Mathematical finance

Heston Model Calibration: Implementation and Empirical Evaluation

Somayeh Fallah*,

Department of Mathematics, Faculty of Mathematical Sciences, Alzahra University, Tehran, Iran
Email: S.Fallah@alzahra.ac.ir

Abstract. This paper offers a detailed and pedagogical examination of the Heston stochastic volatility model, a framework that continues to influence both theoretical research and practical applications in option pricing. By modeling volatility as a stochastic process, the Heston formulation addresses empirical features of financial markets that the classical Black–Scholes model cannot explain, including volatility clustering, skewness, and heavy-tailed return distributions. The discussion is organized around three core components: the model's theoretical foundation, its computational implementation, and its empirical calibration to SPY option data. Particular attention is given to the interaction of model parameters and the numerical challenges involved in estimation. Analytical pricing is introduced through the Fourier-based semi-closed form solution, while the calibration is carried out using Monte Carlo simulation combined with a gradient descent optimization scheme. The paper ultimately seeks to bridge the conceptual structure of the model with its practical interpretation, emphasizing how stochastic volatility theory connects mathematical modeling to observable market behavior.

Keywords: Heston model; stochastic volatility; gradient descent calibration; Monte Carlo simulation

AMS Mathematics Subject Classification [2020]: 91G30, 91G60

1. Introduction

The formulation of the Black–Scholes (BS) model [1] marked a cornerstone in modern option pricing. Its analytical clarity and closed-form solution established the foundation of quantitative finance. However, the model's assumption of constant volatility conflicts with market evidence: real option data, such as SPY options from the Options Industry Council (OIC), display volatility smiles, heavy tails, and time-varying variance. These discrepancies motivated stochastic volatility models, where volatility evolves as a random process.

Among these, the Heston model [2, 3] is particularly valued for combining empirical realism with analytical tractability. Here, the instantaneous variance follows a mean-reverting square-root process, allowing the model to reproduce market phenomena while preserving a semi-closed pricing form through characteristic functions.

*Speaker.

This paper provides a pedagogical overview of the Heston model, emphasizing both theory and implementation. Analytical foundations are integrated with empirical calibration using real SPY option data, demonstrating gradient-based parameter estimation and Monte Carlo simulation for model validation.

2. Model Framework

The Heston model describes asset price dynamics under stochastic volatility. It extends the Black–Scholes framework by allowing time-dependent variance and captures features such as the volatility smile and skew.

2.1. Stochastic Dynamics. Under the risk-neutral measure $\mathbb{Q}$, the asset price S_t and variance v_t evolve as

$$dS_t = rS_t\,dt + \sqrt{v_t}\,S_t\,dW_t^{(1)}, \tag{1}$$

$$dv_t = \kappa(\bar{v} - v_t)\,dt + \eta\sqrt{v_t}\,dW_t^{(2)}, \tag{2}$$

where r is the risk-free rate, κ the rate of mean reversion, $\bar{v}$ the long-run variance, η the volatility of volatility, and $dW_t^{(1)}dW_t^{(2)} = \rho\,dt$. The correlation $\rho \in [-1, 1]$ captures the leverage effect observed in equities.

The process v_t follows a Cox–Ingersoll–Ross (CIR) dynamic, ensuring positivity under the Feller condition $2\kappa\bar{v} > \eta^2$. Together, (S_t, v_t) form a two-dimensional diffusion system capable of generating volatility smiles absent in the Black–Scholes model (Figure 1).

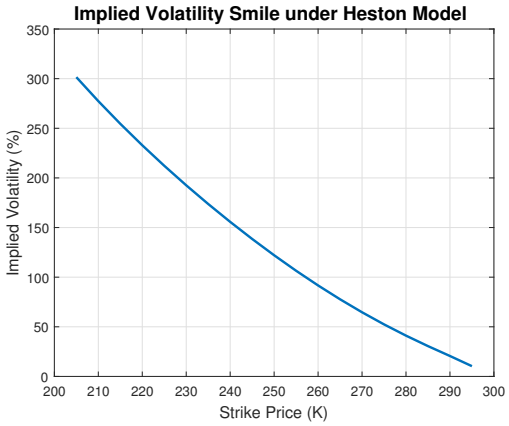


FIGURE 1. Stylized implied volatility smile under the Heston model.

2.2. Analytical Pricing Formulation. Despite its stochastic structure, the Heston model admits a semi-closed form for European call prices:

$$C(S_t, v_t, t) = S_t P_1 - Ke^{-r(T-t)}P_2,$$

where P_1 and P_2 are risk-neutral probabilities computed via Fourier inversion:

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-iu\ln K}\phi_j(u)}{iu}\right] du, \quad j = 1, 2.$$

The characteristic functions $\phi_j(u) = \exp\{C_j + D_j v_t + iu\ln S_t\}$ are obtained from Riccati-type differential equations in parameters κ, $\bar{v}$, η, ρ, and r. This approach ensures numerical stability and efficient pricing, forming a benchmark for calibration.

3. Gradient Descent Calibration Procedure

Model calibration seeks the parameter vector

$$\boldsymbol{\theta} = [v_0, \bar{v}, \kappa, \eta, \rho],$$

that minimizes the mean squared difference between market and model prices:

$$L(\boldsymbol{\theta}^{(t)}) = \frac{1}{N} \sum_{i=1}^{N} \left[C_{\text{model}}(K_i; \boldsymbol{\theta}^{(t)}) - C_{\text{mkt}}(K_i) \right]^2.$$

Gradients are approximated by symmetric finite differences:

$$\frac{\partial L}{\partial \theta_j} \approx \frac{L(\theta_j + \varepsilon) - L(\theta_j - \varepsilon)}{2\varepsilon}, \quad \varepsilon = 10^{-4},$$

and parameters are updated by

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \alpha \nabla L(\boldsymbol{\theta}^{(t)}),$$

with learning rate $\alpha = 0.05$. Parameters are constrained within plausible bounds ($0.04 \leq v_0, \bar{v} \leq 0.25$, $-0.8 \leq \rho \leq 0.2$). The iteration continues until convergence or 40 steps.

4. Numerical Implementation and Results

The procedure was implemented in MATLAB using SPY call options ($T = 72$ days, $S_0 = 287.97$, $r = 0.02$). Monte Carlo pricing employed 200 time steps and 10,000 paths per strike. The initial guess was $\boldsymbol{\theta}_0 = [0.09, 0.09, 2.0, 0.4, -0.7]$. After 40 iterations, the optimized parameters were

$$\boldsymbol{\theta}^* = [0.04, 0.25, 0.5, 0.05, -0.8],$$

yielding a minimum loss $L \approx 1.81$. The algorithm converged smoothly after early oscillations caused by simulation noise.

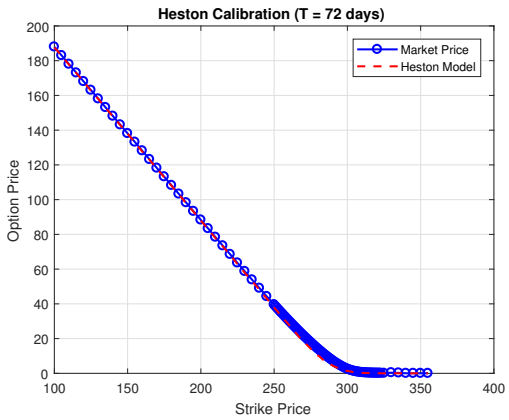


FIGURE 2. Calibrated Heston model vs. market call prices ($T = 72$ days).

The calibrated model closely matches observed market prices across strikes, confirming that the gradient descent method provides a stable and efficient calibration framework.

5. Conclusion

The Heston stochastic volatility model combines analytical rigor and empirical realism. Through stochastic variance, it bridges the gap between the idealized Black–Scholes world and real market behavior. This paper outlined its formulation, calibration, and implementation, demonstrating that a simple gradient descent scheme can effectively estimate parameters from market data. While sensitivity to initialization remains a limitation, the method delivers consistent fits and highlights the models enduring relevance in modern quantitative finance.

References

[1] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *J. Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[2] S. L. Heston, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Rev. Financial Studies*, vol. 6, no. 2, pp. 327–343, 1993.

[3] Z. Cao and X. Lin, "Theoretical and Empirical Validation of the Heston Model," Johns Hopkins University, 2024.

Portfolio Optimization Interdiction: A Novel Risk-Averse Vision for Risk Management

Seyyed Mohammad Reza Kazemi[1,*] Javad Tayyebi[2]

[1]Department of Industrial Engineering, Faculty of Industrial and Computer Engineering, Birjand University of Technology, Birjand, Iran.
Email: kazemi@birjandut.ac.ir

[2]Department of Industrial Engineering, Faculty of Industrial and Computer Engineering, Birjand University of Technology, Birjand, Iran.
Email: javadtayyebi@birjandut.ac.ir

Abstract. Portfolio optimization is a crucial problem in financial management. Since there are several types of risks, decision makers cannot manage their portfolios effectively based solely on historical data. On the other hand, the use of predicted data contains high errors which limits its practical applicability. The most well-known optimization model is Markowitz's model, which minimizes risk level through a quadratic objective function while maintaining a fixed level of profit. However, this approach cannot adequately manage risks arising from different unexpected events.

This paper proposes a bilevel optimization model from the perspective of a risk-averse decision maker. In the first level, different events are modeled under various scenarios. Then, the decision maker operates at the second level, observing these events and determining how to set their portfolio. We utilize concepts from interdiction problems to convert the bilevel problem into a single-level optimization problem that can be solved using optimization solvers such as Gurobi and CPLEX.

Keywords: portfolio optimization, interdiction problems, risk management, Stackelberg game

AMS Mathematics Subject Classification [2020]: 91G10, 90C11, 91A65

1. Introduction

Portfolio optimization has been a fundamental problem in financial management since the pioneering work of Markowitz [1]. The Markowitz mean-variance framework minimizes portfolio risk, measured as variance, while achieving a target expected return. This model has served as the foundation for modern portfolio theory and has been extensively studied and extended in the literature [2, 3].

However, traditional portfolio optimization models face significant limitations in managing risks arising from unexpected events. Historical data alone cannot capture the full spectrum of potential risks, while predicted data often contains substantial errors that

*Speaker.

limit practical applicability. Financial markets are increasingly exposed to various disruptive events such as economic crises, geopolitical conflicts, and pandemics, which can severely impact portfolio performance [4]. Moreover, these events often exhibit complex dependencies where the occurrence of one event may influence the probability or impact of another.

This paper addresses these limitations by proposing a novel bilevel optimization framework that explicitly incorporates event-based risks and their interdependencies. Our approach models different risk events at the upper level while allowing the decision maker to optimize portfolio allocation at the lower level. The resulting interdiction problem captures the strategic interaction between adverse events and portfolio decisions, providing a more robust framework for risk management.

2. Model description

Consider a financial market with n available assets. Let $x = (x_1, x_2, ..., x_n)$ represent the portfolio allocation, where x_i denotes the proportion of wealth invested in asset i. The classical Markowitz model can be formulated as:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x^\top \Sigma x \\ \text{subject to} \quad & \mu^\top x \geq R \\ & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0 \end{aligned} \tag{1}$$

where Σ is the covariance matrix of asset returns, μ is the vector of expected returns, and R is the target return.

In our bilevel framework, we consider K different events that may affect portfolio performance. Each event $k \in \{1, 2, ..., K\}$ can have varying impacts on different assets. Let y_k be a binary variable indicating whether event k occurs, and let δ_k represent the impact magnitude of event k on portfolio returns.

The upper-level problem, from the perspective of a risk-averse decision maker, aims to minimize the worst-case portfolio performance:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & \Phi(y) \\ \text{subject to} \quad & \sum_{k=1}^{K} y_k \leq \Gamma \\ & y_{k_1} \leq y_{k_2} \quad \forall (k_1, k_2) \in \mathcal{C} \\ & y_{k_1} + y_{k_2} \leq 1 \quad \forall (k_1, k_2) \in \mathcal{M} \\ & \sum_{k \in S} y_k \leq r_S \quad \forall S \in \mathcal{S} \\ & y_k \in \{0, 1\}, \quad k = 1, 2, ..., K \end{aligned} \tag{2}$$

where: - $\Phi(y)$ represents the optimal value of the lower-level portfolio optimization problem given event selection y - Γ limits the number of simultaneous events - $\mathcal{C}$ represents the set of complementary events where if event k_1 occurs, then event k_2 must also occur ($y_{k_1} \leq y_{k_2}$) - $\mathcal{M}$ represents the set of mutually exclusive events where at most one of

the two events can occur - $\mathcal{S}$ represents collections of events with cardinality constraints, limiting how many events from set S can occur simultaneously

The lower-level problem represents the portfolio optimization under given events:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & (\mu - \sum_{k=1}^{K} y_k \delta_k)^\top x \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0 \end{aligned} \tag{3}$$

This bilevel structure captures the strategic interaction between adverse events (upper level) and portfolio decisions (lower level), providing a robust framework for risk management that accounts for complex event relationships.

3. Solution approach

To solve the bilevel portfolio optimization problem, we employ the Dualize-and-Combine approach commonly used in interdiction problems. This method involves dualizing the lower-level problem and combining it with the upper-level problem to form a single-level equivalent.

The lower-level linear programming problem has the following dual:

$$\begin{aligned} \underset{\lambda}{\text{minimize}} \quad & \lambda \\ \text{subject to} \quad & \lambda \geq \mu_i - \sum_{k=1}^{K} y_k \delta_{ki}, \quad i = 1, 2, ..., n \end{aligned} \tag{4}$$

By strong duality, the primal and dual objectives are equal at optimality. We can therefore reformulate the bilevel problem as a single-level mixed-integer program:

$$\begin{aligned} \underset{y,\lambda}{\text{minimize}} \quad & \lambda \\ \text{subject to} \quad & \lambda \geq \mu_i - \sum_{k=1}^{K} y_k \delta_{ki}, \quad i = 1, 2, ..., n \\ & \sum_{k=1}^{K} y_k \leq \Gamma \\ & y_{k_1} \leq y_{k_2} \quad \forall (k_1, k_2) \in \mathcal{C} \\ & y_{k_1} + y_{k_2} \leq 1 \quad \forall (k_1, k_2) \in \mathcal{M} \\ & \sum_{k \in S} y_k \leq r_S \quad \forall S \in \mathcal{S} \\ & y_k \in \{0, 1\}, \quad k = 1, 2, ..., K \end{aligned} \tag{5}$$

This single-level reformulation can be efficiently solved using commercial optimization solvers such as Gurobi or CPLEX.

4. Numerical example

To illustrate our approach, we consider a simple example with 5 assets and 4 potential events. The input parameters are as follows:

Expected returns:

$$\mu = \begin{bmatrix} 0.12 & 0.08 & 0.15 & 0.10 & 0.09 \end{bmatrix}^\top \tag{6}$$

Event impacts on assets:

$$\delta = \begin{bmatrix} 0.05 & 0.08 & 0.03 & 0.06 & 0.04 \\ 0.03 & 0.02 & 0.10 & 0.01 & 0.07 \\ 0.06 & 0.04 & 0.02 & 0.09 & 0.05 \\ 0.02 & 0.06 & 0.07 & 0.03 & 0.08 \end{bmatrix} \tag{7}$$

where δ_{ki} represents the impact of event k on asset i.

