}}\right)^{\frac{1}{\alpha}}\right| \to 0, \left|\left(\frac{n_2}{2^{k_2-1}}\right)^{\frac{1}{\beta}} - \left(\frac{n_2-1}{2^{k_2-1}}\right)^{\frac{1}{\beta}}\right| \to 0.$$

This implies

$$\|\mathcal{E}^{m_2,M_2}_{m_1,M_1}\|_{L^2(\Lambda)} \to 0.$$

□

10 Numerical examples

In this section, we investigate the accuracy of the proposed method using six numerical examples. All computations are performed using Wolfram Mathematica 12.0. The resulting linear algebraic systems obtained after applying the proposed method are solved using the LinearSolve command in Mathematica. The computations are handled by Mathematica's default machine-precision arithmetic. In one numerical example, errors as low as 10^{-19} are interpreted as agreement at the level of machine precision. The condition number in that case is 1381.17. A condition number of this magnitude indicates that the system is well-conditioned and that round-off errors are negligibly amplified. We have taken $k_1 = k_2 = k$ and $M_1 = M_2 = M$, and $n = 2^{k-1}M$ denotes the number of basis functions. The formula for absolute error is

$$\mathcal{E}(s,t) = \mid u(s,t) - \mathcal{U}(s,t) \mid, \tag{52}$$

where $\mathcal{U}(s,t)$, $u(s,t)$ denote the exact and approximate solutions, respectively.

Example 1. Consider the following TFCDE:

$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}} - \frac{\partial^{2}\mathcal{U}(s,t)}{\partial s^{2}} = 2\left(\frac{1}{\Gamma(3-\gamma)}t^{2-\gamma} - 1\right), \tag{53}$$

with I.C and B.Cs

$$\mathcal{U}(s,0) = s^{2}, \quad s \in (0,1), \qquad \mathcal{U}(0,t) = t^{2}, \quad \mathcal{U}(1,t) = 1 + t^{2}, \quad t \in (0,1].$$

The exact solution is given as

$$\mathcal{U}(s,t) = s^{2} + t^{2}. \tag{54}$$

Table 1 compares the absolute error of the proposed method with BFM [7] and FWM [38] at $\alpha = \beta = 1$, $\gamma = 0.5$, and $t = 0.25$. In Table 2, we present the absolute error of the proposed method for different values of γ and compare it with FWM [38]. These tables show that the proposed method gives better results with a small number of basis functions. To evaluate the computational efficiency of the proposed method, Table 3 presents the maximum absolute error, CPU time, and memory usage for different values of the resolution parameters. Clearly, CPU time and memory usage increase with higher resolution parameters. To confirm the numerical stability of the proposed method, Table 4 presents the logarithmic condition number and stability measures (p_1, p_2) for different values of n. The condition number increases with an increase in n, indicating greater sensitivity of the numerical system to perturbations. The stability parameters $p_1 \leq 1$ and $p_2 \leq 1$ are satisfied for different n, confirming the stability of the proposed method. Figure 1 shows the plot of exact, approximate, and absolute error with $\alpha = \beta = 0.1$ and $\gamma = 0.8$.

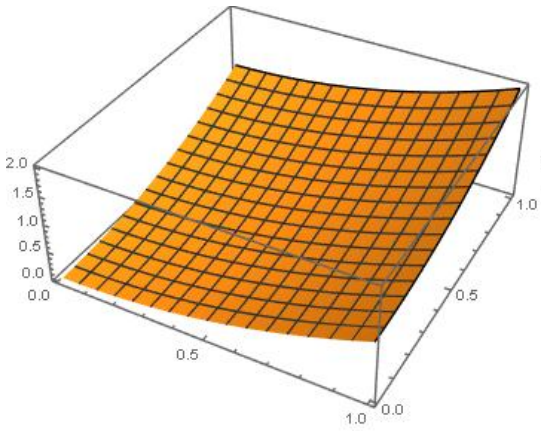
Table 1: Comparison of $\mathcal{E}(s,t)$ of Example 1 at $\alpha = \beta = 1$, $\gamma = 0.5$, and $t = 0.25$.

s	BFM [7]		FWM [38]			Proposed method
	$j = 1, m = 2$	$j = 1, m = 3$	$j = 2, m = 2$	$k = 2, M = 3$	$k = 3, M = 4$	$k = 2,\ M = 3$
0.2	3.3e-02	4.4e-03	8.8e-02	2.0561e-04	6.3909e-05	1.10834e-08
0.4	1.9e-04	5.1e-02	9.8e-02	5.2032e-04	8.6140e-05	4.62512e-08
0.6	1.6e-02	7.1e-02	3.4e-01	9.9285e-05	2.541e-04	1.07939e-07
0.8	1.2e-01	2.8e-02	4.3e-01	9.9285e-05	5.6495e-07	2.05404e-07

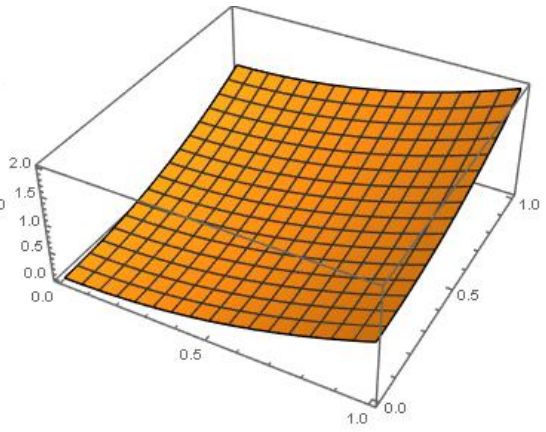
Example 2. Consider the following TFCDE:

$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}} + s\frac{\partial\mathcal{U}(s,t)}{\partial s} + \frac{\partial^{2}\mathcal{U}(s,t)}{\partial s^{2}} = 2t^{\gamma} + 2s^{2} + 2, \tag{55}$$

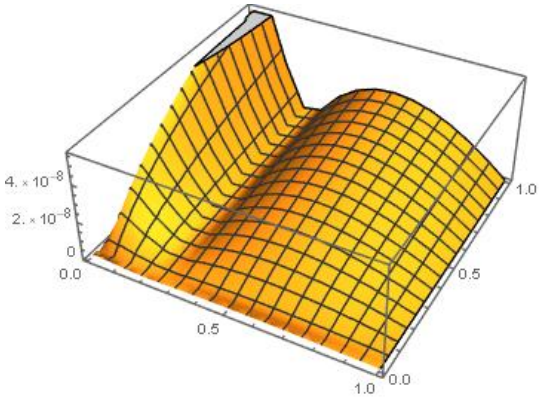
with I.C and B.Cs



(a) Exact plot.



(b) Approximate plot.



(c) Absolute error plot.

Figure 1: Plot of exact, approximate, and absolute error of Example 1 with $\alpha = \beta = 0.1$, $\gamma = 0.8$.

$$\mathcal{U}(s,0) = s^2, \quad s \in (0,1),$$

$$\mathcal{U}(0,t) = \frac{2\Gamma(\gamma+1)}{\Gamma(2\gamma+1)}t^{2\gamma}, \qquad \mathcal{U}(1,t) = 1 + \frac{2\Gamma(\gamma+1)}{\Gamma(2\gamma+1)}t^{2\gamma}, \quad t \in (0,1].$$

The exact solution is given as

$$\mathcal{U}(s,t) = s^2 + \frac{2\Gamma(\gamma+1)}{\Gamma(2\gamma+1)}t^{2\gamma}. \tag{56}$$

Table 5 shows a comparison of the absolute error of the proposed method with the methods in [24], [2], [26], and [38]. This table confirms that the proposed method performs better than the other methods, and the fractional power in the Chelyshkov wavelet makes it more accurate. Table 6 shows the maximum absolute error, CPU time, and memory used for different values

Table 2: Comparison of $\mathcal{E}(s,t)$ of Example 1 for different values of γ.

(s,t)	FWM [38] $(k=3,M=4)$ $\gamma=0.8$	FWM [38] $(k=3,M=4)$ $\gamma=0.9$	Proposed method $(k=2,M=3)$ $\alpha=\beta=0.1$, $\gamma=0.8$	Proposed method $(k=2,M=3)$ $\alpha=\beta=0.1$, γ=0.9
(0.1,0.1)	4.1858e-08	7.4698e-07	7.14461e-09	5.01525e-09
(0.2,0.2)	6.6925e-07	2.8060e-07	3.66583e-09	2.79564e-09
(0.3,0.3)	7.5506e-08	1.2633e-08	4.97482e-09	3.99839e-09
(0.4,0.4)	8.8604e-07	8.1873e-08	1.33681e-08	1.07923e-08
(0.5,0.5)	4.1223e-08	2.4966e-06	1.95202e-08	1.58519e-08
(0.6,0.6)	4.2531e-06	1.3715e-08	2.26026e-08	1.243881e-08
(0.7,0.7)	9.9860e-08	1.4136e-08	2.22673e-08	1.32283e-08
(0.8,0.8)	1.3350e-08	2.1795e-08	1.33388e-08	1.5095e-08
(0.9,0.9)	4.0122e-07	9.3121e-09	1.09523e-08	9.01145e-09

Table 3: Maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha=\beta=0.1$ and $\gamma=0.8$.

	Maximum Absolute Error	CPU Time (seconds)	Memory used (MB)
$k=2,M=3$	4.34443e-07	6.641	2.82407
$k=2,M=4$	1.80993e-07	14.203	9.40698
$k=2,M=5$	1.30899e-07	15.213	21.6597
$k=3,M=4$	1.02445e-08	19.313	50.2342

of k and M with $\alpha=\beta=1,\gamma=0.5$. Table 7 presents the logarithmic condition number and numerical stability parameters for $\alpha=\beta=0.1$ and $\gamma=0.5$, which confirm the numerical stability of the proposed method. Figure 2 shows the plot of exact, approximate, and absolute error for $\alpha=\beta=0.1$, $\gamma=0.5$, $k=2$, and $M=3$.

Example 3. Consider the following TFCDE:

$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}}+\frac{\partial\mathcal{U}(s,t)}{\partial s}-s\frac{\partial^{2}\mathcal{U}(s,t)}{\partial s^{2}}=f(s,t), \tag{57}$$

with I.C and B.Cs

$$\mathcal{U}(s,0)=e^{s},\quad s\in(0,1),$$

$$\mathcal{U}(0,t)=(t^{2}+1)(1-t),\qquad \mathcal{U}(1,t)=(t^{2}+1)(e-t),\quad t\in(0,1],$$

Table 4: The value of logarithmic condition number and numerical stability parameters with $\alpha = \beta = 0.1$, $\gamma = 0.8$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$.

n	$\log_{10}(cond(A))$	p_1	p_2
4	5.7190	3.77442×10^{-02}	3.77422×10^{-02}
6	6.9592	2.18827×10^{-03}	2.18627×10^{-03}
8	9.3294	7.40171×10^{-06}	5.81816×10^{-06}
10	9.4165	5.35671×10^{-06}	5.25612×10^{-06}

Table 5: Comparison of $\mathcal{E}(s,t)$ of Example 2 at $\gamma = 0.5$, $t = 0.5$.

(s,t)	Sinc-Legendre [24]		B-spline Sinc [2]		Chebyshev [26]	FWM [38]	Proposed method	
	$n = 15$	$n = 25$	$n = 5$	$n = 20$	$n = 5$	$n = 16$	$n = 6$ $\alpha = \beta = 1$	$n = 6$ $\alpha = \beta = 0.1$
0.1	6.994e-05	6.462e-06	6.481e-04	1.369e-09	7.964e-06	3.6637e-15	6.1886e-19	3.71677e-19
0.2	1.721e-04	1.578e-05	4.109e-04	7.591e-10	3.912e-06	1.1102e-14	9.07256e-19	1.83842e-19
0.3	2.472e-04	2.272e-05	5.493e-04	1.184e-09	6.162e-06	8.8818e-16	1.39548e-18	4.91185e-19
0.4	2.912e-04	2.674e-05	5.198e-04	1.068e-09	5.953e-06	5.1071e-15	2.08354e-18	6.42139e-19
0.5	3.004e-04	2.759e-05	4.912e-04	9.819e-10	2.103e-06	4.4409e-16	3.56819e-18	6.85571e-19
0.6	2.760e-04	2.534e-05	5.063e-04	1.039e-09	7.639e-06	2.2204e-16	3.35632e-18	6.49804e-19
0.7	2.213e-04	2.035e-05	5.045e-04	1.031e-09	1.967e-06	0	0	0
0.8	1.440e-04	1.320e-05	5.040e-04	1.030e-09	8.103e-06	0	0	0
0.9	5.026e-05	4.653e-06	5.037e-04	1.031e-09	6.019e-06	0	0	0

and $f(s,t) = \frac{2t^{2-\gamma}}{\Gamma(3-\gamma)}e^s - \frac{t^{1-\gamma}}{\Gamma(2-\gamma)} - \frac{6t^{3-\gamma}}{4-\gamma} + (t^2+1)(1-s)e^s$.

The exact solution is given as

$$\mathcal{U}(s,t) = (t^2+1)(e^s - t). \tag{58}$$

We have used the proposed method to solve the above problem for $k = 2$ and $M = 3$. The comparison of the absolute error with the method in [41] for various values of γ is reported in Table 8. Table 9 shows the maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = \gamma = 0.3$. Table 10 confirms the numerical stability of the proposed method. Figure 3 shows the plot of exact and approximate solutions for $\alpha = \beta = 1, \gamma = 0.9$, $k = 2$, and $M = 3$.

Example 4. Consider the following TFCDE:

$$\frac{\partial^\gamma \mathcal{U}(s,t)}{\partial t^\gamma} + t^2 e^s \frac{\partial \mathcal{U}(s,t)}{\partial s} - (s^2+t+1)\frac{\partial^2 \mathcal{U}(s,t)}{\partial s^2} = f(s,t), \tag{59}$$

with I.C and B.Cs

Table 6: Maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = 1, \gamma = 0.5$.

	Maximum Absolute Error	CPU Time (seconds)	Memory used (MB)
$k = 2, M = 3$	3.76595e-18	8.767	1.34804
$k = 2, M = 4$	1.19022e-19	16.531	3.46304
$k = 2, M = 5$	1.03654e-19	35.121	7.53673
$k = 3, M = 3$	1.00454e-19	47.528	8.23176

Table 7: The value of logarithmic condition number and numerical stability parameters with $\alpha = \beta = 0.1$, $\gamma = 0.5$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$.

n	$\log_{10}(cond(A))$	p_1	p_2
4	4.1318	1.3326×10^{-02}	1.21996×10^{-02}
6	5.9837	2.0292×10^{-02}	2.0292×10^{-02}
8	8.2712	1.07868×10^{-04}	1.05834×10^{-04}
10	8.6376	1.61878×10^{-06}	1.53048×10^{-05}

Table 8: Comparison of absolute errors for Example 3 with $\alpha = \beta = \gamma$.

(s,t)	Method in [41] ($k = 2, M = 6$)			Proposed method ($k = 2, M = 3$)		
	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 0.9$	$\gamma = 0.3$	γ=0.5	$\gamma = 0.9$
(0.1,0.1)	7.75e-06	1.20e-05	2.55e-05	2.71477e-06	2.42987e-06	4.63002e-06
(0.2,0.2)	1.70e-05	2.33e-05	3.50e-05	3.80907e-06	1.13041e-05	1.02384e-05
(0.3,0.3)	2.44e-05	3.02e-05	2.55e-05	5.05275e-06	2.67296e-05	1.51364e-05
(0.4,0.4)	2.87e-05	3.37e-05	2.71e-05	5.19807e-06	3.0366e-05	1.70165e-06
(0.5,0.5)	3.03e-05	3.31e-05	2.07e-05	4.7934e-06	8.16393e-06	1.94832e-06
(0.6,0.6)	2.88e-05	3.00e-05	1.51e-05	4.07959e-06	9.31634e-06	1.48382e-06
(0.7,0.7)	2.45e-05	2.45e-05	1.03e-05	3.18235e-06	6.03367e-06	1.33323e-06
(0.8,0.8)	1.79e-05	1.73e-05	6.27e-06	2.175e-06	3.51789e-06	1.22325e-06
(0.9,0.9)	9.58e-06	8.97e-06	2.91e-06	1.10383e-06	1.32524e-06	8.7549e-07

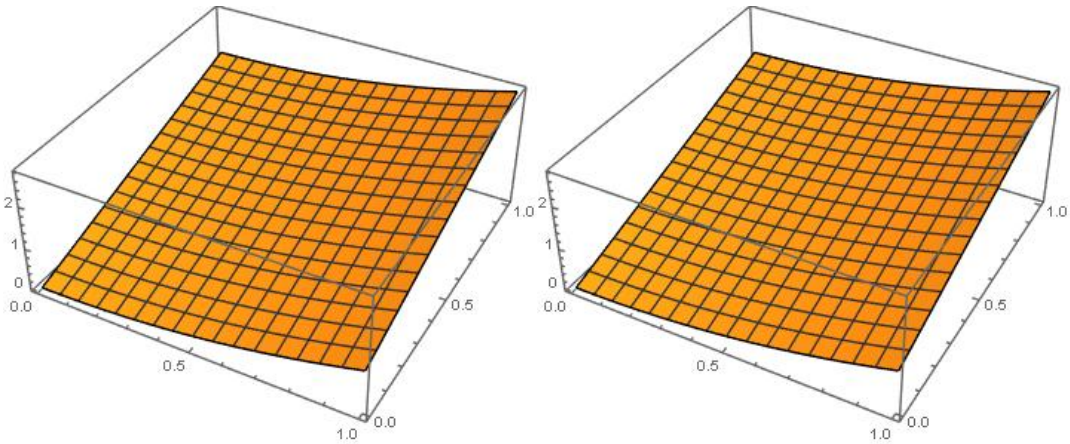
$$\mathcal{U}(s,0) = s^2, \quad s \in (0,1), \qquad \mathcal{U}(0,t) = t^2, \qquad \mathcal{U}(1,t) = 1 + t^2, \quad t \in (0,1],$$

and $f(s,t) = \frac{2t^{2-\gamma}}{\Gamma(3-\gamma)} - 2(s^2 + t + 1) + 2t^2 s e^s$.

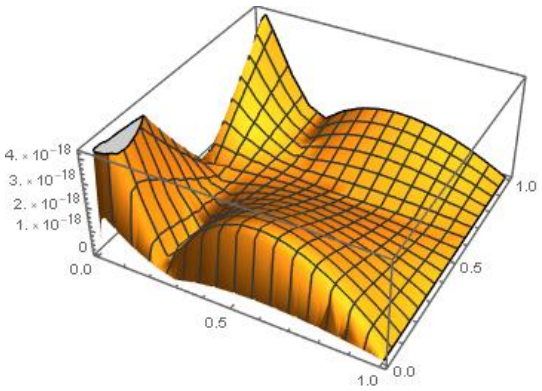
The exact solution is given as

$$\mathcal{U}(s,t) = s^2 + t^2. \tag{60}$$

Table 11 shows the absolute error for different values of α, β, γ, indicating that the fractional power of the Chelyshkov wavelet yields a more accurate result. Table 12 shows the maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = \gamma = 0.1$.



(a) Exact plot. (b) Approximate plot.



(c) Absolute error plot.

Figure 2: Plot of the exact, approximate, and absolute error of Example 2.

Table 13 confirms the numerical stability of the proposed method. Figure 4 shows the graph of exact, approximate, and absolute error for $\alpha = \beta = \gamma = 0.1$, $k = 2$, and $M = 3$; it is clear that by employing a few terms of fractional-order Chelyshkov wavelets, we can obtain a good approximation of the exact solution.

Example 5. Consider the following TFCDE:

$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}} + (1+0.5s)\frac{\partial \mathcal{U}(s,t)}{\partial s} - 0.2(1+s^2)\frac{\partial^2 \mathcal{U}(s,t)}{\partial s^2} = 0, \tag{61}$$

with I.C and B.Cs

Table 9: Maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = \gamma = 0.3$.

	Maximum Absolute Error	CPU Time (seconds)	Memory used (MB)
$k = 2, M = 3$	5.28105e-06	8.126	4.47313
$k = 2, M = 4$	2.3214e-06	9.283	34.618
$k = 2, M = 5$	1.73625e-06	19.818	50.9134
$k = 3, M = 3$	3.26535e-07	39.238	61.2142

Table 10: The value of logarithmic condition number and numerical stability parameters with $\alpha = \beta = \gamma = 0.3$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$.

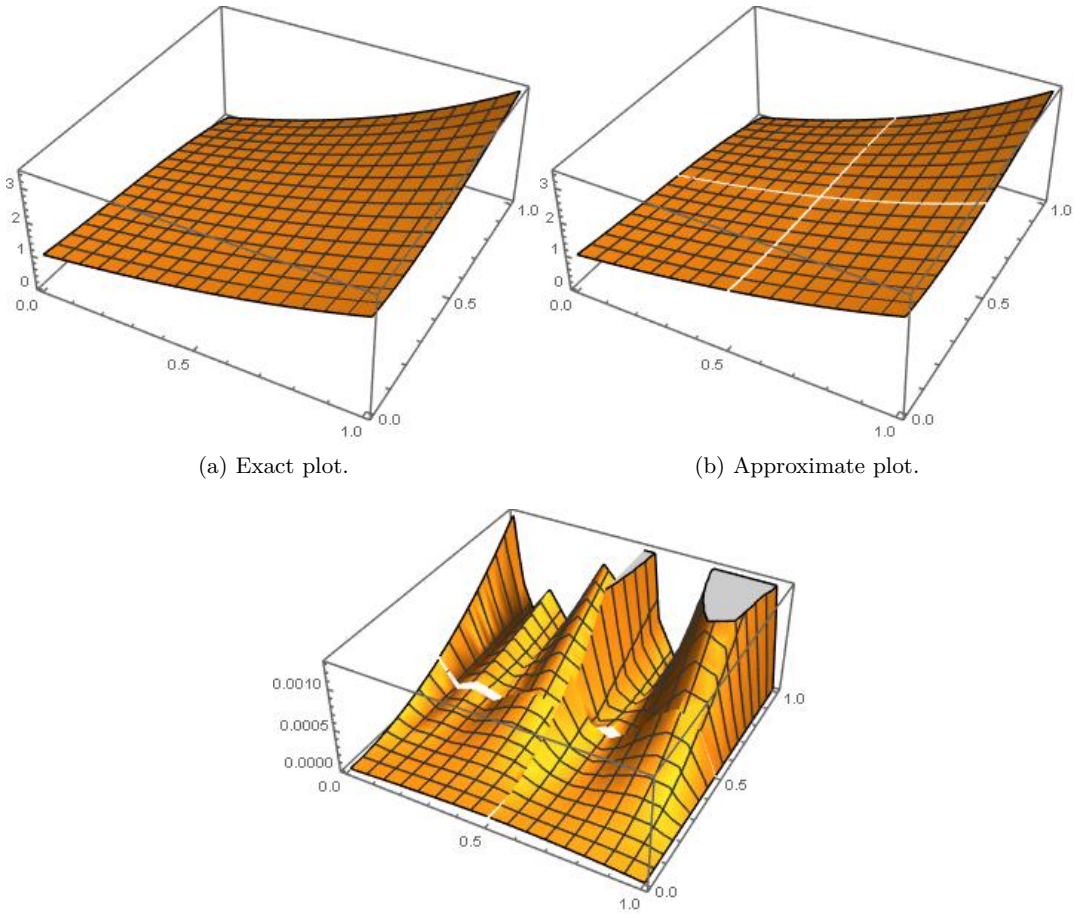
n	$\log_{10}(cond(A))$	p_1	p_2
4	3.8886	1.47854×10^{-01}	1.35767×10^{-01}
6	5.5562	5.55557×10^{-02}	5.55558×10^{-02}
8	7.6073	3.9012×10^{-04}	3.88542×10^{-04}
10	7.8183	5.2348×10^{-06}	5.1574×10^{-06}

Table 11: The absolute error of Example 4 with $k = 2, M = 3$, and $t = 0.1$.

	$\alpha = \beta = 1$		$\alpha = \beta = \gamma$	
s	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.1$	$\gamma = 0.3$
0.1	2.83606e-04	3.02301e-04	4.92041e-05	3.63389e-05
0.2	5.51765e-04	5.86786e-04	2.31822e-05	5.27762e-05
0.3	7.88299e-04	8.35926e-04	6.04444e-05	8.46832e-05
0.4	9.93208e-04	1.04971e-03	4.63942e-05	9.33962e-04
0.5	8.80454e-04	9.11348e-04	1.06622e-04	8.98575e-04
0.6	1.22125e-03	1.27917e-03	1.31595e-05	7.889486e-04
0.7	1.16542e-03	1.21533e-03	1.2884e-05	6.3249e-04
0.8	9.43269e-04	9.80856e-04	1.03605e-05	4.442731e-05
0.9	5.54796e-04	5.75745e-04	5.97019e-05	2.29777e-05

$$\mathcal{U}(s,0) = e^{-10s}, \quad s \in (0,1), \qquad \mathcal{U}(0,t) = 1, \qquad \mathcal{U}(1,t) = e^{-10}, \quad t \in (0,1].$$

Here $\mathcal{U}(s,t)$ denotes the pollutant concentration, $1 + 0.5s$ denotes spatially varying groundwater velocity, and $0.2(1 + s^2)$ is a spatially varying diffusion coefficient. The initial condition shows that the pollutant concentration is highest near the inlet ($s = 0$) and decays exponentially along the flow direction. The boundary condition indicates that a constant pollutant source of concentration 1 is maintained at the inlet, and the outlet concentration is fixed at e^{-10}. The fractional order $0 < \gamma < 1$ denotes anomalous transport with a memory effect due to trapping and heterogeneous pore structures, and $\gamma = 1$ shows the classical transport. The exact solution



(a) Exact plot.

(b) Approximate plot.

(c) Absolute error plot.

Figure 3: Plot of exact, approximate, and absolute error of Example 3.

of the governing equation is not known. Therefore, the proposed wavelet method is employed to obtain an approximate solution. The accuracy of the numerical scheme is assessed by comparing successive approximations obtained with different wavelet resolutions. Table 14 presents the error values obtained by subtracting the numerical solutions corresponding to $k = 1, M = 7$ and $k = 1, M = 8$ at selected spatial and temporal points with $\alpha = \beta = 1$ and $\gamma = 0.8$. Table 15 shows the maximum absolute error, CPU time, and memory used for different values of k and M. The value of logarithmic condition number and numerical stability parameters in Table 16 with $\alpha = \beta = 1, \gamma = 0.8$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$ confirms the stability of the proposed method. Figure 5 shows the plot of the approximate solution for different values of M with $\alpha = \beta = 1$, and $\gamma = 0.8$.

Table 12: Maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = \gamma = 0.1$.

	Maximum Absolute Error	CPU Time (seconds)	Memory used (MB)
$k = 2, M = 3$	4.99892e-04	7.999	1.60871
$k = 2, M = 4$	5.32841e-05	8.016	3.3327
$k = 2, M = 5$	3.79412e-06	8.234	7.2016
$k = 3, M = 3$	1.21131e-07	10.412	8.04546

Table 13: The value of logarithmic condition number and numerical stability parameters with $\alpha = \beta = \gamma = 0.1$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$.

n	$\log_{10}(cond(A))$	p_1	p_2
4	3.4931	1.67239×10^{-01}	1.67219×10^{-01}
6	4.9017	2.87983×10^{-01}	2.8798×10^{-01}
8	6.1741	2.87199×10^{-02}	2.87156×10^{-02}
10	6.9410	2.18385×10^{-03}	2.18194×10^{-03}

Table 14: The value of error at different spatial and temporal points with $\alpha = \beta = 1$ and $\gamma = 0.8$.

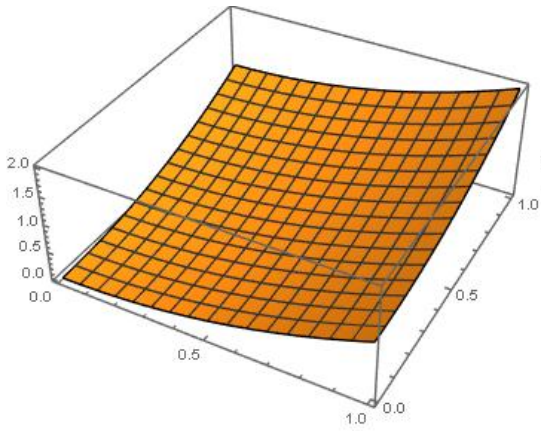
s	$t = 0.25$	$t = 0.50$	$t = 0.75$	$t = 1.00$
0.1	3.52011e-03	3.35035e-03	9.91772e-03	4.82175e-02
0.2	1.380097e-02	1.1367e-04	1.0585599e-02	8.173366e-02
0.3	1.443829e-02	4.84574e-03	1.684423e-02	9.650087e-02
0.4	1.373029e-02	9.63895e-03	2.093999e-02	9.766614e-02
0.5	1.670874e-02	9.20784e-03	1.893917e-02	8.86383e-02
0.6	1.920814e-02	7.58296e-03	1.574153e-02	2.154644e-02
0.7	1.637163e-02	9.18778e-03	1.608100e-02	7.313191e-02
0.8	1.171127e-02	9.72846e-03	1.535096e-02	6.388752e-02
0.9	1,284100e-02	5.0756e-04	4.09452e-03	3.586075e-02

Example 6. To show the practical applicability of the proposed method, we consider the dispersion of sulfur dioxide (SO_2) in the atmosphere. The transport process is governed by the time-fractional convection-diffusion equation given as

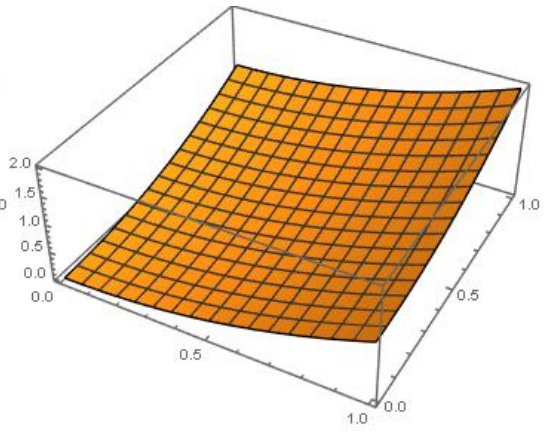
$$\frac{\partial^{\gamma}\mathcal{U}(s,t)}{\partial t^{\gamma}} + v\frac{\partial \mathcal{U}(s,t)}{\partial s} - D\frac{\partial^2 \mathcal{U}(s,t)}{\partial s^2} = 0, \tag{62}$$

with I.C and B.Cs

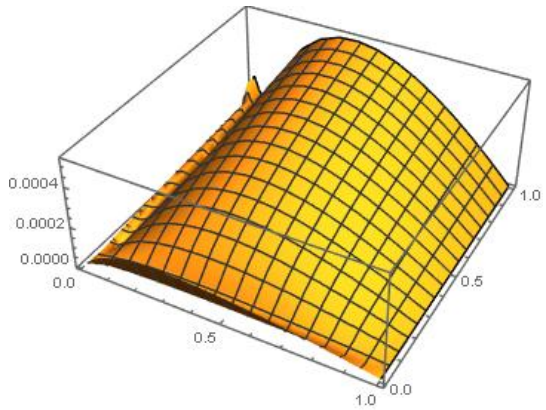
$$\mathcal{U}(s,0) = \sin(\pi s), \quad s \in (0,1), \qquad \mathcal{U}(0,t) = 0, \qquad \mathcal{U}(1,t) = 0, \quad t \in (0,1],$$



(a) Exact plot.



(b) Approximate plot.



(c) Absolute error plot.

Figure 4: Plot of exact, approximate, and absolute error of Example 4.

where $\mathcal{U}(s,t)$ denotes the pollutant concentration, D is the diffusion coefficient, and v represents the transport velocity. We take $D = 3.6 \times 10^{-4} cm^2/s$ and $v = 3.6 \times 10^{-3} cm/s$ from a recent atmospheric pollution study involving (SO_2) dispersion.

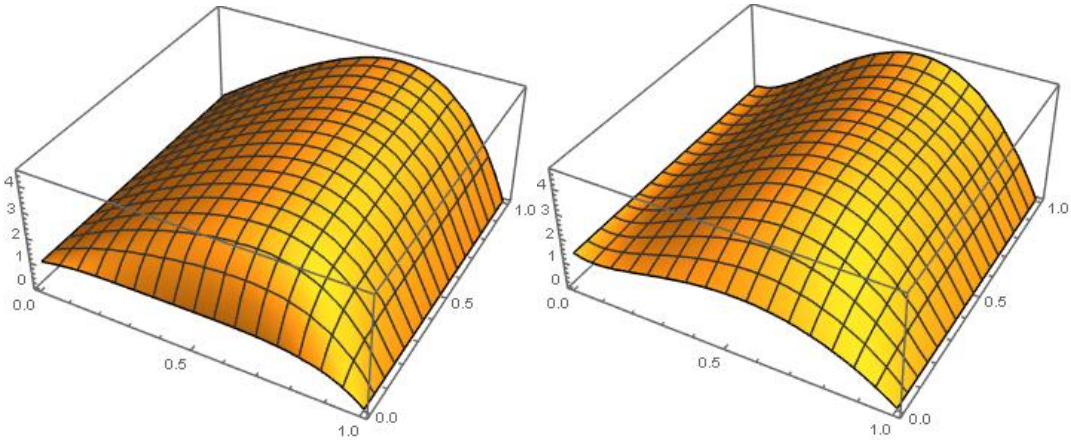
The exact solution of this problem is given as $\mathcal{U}(s,t) = \sin(\pi s)\mathrm{ML}_\gamma\left(-D\pi^2 t^\gamma\right)$, where ML_γ denotes the Mittag-Leffler function. The memory effects in transport processes are captured by the fractional order γ. When $0 < \gamma < 1$, the equation exhibits subdiffusive behavior. In this situation, the pollutants remain trapped for longer periods due to weak atmospheric mixing and stable meteorological conditions. This type of behavior is observed during winter inversions, when pollutants accumulate, leading to poor air quality. When $\gamma = 1$, the transport process is classical diffusion. We have employed FCW to compute the concentration profiles at different times. The

Table 15: Maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = 1, \gamma = 0.8$.