Event dependencies:

- Complementary constraint: If Event 1 occurs, then Event 2 must also occur ($y_1 \leq y_2$)
- Mutually exclusive constraint: Events 3 and 4 cannot occur simultaneously ($y_3 + y_4 \leq 1$)
- Cardinality constraint: At most 2 events can occur simultaneously ($\sum_{k=1}^{4} y_k \leq 2$)

We solved the resulting mixed-integer linear program using Gurobi with the parameters described above. The optimal solution selects events 2 and 3, which together minimize the maximum portfolio return in the worst-case scenario while satisfying all event dependency constraints. Under this worst-case combination of events, the optimal portfolio allocation concentrates entirely on asset 1, yielding an expected worst-case portfolio return of 0.03.

5. Conclusion

This paper has presented a novel bilevel optimization framework for portfolio management that explicitly incorporates event-based risks and their complex dependencies. By modeling the interaction between adverse events and portfolio decisions, and by incorporating realistic event relationship constraints, our approach provides a more robust foundation for risk management.

Future research directions include extending the model to incorporate transaction costs, considering multi-period portfolio optimization, developing specialized algorithms for large-scale instances of the problem, and incorporating probabilistic event dependencies rather than deterministic constraints.

References

1. Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.
2. Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). *Robust portfolio optimization and management*. John Wiley & Sons.
3. Cornuejols, G., and Ttnc, R. (2006). *Optimization methods in finance*. Cambridge University Press.
4. Jorion, P. (2006). *Value at risk: the new benchmark for managing financial risk*. McGraw-Hill.
5. Shan, C., and Zhu, S. (2016). Bilevel programming approaches to portfolio optimization. *European Journal of Operational Research*, 253(3), 671-681.
6. Smith, J. C., and Song, Y. (2014). A survey of network interdiction models and algorithms. *European Journal of Operational Research*, 238(3), 654-674.

12th National Mathematics Conference of Payame Noor University

Sari—Iran

6–7 November 2025

Mathematics Learning

The Impact of AI in Teaching Undergraduate Mathematics

Seyed Amjad Samareh Hashemi [1,*], Zahra Sabri Razm

[1]Department of Mathematics, Payame Noor University (PNU), P.O. Box 19395-4697, Tehran, Iran.

Email: a_hashemi@pnu.ac.ir

Email: zahrasabri79@gmail.com

Abstract. Artificial Intelligence (AI) reshapes undergraduate mathematics education through tools that enable personalized learning, adaptive assessments, and interactive visualizations. These provide real-time feedback, efficient grading, and data insights, helping educators meet diverse student needs. However, AI's limitations in fostering creativity, emotional support, and nuanced reasoning challenge deep understanding. Ethical issues like data privacy, bias, and access equity complicate integration. This article analyzes AI's advantages and disadvantages in undergraduate contexts, such as calculus and algebra, using recent literature and examples. It proposes a hybrid human-AI model to optimize outcomes, emphasizing ethical integration for equitable education.

Keywords: Artificial Intelligence, AI, Mathematics, Education

AMS Mathematics Subject Classification [2020]: 97P80, 97U10, 97U70

1. Introduction

AI emerged as a discipline at the 1956 Dartmouth Conference, organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, coining the term "artificial intelligence" [4]. Alan Turing's foundational work on computability and the Turing Test laid theoretical groundwork [5]. AI revolutionizes undergraduate mathematics education with personalized experiences, adaptive assessments, and interactive tools. Yet, it poses risks like over-reliance and ethical dilemmas. This paper examines AI's multifaceted role, drawing on literature from 2012–2025, to highlight benefits, challenges, and ethics. It contributes by proposing hybrid strategies, such as AI-assisted proofs in discrete math, and calls for empirical studies at institutions like Payame Noor University.

1.1. Methodology. This review synthesizes peer-reviewed articles, position statements (e.g., NCTM), and AI tool documentation from 2012–2025, sourced via databases like Google Scholar and ERIC. Focus areas include advantages, disadvantages, and ethics, with emphasis on undergraduate applications.

*Speaker.

2. Benefits and Challenges

AI transforms undergraduate mathematics but introduces complexities. This section details advantages and disadvantages, supported by examples.

2.1. Advantages of AI in Teaching Undergraduate Mathematics. AI brings transformative benefits to undergraduate mathematics education, including:

- **Personalized and Adaptive Learning:** Platforms like ALEKS[1] assess knowledge and adjust paths [1]. For instance, in calculus, AI identifies derivative misconceptions and provides targeted exercises, boosting engagement and mastery. Recent tools like Khanmigo[2] (2025) use generative AI for interest-aligned problems, e.g., sports-themed integrals.
- **Real-Time Feedback and Interactive Tools:** Photomath[3], Mathway[4] and GeoGebra[5] provide instant feedback and visualizations [3]. In multivariable calculus, Desmos[6] explores of functions like $f(x, y) = x^2+y^2$; newer tools like Tutero AI[7] (2025) simulate Monte Carlo methods for probability courses.
- **Efficiency in Grading and Data Analytics:** AI systems streamline grading of routine assignments, such as algebra exercises, and provide detailed performance analytics [2]. This allows instructors to identify trends—such as widespread difficulty with matrix operations—and adjust their teaching strategies accordingly, ensuring timely support for students.
- **Enhanced Tutoring and Professional Development:** AI-driven tutoring systems adapt to individual learning paces and styles, while also offering customized professional development for educators [2]. For example, an AI tutor might guide a student through a proof step-by-step, while a training module could help instructors integrate technology into their calculus courses effectively.
- **Reinforcement Learning for Math Tutoring.** AI can optimize tutoring through reinforcement learning, adapting strategies based on student responses. For example, an AI tutor might adjust its approach if a student repeatedly struggles with integration by parts, offering alternative explanations or simpler examples. This adaptability enhances tutoring effectiveness, particularly for complex undergraduate topics. [2] notes that reinforcement learning can improve tutoring outcomes by tailoring interactions to individual needs.
- **Online Math Competitions.** AI-powered platforms can host online math competitions with adaptive problem sets, challenging students at their appropriate level. For instance, a platform might present a student with increasingly difficult linear algebra problems based on their performance. These competitions foster a competitive yet supportive environment, encouraging students to deepen their mathematical skills. [2] suggests that such platforms enhance engagement and motivation.

[1] Personalized Learning Platform, Available at: https://www.aleks.com/
[2] An AI-powered personal tutor and teaching assistant, Available at: https://www.khanmigo.ai/
[3] Math Problem Solver App, Available at: https://photomath.com/
[4] Online Math Problem Solver, Available at: https://www.mathway.com/
[5] Interactive Mathematics Software, Available at: https://www.geogebra.org/
[6] Graphing Calculator and Math Tools, Available at: https://www.desmos.com/
[7] Personalized Education for Schools and Families, Avalable at: https://www.tutero.com/

AI tools like ChatGPT[8], can generate diverse assessments and engaging content, reducing the workload on educators. For example, ChatGPT can create multiple versions of a calculus test with different examples, ensuring variety while maintaining learning objectives. This capability saves time and enhances the quality of instructional materials. According to [2], such tools can increase student engagement by providing fresh and relevant content.

TABLE 1. Comparison of AI Tools for Undergraduate Math

Tool	Math Topic Example	Advantage	Limitation
ALEKS	Calculus	Adaptive paths	Limited creativity
GeoGebra	Geometry	3D visualizations	Access inequities
Khanmigo	Probability	Personalized problems	Potential biases
Tutero AI	Algebra	Teacher coaching	Over-simplification

2.2. Disadvantages of AI in Teaching Undergraduate Mathematics. Despite its potential, AI poses several challenges in mathematics education:

- **Limitations in Creativity and Reasoning:** AI struggles to replicate the creative problem-solving and nuanced reasoning human instructors provide [2]. For example, while AI can solve a differential equation, it may not explain the intuition behind selecting a particular method, limiting students' ability to develop independent mathematical insight.
- **Over-Reliance and Skill Degradation:** Excessive dependence on AI tools like equation solvers can erode students' foundational skills and critical thinking [2]. A student who relies on Photomath to factor polynomials might never master the underlying techniques, leaving them unprepared for advanced coursework or exams requiring manual computation.
- **Ethical and Access Issues:** AI systems may perpetuate biases embedded in their training data, and unequal access to technology can widen educational disparities [3]. For instance, an AI tool trained on datasets favoring certain demographics might misjudge the needs of underrepresented students, while rural or low-income students may lack the devices or internet connectivity to use these tools effectively.
- **Lack of Emotional Intelligence:** Unlike human teachers, AI cannot offer emotional encouragement or adapt to students' psychological needs [2]. In mathematics, where frustration is common—such as when grappling with abstract algebra—human empathy and motivation play a vital role that AI cannot yet replicate.
- **Oversimplification of Concepts.** To make concepts accessible, AI might oversimplify them, potentially undermining the complexity of mathematical ideas. For example, an AI tool explaining group theory might focus on basic examples, neglecting the nuances that are critical at the undergraduate level. [2] suggests that this oversimplification can hinder deep understanding.

[8]Optimizing Language Models for Dialogue, Available at: https://openai.com/blog/chatgpt/

3. Ethical Considerations

The deployment of AI in education demands careful attention to ethical principles:

- **Data Privacy:** Robust safeguards must protect student data from misuse or breaches [2]. Institutions should adopt encryption and clear policies to ensure that sensitive information, such as performance records, remains confidential.
- **Bias Mitigation:** Regular audits of AI algorithms are necessary to detect and correct biases that could disadvantage certain groups [3]. This might involve testing tools like ALEKS to ensure they fairly assess students across diverse backgrounds.
- **Inclusive Access:** Bridging the digital divide is critical to ensure all students benefit from AI innovations [2]. Universities could partner with organizations to provide subsidized devices or internet access, making tools like GeoGebra universally available.
- **Stakeholder Collaboration:** Ethical guidelines should emerge from collaboration among educators, technologists, and policymakers [2]. This ensures that AI tools align with educational goals and societal values, such as fairness and transparency in algorithmic decision-making.

4. Conclusion

AI enhances undergraduate mathematics with personalization and visualizations but cannot replace human creativity or support. A balanced hybrid approach—e.g., AI for grading paired with instructor-led discussions—maximizes success. Educators should pilot tools like Tutero in small classes, conduct bias audits, and ensure open-source access. Future research questions: How can AI support proof-based courses? What metrics evaluate hybrid models? Stakeholder collaboration will harness AI responsibly.

References

[1] Hu, X., Craig, S. D., Bargagliotti, A. E., Graesser, A. C., Okwumabua, T., Anderson, C., Cheney, K. R., & Sterbinsky, A. (2012). *The effects of a traditional and technology-based intervention on precalculus students' understanding of the function concept*, Journal of Computers in Mathematics and Science Teaching, 31(3), (2012), 255–279.

[2] Opesemowo, O. O., (2025). *Artificial Intelligence in Mathematics Education: Pros and Cons.*, Handbook of Research on Artificial Intelligence in Education, (2025), (pp. 123–145). IGI Global. https://doi.org/10.4018/978-1-6684-7366-5.ch084

[3] *Artificial Intelligence and Mathematics Teaching*, National Council of Teachers of Mathematics (NCTM). (n.d.), https://www.nctm.org/standards-and-positions/Position-Statements/Artificial-Intelligence-and-Mathematics-Teaching/.

[4] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E., *A proposal for the Dartmouth Summer Research Project on artificial intelligence, August 31, 1955*, AI Magazine, 27(4), (2006), https://doi.org/10.1609/aimag.v27i4.1904.

[5] Turing, A. M., *Computing machinery and intelligence*, Mind, 59(236), (1950), 433–460.

Numerical Analysis

Determinants for certain band matrices by the theory of matrix polynomials

Maryam Shams Solary[1,*]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: shamssolary@pnu.ac.ir or shamssolary@gmail.com

Abstract. This study introduces a novel relationship between Chebyshev polynomials and matrix-less methods, elucidating the role of Chebyshev polynomial roots in the theory of matrix polynomials. This technique facilitate the evaluation of determinants for certain band matrices, with particular emphasis on the polynomial eigenvalue problem.

Keywords: Chebyshev polynomials, Matrix-less method, Eigenvalue

AMS Mathematics Subject Classification [2020]: 65F30, 15A15, 65H04

1. Introduction

Band matrices are a crucial type of matrix frequently encountered in engineering and computational sciences, appearing in areas like splines, partial differential equations, and quantum physics. This paper focuses on the determinant of a specific class of band matrices. The objective is to determine the values of x for which the determinant of these:

$$(1)\qquad det\begin{bmatrix} (a_1+x) & -1 & & & & & \\ -1 & 0 & x & & & & \\ & x & (a_3+a_2x) & -1 & & & \\ & & -1 & 0 & x & & \\ & & & x & (a_5+a_4x) & -1 & \\ & & & \ddots & \ddots & \ddots & \end{bmatrix} = 0,$$

$$(2)\qquad det\ \frac{1}{2}\begin{bmatrix} (2xc_7+c_6) & -1 & -c_7 & 0 & 0 & 0 & 0 \\ (c_5-c_7) & 2x & (c_4-c_6) & -1 & 0 & 0 & 0 \\ -1 & 0 & 2x & 0 & -1 & 0 & 0 \\ 0 & -1 & (c_3-c_5) & 2x & (c_2-c_4) & -1 & 0 \\ 0 & 0 & -1 & 0 & 2x & 0 & -1 \\ 0 & 0 & 0 & -1 & (c_1-c_3) & 2x & (c_0-c_2) \\ 0 & 0 & 0 & 0 & -2 & 0 & 2x \end{bmatrix} = 0,$$

*Speaker.

and

(3)
$$det\ \frac{1}{2}\begin{bmatrix} (2xc_8+c_7) & (c_6-c_8) & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2x & 0 & -1 & 0 & 0 & 0 & 0 \\ -c_8 & (c_5-c_7) & 2x & (c_4-c_6) & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 2x & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & (c_3-c_5) & 2x & (c_2-c_4) & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 2x & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & (c_1-c_3) & 2x & (c_0-c_2) \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 2x \end{bmatrix} = 0.$$

2. Main results

A Hessenberg Toeplitz matrix is a special kind of square matrix with Hessenberg and Toeplitz form. Then, a unit lower Hessenberg Toeplitz matrix is a matrix similar this:

(4)
$$\boldsymbol{T} = \begin{bmatrix} t_1 & 1 & 0 & \cdots & 0 \\ t_2 & t_1 & 1 & \ddots & \ddots \\ t_3 & t_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ t_n & \cdots & t_3 & t_2 & t_1 \end{bmatrix}.$$

Let $p(x) = x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$, be a polynomial with coefficients over an arbitrary field. As is well known, the matrix

(5)
$$\boldsymbol{C} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \ddots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \ddots & \ddots & \ddots & 1 & 0 \end{bmatrix},$$

has the property that $det(xI - \boldsymbol{C}) = p(x)$. The matrix $\boldsymbol{C}$, or some of its modifications, is being called companion matrix of the polynomial $p(x)$:

(6)
$$det\begin{bmatrix} (a_1+x) & -1 & & & & & & \\ -1 & 0 & x & & & & & \\ & x & (a_3+a_2x) & -1 & & & & \\ & & -1 & 0 & x & & & \\ & & & x & (a_5+a_4x) & -1 & & \\ & & & \ddots & \ddots & \ddots & & \end{bmatrix} = 0.$$