	Maximum Absolute Error	CPU Time (seconds)	Memory used (MB)
$k = 1, M = 5$	3.77118e-01	8.468	1.76382
$k = 1, M = 6$	1.02784e-01	8.521	1.99509
$k = 1, M = 7$	9.76661e-02	19.751	2.8343
$k = 2, M = 4$	2.63402e-03	37.469	4.89185
$k = 2, M = 5$	1.09729e-03	38.265	10.5364

Table 16: The value of logarithmic condition number and numerical stability parameters with $\alpha = \beta = 1, \gamma = 0.8$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$.

n	$\log_{10}(cond(A))$	p_1	p_2
5	2.8305	2.95454×10^{-01}	2.95454×10^{-01}
6	3.4326	1.85343×10^{-01}	1.85343×10^{-01}
7	4.0267	2.34766×10^{-02}	2.34766×10^{-02}
8	4.0326	1.02877×10^{-02}	1.02875×10^{-02}
10	5.2664	5.01782×10^{-03}	5.01782×10^{-03}



(a) $k = 1, M = 7$. (b) $k = 1, M = 8$.

Figure 5: Plot of approximate solution for different values of M with $\alpha = \beta = 1$ and $\gamma = 0.8$.

numerical results shown in Table 17 indicate that the pollutant concentration decreases over time as convection and diffusion transport it throughout the domain. In Table 18, we have compared the proposed method with the method in [19]. This table shows that the error increases over time and decreases as the γ increases. However, the errors reported in the table are consistently lower than those reported in [19], demonstrating the proposed method's higher accuracy. As shown in

Table 19 the maximum absolute error decreases with an increase in resolution parameters. Table 20 confirms the stability of the proposed method. Figure 6 shows the plot of exact, approximate, and absolute error with $\alpha = \beta = 1, \gamma = 0.8$, $k = 2$, and $M = 3$.

Table 17: The approximate value of concentration $\mathcal{U}(s,t)$ at different points with $\alpha = \beta = 1, \gamma = 0.8$, $k = 2$, and $M = 3$.

s	Concentration $\mathcal{U}(s,t)$					
	$t = 0.0$	$t = 0.2$	$t = 0.4$	$t = 0.6$	$t = 0.8$	$t = 1.0$
0.0	0	0	0	0	0	0
0.2	0.587736	0.587171	0.586709	0.586298	0.585919	0.585562
0.4	0.950978	0.950072	0.949331	0.948668	0.948056	0.947477
0.6	0.950978	0.950069	0.949326	0.949324	0.948656	0.948038
0.8	0.587736	0.587166	0.586698	0.586276	0.585886	0.585515
1.0	0	0	0	0	0	0

Table 18: Comparison of absolute errors for Example 6 with $\alpha = \beta = 1$, and $s = 0.5$.

t	Method in [19]			Proposed method ($k = 2, M = 3$)		
	$\gamma = 0.4$	$\gamma = 0.7$	$\gamma = 1.0$	$\gamma = 0.4$	γ=0.7	$\gamma = 1.0$
0.2	3.972e-03	2.419e-03	1.421e-03	3.02648e-04	2.73726e-05	1.01261e-05
0.4	5.231e-03	3.931e-03	2.842e-03	3.94373e-04	5.18062e-05	5.80978e-06
0.8	6.918e-03	6.378e-03	5.684e-03	4.97679e-04	1.59196e-04	4.68719e-05
1.0	7.564e-03	7.467e-03	7.564e-03	3.991e-03	3.90017e-03	3.54675e-03

Table 19: Maximum absolute error, CPU time, and memory used for different values of k and M with $\alpha = \beta = 1, \gamma = 0.8$.

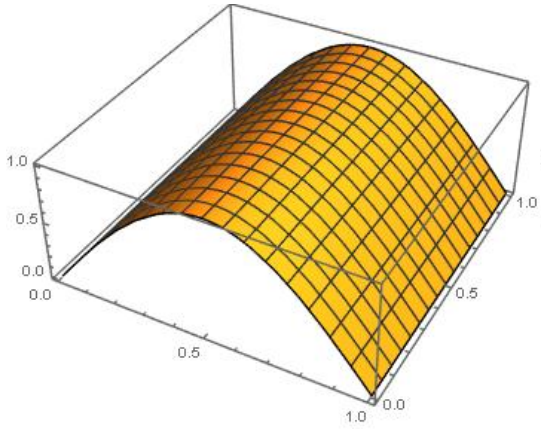
	Maximum Absolute Error	CPU Time (seconds)	Memory used (MB)
$k = 2, M = 3$	2.08857e-03	6.016	3.74405
$k = 2, M = 4$	2.07524e-03	16.031	4.19507
$k = 2, M = 5$	2.04201e-03	36.844	7.62554
$k = 3, M = 3$	1.51867e-04	40.365	8.70413

11 Conclusion

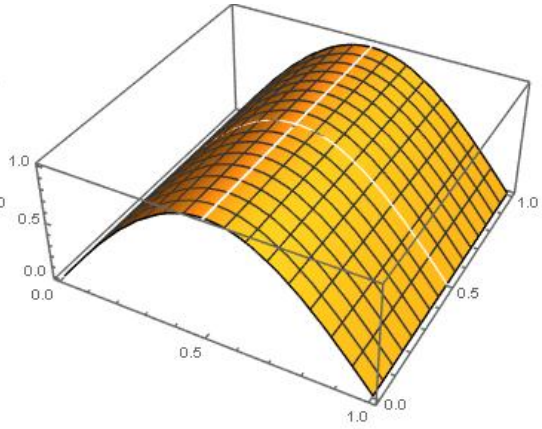
In this work, a numerical method based on the fractional Chelyshkov wavelet was introduced for solving the time fractional convection-diffusion equation with variable coefficients. The method is

Table 20: The value of logarithmic condition number and numerical stability parameters with $\alpha = \beta = 1, \gamma = 0.8$ using relative perturbations $\frac{\|\Delta A\|}{\|A\|} = 10^{-10}$ and $\frac{\|\Delta F\|}{\|F\|} = 10^{-04}$.

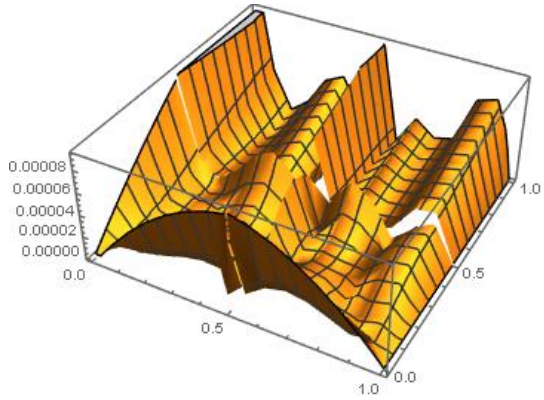
n	$\log_{10}(cond(A))$	p_1	p_2
4	4.4506	7.09284×10^{-01}	7.0928×10^{-01}
6	4.8281	2.9594×10^{-01}	2.95943×10^{-01}
8	5.2831	9.76009×10^{-02}	9.75989×10^{-02}
10	5.7974	3.14977×10^{-02}	3.14957×10^{-02}



(a) Exact plot.



(b) Approximate plot.



(c) Absolute error plot.

Figure 6: Plot of exact, approximate, and absolute error of Example 6.

adopted because of its several inherent properties, such as continuity, compact support, piecewise-defined functions, representation of polynomials to a certain degree, ability to represent functions at different levels of resolution, and ability to transform the fractional partial differential equa-

tions into a system of algebraic equations that can be solved effectively. Furthermore, fractional Chelyshkov wavelet basis functions accurately capture the solution behavior with a relatively small number of basis functions, thereby reducing computational effort. The fractional-order operational matrix for the fractional Chelyshkov wavelet was obtained. The stability analysis is investigated by using perturbation based theory. Convergence and error analyses have also been established. The numerical examples showed that the proposed method yields accurate results for different fractional orders. The error decreases significantly as the wavelet resolution parameters are refined, confirming the method's reliability. A comparison with existing numerical methods is presented, supporting the claim that fractional Chelyshkov wavelets achieve high accuracy. To show the computational efficiency of the proposed method, CPU time and memory usage were reported. Future research may focus on extending the proposed method to nonlinear fractional problems, distributed-order and variable-order fractional models, multidimensional problems, and coupled systems arising from engineering and physical applications.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Abo-Gabal, H., Zaky, M.A. and Doha, E.H. *Fractional Romanovski–Jacobi tau method for time-fractional partial differential equations with nonsmooth solutions,* Appl. Numer. Math. 182 (2022), 214–234.

[2] Adibmanesha, L. and Rashidinia, J. *Sinc and B-spline scaling functions for time-fractional convection-diffusion equations,* J. King Saud Univ. Sci. 33(2) (2021), 101343.

[3] Chen, Y., Wu, Y., Cui, Y., Wang, Z. and Jin, D. *Wavelet method for a class of fractional convection-diffusion equation with variable coefficients,* J. Comput. Sci. 1(3) (2010), 146–149.

[4] Chui, C.K. *Wavelet analysis and its applications,* Wavelets, 725, 1992.

[5] Faheem, M., Khan, A. and Raza, A. *A high resolution Hermite wavelet technique for solving space–time-fractional partial differential equations,* Math. Comput. Simul. 194 (2022), 588–609.

[6] Inc, M. *The approximate and exact solutions of the space- and time-fractional Burgers equations with initial conditions by variational iteration method,* J. Math. Anal. Appl. 345(1) (2008), 476–484.

[7] Irandoust-Pakchin, S., Dehghan, M., Abdi-Mazraeh, S. and Lakestani, M. *Numerical solution for a class of fractional convection–diffusion equations using the flatlet oblique multi-wavelets,* J. Vib. Control 20(6) (2014), 913–924.

[8] Izadkhah, M.M. and Saberi-Nadjafi, J. *Gegenbauer spectral method for time-fractional convection–diffusion equations with variable coefficients,* Math. Methods Appl. Sci. 38(15) (2015), 3183–3194.

[9] Kanna, M.R.R., Kumar, R.P., Nandappa, S. and Cangul, I.N. *On solutions of fractional order telegraph partial differential equation by Crank–Nicholson finite difference method,* Appl. Math. Nonlinear Sci. 5(2) (2020), 85–98.

[10] Kargar, Z. and Saeedi, H. *B-spline wavelet operational method for numerical solution of time-space fractional partial differential equations,* Int. J. Wavelets Multiresolut. Inf. Process. 15(4) (2017), 1750034.

[11] Korkiatsakul, T., Koonprasert, S. and Neamprem, K. *Chebyshev wavelet solutions for time-fractional integro partial differential equation and its application to beam problems,* Compusoft 9(5) (2020), 3677–3684.

[12] Kumbinarasaiah, S. and Mulimani, M. *The Fibonacci wavelets approach for the fractional Rosenau–Hyman equations,* Results Control Optim. 11 (2023), 100221.

[13] Lai, J., Liu, F., Anh, V.V. and Liu, Q. *A space-time finite element method for solving linear Riesz space fractional partial differential equations,* Numer. Algorithms (2021), 1–22.

[14] Luchko, Y. and Gorenflo, R. *An operational method for solving fractional differential equations with the Caputo derivatives,* Acta Math. Vietnam. 24(2) (1999), 207–233.

[15] Md Nasrudin, F.S., Phang, C. and Kanwal, A. *Fractal-fractional advection–diffusion–reaction equations by Ritz approximation approach,* Open Phys. 21(1) (2023), 20220221.

[16] Miller, K.S. *An Introduction to the Fractional Calculus and Fractional Differential Equations*, John Wiley & Sons, 1993.

[17] Ngo, H.T.B., Razzaghi, M. and Vo, T.N. *Fractional-order Chelyshkov wavelet method for solving variable-order fractional differential equations and an application in variable-order fractional relaxation system,* Numer. Algorithms 92(3) (2023), 1571–1588.

[18] Oruç, "O., Esen, A. and Bulut, F. *A Haar wavelet approximation for two-dimensional time fractional reaction–subdiffusion equation,* Eng. Comput. 35(1) (2019), 75–86.

[19] Pariyar, S., Lamichhane, B.P. and Kafle, J. *A time fractional advection–diffusion approach to air pollution: Modeling and analyzing pollutant dispersion dynamics,* Partial Differ. Equ. Appl. Math. 14 (2025), 101149.

[20] Podlubny, I. *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*, Elsevier, 1998.

[21] Rahimkhani, P. and Ordokhani, Y. *A numerical scheme based on Bernoulli wavelets and collocation method for solving fractional partial differential equations with Dirichlet boundary conditions,* Numer. Methods Partial Differ. Equ. 35(1) (2019), 34–59.

[22] Rahimkhani, P., Ordokhani, Y. and Lima, P.M. *An improved composite collocation method for distributed-order fractional differential equations based on fractional Chelyshkov wavelets,* Appl. Numer. Math. 145 (2019), 1–27.

[23] Saadatmandi, A. and Dehghan, M. *A tau approach for solution of the space fractional diffusion equation,* Comput. Math. Appl. 62(3) (2011), 1135–1142.

[24] Saadatmandi, A., Dehghan, M. and Azizi, M.-R. *The Sinc–Legendre collocation method for a class of fractional convection–diffusion equations with variable coefficients,* Commun. Nonlinear Sci. Numer. Simul. 17(11) (2012), 4125–4136.

[25] Safdari, H., Rajabzadeh, M. and Khalighi, M. *Solving a nonlinear fractional convection-diffusion equation using local discontinuous Galerkin method,* Appl. Numer. Math. 165 (2021), 22–34.

[26] Saw, V. and Kumar, S. *The Chebyshev collocation method for a class of time fractional convection-diffusion equation with variable coefficients,* Math. Methods Appl. Sci. 44(8) (2021), 6666–6678.

[27] Singh, A., Kanaujiya, A. and Mohapatra, J. *Chelyshkov wavelet method for solving multi-dimensional variable order fractional optimal control problem,* J. Appl. Math. Comput. 70 (2024), 3135–3160.

[28] Singh, A., Kanaujiya, A. and Mohapatra, J. *Euler wavelets method for optimal control problems of fractional integro-differential equations,* J. Comput. Appl. Math. 454 (2025), 116178.

[29] Singh, A., Kanaujiya, A. and Mohapatra, J. *Fractional Legendre wavelet approach resolving multi-scale optimal control problems involving Caputo–Fabrizio derivative,* Numer. Algorithms 99 (2025), 101–132.

[30] Singh, A., Kanaujiya, A. and Mohapatra, J. *Fractional-order Euler wavelet technique for multidimensional distributed-order optimal control problem,* Commun. Nonlinear Sci. Numer. Simul. 152 (2025), 109316.

[31] Singh, A., Kanaujiya, A. and Mohapatra, J. *Wavelet technique for variable-order modified Caputo–Fabrizio optimal control problems,* Iran J. Sci. 50 (2026), 797–811.

[32] Singh, A., Kanaujiya, A. and Mohapatra, J. *An efficient generalized Lucas wavelet operational matrix method for optimal control problems,* J. Comput. Appl. Math. 471 (2026), 116749.

[33] Singh, A., Kanaujiya, A. and Mohapatra, J. *Fractional order Chelyshkov wavelet method for variable-order time delay optimal control problem,* J. Comput. Appl. Math. 486 (2026), 117703.

[34] Tsuchiya, T. and Kumagai, A. *Numerical simulation of acoustic imaging using a combination of finite difference time domain and boundary integral equation methods,* Jpn. J. Appl. Phys. 48(7S) (2009), 07GN02.

[35] Uddin, M. and Haq, S. *RBFs approximation method for time fractional partial differential equations,* Commun. Nonlinear Sci. Numer. Simul. 16(11) (2011), 4208–4214.

[36] Wang, Y. and Zhu, L. *Solving nonlinear Volterra integro-differential equations of fractional order by using Euler wavelet method,* Adv. Differ. Equ. (2017), 1–16.

[37] Xu, X. and Xu, D. *Legendre wavelets direct method for the numerical solution of time-fractional order telegraph equations,* Mediterr. J. Math. 15 (2018), 1–33.

[38] Yadav, P., Jahan, S. and Nisar, K.S. *Fibonacci wavelet method for time fractional convection–diffusion equations,* Math. Methods Appl. Sci. 47(4) (2024), 2639–2655.

[39] Yousefi, S.A. and Dehghan, M. *The use of He's variational iteration method for solving variational problems,* Int. J. Comput. Math. 87(6) (2010), 1299–1314.

[40] Zada, L. and Aziz, I. *Numerical solution of fractional partial differential equations via Haar wavelet,* Numer. Methods Partial Differ. Equ. 38(2) (2022), 222–242.

[41] Zhou, F. and Xu, X. *The third kind Chebyshev wavelets collocation method for solving the time-fractional convection diffusion equations with variable coefficients,* Appl. Math. Comput. 280 (2016), 11–29.

Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 4, 2026, pp 1591–1621
https://doi.org/10.22067/ijnao.2026.96253.1768
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Queuing analysis of a VoIP call center under retrial policy with re-service, server maintenance and service disruptions

Preena R. and Sumitha D.

Abstract

This paper develops a retrial queuing framework for a voice over internet protocol call center server with two obligatory service phases. Incoming voice packets arrive according to a Poisson process. The model captures caller behavior involving potential re-service requests based on satisfaction after the first phase, introducing dynamic retrial effects. It integrates operational uncertainties such as service disruptions due to malware attacks and immediate repair processes. Additionally, server maintenance activities and post-maintenance buffer searches are integrated to reflect realistic system operations. The supplementary variable technique is employed to derive steady-state solutions for performance evaluation. Numerical analysis highlights the influence of key parameters on system performance. To enhance the study, cost optimization is performed using the particle swarm optimization technique, offering effective strategies to balance service quality and operational cost.

AMS subject classifications (2020): 60K25; 60K30; 68M20; 90B22

Received 30 October 2025; revised 18 February 2026; accepted 2 April 2026

Preena R.
Department of Mathematics, Research Scholar, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore,Tamil Nadu, India. e-mail: preenaravikumar24@gmail.com

Sumitha D.
Department of Science and Humanities, Associate Professor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore,Tamil Nadu, India. e-mail: drsumithad@gmail.com

How to cite this article
Preena R. and Sumitha D., Queuing analysis of a VoIP call center under retrial policy with re-service, server maintenance and service disruptions. *Iran. J. Numer. Anal. Optim.*, 2026; 16(4): 1591-1621. https://doi.org/10.22067/ijnao.2026.96253.1768

Keywords: Retrial; Queue; Repair; Server; Re-service

1 Introduction

The retrial queue is defined as follows: If the arriving customers find the server busy, then they join the retrial group of customers in the orbit to repeat the attempt for their service after a random time. If arriving customers find the server free, then one of the customers joins the service area and the others leave for the orbit. Retrial queuing systems have been extensively applied in telecommunications. Works on general retrial queuing models are found in [12, 3, 10].

In the last two decades, researchers have worked on feedback (re-service) retrial queues to enhance customer satisfaction and the quality of service, where the customers can opt for re-service. Various performance measures of the M/G/1 retrial queue with Bernoulli feedback and single vacation were studied in [21] and the system size distribution at random points was derived using probability generating functions. The work quantifies the impact of feedback on the expected orbit length. Moreover, the inclusion of feedback, starting failures and vacation mechanisms provides a comprehensive framework for analyzing the behavior of retrial customers. Studies in [23, 7, 26] have examined retrial queuing systems with customer feedback and its impact on the performance of the system. Recently, feedback retrial queues in telecommunication systems were discussed [1], along with a sensitivity analysis of performance measures in a telecommunication system. The primary focus of the work is on modeling and analyzing a complex queuing system that reflects real-world scenarios where a batch of customers requiring different types of service rejoin the orbit if unsatisfied with the initial service. The feedback mechanism is a critical component that is embedded to understand the impact of feedback rates on the overall system's efficiency. A retrial queuing system was modeled in [25], analyzing how feedback probability and other system parameters influence server state probabilities and customer satisfaction. The analysis provides insights into how an increase in the feedback probability impacts the system and emphasizes the need to tune system parameters to keep congestion manageable.

In a retrial queuing system, a Bernoulli vacation (server maintenance) is a single vacation where the server is under maintenance for a certain period of time after service completion. During the vacation period, the server may also perform any alternative task. An extensive analysis of a two-phase batch arrival retrial queuing system with Bernoulli vacation schedule was carried out in [8] and the steady-state distribution of the server state and number of customers in the retrial group was obtained. Some important performance measures of the system were derived in [16] by introducing supplementary variables and employing the embedded Markov chain technique in a batch arrival retrial queuing system for essential and optional services with

server breakdown and Bernoulli vacation. A mathematical model of a retrial queuing system featuring two types of customer services and a Bernoulli vacation policy was presented in [2]. The inclusion of the vacation mechanism aims to optimize system performance. A comprehensive literature review of retrial queuing systems dealing with Bernoulli vacation policies was provided in [20], offering a detailed summary of key aspects of the server state probabilities, analytical techniques and practical applications.

The phenomenon of orbital search (buffer search) in retrial queues is incorporated to effectively handle the customers waiting for service in the orbit or buffer. The joint distribution of the server state and orbit length in the steady-state of an M/G/1 retrial queue with non-persistent customers and orbital search is analyzed in [18]. The model is studied with a distinguishing and proactive feature, which is the orbital search mechanism, where the server actively seeks out customers in the orbit to provide service. The work is mathematically modeled to show the effects of orbital search on the steady-state behavior of the queue and the distribution of customers in the orbit. A two-phase retrial queue incorporating Bernoulli feedback and orbital search was investigated in [6], along with analytical derivations for key performance measures such as the expected number of customers in the orbit and the server state probabilities. An M/G/1 feedback retrial queuing system incorporating orbital search was examined in [5]. The authors derived explicit steady-state solutions for the number of customers in the system and in the orbit using probability generating functions and validated the analytical results with numerical examples.

Recently, retrial queues with negative customers (malware attack) such as viruses, have attracted growing interest from many researchers. The arrival of the negative customers is termed as G-queue and it was first introduced in [13]. A survey of negative customer arrivals in queuing systems and their applications were discussed in [27]. The steady-state analysis of an unreliable batch arrival feedback retrial queue with k-stages of service and Bernoulli vacation policy, where the server is subjected to breakdown due to the arrival of negative customers, is analyzed in [22]. A single-server queue with negative arrivals under a retrial policy and a Bernoulli vacation schedule was studied in [19]. The effect of vacation on overall system efficiency is analyzed and important performance indicators such as mean queue length and mean waiting time are calculated. Sensitivity analysis is carried out to illustrate the effects of model parameters on the performance of the system. A complex retrial queuing model integrating both the feedback process and the arrival of the negative customers was structured in [4]. The study provided a stochastic analysis along with explicit expressions for key performance measures of the system and demonstrated how negative arrivals influence the behavior of the system.

In the recent past, there has been a growing interest in the applications of retrial queues in call centers. A comprehensive empirical investigation into the factors driving customers to make repeated calls to call centers was conducted in [15], emphasizing the role of service preferences

and perceived service quality. Their analytical framework helps to quantify these preferences and supports the development of more effective service designs. A queuing network model specifically designed for call centers managing customer retrials was developed and analyzed in [24]. A closed exponential queuing network was proposed as a stochastic framework for modeling call processing in modern call center environments.

Addressing a gap in the existing literature, this paper studies a retrial queuing model for a voice over internet protocol (VoIP) call center server operating under two obligatory service phases. The model incorporates customer-driven re-service requests contingent on satisfaction after the first phase and explicitly accounts for server disruptions caused by malware attacks, followed by immediate repairs. These features are analyzed together with server maintenance and buffer search mechanisms to examine their impact on system performance.

The structure of this paper is outlined in the following sections: Section 2 describes the system model in detail. Section 3 provides the stability condition for the model under study. Section 4 formulates the steady-state balance equations for the proposed model. Section 5 derives the steady-state results by solving these equations using the Supplementary Variable Technique. Section 6 presents the performance measures of the system. Section 7 focuses on the derivation of the mean buffer size and overall system size. Sections 8, 9, and 10 respectively provide the calculation of the reliability indices, mean busy period and mean busy cycle. Section 11 conducts numerical analysis to illustrate the system behavior under varying parameters. Section 12 offers a comprehensive cost analysis using the particle swarm optimization (PSO) technique. Finally, Section 13 concludes the paper with summery remarks and future research.

2 Model description

Multiple incoming voice packets arrive in accordance with a Poisson process with rate λ^t. Let the random variable Y be the packet size with $P\{Y = k\} = C_k$, $k = 1, 2, 3, \ldots, \infty$, and let $C(z)$ be the probability generating function with first two moments m_1 and m_2. The server renders service in two phases: Initial inquiry phase (IIP) and sales assistance phase (SAP).

If the incoming callers find the server busy or inaccessible, then the callers join the buffer to retry for service later with retrial rate η. The retrial time is generally distributed with distribution function $A(x)$, density function $a(x)$, Laplace–Stieltjes transform $A^*(x)$ and retrial completion rate $\eta(x)$.

If the callers find the server available, then one of the callers from the packet gets connected to the IIP and other callers join the buffer to retry for service after some random time. After completion of the IIP, depending upon the quality of the service, the caller either opts for re-

service with probability δ or opts for SAP with the complementary probability $\bar{\delta} = (1-\delta)$. After completion of the SAP, the caller ends the call.

The service time during both phases is generally distributed with distribution function $B_i(x)$, density function $b_i(x)$, Laplace–Stieltjes transform $B_i^*(s)$, and service completion rate $\zeta_i(x)$ with moments ζ_{i1} and ζ_{i2}, where $i = 1, 2$.

After service completion, the server goes for server maintenance with probability ω or remains accessible with the complementary probability $\bar{\omega} = (1-\omega)$. The maintenance time of the server is generally distributed with distribution function $V(x)$, density function $v(x)$, Laplace–Stieltjes transform $V^*(s)$, and maintenance completion rate $g(x)$ with moments ϑ_1 and ϑ_2.

On returning from the maintenance activity, the server searches for the caller in the buffer with probability θ or remains accessible with the complementary probability $\bar{\theta} = (1-\theta)$. The malware attack interrupts the service with negative arrival rate λ^- during both phases of service which ultimately leads to call dropouts.

The repair for the failed server begins immediately. After the repair is completed, the server becomes accessible. The repair time during both phases is generally distributed with distribution function $R_i(x)$, density function $r_i(x)$, Laplace–Stieltjes transform $R_i^*(s)$ and repair completion rate $\beta_i(x)$ with moments β_{i1} and β_{i2}, where $i = 1, 2$. The structure of the described retrial queuing model is illustrated in Figure 1.

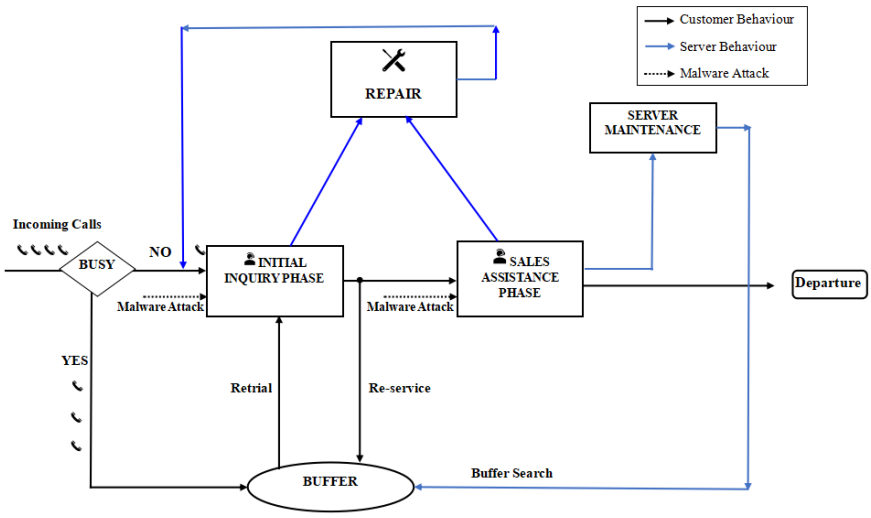


Figure 1: Structure of the described retrial queuing model in VoIP call center

The Markov process for the model can be defined as

$$\{M(t) : t \geq 0\} = \{X(t), C(t) : t \geq 0\},$$

where $X(t)$ is the number of callers in the buffer, $C(t)$ is the server states; that is, 0, 1, 2, 3, 4, 5 depending on whether the server is available, server is busy rendering IIP, server is busy rendering SAP, server is under repair during IIP, server is under repair during SAP, and the server is under maintenance, respectively.

2.1 Steady-state probabilities

- E_0 is defined as the server is in the available state with zero callers in the buffer.
- $E_n(x,t)dx,\ n \geq 1$, is defined as the server is in the available state with n callers in the buffer at time and $\chi_0(t)$ is the elapsed retrial time between x and $x+dx$.
- $W_n(x,t)dx,\ n \geq 0$, is defined as the server is busy serving in IIP with n callers in the buffer and $\chi_1(t)$ is the elapsed IIP service time between x and $x+dx$.
- $P_n(x,t)dx,\ n \geq 0$, is defined as the server is busy serving in SAP with n callers in the buffer and $\chi_2(t)$ is the elapsed SAP service time between x and $x+dx$.
- $R_{1,n}(x,t)dx,\ n \geq 0$, is defined as the server is under repair during IIP due to the negative arrival with n callers in the buffer and $\chi_3(t)$ is the elapsed repair time during IIP between x and $x+dx$.
- $R_{2,n}(x,t)dx,\ n \geq 0$, is defined as the server is under repair during SAP due to the negative arrival with n callers in the buffer and $\chi_4(t)$ is the elapsed repair time during SAP between x and $x+dx$.
- $V_n(x,t)dx,\ n \geq 0$, is defined as the server is on maintenance with n callers in the buffer and $\chi_5(t)$ is the elapsed maintenance time between x and $x+dx$.

3 Stability condition

Let N_n be the orbit length at the time of the nth customer departure, $n \geq 1$. Then $\{N_n, n \geq 1\}$ is ergodic if and only if

$$\delta B_1^*(\lambda^-) + \lambda^+ m_1 \big[1 - \delta B_1^*(\lambda^-) - \bar{\delta} B_1^*(\lambda^-) B_2^*(\lambda^-)\big] / \lambda^-$$
$$+ \lambda^+ m_1 \big[\beta_{11}\big(1 - B_1^*(\lambda^-)\big) + \bar{\delta}\beta_{21} B_1^*(\lambda^-)\big(1 - B_2^*(\lambda^-)\big) + \omega\bar{\delta}\vartheta_1 B_1^*(\lambda^-)$$

$$B_2^*(\lambda^-)\big] + m_1\big(1 - A^*(\lambda^+)\big)\big[1 - \omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big] < 1.$$

It can be proved along similar lines as in [14].

Since the arrival stream is a Poisson process, it can be shown from Burke's theorem [9] that the steady-state probabilities of
$\{X(t), C(t) : t \geq 0\}$ exist and are positive if and only if

$$\begin{aligned}
&\delta B_1^*(\lambda^-) + \lambda^+ m_1\big[1 - \delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big]/\lambda^- \\
&+ \lambda^+ m_1\big[\beta_{11}\big(1 - B_1^*(\lambda^-)\big) + \bar{\delta}\beta_{21}B_1^*(\lambda^-)\big(1 - B_2^*(\lambda^-)\big) + \omega\bar{\delta}\vartheta_1 B_1^*(\lambda^-) \\
&B_2^*(\lambda^-)\big] + m_1\big(1 - A^*(\lambda^+)\big)\big[1 - \omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big] < 1.
\end{aligned}$$

From the mean drift

$$\begin{aligned}
\varrho_j =&\delta B_1^*(\lambda^-) + \lambda^+ m_1\big[1 - \delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big]/\lambda^- \\
&+ \lambda^+ m_1\big[\beta_{11}\big(1 - B_1^*(\lambda^-)\big) + \bar{\delta}\beta_{21}B_1^*(\lambda^-)\big(1 - B_2^*(\lambda^-)\big) \\
&+ \omega\bar{\delta}\vartheta_1 B_1^*(\lambda^-)B_2^*(\lambda^-)\big] \\
&+ m_1\big(1 - A^*(\lambda^+)\big)\big[1 - \omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big] - 1, \qquad \text{for all } j \geq 1.
\end{aligned}$$

We have a reasonable conclusion that the term $\lambda^+ m_1\big[1 - \delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big]/\lambda^-$ represents the number of arriving customers during the service time. The term $\lambda^+ m_1\big[\beta_{11}\big(1 - B_1^*(\lambda^-)\big) + \bar{\delta}\beta_{21}B_1^*(\lambda^-)\big(1 - B_2^*(\lambda^-)\big) + \omega\bar{\delta}\vartheta_1 B_1^*(\lambda^-)B_2^*(\lambda^-)\big]$ represents the number of arriving customers during the repair time and maintenance time of the server.