Also, Chebyshev polynomials of the first kind are of great practical importance

(7)
$$f(x) = \sum_{j=0}^{n} c_j T_j(x) \quad x \in [-1,1],\ c_n \neq 0.$$

The Chebfun system is central to performing accurate numerical computations. Suppose, the input is a symmetric tridiagonal matrix like equation (1) or an unsymmetric pentadiagonal matrix like equation (2) or (3). The goal is to calculate the determinants of these

matrices. We can convert the determinants of the relevant matrices into polynomials as shown in equation (8), that a_i's and c_j 's are in the

$$p(x) = x^n + a_1 x^{n-1} + \ldots + a_{n-1}x + a_n, f(x) = \sum_{j=0}^{n} c_j T_j(x). \tag{8}$$

THEOREM 2.1. *There is a similarity between the Chebyshev polynomials and the companion matrix by the following matrices:*

(9)

$$Q_{n+1} = \begin{pmatrix} 1 & 0 & & & \cdots & & 0 \\ 0 & 1 & 0 & & \cdots & & 0 \\ -1 & 0 & 2 & 0 & \cdots & & 0 \\ 0 & -3 & 0 & 4 & 0 & & \vdots \\ 1 & 0 & -8 & 0 & 8 & 0 & \\ \vdots & & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & & \cdots & & \cdots & 2^{n-1} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \frac{-\beta_1}{\alpha_1} & \frac{1}{\alpha_1} & 0 & \cdots & & & \\ \frac{\gamma_2}{\alpha_2} & \frac{-\beta_2}{\alpha_2} & \frac{1}{\alpha_2} & 0 & \cdots & & \\ \vdots & & \ddots & \ddots & \ddots & & \\ 0 & \cdots & & \cdots & \frac{\gamma_{n-1}}{\alpha_{n-1}} & \frac{-\beta_{n-1}}{\alpha_{n-1}} & \frac{1}{\alpha_{n-1}} \\ \frac{\delta_0}{\alpha_n} & \frac{\delta_1}{\alpha_n} & \cdots & & \cdots & \frac{\delta_{n-2}+\gamma_n}{\alpha_n} & \frac{\delta_{n-1}-\beta_n}{\alpha_n} \end{pmatrix}.$$

Here $\alpha_i = 2,\ \beta_i = 0,\ \gamma_i = 1,$ *and*

$$(\alpha_1\alpha_2\ldots\alpha_n)[\gamma_0, \gamma_1, \ldots, \gamma_{n-1}, 1] = [\delta_0, \delta_1, \ldots, \delta_{n-1}, 1]Q_{n+1}, \tag{10}$$

for all $1 \le i \le n$.

THEOREM 2.2. *The matrix* $\boldsymbol{C}$ *in (5) is similar to* $\boldsymbol{T}$ *in (4):*

$$\boldsymbol{T} = \boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{S}, \tag{11}$$

By these properties:
If $c_1 = 0$, *then: for* $i < j$, $B(i : j, k : l)$ *then* B *whose rows and columns are indexed by* $i : j$ *and* $k : l$, *respectively.*
If $\boldsymbol{S} = (s_{ij})$ *and* $[\boldsymbol{S}^{-1}]_{i,j} = (-1)^{i+j} det\ \hat{\boldsymbol{S}}_{ji}$, *then*
$det\ \hat{\boldsymbol{S}}_{ji} = det\ \boldsymbol{S}(j+1 : i, j : i-1),\ i = 3, \ldots, n,\ j = 1, \ldots, i-2$. *Thus*

$$[S^{-1}]_{i,j} = (-1)^{i+j} det\ S(j+1 : i, j : i-1),\ i = 3, \ldots, n,\ j = 1, \ldots, i-2. \tag{12}$$

$$[\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{S}]_{ij} = s_{(i-j+1)j} + \sum_{k=0}^{i-j+1} (-1)^{i-j-k+3} det\ \boldsymbol{S}(j+k : i, j+k-1 : i-1)s_{kj},$$

$i = 3, \ldots, n-1,\ j = 1, \ldots, n-2$, *with* $s_{0j} = 1,\ s_{1j} = 0,\ s_{pj} = 0$ *for* $p < 0$, $det\ \boldsymbol{S}(1 : i, 0 : i-1) = 0$ *and* $det\ \boldsymbol{S}(h : g, h-1 : g-1) = 0$ *if* $h > g$ *and*

$$[\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{S}]_{nj} = \sum_{k=0}^{n-j+1} ((-1)^{n-j-k+3} det\boldsymbol{S}(j+k : n, j+k-1 : n-1)s_{kj} + c_k s_{(n-j-k+1)j}),$$

$i = n,\ j = 1, \ldots, n-2$, *with* $s_{0j} = 1,\ s_{1j} = 0,\ c_l = 0$ *for* $l \le 1$ *and* $det\ \boldsymbol{S}(i : j; k : l) = 0$ *if* $i > j$. *If* $c_1 \neq 0$, *then* $\alpha = \frac{1}{n}c_1$ *and* $T' = T - \alpha I$, *namely* $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$ *are eigenvalues of matrix* $\boldsymbol{T}$ *and* $\Lambda' = \{\lambda_1 - \alpha, \ldots, \lambda_n - \alpha\}$ *are eigenvalues of matrix* $\boldsymbol{T}'$ *that*

$$\boldsymbol{T} = \begin{bmatrix} t_1 & 1 & 0 & \cdots & 0 \\ t_2 & t_1 & 1 & \ddots & \ddots \\ t_3 & t_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ t_n & \cdots & t_3 & t_2 & t_1 \end{bmatrix}, \quad t_k = \frac{1}{n-k+1}(c_k + g_k), \tag{13}$$

$$g_k = \sum_{2l_2+\cdots+(k-2)l_{k-2}=k} \binom{n-k+l_2+\cdots+l_{k-2}}{n-k, l_2, \ldots, l_{k-2}} (-t_2)^{l_2} \cdots (-t_{k-2})^{l_{k-2}}.$$

From [5, 6], we can prove these results.

Example 2.3. Let

$$p(x) = x^5 - 72x^3 - 648x^2 - 2268x - 3888. \tag{14}$$

Then

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 3888 & 2268 & 648 & 72 & 0 \end{bmatrix}, \quad \boldsymbol{T'} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 18 & 0 & 1 & 0 & 0 \\ 216 & 18 & 0 & 1 & 0 \\ 1620 & 216 & 18 & 0 & 1 \\ 11664 & 1620 & 216 & 18 & 0 \end{bmatrix},$$

and $\boldsymbol{T} = \boldsymbol{T'} + I$ with eigenvalues $\Lambda = \{13, -2+3i, -2-3i, -2+3i, -2-3i\}$.

$$\mathbf{b} = [-3888; -2268; -648; -72; 0; 1]^T, \quad \mathbf{a} = 1.0e+03 \begin{bmatrix} -4.212000000000000 \\ -2.321375000000000 \\ -0.324000000000000 \\ -0.017687500000000 \\ 0 \\ 0.000062500000000 \end{bmatrix}.$$

We run the companion matrix or Toeplitz matrix for finding eigenvalues with Matlab and the roots of Chebyshev polynomials with Chebfun [2]:

$$\begin{bmatrix} -2.999999949630467 - 2.999999948626045i \\ -2.999999949630467 + 2.999999948626045i \\ -3.000000050369539 - 3.000000051373939i \\ -3.000000050369539 + 3.000000051373939i \\ 12.000000000000009 + 0.000000000000000i \end{bmatrix}, \quad \begin{bmatrix} 12.000000000000007 + 0.000000000000000i \\ -3.000000008539148 + 3.000000017608449i \\ -3.000000008539148 - 3.000000017608449i \\ -2.999999991460856 + 2.999999982391550i \\ -2.999999991460856 - 2.999999982391550i \end{bmatrix}.$$

If we use 'eigvals' function in JULIA for the Hessenberg Toeplitz matrix $\boldsymbol{T'}$, we have

$$\begin{bmatrix} -3.000000000000000000009380438790545636605 9 - 3.000000000000000001621757825637227602361im \\ -3.000000000000000000009380438790545636605 9 + 3.000000000000000001621757825637227602361im \\ -2.999999999999999999906195612094543636563 - 2.999999999999999983782421743627724169 29im \\ -2.999999999999999999906195612094543636563 + 2.999999999999999983782421743627724169 29im \\ 12.000000000000000000000000000000000028 + 0.0im \end{bmatrix}.$$

References

[1] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017) *Julia: A Fresh Approach to Numerical Computing*, SIAM Review, **59**, 65–98.

[2] Driscoll, T. A., Hale, N., and Trefethen, L. N. (2014) *Chebfun Guide*, Pafnuty Publications, Oxford.

[3] Ekström, S.-E., and Vassalos, P. A. (2022) *Matrix-Less method to approximate the spectrum and the spectral function of Toeplitz matrices with real eigenvalues*, Numer Algor, **89**, 701–720.

[4] Ekström, S.-E., and Garoni, C. (2019) *A matrix-less and parallel interpolation-extrapolation algorithm for computing the eigenvalues of preconditioned banded symmetric Toeplitz matrices*, Numerical Algorithms, **80**, 819–848.

[5] Good, I.J. (1961) *The colleague matrix, a Chebyshev analogue of the companion matrix*, Quart. J. Math., **12**, 61–68.

[6] Shams Solary, M. (2025) *A review for finding determinants of some band matrices*, Soft Computing, **29**, 3073–3082.

The primary value for matrix Lambert W function

Maryam Shams Solary[1,*]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: shamssolary@pnu.ac.ir or shamssolary@gmail.com

Abstract. In this paper, we aim to find an initial solution for the matrix Lambert W function, defined by $We^W = A$, where $A, W \in \mathbb{H}^{n\times n}$ with $\mathbb{Q} \subseteq \mathbb{H} \subseteq \mathbb{C}$ and $n \in \mathbb{N}$. Our method utilizes only three key formulas within Matlab to obtain this initial solution. The approach primarily leverages power series, along with considering similarities and differences from the derogatory case and canonical forms.

Keywords: Lambert W function, Canonical forms, Companion matrices

AMS Mathematics Subject Classification [2020]: 65F60, 15A15, 15A20

1. Instruction

Finding the roots of a matrix equation or the limit for the roots, or sometimes finding a suitable initial approximation, will be an important step to solve a matrix equation. The most obvious procedure is to take the power series which defines the exponential, which as you surely remember from Calculus is

$$e^x = I + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \ldots + \frac{1}{k!}x^k + \ldots.$$

Recall that the exponential of a matrix can be defined as an infinite sum

$$e^X = I + X + \frac{1}{2}X^2 + \frac{1}{6}X^3 + \ldots + \frac{1}{k!}X^k + \ldots.$$

Then, by direct computation for the Lambert W function $We^W = A$, we have

$$We^W = W + W^2 + \frac{1}{2!}W^3 + \frac{1}{3!}W^4 + \ldots + \frac{1}{k!}W^{k+1} + \ldots = A.$$

$$We^W \simeq W + W^2 + \frac{1}{2!}W^3 + \frac{1}{3!}W^4 + \ldots + \frac{1}{(n-1)!}W^n = A. \tag{1}$$

Then, we can find a bound for $\|W\|$, namely

$$\|W\| \leq \left(\|A\| + \sum_{i=0}^{n-1}\frac{1}{i!}\right)(n-1)!. \tag{2}$$

*Speaker.

2. Main results

Let $\mathbb{H}$ be a field such that $\mathbb{Q} \subseteq \mathbb{H} \subseteq \mathbb{C}$. Here, we collect a summary of several results and some notes of [2, 4] that help us to find three key formulas for solving (1) when A is nonderogatory with entries in $\mathbb{H}$.
In this way, we try to find real roots of

$$P(X) = X + X^2 + \frac{1}{2!}X^3 + \frac{1}{3!}X^4 + \ldots + \frac{1}{(n-1)!}X^n = A, \tag{3}$$

such that $P(\mu_j) = \lambda_j$ and $\mu_j \in \mathbb{H}(\lambda_j)$ for every eigenvalue λ_j of A, and $j = 1, \ldots, n$. Also, the coefficients are determined by the following help vector $P = \left[1,\ 1,\ \frac{1}{2},\ \frac{1}{6},\ \frac{1}{24},\ \ldots,\ \frac{1}{(n-1)!}\right]$.
In fact, this will be the first step of our process for finding the matrix W in (1) or the matrix X in (3). For the second step of our process $P_A(\lambda) = g_1(\lambda)^{d_1}.g_2(\lambda)^{d_2}.\ldots.g_r(\lambda)^{d_r}$ with each $g_j(\lambda)$ a monic irreducible polynomial over $\mathbb{H}$, and suppose that the polynomials $g_j(\lambda)$ are coprime and $T \in \mathbb{H}^{n\times n}$ such that A can be written as
(4)

$$A = T^{-1}(C_1 \oplus C_2 \oplus \ldots \oplus C_r)T,\ \ C_j = I_{d_j} \otimes C_{g_j} + N \otimes I_{k_j} = \begin{bmatrix} C_{g_j} & I_{k_j} & 0 & \cdots & 0 \\ 0 & C_{g_j} & I_{k_j} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & C_{g_j} & I_{k_j} \\ 0 & \cdots & & 0 & C_{g_j} \end{bmatrix}.$$

Here, C_j will be called the companion-Jordan form of A and the generalized Jordan matrix [6], and N is the $d_j \times d_j$ upper triangular matrix with zeroes everywhere except for ones in the $(j, j+1)$ entries. Then, we try to find matrix X, such that

$$X = T^{-1}(X_1 \oplus X_2 \oplus \ldots \oplus X_r)T, \tag{5}$$

with X_j commuting with C_j. This forces us to have X_j the form of a block upper triangular Toeplitz matrix.