The term $\delta B_1^*(\lambda^-)$ corresponds to the number of customers opting for re-service. The term $m_1\big(1 - A^*(\lambda^+)\big)\big[1 - \omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big] - 1$ refers to the contribution to the buffer size due to a batch arrival during the retrial time excluding the arbitrary customer of the arriving batch whose service commences so that he no longer belongs to the buffer.

Similar interpretation can be provided for $j = 0$. The condition $j < 0$ ensures that the buffer size does not grow indefinitely in course of time.

Accordingly the stability condition is

$$\begin{aligned}
&\delta B_1^*(\lambda^-) + \lambda^+ m_1\big[1 - \delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big]/\lambda^- \\
&+ \lambda^+ m_1\big[\beta_{11}\big(1 - B_1^*(\lambda^-)\big) + \bar{\delta}\beta_{21}B_1^*(\lambda^-)\big(1 - B_2^*(\lambda^-)\big) \\
&+ \omega\bar{\delta}\vartheta_1 B_1^*(\lambda^-)B_2^*(\lambda^-)\big] + m_1\big(1 - A^*(\lambda^+)\big)\big[1 - \omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)\big] < 1.
\end{aligned}$$

4 Steady-state equations

The steady-state equations are framed for

$$\lambda^+ E_0 = \bar{\omega} \int_0^\infty P_0(x)\zeta_2(x)dx + \int_0^\infty V_0(x)\vartheta(x)dx + \int_0^\infty R_1(x)\beta_1(x)dx + \int_0^\infty R_2(x)\beta_2(x)dx, \tag{1}$$

$$\frac{d}{dx}E_n(x) = -(\lambda^+ + \eta(x))E_n(x), \quad n \geq 1, \tag{2}$$

$$\frac{d}{dx}W_n(x) = -(\lambda^+ + \lambda^- + \zeta_1(x))W_n(x) + \lambda^+ \sum_{k=1}^{n} C_k W_{n-k}(x), \quad n \geq 0, \tag{3}$$

$$\frac{d}{dx}P_n(x) = -(\lambda^+ + \lambda^- + \zeta_2(x))P_n(x) + \lambda^+ \sum_{k=1}^{n} C_k P_{n-k}(x), \quad n \geq 0, \tag{4}$$

$$\frac{d}{dx}R_{1,n}(x) = -(\lambda^+ + \beta_1(x))R_{1,n}(x) + \lambda^+ \sum_{k=1}^{n} C_k R_{1,n-k}(x), \quad n \geq 0, \tag{5}$$

$$\frac{d}{dx}R_{2,n}(x) = -(\lambda^+ + \beta_2(x))R_{2,n}(x) + \lambda^+ \sum_{k=1}^{n} C_k R_{2,n-k}(x), \quad n \geq 0, \tag{6}$$

$$\frac{d}{dx}V_n(x) = -(\lambda^+ + \vartheta(x))V_n(x) + \lambda^+ \sum_{k=1}^{n} C_k V_{n-k}(x), \quad n \geq 0, \tag{7}$$

with boundary conditions

$$E_n(0) = \delta \int_0^\infty W_{n-1}(x)\zeta_1(x)\,dx + \bar{\omega} \int_0^\infty P_n(x)\zeta_2(x)\,dx + \bar{\theta} \int_0^\infty V_n(x)\vartheta(x)\,dx + \int_0^\infty R_{1,n}(x)\beta_1(x)\,dx + \int_0^\infty R_{2,n}(x)\beta_2(x)\,dx, \quad n \geq 1, \tag{8}$$

$$W_0(0) = \lambda^+ C_1 E_0 + \int_0^\infty E_1(x)\eta(x)\,dx + \theta \int_0^\infty V_1(x)\vartheta(x)\,dx, \tag{9}$$

$$\begin{aligned} W_n(0) = {} & \lambda^+ C_{n+1} E_0 + \int_0^\infty E_{n+1}(x)\eta(x)\,dx \\ & + \lambda^+ \sum_{k=1}^{n} C_k \int_0^\infty E_{n-k+1}(x)\,dx \\ & + \theta \int_0^\infty V_{n+1}(x)\vartheta(x)\,dx, \quad n \geq 1, \end{aligned} \tag{10}$$

$$P_n(0) = \bar{\delta} \int_0^\infty W_n(x)\zeta_1(x)\,dx, \quad n \geq 0, \tag{11}$$

$$R_{1,n}(0) = \lambda^- \int_0^\infty W_n(x)\,dx, \quad n \geq 0, \tag{12}$$

$$R_{2,n}(0) = \lambda^- \int_0^\infty P_n(x)\,dx, \quad n \geq 0, \tag{13}$$

$$V_n(0) = \omega \int_0^\infty P_n(x)\zeta_2(x)\,dx, \quad n \geq 0. \tag{14}$$

The normalizing equation is

$$\begin{aligned} & E_0 + \sum_{n=1}^{\infty} \int_0^\infty E_n(x)\,dx + \sum_{n=0}^{\infty} \int_0^\infty W_n(x)\,dx + \sum_{n=0}^{\infty} \int_0^\infty P_n(x)\,dx \\ & + \sum_{n=0}^{\infty} \int_0^\infty R_{1,n}(x)\,dx + \sum_{n=0}^{\infty} \int_0^\infty R_{2,n}(x)\,dx + \sum_{n=0}^{\infty} \int_0^\infty V_n(x)\,dx = 1. \end{aligned} \tag{15}$$

Define the probability generating functions

$$\begin{aligned} \mathrm{E}(x,z) &= \sum_{n=1}^{\infty} E_n(x) z^n, & \mathrm{R}_1(x,z) &= \sum_{n=0}^{\infty} R_{1,n}(x) z^n, \\ \mathrm{W}(x,z) &= \sum_{n=0}^{\infty} W_n(x) z^n, & \mathrm{R}_2(x,z) &= \sum_{n=0}^{\infty} R_{2,n}(x) z^n, \\ \mathrm{P}(x,z) &= \sum_{n=0}^{\infty} P_n(x) z^n, & \mathrm{V}(x,z) &= \sum_{n=0}^{\infty} V_n(x) z^n. \end{aligned}$$

By multiplying equation (2) by z^n and summing over n, we get

$$\left(\frac{d}{dx}+\lambda^{+}+\eta(x)\right)E(x,z)=0 \tag{16}$$

Solving the partial differential equation (16), we get

$$\begin{aligned}E(x,z)=&Ce^{-\int[\lambda^{+}+\eta(x)]\,dx}\\=&Ce^{-\lambda^{+}x}[1-A(x)].\end{aligned}$$

Eliminating C by taking $x=0$, we obtain

$$\therefore\ E(x,z)=E(0,z)e^{-\lambda^{+}x}[1-A(x)]. \tag{17}$$

Similarly, by multiplying (3)-(7) by z^n, summing over n and solving, we get

$$W(x,z)=W(0,z)e^{-\left[\lambda^{-}+\lambda^{+}-\lambda^{+}C(z)\right]x}\left[1-B_1(x)\right], \tag{18}$$

$$P(x,z)=P(0,z)e^{-\left[\lambda^{-}+\lambda^{+}-\lambda^{+}C(z)\right]x}\left[1-B_2(x)\right], \tag{19}$$

$$R_1(x,z)=R_1(0,z)e^{-\left[\lambda^{+}-\lambda^{+}C(z)\right]x}\left[1-R_1(x)\right], \tag{20}$$

$$R_2(x,z)=R_2(0,z)e^{-\left[\lambda^{+}-\lambda^{+}C(z)\right]x}\left[1-R_2(x)\right], \tag{21}$$

$$V(x,z)=V(0,z)e^{-\left[\lambda^{+}-\lambda^{+}C(z)\right]x}\left[1-V(x)\right]. \tag{22}$$

Multiplying (8) to (14) by z^n and summing over n, we get

$$\begin{aligned}E(0,z)=&\,\delta z\int_0^{\infty}W(x,z)\zeta_1(x)\,dx+\bar{\omega}\int_0^{\infty}P(x,z)\zeta_2(x)\,dx\\&+\bar{\theta}\int_0^{\infty}V(x,z)\vartheta_(x)\,dx+\int_0^{\infty}R_1(x,z)\beta_1(x)\,dx\\&+\int_0^{\infty}R_2(x,z)\beta_2(x)\,dx-\lambda^{+}E_0,\end{aligned} \tag{23}$$

$$W(0,z)=\frac{\lambda^{+}C(z)}{z}E_0+\frac{1}{z}\int_0^{\infty}E(x,z)\eta(x)\,dx$$

$$+\frac{\lambda^{+}C(z)}{z}\int_0^{\infty}E(x,z)\,dx+\frac{\theta}{z}\int_0^{\infty}V(x,z)\vartheta(x)\,dx, \tag{24}$$

$$P(0,z)=\bar{\delta}\int_0^{\infty}W(x,z)\zeta_1(x)dx, \tag{25}$$

$$R_1(0,z)=\lambda^{-}\int_0^{\infty}W(x,z)dx, \tag{26}$$

$$R_2(0,z)=\lambda^{-}\int_0^{\infty}P(x,z)dx, \tag{27}$$

$$V(0,z)=\omega\int_0^{\infty}P(x,z)\zeta_2(x)dx. \tag{28}$$

Substituting (17)-(22) into (23)-(28), we get the following equations (29) to (34):

$$E(0,z)=\frac{\lambda^{+}E_0[[\theta\omega\bar{\delta}B_1^{*}(u_1(z))B_2^{*}(u_1(z))V^{*}(u_2(z))-z]u_1(z)+C(z)G_2(z)]}{K_1(z)}, \tag{29}$$

$$\mathrm{W}(0,z)=\frac{\lambda^{+}E_0A^{*}(\lambda^{+})(C(z)-1)}{K(z)}, \tag{30}$$

$$\mathrm{P}(0,z)=\frac{\lambda^{+}E_0A^{*}(\lambda^{+})\bar{\delta}(C(z)-1)B_1^{*}(u_1(z))}{K(z)}, \tag{31}$$

$$R_1(0,z)=\frac{\lambda^{+}\lambda^{-}E_0A^{*}(\lambda^{+})(C(z)-1)[1-B_1^{*}(u_1(z))]}{K_1(z)}, \tag{32}$$

$$R_2(0,z)=\frac{\lambda^{+}\lambda^{-}E_0A^{*}(\lambda^{+})\bar{\delta}(C(z)-1)B_1^{*}(u_1(z))[1-B_2^{*}(u_1(z))]}{K_1(z)}, \tag{33}$$

$$V(0,z)=\frac{\lambda^{+}E_0A^{*}(\lambda^{+})\omega\bar{\delta}(C(z)-1)B_1^{*}(u_1(z))B_2^{*}(u_1(z))}{K(z)}, \tag{34}$$

where

$$
\begin{aligned}
u_1(z) &= \lambda^- + \lambda^+ - \lambda^+ C(z),\\
u_2(z) &= \lambda^+ - \lambda^+ C(z),\\
K(z) &= [[z - \theta\omega\bar{\delta}B_1^*(u_1(z))B_2^*(u_1(z))V^*(u_2(z))]u_1(z) - G_1(z)G_3(z)]/u_1(z),\\
K_1(z) &= [z - \theta\omega\bar{\delta}B_1^*(u_1(z))B_2^*(u_1(z))V^*(u_2(z))]u_1(z) - G_1(z)G_3(z),\\
G_1(z) &= A^*(\lambda^+) + C(z)(1 - A^*(\lambda^+)),\\
G_2(z) &= [\delta z B_1^*(u_1(z)) + \bar{\delta}B_1^*(u_1(z))B_2^*(u_1(z))(\bar{\omega} + \omega\bar{\theta}V^*(u_2(z))]u_1(z)\\
&\quad + \lambda^- R_1^*(u_2(z))[1 - B_1^*(u_1(z))]\\
&\quad + \lambda^-\bar{\delta}R_2^*(u_2(z))B_1^*(u_1(z))[1 - B_2^*(u_1(z))],\\
G_3(z) &= [\delta z B_1^*(u_1(z)) + \bar{\delta}B_1^*(u_1(z))\ B_2^*(u_1(z))(\bar{\omega} + \omega\bar{\theta})]u_1(z)\\
&\quad + \lambda^- R_1^*(u_2(z))[1 - B_1^*(u_1(z))]\\
&\quad + \lambda^-\bar{\delta}R_2^*(u_2(z))B_1^*(u_1(z))[1 - B_2^*(u_1(z))].
\end{aligned}
$$

5 Steady-state results

i. The partial probability generating function of the buffer size when the server is available is

$$
\begin{aligned}
E(z) &= \int_0^\infty E(x,z)\,dx\\
&= \int_0^\infty E(0,z)e^{-\lambda^+ x}[1 - A(x)]\,dx\\
&\quad E_0\,[1 - A^*(\lambda)]\,\Big\{C(z)G_z(z)\\
&= \frac{-\,[z - \theta\omega\bar{\delta}B_1^*(u_1(z))B_2^*(u_1(z))V^*(u_2(z))]u_1(z)\Big\}}{K_1(z)}.
\end{aligned}
\tag{35}
$$

ii. The partial probability generating function of the buffer size when the server is busy serving in IIP is

$$
\begin{aligned}
W(z) &= \int_0^\infty W(x,z)\,dx\\
&= \int_0^\infty W(0,z)e^{-(u_1(z))x}\,[1 - B_1(x)]\ dx\\
&= \frac{\lambda^+ E_0 A^*(\lambda^+)\{[C(z) - 1][1 - B_1^*(u_1(z))]\}}{K_1(z)}.
\end{aligned}
\tag{36}
$$

iii. The partial probability generating function of the buffer size when the server is busy serving in SAP is

$$\begin{aligned} P(z) &= \int_0^\infty P(x,z)\,dx \\ &= \int_0^\infty P(0,z)e^{-(u_1(z))x}\,[1-B_2(x)]\;dx \\ &= \frac{\lambda^+ E_0 A^*(\lambda^+)\{\bar{\delta}[C(z)-1]B_1^*(u_1(z))[1-B_2^*(u_1(z))]\}}{K_1(z)}. \end{aligned} \tag{37}$$

iv. The partial probability generating function of the buffer size when the server is under repair due to negative arrival during IIP is

$$\begin{aligned} R_1(z) &= \int_0^\infty R_1(x,z)\,dx \\ &= \int_0^\infty R_1(0,z)e^{-(u_2(z))x}\,[1-R_1(x)]\;dx \\ &= \frac{\lambda^- E_0 A^*(\lambda^+)\{[1-B_1^*(u_1(z))][R_1^*(u_2(z))-1]\}}{K_1(z)}. \end{aligned} \tag{38}$$

v. The partial probability generating function of the buffer size when the server is under repair due to negative arrival during SAP is

$$\begin{aligned} R_2(z) &= \int_0^\infty R_2(x,z)\,dx \\ &= \int_0^\infty R_2(0,z)e^{-(u_2(z))x}\,[1-R_2(x)]\;dx \\ &= \frac{\lambda^- E_0 A^*(\lambda^+)\{\bar{\delta}B_1^*(u_1(z))[1-B_2^*(u_1(z))][R_2^*(u_2(z))-1]\}}{K_1(z)}. \end{aligned} \tag{39}$$

vi. The partial probability generating function of the buffer size when the server is under maintenance is

$$\begin{aligned} V(z) &= \int_0^\infty V(x,z)\,dx \\ &= \int_0^\infty V(0,z)e^{-(u_2(z))x}\,[1-V(x)]\;dx \\ &= \frac{E_0 A^*(\lambda^+)\{\omega\bar{\delta}B_1^*(u_1(z))B_2^*(u_1(z))[V^*(u_2(z))-1]\}u_1(z)}{K_1(z)}. \end{aligned} \tag{40}$$

Using (35) to (40), we get the following result:

- The probability generating function of the number of callers in the buffer is

$$\begin{aligned}\phi_q(z) = E_0 + E(z) + W(z) + P(z) + R_1(z) + R_2(z) + V(z)\\ \{E_0 A^*(\lambda^+)\{[z - \delta z\, B_1^*(u_1(z)) - \bar{\delta}\, B_1^*(u_1(z))B_2^*(u_1(z))]\\ u_1(z) + \lambda^+(C(z) - 1)[1 - B_1^*(u_1(z))[\delta + \bar{\delta}\, B_2^*(u_1(z))]]\\ = \frac{-\lambda^-[1 - B_1^*(u_1(z))] - \lambda^-\bar{\delta}\, B_1^*(u_1(z))[1 - B_2^*(u_1(z))]\}\}}{K_1(z)}. \end{aligned} \tag{41}$$