Theorem 2.1. *Let X be only a 2×2 block upper triangular matrix, and for a polynomial $P(\lambda)$, the form of $P(X)$ are given by*

$$X = \begin{bmatrix} W & Z \\ 0 & Y \end{bmatrix},\ \ P(X) = \begin{bmatrix} P(W) & P(Z) \\ 0 & P(Y) \end{bmatrix},$$

for some matrix $\tilde{Z}$ which depends on W, Y, Z and the polynomial $P(\lambda)$ in an intricate manner. Introducing the notation $\tilde{Z} = \Delta P(W, Y)(Z)$ it turns out that this is equal to the finite difference operator, i.e., $P(X_1) - P(X_2) = \Delta P(X_1, X_2)(X_1 - X_2)$:

$$A = \begin{bmatrix} C_g & I_k & 0 & \cdots & 0 \\ 0 & C_g & I_k & \ddots & \ddots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & C_g & I_k \\ 0 & \cdots & & 0 & C_g \end{bmatrix},\ \ X = \begin{bmatrix} X_1 & X_2 & \cdots & \cdots & X_d \\ 0 & X_1 & X_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & X_1 & X_2 \\ 0 & \cdots & & 0 & X_1 \end{bmatrix}.$$

Here, $g(\lambda) = g_0 + g_1\lambda + \ldots + g_{k-1}\lambda^{k-1} + \lambda^k$ is a monic and irreducible polynomial of degree k over $\mathbb{H}$ that help us to solve (1), such that $n = kd$ and X is the solution of the matrix

equation in question. Observe that, $P(X_1) = C_g$ *and*

$$P\left(\begin{bmatrix} X_1 & X_2 \\ 0 & X_1 \end{bmatrix}\right) = \begin{bmatrix} C_g & I_k \\ 0 & C_g \end{bmatrix} = \begin{bmatrix} P(X_1) & \Delta P(X_1, X_1)(X_2) \\ 0 & P(X_1) \end{bmatrix}.$$

So, X_2 *may be determined from the linear matrix equation*

$$\Delta P(X_1, X_1)(X_2) = \sum_{m=1}^{n} \frac{1}{(m-1)!} \sum_{j=0}^{m-1} X_1^{m-1-j} X_2 X_1^j = I_k,$$

we can replace the above equation with the following equation in terms of Kronecker product and by the mapping $vec : \mathbb{H}^{m\times n} \to \mathbb{H}^{mn}$*:*

$$\left(\sum_{m=1}^{n} \frac{1}{(m-1)!} \sum_{i=0}^{m-1} (X_1^T)^i \otimes X_1^{m-1-i}\right) vec(X_2) = vec(I_k). \tag{6}$$

Next, to find $j = 3,\ 4,\ \ldots,\ d$ *we consider the* $jd \times jd$ *block in the upper right hand corner, and the equation*

$$P\left(\begin{bmatrix} X_1 & X_2 & \cdots & \cdots & X_j \\ 0 & X_1 & \ddots & \ddots & X_{j-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & X_1 & X_2 \\ 0 & \cdots & & 0 & X_1 \end{bmatrix}\right) = \begin{bmatrix} C_g & I_k & 0 & \cdots & 0 \\ 0 & C_g & I_k & \ddots & \ddots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & C_g & I_k \\ 0 & \cdots & & 0 & C_g \end{bmatrix},$$

that is uniquely solvable for each j*, and so* $X_2, \ldots, X_d$ *depend uniquely on* X_1*.*

The output of our process is shown with the output lambertwmatrix function [3]. We should say diagonalization approach that computes $A = Mdiag(\lambda_1, \ldots, \lambda_n)M^{-1}$ with forms $W_k(A) = Mdiag(W_k(\lambda_1), ..., W_k(\lambda_n))M^{-1}$.

(7)

$$E1 := \rho(\hat{W}, A) = \frac{\|\hat{W}exp(\hat{W}) - A\|_F}{\|\hat{W}exp(\hat{W})\|_F + \|A\|_F},\ E2 := \rho(\hat{W}, A) = \frac{\|\hat{W}expm(\hat{W}) - A\|_F}{\|\hat{W}expm(\hat{W})\|_F + \|A\|_F},$$

as pointed out by Deadman and Higham [1,5], by rounding the result to double precision.

Example 2.2. Consider the companion matrix $A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -4 & 0 & 0 & -4 & 0 & 0 \end{bmatrix}$, with characteristic polynomial $P_A(\lambda) = \lambda^6 + 4\lambda^3 + 4 = (\lambda^3 + 2)^2$. We try to find an initial solution for the primary matrix Lambert W function W, such that $Wexp(W) = A$.
As we have mentioned in the process of Section 2, we must find the matrix T such that $T^{-1}AT = \begin{bmatrix} A_1 & I_3 \\ 0 & A_1 \end{bmatrix}$, where $A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 0 & 0 \end{bmatrix}$. We have $g(\lambda) = \lambda^3 + 2$, i.e., $d = 2$

and $k = 3$. Then $\lambda_0 = -\sqrt[3]{2}$ is one of the roots of $P_A(\lambda)$ with associated eigenvector

$$v_6(-\sqrt[3]{2}) = \begin{bmatrix} 1 \\ -\sqrt[3]{2} \\ \sqrt[3]{4} \\ -2 \\ 2\sqrt[3]{2} \\ -2\sqrt[3]{4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -\sqrt[3]{2} \\ \sqrt[3]{4} \end{bmatrix} = Y v_3(-\sqrt[3]{2}),$$

$$S_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \end{bmatrix}, \quad T = [Y \ S_1 Y] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 0 & 0 & 3 \\ 0 & -2 & 0 & -8 & 0 & 0 \\ 0 & 0 & -2 & 0 & -10 & 0 \end{bmatrix}.$$

Now, we try to build matrix M such that $\mu_1 v_3(\lambda_1) = M v_3(\lambda_1)$, we find here μ_1 by roots function in Matlab for (3), such that μ_1 be a real value:

$$\begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{bmatrix} \begin{bmatrix} 1 \\ -\sqrt[3]{2} \\ \sqrt[3]{4} \end{bmatrix} = \mu \begin{bmatrix} 1 \\ -\sqrt[3]{2} \\ \sqrt[3]{4} \end{bmatrix}.$$

Then $M = X_1 = \begin{bmatrix} 8.624450300829338 & 5.051530214107873 & -3.803102713946090 \\ 7.606205427892180 & 8.624450300829338 & 5.051530214107873 \\ -10.103060428215747 & 7.606205427892180 & 8.624450300829338 \end{bmatrix}$,

and from (6), we write

$$X_2 = \begin{bmatrix} 0.1705414154610 & -0.1353580294458 & 0.1074336290667 \\ -0.2148672581335 & 0.1705414154610 & -0.1353580294458 \\ 0.2707160588917 & -0.2148672581335 & 0.1705414154610 \end{bmatrix}, X = T^{-1} \begin{bmatrix} X_1 & X_2 \\ 0 & X_1 \end{bmatrix} T.$$

We obtain after some calculation

$$X = \begin{bmatrix} 8.696072720207200 & 6.678526480323479 & -6.293385180094861 & 0.035811209688931 & 0.813498133107803 & -1.245141233074385 \\ 4.980564932297542 & 8.696072720207198 & 6.678526480323479 & -1.312820247797319 & 0.035811209688930 & 0.813498133107803 \\ -3.253992532431213 & 4.980564932297542 & 8.696072720207198 & 3.424533947892267 & -1.312820247797319 & 0.035811209688930 \\ -0.143244838755724 & -3.253992532431213 & 4.980564932297541 & 8.552827881451476 & 3.424533947892267 & -1.312820247797320 \\ 5.251280991189278 & -0.143244838755719 & -3.253992532431211 & 10.231845923486819 & 8.552827881451478 & 3.424533947892268 \\ -13.698135791569069 & 5.251280991189279 & -0.143244838755725 & -16.952128324000281 & 10.231845923486819 & 8.552827881451474 \end{bmatrix},$$

References

[1] Corless, R. M. and Jeffrey, D. J. (2015) *The Lambert W function*, Princeton Companion to Applied Mathematics, Princeton University Press, Princeton, NJ.

[2] Drazin, M. P. (2007) *Exact rational solutions of the matrix equation A = p(X) by linearization*, Linear Algebra Appl., **426**, 502–515.

[3] Fasi, M., Higham, N. J. and Iannazzo, B. (2015) *An algorithm for the matrix Lambert w function*, SIAM J. Matrix Anal. Appl., **36**, 669–685.

[4] Groenewald, G. J., Janse van Rensburg, D. B., Ran, A. C. M., Theron, F. and van Straaten, M. (2023) *Solutions of the matrix equation $p(X) = A$, with polynomial function $p(\lambda)$ over field extensions of $\mathbb{Q}$*, Linear Algebra Appl., **665**, 107–138.

[5] Higham, N. J. (2008) *Functions of Matrices: Theory and Computation*, SIAM Publications, Philadelphia.

[6] Horn, R. A. and Johnson, C. R. (2013) *Matrix Analysis, 2nd ed.*, Cambridge University Press.

A two-dimensional Boubaker polynomial expansion scheme for the numerical solution of the nonlinear Schrödinger equation in (2+1) dimensions

Farshad Mehdifar[1,*]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: farshad.mehdifar@pnu.ac.ir

Abstract. This paper introduces an approximation method for the analytical solution of the (2+1)-dimensional nonlinear Schrödinger equation in a stationary pulsed regime. The approach is based on the two-dimensional Boubaker polynomials expansion scheme. Exploiting the analytical solutions in three-dimensional space that correspond to stationary states akin to standing waves invariant over time, we derive the probability density function to eliminate the pulsing component. The discussion includes error analysis of the approximate solutions at zero-time plane for various separation constants ($\omega > 0$). These results highlight the optimal accuracy of time-independent solutions, especially when the separation constant approaches zero. The findings are presented clearly and comprehensively, ensuring their validity and reliability.

Keywords: two-dimensions Boubaker polynomials expansion scheme method, (N+1)-dimensions nonlinear Schrödinger equation, Mathieu functions, quantum mechanics.

AMS Mathematics Subject Classification [2020]: 65M80, 35Q40, 35D30

1. Introduction

Partial differential equations (PDEs) are fundamental in various scientific disciplines—including mathematics, computer science, physics—and engineering fields such as mechanics, electrical engineering, and materials science. Due to the complex and nonlinear nature of many real-world phenomena, analytical solutions are often unattainable or difficult to obtain, prompting the development of numerical methods. Examples of analytical approaches include the Differential Transform Method (DTM) [5] and the B-Spline finite element method [3]. In this study, we employ the two-dimensional Boubaker polynomials expansion scheme (2DBPES) [1] to approximate solutions to the (2+1)-dimensional Schrödinger equation (SchE) [2].

The (2+1)-dimensional nonlinear Schrödinger equation (NLSE) considered here is given by:

$$i\,\frac{\partial\psi}{\partial t}+\frac{1}{2}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}\right)-\tau\,\psi-\sigma|\psi|^2\,\psi=0\,, \tag{1}$$

where $\sigma = 1$ and subject to the initial and boundary conditions (***IBC***):

$$(IC):\quad \psi(x,y,t)\Big|_{t=0}=f(x,y)\quad,\quad (x,y)\in\mathbb{R}^+\times\mathbb{R}^+\;;\;t\in[0,T]\,, \tag{2}$$

*Speaker.

$$(3) \qquad (BC): \quad \frac{\partial^2\psi(x,y,t)}{\partial x^2}\Big|_{x=0} = 0 \quad , \quad \frac{\partial^2\psi(x,y,t)}{\partial y^2}\Big|_{y=0} = 0\,,$$

with $\psi(x,y,t)$ representing the complex-valued wave function in a Hilbert space, where T is the characteristic time of the system, and $|\psi|^2$ corresponds to the probability density. The function $f(x,y)$ is specified and complex-valued. The potential term $\tau(x,y)$ is periodic on the domain $[0,2\pi]\times[0,2\pi]$, representing a spatially varying potential energy component. This study focuses on obtaining analytical solutions for the (2+1)-dimensional NLSE in a stationary, pulsed regime.

2. Two-dimensional Separated Boubaker Polynomials

Definition 2.1 ([1]). The first monomial definition of the Boubaker polynomials is derived from physical considerations, originally used to solve heat equations in pyrolysis models:

$$B_0(x) = 1\,, B_n(x) = \sum_{i=0}^{\left[\frac{n}{2}\right]} (-1)^i \binom{n-i}{i} \frac{n-4i}{n-i}\, x^{n-2i}, \quad n \geq 1\,,$$

where $\left[\frac{n}{2}\right] = \frac{2\,n+((-1)^n-1)}{4}$ (The symbol: [.] designates the floor function). The two-dimensional Boubaker polynomials are defined as: $\mathbb{B}_{0,0}(x,y) = 1\,, \mathbb{B}_{n,m}(x,y) = B_n(x)\,B_m(y)\,.$

Each $4q$-order Boubaker polynomial possesses exactly $2q-1$ positive roots within the interval $(0,2)$. The minimal positive roots, denoted by α_n, underpin the polynomials' orthogonal properties and facilitate their application in various physical problems. For a complex function $f(x,y)$ defined on $[-a,a]\times[-a,a]$, the $4n$-order BPES expansion is given by:

$$(4) \qquad F(x,y) \approx \frac{1}{4N_0M_0} \sum_{q_1=1}^{N_0}\sum_{q_2=1}^{M_0} \zeta_{q_1}^{q_2}\, \mathbb{B}_{4q_1,4q_2}\left(x\frac{\alpha_{q_1}}{a}, y\frac{\alpha_{q_2}}{a}\right),$$

where $\zeta_{q_1}^{q_2}$ are complex coefficients, and N_0, M_0 are fixed integers. Approximate solutions are obtained by minimizing the residual of the linear operator applied to $F(x,y)$ against a target value.

3. Solution Derivation for the (2+1)-Dimensional NLSE

3.1. Application of 2DBPES. The resolution methodology is based on the two-dimensional **BPES**, utilizing the following pulsed representation:

$$(5) \qquad \psi(x,y,t) = \frac{1}{4N_0M_0}\left(\sum_{k=1}^{N_0}\sum_{l=1}^{M_0} \lambda_k^l \,.\, \mathbb{B}_{4k\,,\,4l}\,(x\times r_k\,,\, y\times r_l)\right).e^{-i\omega t} = p(x,y)e^{-i\omega t}\,;\ \omega>0\,,$$

where B_{4k} and B_{4l} denote orthogonal 4k-order and 4l-order Boubaker polynomials, respectively; r_k and r_l are their corresponding minimal positive roots; N_0 and M_0 are predetermined integers; and ω represents the stationary pulsation related to separation constants. The coefficients λ_k^l are real parameters to be determined. Since $\psi(x,y,t) = p(x,y)\big(\cos(\omega\,t) - i\,\sin(\omega\,t)\big)$ is a complex-value function, We clearly have; $|\psi(x,y,t)| = |p(x,y)|$, therefore, Eq.(1) is hence written:

(6)
$$\left[(\omega-\tau)p(x,y) + \frac{1}{8N_0M_0}\left(\sum_{k=1}^{N_0}\sum_{l=1}^{M_0}\lambda_k^l\big(r_k^2\frac{\partial^2 B_{4k}\,(x\times r_k)}{\partial x^2} + r_l^2\frac{\partial^2 B_{4l}\,(y\times r_l)}{\partial y^2}\big)\right)\right]e^{-i\omega t} = (p(x,y))^3 e^{-i\omega t}\,.$$

where, $\tau \equiv \tau(x,y)$ is a known real function of two variables. Simplifying, the main equation becomes:

$$(7) \quad \frac{(\omega-\tau)}{4N_0M_0}\left(\sum_{k=1}^{N_0}\sum_{l=1}^{M_0}\lambda_k^l.\mathbb{B}_{4k\,,\,4l}\,(x\times r_k\,,\, y\times r_l)\right) + \frac{1}{8N_0M_0}\left(\sum_{k=1}^{N_0}\sum_{l=1}^{M_0}\lambda_k^l\Big(r_k^2\frac{\partial^2 B_{4k}\,(x\times r_k)}{\partial x^2}B_{4l}\,(y\times r_l)\right.$$
$$\left. + r_l^2\frac{\partial^2 B_{4l}\,(y\times r_l)}{\partial y^2}B_{4k}\,(x\times r_k)\Big)\right) = (p(x,y))^3\,.$$

3.2. Initial Conditions and Coefficient Determination. Using a standard form and the initial condition (2), which states $\psi(x,y,0) = f(x,y) \equiv p(x,y)$, the problem reduces to solving:

(8)
$$\begin{cases} \displaystyle\sum_{k=1}^{N_0}\sum_{l=1}^{M_0} \lambda_k^l \,.\, \Theta_k^l(x,y) = \left(\, f(x,y)\,\right)^3 \,, \\ \Theta_k^l(x,y) = \dfrac{1}{4N_0M_0}\Big((\omega-\tau)\mathbb{B}_{4k\,,\,4l}\,(x\times r_k, y\times r_l) + \dfrac{1}{2}\Big(r_k^2\dfrac{\partial^2 B_{4k}\,(x\times r_k)}{\partial x^2}B_{4l}\,(y\times r_l) + r_l^2\dfrac{\partial^2 B_{4l}\,(y\times r_l)}{\partial y^2}B_{4k}\,(x\times r_k)\Big)\Big)\,. \end{cases}$$

The following step consists of evaluating the coefficients $\lambda'_{k,l}\big|_{k=1,\dots,N_0}^{l=1,\dots,M_0}$ that verify:

$\left(\, f(x,y)\,\right)^3 = \frac{1}{4N_0M_0}\left(\sum_{k=1}^{N_0}\sum_{l=1}^{M_0}\lambda'_{k,l}.\mathbb{B}_{4k\,,\,4l}\,(x\times r_k\,,\,y\times r_l)\right)$. This operation leads to the weak solution defined by the system Eq.(9)

(9)
$$\begin{cases} \displaystyle\sum_{k=1}^{N_0}\sum_{l=1}^{M_0} \lambda_k^l \,.\, I_{k,l} = \sum_{k=1}^{N_0}\sum_{l=1}^{M_0} \lambda'_{k,l} \,.\, J_{k,l}\,, \\ I_{k,l} = 4\displaystyle\int_0^1\int_0^1 \Theta_k^l(x,y)\,\mathrm{d}x\mathrm{d}y\,, \qquad J_{k,l} = \dfrac{1}{N_0M_0}\int_0^1\int_0^1 \mathbb{B}_{4k\,,\,4l}\,(x\times r_k\,,\,y\times r_l)\,\mathrm{d}x\mathrm{d}y\,. \end{cases}$$

The values of $I_{k,l}\big|_{k=1,\dots,N_0}^{l=1,\dots,M_0}$ and $J_{k,l}\big|_{k=1,\dots,N_0}^{l=1,\dots,M_0}$ are calculated using Eq.(9) along with the two-dimensions 4q-Boubaker polynomial properties [1] of the itself and second derivative value at boundary conditions (3), in the reduced real domain $[0,r_k]\times[0,r_l]$. Finally, the coefficients $\lambda_{k,l}^{sol.}\big|_{k=1,\dots,N_0}^{l=1,\dots,M_0}$ are deduced by identification:

(10)
$$\lambda_{k,l}^{sol.} = \lambda'_{k,l}\,.\frac{J_{k,l}}{I_{k,l}}\,, \quad k=1,\dots,N_0\,, \quad l=1,\dots,M_0\,.$$

A final solution $\psi_{BPES}^{sol}(x,y,t)$ is consequently:

(11)
$$\psi_{BPES}^{sol}(x,y,t) = \frac{1}{4N_0M_0}\left(\sum_{k=1}^{N_0}\sum_{l=1}^{M_0}\lambda_{k,l}^{sol.}\,.\,\mathbb{B}_{4k\,,\,4l}\,(x\times r_k\,,\,y\times r_l)\right).e^{-i\omega t}\,.$$

3.3. Exact Solution for the Initial Boundary Condition. The exact solution can be obtained by combining Eqs.(1) and (5) as follows:

(12)
$$\frac{\partial^2 p(x,y)}{\partial x^2} + \frac{\partial^2 p(x,y)}{\partial y^2} + 2\left(\big(\omega-\tau(x,y)\big) - \big(p(x,y)\big)^2\right)p(x,y) = 0\,.$$

Let $\tau(x,y) = -\sin^2(\omega\,x)\,\sin^2(\omega\,y)$, Considering that this nonlinear partial differential equation (PDE) is elliptic and in canonical form, its general solution will primarily involve trigonometric functions. To handle the nonlinear term $\big(p(x,y)\big)^2$, we employ a double Fourier series expansion:

(13)
$$\tilde{p}(x,y) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{4}{\pi^2}\left(\int_0^\pi\int_0^\pi \big(p(x,y)\big)^2\cos(mx)\cos(ny)\mathrm{d}y\mathrm{d}x\right)\cos(mx)\cos(ny)$$

Substituting this expansion into the PDE transforms it into a linear PDE:

(14)
$$\frac{\partial^2 p(x,y)}{\partial x^2} + \frac{\partial^2 p(x,y)}{\partial y^2} + 2\Big[\big(\omega-\tau(x,y)\big) - \tilde{p}(x,y)\Big]\,p(x,y) = 0\,.$$

Applying the initial condition $p(x,y) \equiv f(x,y) = \sin(\omega x)\sin(\omega y)$ to Eqs. (12)–(13) and incorporating it into the modified PDE (14), where $\tilde{p}(x,y) = \frac{1}{4}\cos(2\omega x)\cos(2\omega y)$, allows us to solve the problem via the separation of variables method. The solution involves Mathieu functions [4]:

(15)
$$\begin{aligned} p^{ex.}(x,y) \;=\; & \mathrm{e}^{c_3}\left[\mathsf{MathieuC}\left(\frac{c_1}{\omega^2},\frac{1}{4\omega^2},\omega\,x\right) + c_2\,\mathsf{MathieuS}\left(\frac{c_1}{\omega^2},\frac{1}{4\omega^2},\omega\,x\right)\right]\times \\ & \left[c_2\,\mathsf{MathieuC}\left(\frac{4\,\omega+1}{2\omega^2}-c_1,\frac{1}{4\omega^2},\omega\,y\right) + c_3\,\mathsf{MathieuS}\left(\frac{4\,\omega+1}{2\omega^2}-c_1,\frac{1}{4\omega^2},\omega\,y\right)\right]\,, \end{aligned}$$

where MathieuC and MathieuS denote the even and odd Mathieu functions, respectively, and c_1, c_2, c_3 are arbitrary constants.

Finally, the complete exact solution for the wave function $\psi^{ex.}(x, y, t)$ is given by

$$\psi^{ex.}(x, y, t) = p^{ex.}(x, y)\, e^{-i\omega t}\,. \tag{16}$$

4. Discussion and Numerical Results

For the function $f(x, y) = \sin(\omega\, x)\sin(\omega\, y)$ in Equation (2), and by setting Equation (8) and solving Equation (9) for its parameters, the 2D boundary value problem with exact solutions (2DBPES) corresponding to the real part of the approximate solution $\langle \psi^{sol.}_{BPES}(x, y, t)\rangle$ at $t = 0$ has been estimated. These results are compared with the exact solution $\langle \psi^{ex.}(x, y, t)\rangle$ in Table 1. The parameter values for both the 2DBPES-related and exact solutions are $N_0 = M_0 = 5$, with

TABLE 1. Numerical results of the present method in uniform nodes for different ω.

Nodes $(0.2\,\mathrm{i}, 0.2\,\mathrm{j}, 0)$		$\omega = \frac{1}{8}$		$\omega = \frac{1}{2}$		$\omega = 1$	
		$Re\,\langle \psi^{sol.}_{BPES}\rangle$	$\mathbf{E}^{\psi}_{\mathbf{5},\mathbf{5}}$	$Re\,\langle \psi^{sol.}_{BPES}\rangle$	$\mathbf{E}^{\psi}_{\mathbf{5},\mathbf{5}}$	$Re\,\langle \psi^{sol.}_{BPES}\rangle$	$\mathbf{E}^{\psi}_{\mathbf{5},\mathbf{5}}$
i= j= 0	$(0, 0, 0)$	6.06×10^{-7}	6.06×10^{-7}	4.87×10^{-4}	4.87×10^{-4}	3.29×10^{-3}	3.29×10^{-3}
i= j= 1	$(0.2, 0.2, 0)$	5.97×10^{-7}	6.24×10^{-4}	4.79×10^{-4}	9.49×10^{-3}	3.23×10^{-3}	3.62×10^{-2}
i= j= 2	$(0.4, 0.4, 0)$	5.70×10^{-7}	2.50×10^{-3}	4.55×10^{-4}	3.90×10^{-2}	3.07×10^{-3}	1.49×10^{-1}
i= j= 3	$(0.6, 0.6, 0)$	5.24×10^{-7}	5.61×10^{-3}	4.16×10^{-4}	8.69×10^{-2}	2.80×10^{-3}	3.16×10^{-1}
i= j= 4	$(0.8, 0.8, 0)$	4.62×10^{-7}	9.97×10^{-3}	3.63×10^{-4}	1.51×10^{-1}	2.44×10^{-3}	5.12×10^{-1}
i= j= 5	$(1.0, 1.0, 0)$	3.84×10^{-7}	1.55×10^{-2}	2.98×10^{-4}	2.30×10^{-1}	1.99×10^{-3}	7.06×10^{-1}

ω taking values of $\frac{1}{8}$, $\frac{1}{2}$, and 1. Additionally, $\mathbf{E}^{\psi}_{\mathbf{N_0},\mathbf{M_0}} = \left\|\mathrm{Re}\,\langle \psi^{\mathrm{ex}}\rangle - \mathrm{Re}\,\langle \psi^{\mathrm{sol}}_{BPES}\rangle\right\|$ represents an estimate of the absolute error of the exact solution $\mathrm{Re}\,\langle \psi^{\mathrm{ex}}\rangle$. The findings suggest that as the value of ω approaches zero, the convergence of the approximate solution to the exact solution improves, and the calculation error decreases notably.

The results demonstrate a strong agreement between the exact and proposed solutions along the $t = 0$ plane (see Table 1). The consistency for higher values of t is expected, since the t-dependent terms—though eliminated at the beginning of the solution process (Subsection 3.1)—remain unchanged for both the exact and approximate solutions. This observation indicates that analytical solutions can be derived even when exact solutions are challenging or infeasible.

The analytical approach employed in this method provides a solid foundation for the **2DBPES** protocol, especially when exact solutions are difficult to obtain. The computational implementation and tabulation were performed using **Maple 24** software, with calculations carried out to five significant digits.

References

1. M. Agida, A. S. Kumar, *A Boubaker polynomials expansion scheme solution to random Love equation in the case of a rational kernel*, El. J. Theor. Phys., **7** (2010), 319–326.
2. U. Al Khawaja, L. Al Sakkaf, *Handbook of Exact Solutions to the Nonlinear Schrödinger Equations*, IOP Publishing, Bristol, UK, 2020.
3. A. Chakraborty, B.V. Rathish Kumar, *Non Uniform Weighted Extended B-Spline Finite Element Analysis of Non Linear Elliptic Partial Differential Equations*, Differential Equations and Dynamical Systems, **30**(3) (2022), 485–497.
4. N. W. McLachlan, *Theory and Applications of Mathieu Functions*, Oxford University Press, 1964.
5. Asal B. Saleh, Abdulghafor M. Al-Rozbayani, *Analytical Solution of the Klein-Gordon coupled Equations via the Differential Transform Method*, Advances in Nonlinear Variational Inequalities, **27**(3) (2024), 393–403.

A Hybrid Fast Numerical Method for the Lane-Emden Differential Equation Using GHFs

Seyed Amjad Samareh Hashemi[1,*], Rasoul Hatamian[1]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: a_hashemi@pnu.ac.ir
Email: hatamianr@pnu.ac.ir

Abstract. This paper introduces a novel hybrid numerical method for solving the Lane-Emden equation, leveraging Generalized Hat Functions (GHFs) of different degrees to achieve exceptional computational efficiency. By using linear GHFs for converting the equation into a block-structured nonlinear system solved via forward substitution, followed by cubic GHFs for refined approximation, the approach delivers up to $1000x$ speedup over direct cubic methods while maintaining L_∞ errors around 10^{-4}. Adaptable to various nonlinear differential equations, it ensures consistent accuracy across interval lengths and extends seamlessly to fractional-order cases with minimal adjustments.

Keywords: Generalized Hat Functions, Lane-Emden Equation, Operational Matrix of Integration, Numerical Differential Equations.

AMS Mathematics Subject Classification [2020]: 65T60, 65N06, 35R11

1. Introduction

The Lane-Emden equation, vital in astrophysics for modeling stellar structures [1], is expressed as:

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right) = -\theta^n$$

where ξ is a dimensionless radius, θ relates to density, and n is the polytropic index. Exact solutions exist for $n = 0, 1, 5$; otherwise, numerical methods are employed, such as the Taylor wavelet method [4], B-spline collocation [3], and machine learning with Rational Chebyshev polynomials [2].

Generalized Hat functions (GHFs) are continuous functions with shape of hats (in case of degree one). Suppose that the interval $[0, T]$, for $T > 0$, is divided into N subintervals $[ih, (i+1)h]$, $i = 0, 1, \cdots, N-1$ of equal lengths h where $h = \frac{T}{N}$.

Definition 1.1. **First-degree (Linear) GHFs** were the building blocks upon which the broader concept of GHFs was constructed. The GHFs of degree one are defined as: [5]

$$\psi_0(t) = \begin{cases} \frac{h-t}{h}, & 0 \le t < h, \\ 0, & \text{otherwise.} \end{cases}$$

*Speaker.

$$\psi_i(t) = \begin{cases} \frac{t-(i-1)h}{h}, & (i-1)h \le t < ih, \\ \frac{(i+1)h-t}{h}, & ih \le t < (i+1)h, \\ 0, & \text{otherwise.} \end{cases} \quad i = 1, 2, \cdots, N-1$$

$$\psi_N(t) = \begin{cases} \frac{t-(T-h)}{h}, & T-h \le t \le T, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 1.2. **(Third degree (Cubic) GHFs)** Suppose N is a positive integer of multiple three ($N = 3m,\ m \in \mathbb{N}$) and $h = \frac{T}{N}$. A set of adjusted GHFs of degree 3 is defined on $[0, T]$ as:

$$\psi_0(t) = \begin{cases} \frac{-1}{6h^3}(t-h)(t-2h)(t-3h), & 0 \le t < 3h, \\ 0, & \text{otherwise.} \end{cases}$$

If $i = 3k+1,\ k = 0, 1, \cdots, m-1$:

$$\psi_i(t) = \begin{cases} \frac{1}{2h^3}(t-(i-1)h)(t-(i+1)h)(t-(i+2)h), & (i-1)h \le t < (i+2)h, \\ 0, & \text{otherwise.} \end{cases}$$

If $i = 3k+2,\ k = 0, 1, \cdots, m-1$:

$$\psi_i(t) = \begin{cases} \frac{-1}{2h^3}(t-(i-2)h)(t-(i-1)h)(t-(i+1)h), & (i-2)h \le t < (i+1)h, \\ 0, & \text{otherwise.} \end{cases}$$

If $i = 3k,\ k = 0, 1, \cdots, m-1$:

$$\psi_i(t) = \begin{cases} \frac{1}{6h^3}(t-(i-3)h)(t-(i-2)h)(t-(i-1)h), & (i-3)h \le t < ih, \\ \frac{-1}{6h^3}(t-(i+1)h)(t-(i+2)h)(t-(i+3)h), & ih \le t < (i+3)h, \\ 0, & \text{otherwise.} \end{cases}$$

$$\psi_N(t) = \begin{cases} \frac{1}{6h^3}(t-(T-h))(t-(T-2h))(t-(T-3h)), & T-3h \le t \le T, \\ 0, & \text{otherwise.} \end{cases}$$

GHFs properties and their operational matrix of integration. A function $y(t) \in L^2[0, T]$ is approximated as:

$$y(t) \approx \sum_{k=0}^{N} y(kh)\psi_k(t) = C_y^\top \mathbf{\Psi}(t),$$

where $\mathbf{\Psi(t)} = [\psi_0(t), \ldots, \psi_N(t)]^\top$. The operational matrix of integration $\mathbf{P}$ approximates:

$$\int_0^t \mathbf{\Psi(s)}\, ds \approx \mathbf{P\Psi(t)},$$

with $\mathbf{P}$ being lower triangular for linear GHFs. For $N \in \mathbb{N}$, $\mathbf{P}$ is obtained as follows:

$$\mathbf{P} = \frac{h}{2}\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 & \ldots & 2 & 2 \\ 0 & 0 & 1 & 2 & 2 & \ldots & 2 & 2 \\ 0 & 0 & 0 & 1 & 2 & \ldots & 2 & 2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}.$$

Unlike recent approaches such as eighth-order boundary value solvers or physics-informed neural networks, our method uniquely combines linear and cubic GHFs in a

two-stage process, yielding significant computational savings (e.g., 0.42s vs. 452s CPU time) without substantial accuracy loss, as validated through examples.

2. Main Results

The Lane-Emden equation is:

$$y''(x) + \frac{\alpha}{t}y'(x) + y^n(x) = 0, \qquad 0 \le x \le T \tag{1}$$

$$y(0) = \beta, \quad y'(0) = \eta, \tag{2}$$

The method approximates $y''(x) \approx C_{y''}^{\top}\Psi(x)$ using linear GHFs, then:

$$y'(x) \approx C_{y''}^{\top}\mathbf{P}\mathbf{\Psi}(x) + \eta, \quad y(x) \approx C_{y''}^{\top}\mathbf{P}^2\mathbf{\Psi}(x) + \eta x + \beta.$$

Expanding $x \approx E^{\top}\mathbf{\Psi}(x)$ and $1 \approx J^{\top}\mathbf{\Psi}(x)$, and substituting, yields a nonlinear system:

$$(E \odot C_{y''}) + \alpha(\mathbf{P}^{\top}C_{y''} + \eta J) + E \odot ((\mathbf{P}^{\top})^2 C_{y''} + \eta E + \beta J)_{\odot}^{n} = 0, \tag{3}$$

solved via forward substitution. The solution $y(x)$ is then constructed using cubic GHFs for improved accuracy. An example with $n = 5$, $T = 4$, $\alpha = 2$, $\beta = 0$, $\eta = 1$ shows the method's efficiency over a direct cubic GHF approach.

In addition to the theoretical analysis of error, solved examples also support the fact that this method yields a more accurate solution than the conventional method (using degree 1 GHFs). Besides, the method has the advantage of less computational complexity and subsequently less time-consuming relative to the method using 3rd-degree GHFs at the initial step. A test case using the Lane-Emden equation with $n = 5$ highlights the method's primary advantage: speed.

Example 2.1. In this problem we consider the Lane-Emden equation (1) and initial conditions (2) with $n = 5$, $T = 4$, $\alpha = 2$, $\beta = 0$ and $\eta = 1$ whose analytical solution is $y(x) = \frac{1}{\sqrt{1+\frac{1}{3}x^2}}$. The proposed method and the direct method with degree 3 GHFs are applied with $N = 60$ and the results are summarized in figure 1 and table 1 to comparison.

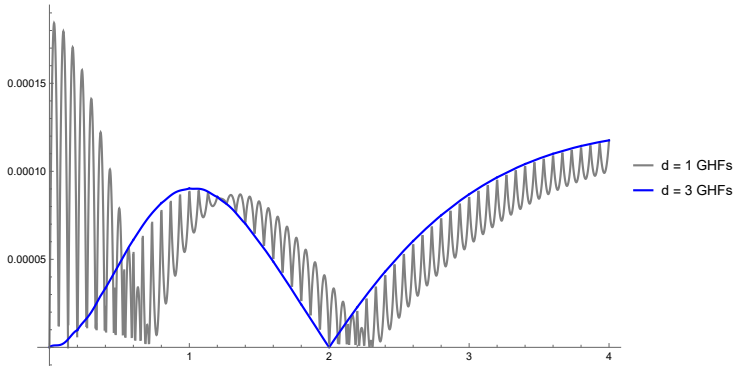


Figure 1. Error of produced solutions using GHFs of degree 1 and the proposed method with $N = 60$

As shown in Table 1, the proposed method is orders of magnitude faster than the direct method using 3rd-degree GHFs, trading a small amount of accuracy for a significant gain in computational speed.

TABLE 1. CPU Time and L_∞ Error comparison for $N = 60$

	Proposed Method	Direct Method (3rd-degree GHFs)
CPU time (s.)	**0.421875**	451.719
L_∞ error	1.17461×10^{-4}	8.0×10^{-8}

The main advantages of the algorithm are:

- **Speed:** It is exceptionally fast due to the use of 1st-degree GHFs for the system-solving step and the forward substitution approach.
- **Efficiency:** The method maintains a low runtime even for large values of N, making it highly scalable.
- **Adaptability:** The framework can be readily extended to solve fractional-order Lane-Emden equations and other types of nonlinear differential equations with minimal changes.
- **Simplicity:** The algorithm's structure is straightforward, allowing for easy implementation in various programming environments.

Conclusion

The proposed numerical method offers a powerful and pragmatic approach for solving the Lane-Emden equation. By combining the strengths of different degrees of Generalized Hat Functions, it achieves an outstanding balance of speed, efficiency, and accuracy. Its simple structure and adaptability make it a versatile tool for researchers and practitioners tackling a wide array of nonlinear differential equations in science and engineering. The primary unique contribution lies in the hybrid GHF framework, which not only accelerates solving by orders of magnitude compared to uniform-degree methods but also enhances adaptability for fractional and nonlinear DEs in astrophysics and engineering. As future research, we can work on improving the proposed method to reduce errors while maintaining speed.

References

[1] Singh, R., Singh, G. and Singh, M. *Numerical Algorithm for Solution of the System of Emden–Fowler Type Equations*, Int. J. Appl. Comput. Math, *7*(136), (2021), https://doi.org/10.1007/s40819-021-01066-7.

[2] Jilong He, Zhoushun Zheng and Changfa Du, *A New Constructing Rational Functions Method For Solving Lane-Emden Type Equations*, Neural Processing Letters, *55*, (2023), 1889–1918, https://doi.org/10.1007/s11063-022-10968-6.

[3] Mohammad Prawesh Alam, Tahera Begum and Arshad Khan, *A high-order numerical algorithm for solving Lane–Emden equations with various types of boundary conditions*, Computational and Applied Mathematics, *40*(204), (2021), https://doi.org/10.1007/s40314-021-01591-7.

[4] Sevin Gümgüm, *Taylor wavelet solution of linear and nonlinear Lane-Emden equations*, Appl. Numer. Math., (2020), https://doi.org/10.1016/j.apnum.2020.07.019.

[5] Manoj P. Tripathi, Vipul K. Baranwal, Ram K. Pandey and Om P. Singh, *A new numerical algorithm to solve fractional differential equations based on operational matrix of generalized hat functions*, Communications in Nonlinear Science and Numerical Simulation, *18*(6), (2013), Pages 1327–1340, https://doi.org/10.1016/j.cnsns.2012.10.014.

Numerical solution of pantograph differential equations using the double Sinc method

Mansour Shiralizadeh[1,*], Kaivan Mohammadi[2]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m.shiralizadeh@pnu.ac.ir
[2]Department of Mathematics, University of Kurdistan, Sanandaj, Iran.
Email: ki.mohammadi@uok.ac.ir

Abstract. In this work, the double Sinc method has been used to numerically solve the pantograph differential equations. One of the characteristics of the presented method for solving these equations is that the problem's solution is transformed into the solution of a system of linear algebraic equations. To ensure the validity of the proposed method and to compare it with the Sinc method, a numerical example is presented.

Keywords: pantograph differential equations, double Sinc

AMS Mathematics Subject Classification [2020]: 97N40, 41A30, 65L60

1. Introduction

We consider the following class of pantograph-type differential equations:

$$z'(t) + az(t) + d(t)z(ct) = q(t), \quad t \in (0,1), \tag{1}$$

with the boundary condition

$$z(0) = A, \tag{2}$$

where a is a constant, $0 < c < 1$ and $d(t)$ and $q(t)$ are continuous functions. Some problems in science and engineering are modeled as ordinary pantograph differential equations. Some of the methods that have been used by researchers to solve pantograph differential equations are: the Legendre Ritz-least squares method [4], various wavelet techniques [3], the homotopy perturbation method [1] and the Laplace transform method [2]. In 1997, double exponential transformations (DE) was applied by Sugihara in the Sinc method and for n points reached the convergence rate $O\left(e^{\left(\frac{-kn}{\log n}\right)}\right)$, which was much faster than the convergence rate achieved by typical mapping [8].

2. Sinc function preliminaries

To use in the next sections, we recall the following results, taken from [5, 7].
The Sinc basis functions are defined on the infinite domain $D_s = \{w = u + iv : |v| < d \leq \pi/2\}$, but the domain of the equation (1) is finite, so to convert the interval $(0,1)$ to the interval $(-\infty, \infty)$,

*Speaker.

we use suitable conformal mappings $w = \phi(y)$. The Sinc basis functions on the finite interval are as follows

$$S_l(y) = Sinc\left(\frac{\phi(y) - lh}{h}\right), \tag{3}$$

We apply the transformation mapping $\phi(y)$ that is suitable for the double exponential transformation as follows:

$$\phi(y) = ln\left[\frac{1}{\pi}ln\left(\frac{y}{1-y}\right) + \sqrt{1 + \left(\frac{1}{\pi}ln\left(\frac{y}{1-y}\right)\right)^2}\right],$$

Which ϕ maps the region

$$D_E = \left\{y \in \mathbb{C} : \left|arg\left[\frac{1}{\pi}ln\left(\frac{y}{1-y}\right) + \sqrt{1 + \left(\frac{1}{\pi}ln\left(\frac{y}{1-y}\right)\right)^2}\right]\right| < d\right\},$$

into D_s . The Sinc grid points in the domain $(0, 1)$ are as follows

$$t_j = \phi_{DE}^{-1}(jh) = \frac{1}{2}\tanh\left(\frac{\pi}{2}\sinh(jh)\right) + \frac{1}{2}, \quad j = 0, \pm1, \pm2, \ldots. \tag{4}$$

The transferred Sinc functions $S_l(t)$ at the points $t_j = \phi^{-1}(jh)$ are cardinal, i.e.

$$[S_l(t)]\Big|_{t=t_j} = \delta_{l,j}^{(0)} = \begin{cases} 1, & l = j, \\ 0, & l \neq j, \end{cases} \tag{5}$$

Also, the derivative of the transferred Sinc functions relative to the transform mapping ϕ is as follows:

$$\frac{d}{d\phi}[S_l(t)]\Big|_{t=t_j} = \frac{1}{h}\delta_{l,j}^{(1)} = \frac{1}{h}\begin{cases} 0, & l = j, \\ \frac{(-1)^{j-l}}{j-l}, & l \neq j, \end{cases} \tag{6}$$

Definition 2.1. The approximation of the function $z(t)$ with Sinc functions on the interval $(0, 1)$ is defined as follows

$$z(t) \simeq z_m(t) = \sum_{l=-m}^{m} z(lh)S_l(t). \tag{7}$$

Definition 2.2. Let $B(D_E)$ be the class of all functions z which are analytic in D_E, and

$$\int_{\phi^{-1}(L+u)} |z(y)dy| \to 0, \quad u \to \pm\infty,$$

where $L = \{iv : |v| < d\}$ and on the boundary of D_E (denoted by ∂D_E)

$$N(z, D_E) \equiv \int_{\partial D_E} |z(y)dy| < \infty.$$

Theorem 2.3. *Let $\phi' z \in B(D_E)$ and there are positive constants α, β, and c_1 such that*

$$|z(t)| \leq c_1 \begin{cases} e^{-\alpha e^{|\phi(t)|}}, & t \in \Gamma_a, \\ e^{-\beta e^{|\phi(t)|}}, & t \in \Gamma_b, \end{cases}$$

where

$$\Gamma_a = \{t \in \Gamma : \phi(t) \in (-\infty, 0)\}, \qquad \Gamma_b = \{t \in \Gamma : \phi(t) \in [0, \infty)\}.$$

Also let the mesh-size h be chosen as

$$h = \frac{\log\left(\frac{\pi md}{\alpha}\right)}{m}, \tag{8}$$

then for all $t \in \Gamma$ we have

$$\left|z(t) - \sum_{l=-m}^{m} z(t_l)Sinc\left(\frac{\phi(t) - lh}{h}\right)\right| \leq d_1 e^{\frac{-\pi md}{\log\left(\frac{\pi md}{\alpha}\right)}}. \tag{9}$$

3. The Sinc method

We define the approximate solution for equations (1)-(2) as follows

$$z_m(t) = r_m(t) + p(t), \tag{10}$$

where

$$r_m(t) = \sum_{l=-m}^{m} b_l S_l(t),$$

and because the boundary condition of (2) is not homogeneous, we have added the polynomial $p(t)$ to the approximate solution, which is a combination of Hermite interpolation at zero and one boundary point.

$$p(t) = A(2t+1)(1-t)^2 + b_{-m-1}t(1-t)^2 + b_{m+1}t^2(3-2t) + b_{m+2}t^2(t-1). \tag{11}$$

Now, the approximate solution $z_m(t)$ satisfies the boundary condition (2), i.e.,

$$z_m(0) = r_m(0) + p(0) = A. \tag{12}$$

By substituting the approximate solution $z_m(t)$ into equation (1) and multiplying both sides of the equation by $\frac{h}{\phi'(t)}$, we obtain

$$\begin{aligned} &h\sum_{l=-m}^{m}\left(\frac{1}{\phi'(t)}\frac{d}{dt}S_l(t) + \frac{a}{\phi'(t)}S_l(t) + \frac{d(t)}{\phi'(t)}S_l(ct)\right)b_l \\ &+\frac{h}{\phi'(t)}\left(p'(t) + ap(t) + d(t)p(ct)\right) = \frac{h}{\phi'(t)}q(t). \end{aligned} \tag{13}$$

by discretizing (13) at Sinc grid points $t_j, \quad j=-m,\cdots,m$ and using (5)-(6), the following system is obtained:

$$\begin{aligned} &\sum_{l=-m}^{m}\left(\delta_{lj}^{(1)} + h\left(\frac{a}{\phi'(t_j)}\right)\delta_{lj}^{(0)} + h\left(\frac{d(t_j)}{\phi'(t_j)}\right)S_l(ct_j)\right)b_l \\ &+\frac{h}{\phi'(t_j)}\left(p'(t_j) + ap(t_j) + d(t_j)p(ct_j)\right) = \frac{h}{\phi'(t_j)}q(t_j), \quad j=-m,\cdots,m. \end{aligned} \tag{14}$$

Since the number of equations is three less than the number of unknowns, we set $b_{-m} = b_{m-1} = b_m = 0$. Now, by solving System (14), the unknowns b_ls and $z_m(t)$ are obtained.