- The probability generating function of the number of callers in the system is

$$\begin{aligned}\phi_s(z) = E_0 + E(z) + z(W(z) + P(z)) + R_1(z) + R_2(z) + V(z)\\ \{E_0 A^*(\lambda^+)\{[z - \delta z\, B_1^*(u_1(z)) - \bar{\delta}\, B_1^*(u_1(z))B_2^*(u_1(z))]\\ u_1(z) + \lambda^+ z(C(z) - 1)[1 - B_1^*(u_1(z))[\delta + \bar{\delta}\, B_2^*(u_1(z))]]\\ = \frac{-\lambda^-[1 - B_1^*(u_1(z))] - \lambda^-\bar{\delta}\, B_1^*(u_1(z))[1 - B_2^*(u_1(z))]\}\}}{K_1(z)}, \end{aligned} \tag{42}$$

where

$$E_0 = \frac{K_2}{\lambda^- A^*(\lambda^+)[1 - \delta B_1^*(\lambda^-)]},$$

$$\begin{aligned}K_2 = \lambda^+\{1 - \delta B_1^*(\lambda^-) + m_1(1 - A^*(\lambda^+))[\omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-) - 1]\\ - S_5 - \lambda^+ m_1[1 - \delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)]\}.\end{aligned}$$

6 Performance measures

Performance measures of the system are analyzed based on the following system characteristics:

6.1 Server state probabilities

i. The steady-state probability when the server is available

$$E = \lim_{z\to 1} E(z) = \frac{\begin{array}{c}\{E_0[1-A^*(\lambda^-)][\,\lambda^- m_1[1-\omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)] \\ -\lambda^-[1-\delta B_1^*(\lambda^-)] - S_5 - \lambda^+ m_1[1-\delta B_1^*(\lambda^-) \\ -\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)]]\}\end{array}}{K_2}. \tag{43}$$

ii. The steady-state probability when the server is busy in the IIP

$$W = \lim_{z\to 1} W(z) = \frac{\lambda^+ E_0 A^*(\lambda^+) m_1[1-B_1^*(\lambda^-)]}{K_2}. \tag{44}$$

iii. The steady-state probability when the server is busy in the SAP

$$P = \lim_{z\to 1} P(z) = \frac{\lambda^+ E_0 A^*(\lambda^+)\bar{\delta} m_1 B_1^*(\lambda^-)[1-B_2^*(\lambda^-)]}{K_2}. \tag{45}$$

iv. The steady-state probability when the server is under repair during IIP

$$R_1 = \lim_{z\to 1} R_1(z) = \frac{\lambda^-\lambda^+ E_0 A^*(\lambda^+) m_1\beta_{11}[1-B_1^*(\lambda^-)]}{K_2}. \tag{46}$$

v. The steady-state probability when the server is under repair during SAP

$$R_2 = \lim_{z\to 1} R_2(z) = \frac{\lambda^-\lambda^+ E_0 A^*(\lambda^+) m_1\bar{\delta}\beta_{21} B_1^*(\lambda^-)[1-B_2^*(\lambda^-)]}{K_2}. \tag{47}$$

vi. The steady-state probability when the server is under maintenance

$$V = \lim_{z\to 1} V(z) = \frac{\lambda^-\lambda^+ E_0 A^*(\lambda^+)\omega\bar{\delta}\vartheta_1 m_1 B_1^*(\lambda^-)B_2^*(\lambda^-)}{K_2}. \tag{48}$$

6.2 Expected buffer size

The expected size of the buffer for the server states under steady-state conditions are calculated and given in Table 1.

In Table 1, we have

$$\begin{aligned} K_3 = &-\lambda^+ m_2[1-\delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)(1-B_2^*(\lambda^-))] \\ &- \lambda^- m_2(1-A^*(\lambda^+))[1-\theta\omega\delta B_1^*(\lambda^-)B_2^*(\lambda^-)] \\ &- 2\lambda^+ m_1[1-\delta B_1^*(\lambda^-)] + 2(\lambda^+ m_1)^2 S_1 - 2m_1(1-A^*(\lambda^+))\{\lambda^-\bar{\delta}B_1^*(\lambda^-) \end{aligned}$$

Table 1: Expected buffer size

S. No.	Server States	Expected Buffer Size
1.	Available	$M_E = \lim_{z\to 1} \frac{d}{dz} E(z) = \frac{H_2K_2 - H_1K_3}{2[K_2]^2}$
2.	Busy	$M_W = \lim_{z\to 1} \frac{d}{dz} W(z) = \frac{H_4K_2 - H_3K_3}{2[K_2]^2}$
		$M_P = \lim_{z\to 1} \frac{d}{dz} P(z) = \frac{H_6K_2 - H_5K_3}{2[K_2]^2}$
3.	Repair	$M_{R_1} = \lim_{z\to 1} \frac{d}{dz} R_1(z) = \frac{H_8K_2 - H_7K_3}{2[K_2]^2}$
		$M_{R_2} = \lim_{z\to 1} \frac{d}{dz} R_2(z) = \frac{H_{10}K_2 - H_9K_3}{2[K_2]^2}$
4.	Maintenance	$M_V = \lim_{z\to 1} \frac{d}{dz} V(z) = \frac{H_{12}K_2 - H_{11}K_3}{2[K_2]^2}$

$$
\begin{aligned}
&- \lambda^+ m_1[\delta B_1^*(\lambda^-) + \bar{\delta}(1-\theta\omega)B_1^*(\lambda^-)B_2^*(\lambda^-)] + S_2\big\} \\
&- \lambda^-[S_3 + 2\lambda^+ m_1\delta\zeta_{11} + 2(\lambda^+ m_1)^2 S_4], \\
H_1 &= E_0(1-A^*(\lambda^+))\{\lambda^- m_1[1-\omega\theta\bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)] - \lambda^+ m_1[1-\delta B_1^*(\lambda^-) \\
&- \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)] - S_5 - \lambda^-(1-\delta B_1^*(\lambda^-))\}, \\
H_2 &= E_0(1-A^*(\lambda^+))\big\{\lambda^+ m_2[1-\delta B_1^*(\lambda^-) - \bar{\delta}B_1^*(\lambda^-)B_2^*(\lambda^-)] \\
&+ \lambda^- m_2[1-\theta\omega\delta B_1^*(\lambda^-)B_2^*(\lambda^-)] + 2\lambda^+ m_1[1-\delta B_1^*(\lambda^-)] - 2(\lambda^+ m_1)^2 S_1 \\
&+ 2m_1\big\{\lambda^-\delta B_1^*(\lambda^-) - \lambda^+ m_1[\delta B_1^*(\lambda^-) + \bar{\delta}(1-\theta\omega)B_1^*(\lambda^-)B_2^*(\lambda^-)] + S_2\big\} \\
&+ \lambda^-[S_3 + 2\lambda^+ m_1\delta\zeta_{11} + 2(\lambda^+ m_1)^2 S_4]\big\}, \\
H_3 &= \lambda^+ m_1 A^*(\lambda^+)E_0[1-B_1^*(\lambda^-)], \\
H_4 &= \lambda^+ A^*(\lambda^+)E_0\{[1-B_1^*(\lambda^-)] - 2\lambda^+ m_1^2\zeta_{11}\}, \\
H_5 &= \lambda^+\bar{\delta}m_1 A^*(\lambda^+)E_0 B_1^*(\lambda^-)[1-B_2^*(\lambda^-)], \\
H_6 &= \lambda^+\bar{\delta}A^*(\lambda^+)E_0\big\{m_2 B_1^*(\lambda^-)[1-B_2^*(\lambda^-)] + 2m_1[\lambda^+ m_1\zeta_{11}(1-B_2^*(\lambda^-)) \\
&- \lambda^+ m_1\zeta_{21}B_1^*(\lambda^-)]\big\}, \\
H_7 &= \lambda^-\lambda^+ m_1 A^*(\lambda^+)E_0\beta_{11}[1-B_1^*(\lambda^-)], \\
H_8 &= -\lambda^- A^*(\lambda^+)E_0\big\{2(\lambda^+ m_1)^2\zeta_{11}\beta_{11} - [1-B_1^*(\lambda^-)]h_3\big\}, \\
H_9 &= \lambda^-\lambda^+\bar{\delta}m_1 A^*(\lambda^+)E_0\beta_{21}B_1^*(\lambda^-)[1-B_2^*(\lambda^-)],
\end{aligned}
$$

$$
\begin{aligned}
H_{10} &= \lambda^{-}\bar{\delta}A^{*}(\lambda^{+})E_0\big\{2(\lambda^{+}m_1)^2\beta_{21}[\lambda^{+}m_1\zeta_{11}(1-B_2^{*}(\lambda^{-}))\\
&\quad -\lambda^{+}m_1\zeta_{21}B_1^{*}(\lambda^{-})]+B_1^{*}(\lambda^{-})[1-B_2^{*}(\lambda^{-})]h_4\big\},\\
H_{11} &= \lambda^{-}\lambda^{+}\omega\bar{\delta}m_1A^{*}(\lambda^{+})E_0\vartheta_1B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-}),\\
H_{12} &= \omega\bar{\delta}A^{*}(\lambda^{+})E_0\big\{2(\lambda^{+}m_1)^2\vartheta_1[\lambda^{-}(\zeta_{11}B_2^{*}(\lambda^{-})-\zeta_{21}B_1^{*}(\lambda^{-}))\\
&\quad -B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-})]+\lambda^{-}B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-})h_5\big\},\\
h_1 &= \lambda^{+}m_2\zeta_{11}+(\lambda^{+}m_1)^2\zeta_{12},\\
h_2 &= \lambda^{+}m_2\zeta_{21}+(\lambda^{+}m_1)^2\zeta_{22},\\
h_3 &= \lambda^{+}m_2\beta_{11}+(\lambda^{+}m_1)^2\beta_{12},\\
h_4 &= \lambda^{+}m_2\beta_{21}+(\lambda^{+}m_1)^2\beta_{22},\\
h_5 &= \lambda^{+}m_2\vartheta_1+(\lambda^{+}m_1)^2\vartheta_2,\\
S_1 &= \delta\zeta_{11}+\bar{\delta}[\zeta_{11}B_2^{*}(\lambda^{-})+\zeta_{21}B_1^{*}(\lambda^{-})+\omega\vartheta_1B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-})],\\
S_2 &= \lambda^{-}\lambda^{+}m_1\big\{\beta_{11}(1-B_1^{*}(\lambda^{-}))+\bar{\delta}\beta_{21}B_1^{*}(\lambda^{-})(1-B_2^{*}(\lambda^{-}))\\
&\quad +\omega\theta\bar{\delta}[\zeta_{11}B_2^{*}(\lambda^{-})+\zeta_{21}B_1^{*}(\lambda^{-})]+\omega\bar{\theta}\bar{\delta}\vartheta_1B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-})\big\},\\
S_3 &= h_3(1-B_1^{*}(\lambda^{-}))+\bar{\delta}h_4B_1^{*}(\lambda^{-})(1-B_2^{*}(\lambda^{-}))+(1-\theta\delta)\omega h_5B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-}),\\
S_4 &= (1-\theta\delta)\omega\vartheta_1[\zeta_{11}B_2^{*}(\lambda^{-})+\zeta_{21}B_1^{*}(\lambda^{-})]-\zeta_{11}\beta_{11}-\zeta_{21}\beta_{21}B_1^{*}(\lambda^{-})\\
&\quad +\bar{\delta}\zeta_{11}\beta_{21}(1-B_2^{*}(\lambda^{-})),\\
S_5 &= \lambda^{-}\lambda^{+}m_1\big\{\beta_{11}(1-B_1^{*}(\lambda^{-}))+\bar{\delta}\beta_{21}B_1^{*}(\lambda^{-})(1-B_2^{*}(\lambda^{-}))\\
&\quad +\omega\bar{\delta}\vartheta_1B_1^{*}(\lambda^{-})B_2^{*}(\lambda^{-})\big\}.
\end{aligned}
$$

7 Mean buffer size and system size

- The mean number of callers in the buffer is given by

$$L_q = \lim_{z\to 1}\frac{d}{dz}\,\phi_q(z).$$

Let $\mathrm{Nr}(\phi_q(z))$ and $\mathrm{Dr}(\phi_q(z))$ be the numerator and denominator of $\phi_q(z)$.

As $\mathrm{Nr}(\phi_q(1)) = \mathrm{Dr}(\phi_q(1)) = 0$, using L Hospital's rule, we have

$$L_q = \frac{\mathrm{Dr}'(\phi_q(1))\,\mathrm{Nr}''(\phi_q(1)) - \mathrm{Nr}'(\phi_q(1))\,\mathrm{Dr}''(\phi_q(1))}{2\left[\mathrm{Dr}'(\phi_q(1))\right]^2}, \tag{49}$$

where

$$\mathrm{Nr}'(\phi_q(1)) = \lambda^- E_0 A^*(\lambda^+)[1 - \delta B_1^*(\lambda^-)],$$

$$\mathrm{Nr}''(\phi_q(1)) = -2\lambda^+ m_1 E_0 A^*(\lambda^+)[1 - \delta B_1^*(\lambda^-) + \lambda^- \delta \mu_{11}],$$

$$\mathrm{Dr}'(\phi_q(1)) = K_2,$$

$$\mathrm{Dr}''(\phi_q(1)) = K_3.$$

- The mean number of callers in the system is given by

$$L_s = \lim_{z \to 1} \frac{d}{dz} \phi_s(z).$$

Let $\mathrm{Nr}(\phi_s(z))$ and $\mathrm{Dr}(\phi_s(z))$ be the numerator and denominator of $\phi_s(z)$.

As $\mathrm{Nr}(\phi_s(1)) = \mathrm{Dr}(\phi_s(1)) = 0$, using L Hospital's rule, we have

$$L_s = \frac{\mathrm{Dr}'(\phi_s(1))\,\mathrm{Nr}''(\phi_s(1)) - \mathrm{Nr}'(\phi_s(1))\,\mathrm{Dr}''(\phi_s(1))}{2\left[\mathrm{Dr}'(\phi_s(1))\right]^2}, \tag{50}$$

where

$$\begin{aligned}\mathrm{Nr}'(\phi_s(1)) = E_0 A^*(\lambda^+)\big\{&\lambda^-[1 - \delta B_1^*(\lambda)] \\ &- \lambda^+ m_1[1 - \delta B_1^*(\lambda^-) - \bar{\delta} B_1^*(\lambda^-) B_2^*(\lambda^-)]\big\},\end{aligned}$$

$$\begin{aligned}\mathrm{Nr}''(\phi_s(1)) =& E_0 A^*(\lambda^+)\big\{\lambda^-(1 - \delta h_1) \\ &+ -2\lambda^+ m_1[\lambda^- \bar{\delta} \mu_{11} \bar{\delta} B_1^*(\lambda^-) B_2^*(\lambda^-)]\big\},\end{aligned}$$

$$\mathrm{Dr}'(\phi_s(1)) = K_2,$$

$$\mathrm{Dr}''(\phi_s(1)) = K_3.$$

8 Reliability indices

8.1 Availability of the server

The availability of the server under steady-state condition is

$$J(t) = E_0 + \lim_{z\to 1} E(z) + \lim_{z\to 1} W(z) + \lim_{z\to 1} P(z)$$
$$= \frac{\lambda^-[1-\delta B_1^*(\lambda^-)] - S_5}{\lambda^-[1-\delta B_1^*(\lambda^-)]}.$$

8.2 Failure frequency

The failure frequency of the server under steady-state condition is

$$U(t) = \lambda^-\big[\lim_{z\to 1} W(z) + \lim_{z\to 1} P(z)\big]$$
$$= \frac{\lambda^+ m_1[1-\delta B_1^*(\lambda^-) - \bar{\delta} B_1^*(\lambda^-)B_2^*(\lambda^-)]}{1-\delta B_1^*(\lambda^-)}.$$

9 Mean busy period

Mean busy period is the average time of the server when it is busy serving the callers from the arrival time of the first caller until the server is available for the next caller.

$$\mathrm{E}[S_B] = \frac{1}{\lambda^+ m_1}\left(\frac{1}{E_0} - 1\right)$$
$$\lambda^-[1-A^*(\lambda^+)][\delta B_1^*(\lambda^-) - 1 - m_1[\omega\theta\bar{\delta} B_1^*(\lambda^-)B_2^*(\lambda^-) - 1]]$$
$$= \frac{+\lambda^+ m_1[1-\delta B_1^*(\lambda^-) - \bar{\delta} B_1^*(\lambda^-)B_2^*(\lambda^-)] + S_5}{\lambda^+ m_1 K_2}.$$

10 Mean busy cycle

Mean busy cycle is the average time of a complete cycle which includes both idle period and busy period of the server.

$$\mathrm{E}[S_C] = \frac{1}{\lambda^+ E_0} = \frac{\lambda^- A^*(\lambda^+)[1-\delta B_1^*(\lambda^-)]}{\lambda^+ K_2}.$$

11 Numerical analysis

Numerical analysis of the system is carried out using MATLAB by assuming that the retrial time, IIP service time, SAP service time, repair time during both phases and vacation time

follow exponential distribution with respective rates $\eta, \zeta_1, \zeta_2, \beta_1, \beta_2$ and ϑ. The default values of the parameters are set as $\lambda^+ = 0.8$, $\lambda^- = 0.4$, $\delta = 0.8$, $\omega = 0.3$, $\theta = 0.7$, $\eta = 15$, $\zeta_1 = 10$, $\zeta_2 = 8$, $\beta_1 = 10$, $\beta_2 = 5$, $\vartheta = 25$, $C_1 = 0.5$ and $C_2 = 0.5$[17].

The effect of the positive arrival rate λ^+ on all the server state probabilities when the server is available in the empty state (E_0), when the server is available in the non-empty system (E), when the server is busy serving the caller in two phases of service (W & P), when the server is under repair during two phases of service (R_1 & R_2) and when the server is on maintenance activity (V) is analyzed and given in Table 2.

Table 2: Effect of λ^+ on the performance measures

λ^+	E_0	E	W	P	R_1	R_2	V	Z
0.1	0.8820	0.0355	0.0625	0.0149	0.0025	0.0012	0.0014	1.0000
0.2	0.7628	0.0722	0.1250	0.0298	0.0050	0.0024	0.0029	1.0000
0.3	0.6425	0.1100	0.1875	0.0446	0.0075	0.0036	0.0043	1.0000
0.4	0.5212	0.1488	0.2500	0.0595	0.0100	0.0048	0.0057	1.0000
0.5	0.3987	0.1887	0.3125	0.0744	0.0125	0.0060	0.0071	1.0000
0.6	0.2752	0.2298	0.3750	0.0893	0.0150	0.0071	0.0086	1.0000
0.7	0.1506	0.2719	0.4375	0.1042	0.0175	0.0083	0.0100	1.0000
0.8	0.0248	0.3152	0.5000	0.1190	0.0200	0.0095	0.0114	1.0000

The results indicate that as the arrival rate increases, the probability of the server being available in the empty state decreases, while the remaining server state probabilities increase. Several of these measures exhibit near-linear trends over the selected arrival-rate range, which is expected due to their analytical dependence on the arrival rate when other system parameters are held fixed and the system operates under moderate traffic conditions. Furthermore, the normalization condition given in (15) is satisfied, that is,

$$Z = E_0 + E + W + P + R_1 + R_2 + V = 1.$$

The effects of various parameters on the server state probability when the server is available in the empty state (E_0) and the mean number of callers in the system (L_S) are shown in the Figures 2 to 8.

It is evident from the analysis that

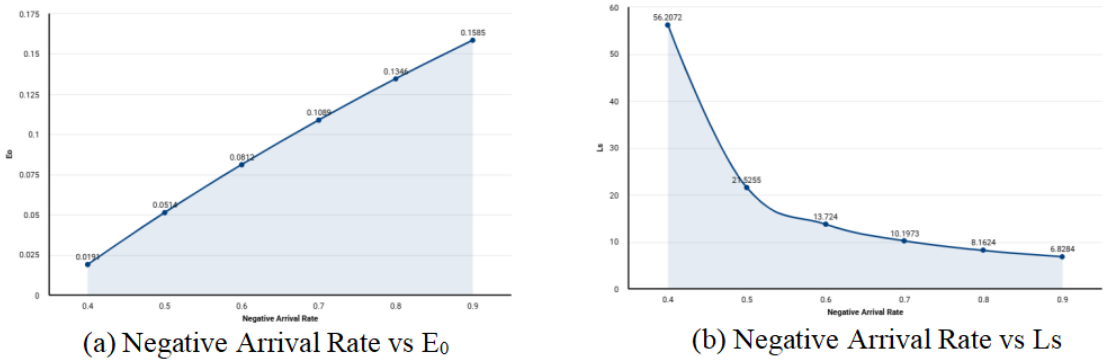
- Figure 2 shows that, with the increase of the negative arrival rate, E_0 rises and L_S reduces.
- In Figure 3, as the retrial rate rises, there is an escalation in E_0 and reduction in L_S.

- In Figure 4, as the re-service probability increases, E_0 reduces and L_S rises.

- In Figure 5, as the buffer search probability rises, there is an escalation in E_0 and a reduction in L_S.

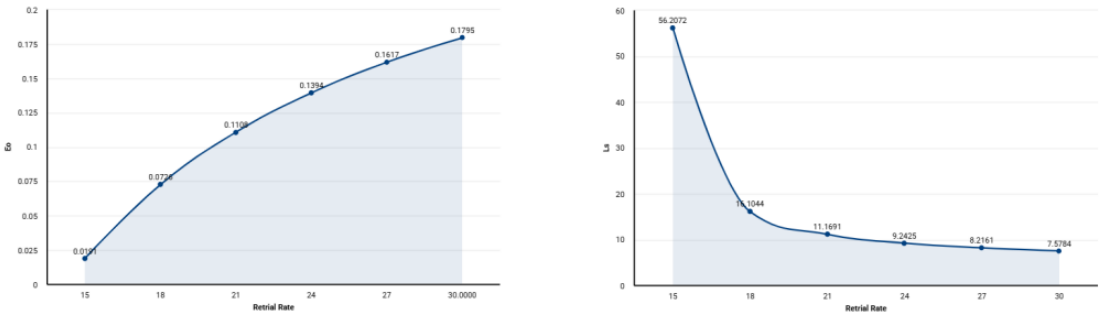
- Figures 6 and 7 show that, with increases of the both IIP and SAP service completion rate, there is an escalation in E_0 and reduction in L_S.

- In Figure 8, as the server maintenance completion rate increases, E_0 rises and L_S reduces.



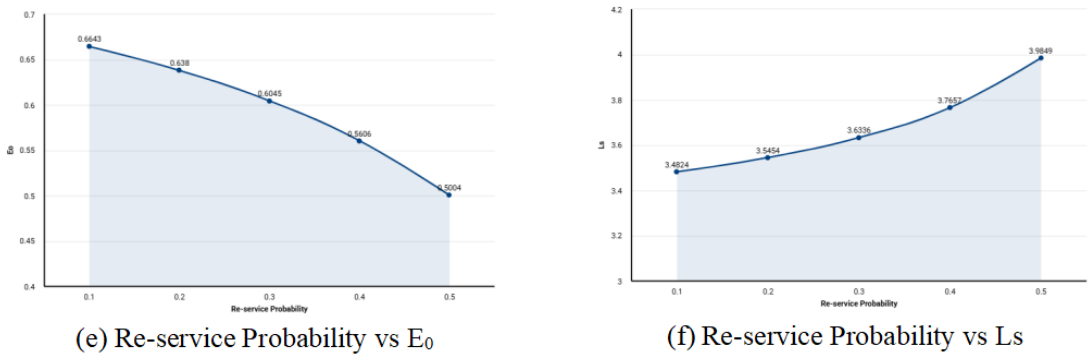
(a) Negative Arrival Rate vs E_0 (b) Negative Arrival Rate vs Ls

Figure 2: Effect of the negative arrival rate on the performance measures



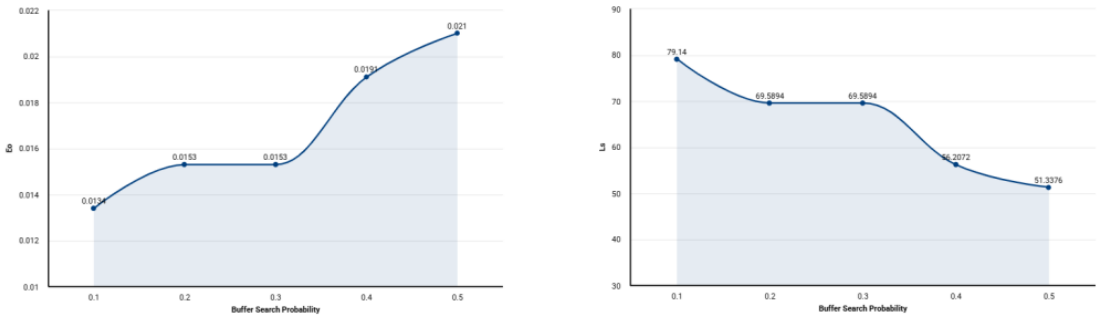
(c) Retrial Rate vs E_0 (d) Retrial Rate vs Ls

Figure 3: Effect of the retrial rate on the performance measures



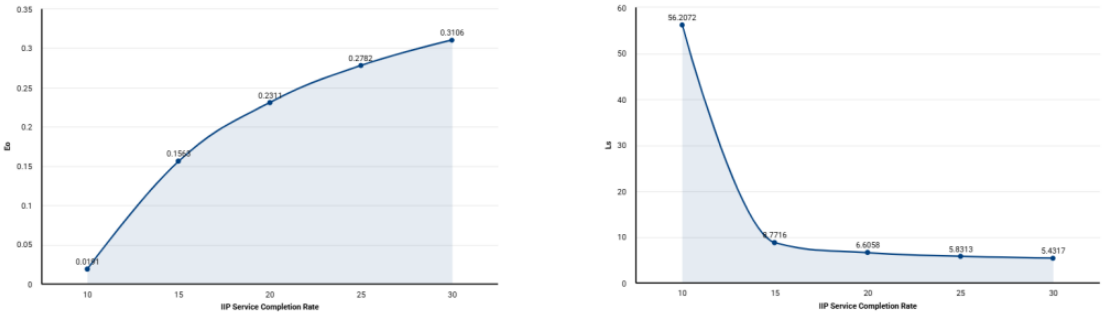
(e) Re-service Probability vs E_0 (f) Re-service Probability vs Ls

Figure 4: Effect of the re-service probability on the performance measures



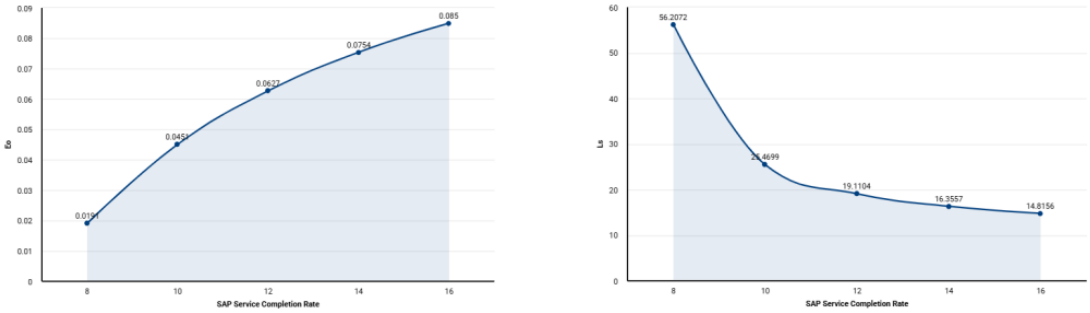
(g) Buffer Search Probability vs E_0 (h) Buffer Search Probability vs Ls

Figure 5: Effect of the buffer search probability on the performance measures



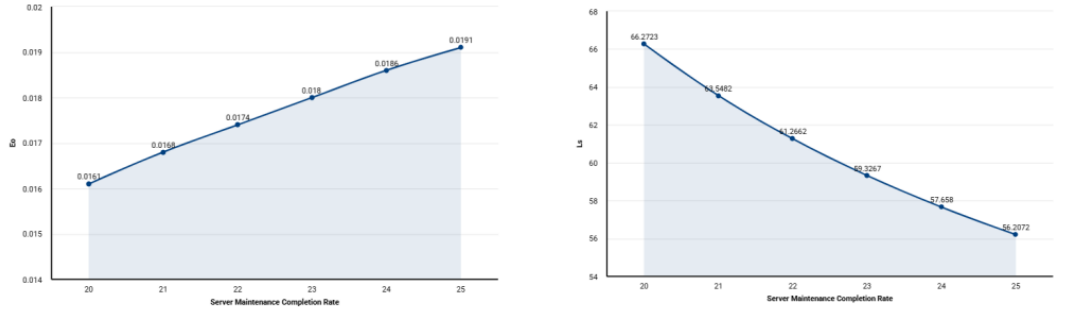
(i) IIP Service Completion Rate vs E_0 (j) IIP Service Completion Rate vs Ls

Figure 6: Effect of the IIP service completion rate on the performance measures



(k) SAP Service Completion Rate vs E_0 (l) SAP Service Completion Rate vs Ls

Figure 7: Effect of the SAP service completion rate on the performance measures



(m) Server Maintenance Completion Rate vs E_0 (n) Server Maintenance Completion Rate vs Ls

Figure 8: Effect of the server maintenance completion rate on the performance measures

12 Cost optimization

PSO is a computational technique that draws inspiration from the collective behavior of bird flocks [11]. PSO is an optimization algorithm that is population-based in nature, where a group of particles will iteratively try to find an optimal solution to a specific problem. PSO is applied for the optimization of retrial queuing systems by formulating a cost function to enhance the system performance, which is a critical problem in VoIP call centers where resource utilization and efficiency of cost are significant.

The expected total cost (ETC) for the system under study has been calculated by selecting parameter values based on standard practices reported in the literature [17] are set as $\lambda^+ = 0.8$, $\lambda^- = 0.4$, $\delta = 0.8$, $\omega = 0.3$, $\theta = 0.7$, $\eta = 15$, $\zeta_1 = 10$, $\zeta_2 = 8$, $\beta_1 = 10$, $\beta_2 = 5$, $\vartheta = 25$, $C_1 = 0.5$, $C_2 = 0.5$ and the cost values as

- Server Cost: Server cost is often calculated based on the resources used and the bandwidth consumption for each call, $Cs = \$0.20$.
- Holding Cost: Holding cost is associated with the time that a caller is on hold, $C_H = \$0.80$
- Service Cost: Service cost includes the cost for providing two phases of service, $C_E = \$0.20$.
- Repair Cost: Repair cost includes the cost for providing repair to the failed server per unit time, $C_R = \$0.20$.
- Maintenance Cost: Maintenance cost per unit time, $C_V = \$0.10$.

Using the above cost components and the steady-state performance measures derived from the retrial queuing model, the ETC function is formulated as

$$ETC = C_S + C_H L_S + C_E(W + P) + C_R(R_1 + R_2) + C_V V.$$

Objective: The objective of the optimization is to minimize the ETC by optimally selecting system parameters that significantly influence the performance of the retrial queuing system. These parameters include the incoming call rate (λ^+), re-service probability (δ), server maintenance probability (ω) and buffer search probability (θ) by varying the input learning parameters retrial rate η and IIP service completion rate $\zeta_1 = 10$.

To minimize the total cost, ETC function is formulated as

$$ETC\,(\eta^*, \zeta_1^*) = \min ETC\,(\eta, \zeta_1)$$

12.1 PSO algorithm

Step 1

Input parameters: The input learning parameters (η, ζ_1) are given.

Step 2

Initialization: Randomly initialize a swarm of particles with random values for the decision variables (η, ζ_1) within the bounds; lower bound = [10, 10] and upper bound = [40, 40]

- Number of particles = 25
- Maximum number of iterations = 50

- Inertia weight = 0.9

- Cognitive coefficient = 1.5 and Social coefficient = 1.5. The inertia weight close to unity promotes broader search capability in the initial stages, while equal cognitive and social coefficients ensure balanced influence between a particle's personal best position and the global best position of the swarm.

Step 3

Evaluation: Evaluate the cost function at each particle's current position.

Step 4

Iteration: Each particle iteratively updates its position based on its own best-known position (personal best) and the best-known position of the entire swarm (global best).

Step 5

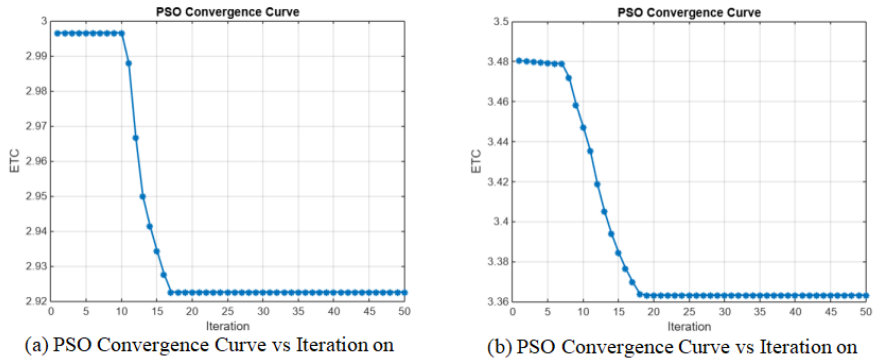
Position Update: The personal best and global best positions are updated based on the cost evaluations.

Step 6

Optimal Solution: The global best position at the end of the optimization process represents the optimal set of parameters of the retrial queuing system, minimizing the cost.

Output: The minimum cost $ETC\,(\eta^*, \zeta_1^*)$ is obtained.