4. Error analysis

THEOREM 4.1. *Let z and z_m be the exact and approximate solutions of Equation (1), respectively. then we have the following bound*

$$\left|z(t) - z_m(t)\right| \le d_2 m^{\frac{1}{2}} e^{\frac{-\pi m d}{\log\left(\frac{\pi m d}{\alpha}\right)}}, \qquad t \in \Gamma$$

PROOF. The proof is similar to Theorem 5.4 in [6]. □

5. Numerical Example

In this section, we present an example to better demonstrate the efficiency of the presented method and to show in practice that the convergence rate of the method is exponential. In the example, we select $\alpha = \pi$, $d = \frac{\pi}{8}$, and $h = \frac{ln\left(\frac{2\pi dm}{\alpha}\right)}{m}$.
We obtain the maximum absolute error in the following form

$$E = \max\left\{\left|z_m\left(\frac{j}{100}\right) - z\left(\frac{j}{100}\right)\right|,\ j=0,1,\cdots,100\right\}$$

Example 1. Consider the following equation

$$z'(t) + z(t) - \frac{1}{2}z\left(\frac{t}{2}\right) = -\frac{1}{2}e^{\frac{-t}{2}},$$

$$z(0) = 1,$$

the exact solution is $z(t) = e^{-t}$. The maximum absolute errors of the proposed method and the Sinc method are presented in Table 1, which shows that the convergence order of the method is exponential. Also, the CPU execution time is presented in Table 2. These two tables show that the convergence speed of the proposed method is faster than the Sinc method, but the CPU execution time for the proposed method is longer than the Sinc method.

TABLE 1. The maximum absolute errors for Example 1

m	5	10	20	30	40	50	60
Sinc method	1.8(−5)	2.0(−6)	2.6(−8)	7.9(−10)	3.7(−11)	2.3(−12)	1.8(−13)
E	4.1(−6)	1.6(−7)	3.7(−12)	7.9(−17)	2.8(−20)	3.8(−24)	5.5(−28)

TABLE 2. CPU times for Example 1

m	5	10	20	30	40	50	60
Sinc method	2.2	5.4	12.8	21.1	31.5	46.8	58.8
double Sinc	13.8	30.9	74.3	128.9	213.0	319.6	463.2

6. Conclusion

In this work, we successfully employed the double Sinc method to approximate the solution for pantograph differential equations. The double Sinc method demonstrated its high robustness and accuracy by achieving exponential convergence rates. A numerical example was presented to demonstrate the efficiency of the proposed method for solving these equations, which shows the superiority of the proposed method compared to the Sinc method.

References

1. Albidah. A.B, Kanaan. N.E, Ebaid. A, Al-Jeaid. H.K, *Exact and Numerical Anal- ysis of the Pantograph Delay Differential Equation via the Homotopy Perturbation Method,* Mathematics. 11(2023) 944.
2. Alrebdi. R, Al-Jeaid. H.K, *Accurate Solution for the Pantograph Delay Differential Equation via Laplace Transform,* Mathematics. 11(2023) 2031.
3. Jaiswal. J.P, Yadav. K, *A comparative study of numerical solution of pantograph equations using various wavelets techniques,* App. Eng. Math. 11(2021) 772-788.
4. Khorshidi. M.N, Firoozjaee. M.A, *Legendre Ritz-Least squares method for the nu- merical solution of delay differential equations of the multi-pantograph type,* Math. Comput. Sci. 5(1)(2024) 20-29.
5. Lund. J, Bowers. K, *Sinc methods for quadrature and differential equations,* SIAM Philadelphia, 1992.
6. Mohammadi. K, Alipanah. A, and Ghasemi. M , *A non-classical Sinc-collocation method for the solution of singular boundary value problems arising in physiology,* Int. J. Comput. Math. 99, (2022) 1941-1967.
7. Nabati. M, jalalvand. M, *Solution of troeschs problem through double exponential Sinc-Galerkin method,* Comput. Differ. Equ. 5, (2017) 141-157.
8. Sugihara. M, *Optimality of the double exponential formula-functional analysis approach,* Numer. Math. 75, (1997) 379-395.

Numerical solution of singular Volterra-Fredholm integro-differential equations using the Sinc method

Mansour Shiralizadeh[1,*], Kaivan Mohammadi[2]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m.shiralizadeh@pnu.ac.ir
[2]Department of Mathematics, University of Kurdistan, Sanandaj, Iran.
Email: ki.mohammadi@uok.ac.ir

Abstract. In this work, the Sinc method has been used to numerically solve the singular Volterra-Fredholm integro-differential equations. One of the characteristics of the method presented for solving these equations is that their solution is transformed into the solution of a system of linear algebraic equations. The ability of the Sinc approximation to overcome the singularity makes it an efficient method. A numerical example is provided to ensure the validity of the proposed method.

Keywords: singular, integro-differential equation, Sinc approximation
AMS Mathematics Subject Classification [2020]: 45J05, 45E05, 41A30

1. Introduction

We consider the following singular Volterra-Fredholm integro-differential equation:

$$y'' + \frac{1}{x}y' + \frac{1}{x^2}y + \gamma_1 \int_0^x k(x,t)y(t)dt + \gamma_2 \int_0^1 G(x,t)y(t)dt = f(x), \quad x,t \in (0,1), \tag{1}$$

with the boundary conditions

$$y(0) = A, \qquad y(1) = B \tag{2}$$

where γ_1, γ_2, A and B are given constants and $f(x)$, $k(x,t)$, $G(x,t)$ are sufficiently smooth functions. Many problems in science and engineering, such as those in financial mathematics, fluid mechanics, electromagnetic theory, and plasma physics, are modeled by integro-differential equations [1, 2]. Our method is more efficient in controlling the singularity of the equation than many other methods and works well in dealing with this type of problem. This is because the singularity of the equation occurs at the end point of the interval, and the Sinc functions do not need to be smooth near the boundaries; continuity is sufficient. In addition, most of the Sinc grid points are gathered near the endpoints of the interval, which help us to control the singularity as well [4].

2. Sinc function preliminaries

For use in the next sections, we recall the following results, taken from [3, 5].
The Sinc basis functions are defined on the infinite domain $D_s = \{w = u + iv : |v| < d \leq \pi/2\}$,

*Speaker.

but the domain of Equation (1) is finite, so to convert the interval $(0,1)$ to the interval $(-\infty,\infty)$, we use a suitable conformal mapping $w=\phi(z)$.

$$\phi(z)=ln\left(\frac{z}{1-z}\right),$$

which maps the eye-shaped region

$$D_E=\left\{z=x+iy:\left|arg\left(\frac{z}{1-z}\right)\right|<d\leq\pi/2\right\},$$

onto D_s.
The Sinc basis functions on the finite interval are defined as follows

$$S_l(z)=S(l,h)o\phi(z)=Sinc\left(\frac{\phi(z)-lh}{h}\right),$$

for $z\in D_E$.
The Sinc grid points in $(0,1)$ are defined as follows

$$x_j=\phi^{-1}(jh)=\frac{e^{jh}}{1+e^{jh}},\quad j=0,\pm1,\pm2,\cdots.$$

Definition 2.1. The approximation of the function $y(x)$ with Sinc functions on the interval $(0,1)$ is defined as follows

$$y(x)\simeq y_m(x)=\sum_{l=-m}^{m}y(lh)S_l(x).$$

Definition 2.2. Let $B(D_E)$ be the class of all functions y which are analytic in D_E, and

$$\int_{\phi^{-1}(L+u)}|y(z)dy|\to 0,\quad u\to\pm\infty,$$

where $L=\{iv:|v|<d\}$ and on the boundary of D_E (denoted by ∂D_E)

$$N(y,D_E)\equiv\int_{\partial D_E}|y(z)dz|<\infty.$$

Theorem 2.3. *Let $\phi' z\in B(D_E)$ and there are positive constants α, β, and c such that*

$$|y(x)|\leq c_1\begin{cases}e^{-\alpha e^{|\phi(x)|}}, & x\in\Gamma_a,\\ e^{-\beta e^{|\phi(x)|}}, & x\in\Gamma_b,\end{cases}$$

where

$$\Gamma_a=\{x\in\Gamma:\phi(x)\in(-\infty,0)\},\qquad \Gamma_b=\{x\in\Gamma:\phi(x)\in[0,\infty)\}. \tag{3}$$

Then for all $x\in\Gamma$ we have,

$$\left|y(x)-\sum_{l=-m}^{m}y(x_l)S_l(x)\right|\leq d_1m^{1/2}e^{-\sqrt{\pi d\alpha m}}.$$

Where the mesh-size h is chosen as:

$$h=\sqrt{\frac{\pi d}{\alpha m}}. \tag{4}$$

Theorem 2.4. *Let $y\in B(D_E)$ and there are positive constants α, β and c such that*

$$\left|\frac{y(x)}{\phi'(x)}\right|\leq c\begin{cases}e^{-\alpha|\phi(x)|}, & x\in\Gamma_a,\\ e^{-\beta|\phi(x)|}, & x\in\Gamma_b,\end{cases}$$

where Γ_a and Γ_b are defined in (3). If h is selected as (4), then for all $x\in\Gamma$,

$$\left|\int_0^1 y(s)ds-h\sum_{k=-m}^{m}\frac{y(x_k)}{\phi'(x_k)}\right|\leq d_2e^{-\sqrt{\pi d\alpha m}}, \tag{5}$$

$$\left| \int_0^{x_l} y(s)ds - h \sum_{k=-m}^{m} \delta_{lk}^{(-1)} \frac{y(x_k)}{\phi'(x_k)} \right| \leq d_3 e^{-\sqrt{\pi d \alpha m}}, \tag{6}$$

where

$$\delta_{lk}^{(-1)} = \frac{1}{2} + \int_0^{l-k} \frac{\sin(\pi x)}{\pi x} dx. \tag{7}$$

LEMMA 2.5. *Let Φ be a one-to-one conformal mapping from the simply connected domain D_E onto D_s, then*

$$\delta_{l,k}^{(0)} = [S(l,h)o\phi(x)]\Big|_{x=x_k} = \begin{cases} 1, & l=k, \\ 0, & l \neq k, \end{cases} \tag{8}$$

$$\delta_{l,k}^{(1)} = h\frac{d}{d\phi}[S(l,h)o\phi(x)]\Big|_{x=x_k} = \begin{cases} 0, & l=k, \\ \frac{(-1)^{k-l}}{k-l}, & l \neq k, \end{cases} \tag{9}$$

$$\delta_{l,k}^{(2)} = h^2\frac{d^2}{d\phi^2}[S(l,h)o\phi(x)]\Big|_{x=x_k} = \begin{cases} \frac{-\pi^2}{3}, & l=k, \\ \frac{-2(-1)^{k-l}}{(k-l)^2}, & l \neq k. \end{cases} \tag{10}$$

3. The Sinc method

We define the approximate solution for equations (1) and (2) as follows

$$y_m(x) = u_m(x) + p(x), \tag{11}$$

where

$$u_m(x) = \sum_{l=-m}^{m} b_l S_l(x),$$

and because the boundary conditions in (2) are not homogeneous, we have added the polynomial $p(t)$ to the approximate solution, which is a combination of Hermite interpolation at zero and one boundary points.

$$p(x) = A(2x+1)(1-x)^2 + Bx(1-x)^2 + b_{-m-1}x^2(3-2x) + b_{m+1}x^2(x-1). \tag{12}$$

Now the approximate solution $y_m(x)$ satisfies the boundary condition (2), that is:

$$y_m(0) = u_m(0) + p(0) = A, \qquad y_m(1) = u_m(1) + p(1) = B \tag{13}$$

We substitute the approximate solution $y_m(x)$ into Equation (1) and multiply the both sides of the equation by $\frac{h^2}{\phi'^2}$. By discretizing the result at the Sinc grid points x_j, $j = -m, \cdots, m$ and using (5)-(10) the following system is obtained:

$$\begin{aligned} &\sum_{l=-m}^{m} \left[\delta_{lj}^{(2)} + h\left(\frac{\phi''(x_j))}{\phi'^2(x_j)} + \frac{1}{x_j\phi'(x_j)} \right) \delta_{lj}^{(1)} + h^2 \left(\frac{1}{x_j^2\phi'^2} \right) \delta_{lj}^{(0)} \right] b_l \\ &+ \frac{h^2}{\phi'^2(x_j)} \left(p''(x_j) + \frac{1}{x_j}p'(x_j) + \frac{1}{x_j^2}p(x_j) \right) + \gamma_1 \frac{h^3}{\phi'^2(x_j)} \sum_{g=-m}^{m} \delta_{gj}^{(-1)} \frac{1}{\phi'(s_g)} k(x_j, s_g) \sum_{l=-m}^{m} b_l \delta_{lg}^{(0)} \\ &+ \gamma_1 \frac{h^3}{\phi'^2(x_j)} \sum_{g=-m}^{m} \delta_{gj}^{(-1)} \frac{1}{\phi'(s_g)} G(x_j, s_g) p(s_g) + \gamma_2 \frac{h^3}{\phi'^2(x_j)} \sum_{k=-m}^{m} \frac{1}{\phi'(s_k)} G(x_j, s_k) \sum_{l=-m}^{m} b_l \delta_{lk}^{(0)} \\ &+ \gamma_2 \frac{h^3}{\phi'^2(x_j)} \sum_{k=-m}^{m} \frac{1}{\phi'(s_k)} G(x_j, s_k) p(s_k) = \frac{h^2}{\phi'^2(x_j)} f(x_j), \; j = -m, \cdots, m \end{aligned} \tag{14}$$

Because the number of equations is less than the number of unknowns, we set $b_{-m} = b_m = 0$. By solving the system (14) with iterative methods, the unknowns b_ls and $y_m(x)$ are obtained.

4. Numerical Example

In this section, we present an example to better demonstrate the efficiency of the presented method and to show in practice that the convergence rate of the method is exponential. In the example, we select $\alpha = 1$, $d = \frac{\pi}{2}$, and $h = \sqrt{\frac{\pi d}{\alpha m}}$.
We obtain maximum absolute error in the following form

$$E = \max\left\{\left|y_m\left(\frac{j}{100}\right) - y\left(\frac{j}{100}\right)\right|,\ j = 0, 1, \cdots, 100\right\}.$$

If we denote by E_1 and E_2 the maximum absolute errors obtained with $m = m_1$ and m_2, respectively, then the practical order of convergence can be calculated by the formula $ln\left(\frac{E_1}{E_2}\right)/ln\left(\frac{m_2}{m_1}\right)$.

Example 1. Consider the following equation

$$y'' + \frac{1}{x}y' + \frac{1}{x^2}y - \int_0^x sin(x-t)y(t)dt - \int_0^1 cos(x-t)y(t)dt = f(x),$$
$$y(0) = 0, \qquad y(1) = 0$$

where $f(x) = \frac{2}{x} - 7 - x + x^2 - \sin(x) + 3\cos(x) + \cos(x-1) + 2\sin(x-1)$ and the exact solution is $y(x) = x - x^2$.

The maximum absolute errors and practical convergence orders of the proposed method are presented in Tables 1 and 2. The results of these tables show that the convergence order of the proposed method is exponential.