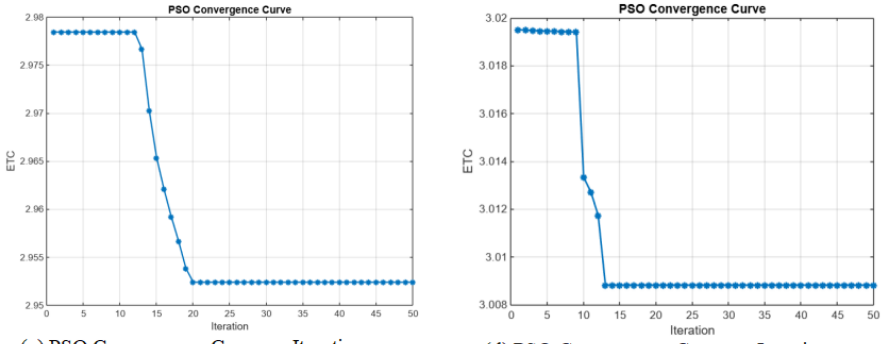
Results and Discussion: The global best position obtained at the end of the PSO process represents the optimal set of system parameters that minimize the ETC of the retrial queuing system. The convergence curves presented in Figures 9–12 show smooth and stable behavior, with the optimal ETC values reached within a small number of iterations.
To examine the sensitivity of the optimization results, key system parameters were varied within reasonable operational ranges. It was observed that the optimal ETC values changed gradually and remained within a narrow range, indicating that the obtained solutions are not highly sensitive to moderate parameter variations. This behavior confirms the robustness and reliability of the PSO-based optimization for the proposed retrial queuing model.



(a) PSO Convergence Curve vs Iteration on $\lambda^+ = 0.7$ (b) PSO Convergence Curve vs Iteration on $\lambda^+ = 0.8$

Figure 9: PSO convergence curve vs iteration on positive arrival rate

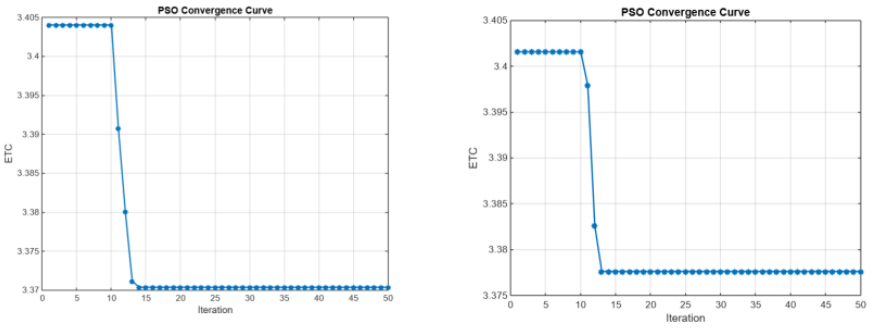
- Figure 9 illustrates that, at a lower positive arrival rate $\lambda^+ = 0.7$, the optimal value of the ETC is \$2.9226, achieved in the 17th iteration. In contrast, at a higher arrival rate $\lambda^+ = 0.8$, the optimal ETC increases to \$3.3632, obtained in the 19th iteration. This indicates that as the arrival rate increases, the optimal ETC also rises, reflecting that higher call volumes lead to increased operational costs.



(c) PSO Convergence Curve vs Iteration on $\delta = 0.5$ (d) PSO Convergence Curve vs Iteration on $\delta = 0.6$

Figure 10: PSO convergence curve vs iteration on re-service probability

- Figure 10 shows that, at a re-service probability $\delta = 0.5$, the optimal ETC is \$2.9524, achieved in the 20th iteration. As the re-service probability increases to $\delta = 0.6$, the optimal ETC also rises to \$3.0088, reached in the 13th iteration. This suggests that when more customers require additional service after their IIP, the overall system cost increases. This highlight the impact of re-service on operational efficiency and cost management in call centers.

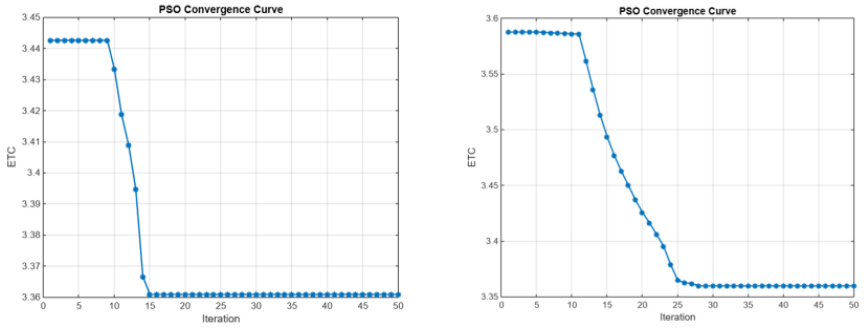


(e) PSO Convergence Curve vs Iteration on $\omega = 0.4$

(f) PSO Convergence Curve vs Iteration on $\omega = 0.5$

Figure 11: PSO convergence curve vs iteration on server maintenance probability

- Figure 11 shows that, at the server maintenance probability $\omega = 0.4$, the optimal value of the ETC, \$3.3704, is obtained in the 14th iteration. When the server maintenance probability increases to $\omega = 0.5$, the optimal ETC slightly rises to \$3.3776, reached in the 13th iteration. This marginal increase in cost shows that more frequent maintenance activities can lead to minor operational expenses in the call center.



(g) PSO Convergence Curve vs Iteration on $\theta = 0.6$

(h) PSO Convergence Curve vs Iteration on $\theta = 0.7$

Figure 12: PSO convergence curve vs iteration on the buffer search probability

- Figure 12 shows that, at the buffer search probability $\theta = 0.6$, the optimal value of the ETC, \$3.3608 is obtained in the 15th iteration. As the buffer search probability increases to $\theta = 0.7$, the optimal ETC slightly decreases to \$3.3596, obtained in the 28th iteration. This slight reduction in cost suggests that the buffer search mechanism reduces the system's overall cost.

13 Conclusion and future work

In this study, a comprehensive analytical model for a single-server VoIP call center with two obligatory service phases has been developed and analyzed. The model incorporates critical real-world factors such as re-service based on caller satisfaction, service interruptions caused by malware attacks, maintenance and buffer search activities. Key performance metrics, including server state probabilities, expected buffer size, mean number of callers in the system and buffer, reliability indices, mean busy period and mean busy cycle have been rigorously derived. The verification of the stochastic decomposition property validates the structural soundness of the model. Numerical analysis reveals the significant impact of system parameters on the performance measures. Furthermore, the PSO technique has been successfully employed to optimize the ETC of operating the proposed system. This research provides valuable retrial queuing insights in VoIP call center systems under practical constraints.

Future research could extend the model by incorporating a standby server operating at a reduced service rate during malware-induced interruptions, thereby enhancing service continuity. Exponential distributions are assumed for all time variables to preserve analytical solvability; extending to general distributions would increase model complexity and offer a direction for future research. Multi-server configurations can be investigated further, which would involve tracking multiple server states and require multi-dimensional Markov chains or matrix-analytic techniques to capture interactions among servers. Incorporating advanced customer behavior, such as impatience, alongside exploring alternative optimization techniques, would enrich the model's applicability.

Acknowledgements

The authors are grateful to their anonymous referees and editor for their constructive comments.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Abdollahi, S. and Salehi Rad, M.R. *Analysis of a batch arrival retrial queue with two-phase services, feedback and admission*, *Bull. Iran. Math. Soc.* (2022), 1–14.

[2] Arivudainambi, D. and Gowsalya, M. *A single server non-Markovian retrial queue with two types of service and Bernoulli vacation*, *Int. J. Oper. Res.* 33(1) (2018), 55–81.

[3] Artalejo, J.R. *Accessible bibliography on retrial queues: Progress in 2000–2009*, *Math. Comput. Model.* 51(9–10) (2010), 1071–1081.

[4] Boualem, M., Bouchentouf, A.A., Bareche, A. and Cherfaoui, M. *Stochastic interpretation for a single server retrial queue with Bernoulli feedback and negative customers*, *Appl. Math. J. Chin. Univ.* 40(1) (2025), 1–19.

[5] Boussaha, Z., Oukid, N., Zeghdoudi, H., Soualhi, S. and Djellab, N. *On the M/G/1 feedback retrial queuing with orbital search of customers*, *Adv. Math. Sci. J.* 11(8) (2022), 723–739.

[6] Chandrika, K.U. and Sumitha, D. *Performance measures of feedback batch arrival retrial queue with optional second service and orbital search*, *Adv. Appl. Res.* 3(2) (2011), 179–186.

[7] Chang, F.M., Liu, T.H. and Ke, J.C. *On an unreliable-server retrial queue with customer feedback and impatience*, *Appl. Math. Model.* 55 (2018), 171–182.

[8] Choudhury, G. *A two phase batch arrival retrial queuing system with Bernoulli vacation schedule*, *Appl. Math. Comput.* 188(2) (2007), 1455–1466.

[9] Cooper, R.B. *Introduction to Queueing Theory*, Elsevier (North-Holland), Amsterdam, 1981.

[10] Dimitriou, I. *A single server retrial queue with event-dependent arrival rates*, *Ann. Oper. Res.* 331(2) (2023), 1053–1088.

[11] Eberhart, R. and Kennedy, J. *Particle swarm optimization*, *Proc. IEEE Int. Conf. Neural Netw.* 4 (1995), 1942–1948.

[12] Falin, G.I. *A survey of retrial queues*, *Queueing Syst.* 7 (1990), 127–167.

[13] Gelenbe, E. *Random neural networks with negative and positive signals and product form solution*, *Neural Comput.* 1(4) (1989), 502–510.

[14] Gómez-Corral, A. *Stochastic analysis of a single server retrial queue with general retrial times*, *Naval Res. Logist.* 46(5) (1999), 561–581.

[15] Hu, K., Allon, G. and Bassamboo, A. *Understanding customer retrials in call centers: Preferences for service quality and service speed*, *Manuf. Serv. Oper. Manag.* 24(2) (2022), 1002–1020.

[16] Jain, M., Sharma, G.C. and Sharma, R. *A batch arrival retrial queuing system for essential and optional services with server breakdown and Bernoulli vacation*, *Int. J. Internet Enterp. Manag.* 8 (2012), 16–45.

[17] Keerthiga, S. and Indhira, K. *Analysis of single server feedback retrial queue with Bernoulli working vacation and starting failure*, *Reliab. Theory Appl.* 19(3(79)) (2024), 310–326.

[18] Krishnamoorthy, A., Deepak, T.G. and Joshua, V.C. *An M/G/1 retrial queue with non-persistent customers and orbital search*, *Stoch. Anal. Appl.* 23(5) (2005), 975–997.

[19] Kumar, B. *A retrial queuing system with negative arrival under Bernoulli vacation schedule and optional repair*, (2024) PREPRINT (Version 1) available at Research Square https://doi.org/10.21203/rs.3.rs-4591933/v1

[20] Micheal, M.N.G. and Indhira, K. *A literature review on retrial queuing system with Bernoulli vacation*, *Yugosl. J. Oper. Res.* 34(1) (2024), 109–134.

[21] Mokaddis, G.S., Metwally, S.A. and Zaki, B.M. *A feedback retrial queuing system with starting failures and single vacation*, *J. Appl. Sci. Eng.* 10(3) (2007), 183–192.

[22] Radha, J., Indhira, K. and Chandrasekaran, V.M. *A repairable M[X]/Gk/1 feedback retrial G-queue, balking and reneging, search of customers with vacation*, *Int. J. Knowl. Manag. Tour. Hosp.* 1(4) (2017), 457–476.

[23] Rajadurai, P., Saravanarajan, M.C. and Chandrasekaran, V.M. *Analysis of an M/G/1 feedback retrial queue with unreliable server, non-persistent customers, single working vacation and vacation interruption*, *Int. J. Serv. Oper. Manag.* 24(2) (2016), 235–266.

[24] Rusilko, T.V. and Pankov, A.V. *Queuing network model of a call center with customer retrials and impatient customers*, *Izv. Saratov Univ. Math. Mech. Inform.* 24(2) (2024), 287–297.

[25] Seenivasan, M., Ramesh, R., Chandralekha, S. and Pandisaranya, S. *Single server retrial queuing system with reneging, feedback and balking customers*, *Proc. 1st Int. Conf. Innov. Commun. Electr. Comput. Eng. (ICICEC)* (2024), 1–7.

[26] Singla, N. and Kaur, H. *A two-state retrial queuing model with feedback and balking*, *Reliab. Theory Appl.* 16(SI 2(64)) (2021), 142–155.

[27] Van Do, T. *Bibliography on G-networks, negative customers and applications*, *Math. Comput. Model.* 53(1–2) (2011), 205–212.

Aims and scope

Iranian Journal of Numerical Analysis and Optimization (IJNAO) is published by the Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad. Papers dealing with different aspects of numerical analysis and optimization, theories and their applications in engineering and industry are considered for publication.

Journal Policy

All submissions to IJNAO are first evaluated by the journal's Editor-in-Chief or one of the journal's Associate Editors for their appropriateness to the scope and objectives of IJNAO. If deemed appropriate, the paper is sent out for review using a single blind process. Manuscripts are reviewed simultaneously by reviewers who are experts in their respective fields. The first review of every manuscript is performed by at least two anonymous referees. Upon the receipt of the referee's reports, the paper is accepted, rejected, or sent back to the author(s) for revision. Revised papers are assigned to an Associate Editor who makes an evaluation of the acceptability of the revision. Based upon the Associate Editor's evaluation, the paper is accepted, rejected, or returned to the author(s) for another revision. The second revision is then evaluated by the Editor-in-Chief, possibly in consultation with the Associate Editor who handled the original paper and the first revision, for a usually final resolution.

The authors can track their submissions and the process of peer review via: http://ijnao.um.ac.ir

All manuscripts submitted to IJNAO are tracked by using "iThenticate" for possible plagiarism before acceptance.

Instruction for Authors

The Journal publishes all papers in the fields of numerical analysis and optimization. Articles must be written in English.
All submitted papers will be refereed and the authors may be asked to revise their manuscripts according to the referee's reports. The Editorial Board of the Journal keeps the right to accept or reject the papers for publication.
The papers with more than one authors, should determine the corresponding author. The e-mail address of the corresponding author must appear at the end of the manuscript or as a footnote of the first page.

It is strongly recommended to set up the manuscript by Latex or Tex, using the template provided in the web site of the Journal. Manuscripts should be typed double-spaced with wide margins to provide enough room for editorial remarks.

References should be arranged in alphabetical order by the surname of the first author as examples below:

[1] Brunner, H. *A survey of recent advances in the numerical treatment of Volterra integral and integro-differential equations*, J. Comput. Appl. Math. 8 (1982), 213-229.

[2] Stoer, J. and Bulirsch, R. *Introduction to Numerical Analysis*, Springer-verlag, New York, 2002.

Iranian Journal of Numerical Analysis and Optimization

CONTENTS Vol. 16, No. 4, pp 1189-1621

A semi-analytical study on Hepatitis A virus transmission in Indonesia 1189
A. Maheswari, V. Ananthaswamy and J. Anantha Jothi

Second-kind Chebyshev spectral Petrov–Galerkin scheme for solving the time-fractional Euler–Bernoulli beam equation 1221
Y.H. Youssri and M.M. Muttardi

Analysis and optimal control of a proliferating-quiescent Glioma model with combined therapy 1239
T. Jaber and O. Balatif

Traveling-wave patterns and bifurcation dynamics in a spatial predator-prey model 1271
S. Hussain and M. Sen

Practical early stopping for adaptive barrier SGD: Balancing stochastic speed with validation accuracy 1295
A., Fillali, S., Addoune and R., Zeghdane

A novel collocation-based numerical method for higher-order linear ODE systems with pharmacokinetic applications 1319
F. Kaci

Conformable power series method for solving Fisher–Kolmogorov–Petrovskii–Piskunov equation 1341
D.L. Suthar, F. Desalew, M. Aychluh and G. Tirite

Impact of lifestyle interventions and adequate medical infrastructure on diabetes prevention: Dynamics analysis and optimal control 1366
M. Kalita, Z. Mondal, B. Bhuyan, B.J. Nath, H.K. Sarmah

Mathematical modeling and control strategies for meningitis transmission . . . 1401
M. Baroudi, B. Bettioui, M. Belam and A. Labzai

Design of a multi-stage hybrid optimizer combining Bayesian and evolutionary algorithms for modeling and optimizing seizure-related information in EEG signals 1432
M. Khansari and H. Afshari Rad

Coupled fixed points for a class of noncyclic contractions in uniformly convex Banach spaces with error estimates 1464
A. Safari-Hafshejani and Kh. Amiri

A multi-layer SEIRS-V network epidemic model for IoT malware propagation with heterogeneous device types and patch dynamics 1484
A. Tajari Siahmarzkooh

A collocation framework for numerical reconstruction of the source function in a two-term fractional diffusion problem 1518
S.S. Jalalian, A. Babaei and M. Gachpazan

Fractional order Chelyshkov wavelet for solving variable coefficients Caputo time-fractional convection-diffusion equation 1550
S. Sabir and A. Ahmad

Queuing analysis of a VoIP call center under retrial policy with re-service, server maintenance and service disruptions 1591
Preena R. and Sumitha D.

web site: https://ijnao.um.ac.ir
Email: ijnao@um.ac.ir
ISSN-Print: 2423-6977
ISSN-Online: 2423-6969