TABLE 1. The maximum absolute errors for Example 1

m	5	10	20	30	40
E	2.1(−5)	9.0(−8)	3.2(−10)	9.2(−12)	1.5(−13)

TABLE 2. Orders of convergence for Example 1

m	5	10	20	30	40
Order	–	7.9	8.1	8.8	14.3

5. Conclusion

In this work, we successfully employed the Sinc method to approximate the solution for singular Volterra-Fredholm integro-differential equations. A numerical example was presented to demonstrate the efficiency of the proposed method for solving these equations, the results of which showed that the convergence rate of the method is exponential.

References

1. J. Chen, M. He, Y. Huang, *A fast multiscale Galerkin method for solving second order linear Fredholm integro-differential equation with Dirichlet boundary conditions,* J. Comput. Appl. Math, 364(2020) 112352.
2. M.E. Durmaz, G.M. Amiraliyev, M. Kudu, *Numerical solution of a singularly perturbed Fredholm integro differential equation with Robin boundary condition,* Turk. J. Math. 46 (2022) 207-224.
3. J. Lund, K. Bowers, *Sinc methods for quadrature and differential equations,* SIAM Philadelphia, 1992.
4. K. Mohammadi, A. Alipanah, and M. Ghasemi, *A non-classical Sinc-collocation method for the solution of singular boundary value problems arising in physiology,* Int. J. Comput. Math. 99, (2022) 1941-1967.
5. M. Nabati, M. Jalalvand, *Solution of troeschs problem through double exponential Sinc-Galerkin method,* Comput. Differ. Equ. 5, (2017) 141-157.

Solutions of modified Fornberg–Whitham equation via the $\tan(\phi/2)$-expansion method

Mehdi Fazli Aghdaei[1,*] and Jalil Manafian[2,3]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m.fazliaghdaei@gmail.com
[2]Department of Applied Mathematics, Faculty of Mathematical Science, University of Tabriz, Tabriz, Iran.
Email: manafeian2@gmail.com
[3]Natural Sciences Faculty, Lankaran State University, 50, H. Aslanov str., Lankaran, Azerbaijan.

Abstract. The aim of present paper is to obtain the analytical solution of modified Fornberg–Whitham equation by using $\tan(\phi/2)$-expansion method. These methods are used to construct solitary and soliton solutions of nonlinear evolution equation. The method is straightforward and concise, and its applications are promising. This method is developed for searching exact travelling wave solutions of nonlinear partial differential equations. Also, it is shown that the method, with the help of symbolic computation, provides a straightforward and powerful mathematical tool for solving nonlinear evolution equations.

Keywords: $\tan(\phi/2)$-expansion method, modified Fornberg–Whitham equation; Solitary and soliton solutions

AMS Mathematics Subject Classification [2020]: 58E11, 53B30, 53C50

1. Introduction

The Fornberg-Whitham equation (FWE) was first proposed for studying the qualitative behavior of wave breaking [1]. The FWE can be written as

$$u_t - u_{xxt} + u_x = uu_{xxx} - uu_x + 3u_xu_{xx}, \tag{1}$$

which has a type of travelling wave solution called a kink-like wave solution and anti kink-like wave solutions. Eq. (1) was used to study the qualitative behaviour of wave-breaking [1]. It is a nonlinear dispersive wave equation. Since Eq. (1) was derived, little attention has been paid to studying it. Fornberg and Whitham obtained a peaked solution of the form $u(x,t) = A\exp\left(-\frac{1}{2}|x-\frac{4}{3}t|\right)$, where A is an arbitrary constant. Conservation laws and exact solutions of the Whitham-type equations have studied by [2]. Authors of [3] have studied the wave breaking behavior by homotopy perturbation and variational iteration methods. It is significantly important in mathematical physics to search for exact

*Speaker.

solutions of nonlinear differential equations. The modified Fornberg–Whitham equation can be written as

$$u_t - u_{xxt} + u_x = uu_{xxx} - u^2u_x + 3u_xu_{xx}. \tag{2}$$

The investigation of the travelling wave solutions plays an important role in nonlinear sciences. With the development of solitary theory, many powerful methods were established for obtaining the exact solutions of NLPDEs, such as the sine-cosine method [4], homotopy perturbation method [5], homotopy analysis method [6], the $tan(\phi/2)$-expansion method [7], the Exp-function method [8], the G'/G-expansion method [9], Hirota's bilinear operator [10] and finite element method [11]. The aim of this study is to establish the exact soliton solutions to the (1+1) modified Fornberg-Whitham equation by $\tan(\phi/2)$-expansion approach.

2. The $\tan(\phi/2)$-expansion technique

The objective of this Section is to outline the use of the $\tan(\phi/2)$-expansion for solving certain nonlinear PDE.

Step 1. We suppose that the given nonlinear fractional partial differential equation for $u(x,t)$ to be in the form

$$\mathcal{N}(u, u_x, u_t, u_{xx}, u_{tt}, ...) = 0, \tag{3}$$

which can be converted to an ODE

$$\mathcal{Q}(u, u', -\mu u', u'', \mu^2 u'', ...) = 0, \tag{4}$$

the transformation $\xi = \mathrm{x} - \mu\mathrm{t}$, is wave variable. Also, μ is constant to be determined later.

Step 2. Suppose the traveling wave solution of Eq. (4) can be expressed as follows:

$$u(\xi) = \sum_{k=0}^{m} A_k \left[p + \tan\left(\frac{\Phi(\xi)}{2}\right)\right]^k + \sum_{k=1}^{m} B_k \left[p + \tan\left(\frac{\Phi(\xi)}{2}\right)\right]^{-k}, \tag{5}$$

where $A_k(0 \le k \le m)$ and $B_k(1 \le k \le m)$ are constants to be determined, such that $A_m \neq 0, B_m \neq 0$ and $\Phi = \Phi(\xi)$ satisfies the following ordinary differential equation:

$$\Phi'(\xi) = \mathrm{a}\sin(\Phi(\xi)) + \mathrm{b}\cos(\Phi(\xi)) + \mathrm{c}. \tag{6}$$

We will consider the following special solutions of equation (6):

Family 1: **When** $K = a^2 + b^2 - c^2 < 0$ **and** $\mathrm{b} - \mathrm{c} \neq 0$, then
$\Phi(\xi) = 2\tan^{-1}\left[\frac{a}{b-c} - \frac{\sqrt{-K}}{b-c}\tan\left(\frac{\sqrt{-K}}{2}(\xi + C)\right)\right]$.

Family 2: **When** $K = a^2 + b^2 - c^2 > 0$ **and** $\mathrm{b} - \mathrm{c} \neq 0$,
then $\Phi(\xi) = 2\tan^{-1}\left[\frac{a}{b-c} + \frac{\sqrt{K}}{b-c}\tanh\left(\frac{\sqrt{K}}{2}(\xi + C)\right)\right]$.

Family 3: **When** $b-c \neq 0$ **and** $\mathrm{a} = 0$, then $\Phi(\xi) = 2\tan^{-1}\left[\sqrt{\frac{b+c}{b-c}}\tanh\left(\frac{\sqrt{b^2-c^2}}{2}(\xi + C)\right)\right]$.

Family 4: **When** $\mathrm{a} = 0$ **and** $\mathrm{c} = 0$, then $\Phi(\xi) = \tan^{-1}\left[\frac{\mathrm{e}^{2\mathrm{b}(\xi+\mathrm{C})}-1}{\mathrm{e}^{2\mathrm{b}(\xi+\mathrm{C})}+1}, \frac{2\mathrm{e}^{\mathrm{b}(\xi+\mathrm{C})}}{\mathrm{e}^{2\mathrm{b}(\xi+\mathrm{C})}+1}\right]$.

Family 5: **When** $\mathrm{b} = 0$ **and** $\mathrm{c} = 0$, then $\Phi(\xi) = \tan^{-1}\left[\frac{2\mathrm{e}^{\mathrm{a}(\xi+\mathrm{C})}}{\mathrm{e}^{2\mathrm{a}(\xi+\mathrm{C})}+1}, \frac{\mathrm{e}^{2\mathrm{a}(\xi+\mathrm{C})}-1}{\mathrm{e}^{2\mathrm{a}(\xi+\mathrm{C})}+1}\right]$.

Family 6: **When** $\mathrm{a}^2 + \mathrm{b}^2 = \mathrm{c}^2$, then $\Phi(\xi) = -2\tan^{-1}\left[\frac{(\mathrm{b}+\mathrm{c})(\mathrm{a}(\xi+\mathrm{C})+2)}{\mathrm{a}^2(\xi+\mathrm{C})}\right]$.

Family 7: **When** $\mathrm{a} = \mathrm{c}$, then $\Phi(\xi) = 2\tan^{-1}\left[\frac{(\mathrm{b}+\mathrm{c})\mathrm{e}^{\mathrm{b}(\xi+\mathrm{C})}+1}{(\mathrm{b}-\mathrm{c})\mathrm{e}^{\mathrm{b}(\xi+\mathrm{C})}-1}\right]$.

Family 8: **When** $\mathrm{b} = \mathrm{c}$ then $\Phi(\xi) = 2\tan^{-1}\left[\frac{\mathrm{a}\mathrm{e}^{\mathrm{a}(\xi+\mathrm{C})-\mathrm{c}}}{\mathrm{a}}\right]$, where $\mathrm{A_k}, \mathrm{B_k}(\mathrm{k} = 1, 2, ..., \mathrm{m}), \mathrm{a}, \mathrm{b}$ and c are constants to be determined later. But, the positive integer m can be determined

by considering the homogeneous balance between the highest order derivatives and nonlinear terms appearing in Eq. (6).
Step 3. Substituting (5) into Eq. (4) with the value of m obtained in Step 2. Collecting the coefficients of functions, then setting each coefficient to zero, we can get a set nonlinear algebra equations.
Step 4. Solving the algebraic equations in Step 3, then we get $A_0, A_1, B_1, ..., A_m, B_m, \mu, p$.

3. Application of $\tan(\phi/2)$-expansion technique

The modified Fornberg–Whitham equation can be written as

$$u_t - u_{xxt} + u_x = uu_{xxx} - u^2u_x + 3u_xu_{xx}. \tag{7}$$

Using the wave variables as follow $\xi = \mathrm{kx} + \mathrm{wt}$, Eq. (7) becomes

$$(m+k)u' - k^2wu''' - k^3uu''' + ku^2u' - 3k^3u'u'' = 0. \tag{8}$$

In order to determine the values of m, we balance the linear term of the highest order u‴ with the highest order nonlinear term u^2u' in Eq. (8), to get

$$M + 3 = 3M + 1, \qquad \Rightarrow M = 1. \tag{9}$$

Then the trail solution is

$$v(\xi) = A_0 + A_1 \tan\left(\frac{\Phi(\xi)}{2}\right) + B_1 \cot\left(\frac{\Phi(\xi)}{2}\right). \tag{10}$$

Substituting (10) and (6) into Eq. (8) and by using the well-known Maple software, we obtain the following sets:
Set I:
(11)

$$b = c,\ c = c,\ k = k,\ a = \tfrac{kc^2}{\alpha},\ w = \tfrac{k(\alpha^4 - 2c^4k^4)}{\alpha^2},\ A_0 = \tfrac{2c^4k^4}{\alpha^2},\ A_1 = ck\alpha,\ B_1 = 0,$$
$$\beta = \left(3c^4k^4 - 66c^8k^8 + 64c^{12}k^{12} - 1 + 3\sqrt{-3c^8k^8(2c^4k^4-1)(128c^8k^8 - 6c^4k^4 + 3)}\right)^{\frac{1}{3}},$$
$$\alpha = \pm\tfrac{1}{\sqrt{3}\beta}\left[\beta(\beta^2 - (1 - 4c^4k^4)\beta + 16c^8k^8 - 2c^4k^4 + 1)\right]^{\frac{1}{2}},\quad u(\xi) = A_0 + A_1\tan\left(\tfrac{\Phi(\xi)}{2}\right),$$

where a, b and c are arbitrary constants. Using the (11) and **Family 8**, we get

$$u_1(\xi) = \frac{2c^4k^4}{\alpha^2} + ck\alpha\left[e^{\frac{kc^2}{\alpha}(\xi + C)} - \frac{\alpha}{kc}\right],\quad \xi = kx + \frac{k(\alpha^4 - 2c^4k^4)}{\alpha^2}t. \tag{12}$$

Set II:
(13)

$$b = c,\quad c = c,\quad k = k,\quad a = \tfrac{kc^2}{\alpha},\quad w = \tfrac{k(\alpha^4 - 2c^4k^4)}{\alpha^2},\quad A_0 = \tfrac{2c^4k^4}{\alpha^2},\quad A_1 = ck\alpha,\quad B_1 = 0$$
$$\beta = \left(3c^4k^4 - 66c^8k^8 + 64c^{12}k^{12} - 1 + 3\sqrt{-3c^8k^8(2c^4k^4-1)(128c^8k^8 - 6c^4k^4 + 3)}\right)^{\frac{1}{3}},$$
$$\alpha = \pm\tfrac{1}{\sqrt{6}\beta}\left[\beta(\beta^2(\sqrt{3}i - 1) + 2(4c^4k^4 - 1)\beta - 16c^8k^8(1 + \sqrt{3}i) + 2(1 + \sqrt{3}i)c^4k^4 - 1 - \sqrt{3}i)\right]^{\frac{1}{2}},$$
$$u(\xi) = A_0 + A_1\tan\left(\tfrac{\Phi(\xi)}{2}\right),$$

where a, b and c are arbitrary constants. Using the (11) and **Family 8**, we receive

$$\mathrm{u_2}(\xi) = \frac{2\mathrm{c}^4\mathrm{k}^4}{\alpha^2} + \mathrm{ck}\alpha\left[\mathrm{e}^{\frac{\mathrm{kc}^2}{\alpha}(\xi + \mathrm{C})} - \frac{\alpha}{\mathrm{kc}}\right],\qquad \xi = \mathrm{kx} + \frac{\mathrm{k}(\alpha^4 - 2\mathrm{c}^4\mathrm{k}^4)}{\alpha^2}\mathrm{t}. \tag{14}$$

Set III:
(15)

$$b = c,\ c = c,\ k = k,\ a = \frac{kc^2}{\alpha},\ w = \frac{k(\alpha^4 - 2c^4k^4)}{\alpha^2},\ A_0 = \frac{2c^4k^4}{\alpha^2},\ A_1 = ck\alpha,\ B_1 = 0,$$
$$\beta = \left(3c^4k^4 - 66c^8k^8 + 64c^{12}k^{12} - 1 + 3\sqrt{-3c^8k^8(2c^4k^4 - 1)(128c^8k^8 - 6c^4k^4 + 3)}\right)^{\frac{1}{3}},$$
$$\alpha = \pm\frac{1}{\beta}\left[-\beta(\beta^2(\sqrt{3}i + 1) - 2(4c^4k^4 - 1)\beta + 16c^8k^8(1 - \sqrt{3}i) + 2(\sqrt{3}i - 1)c^4k^4 + 1 - \sqrt{3}i)\right]^{\frac{1}{2}},$$
$$u(\xi) = A_0 + A_1 \tan\left(\frac{\Phi(\xi)}{2}\right),$$

where a, b and c are arbitrary constants. Via the (11) and **Family 8**, we get

$$u_3(\xi) = \frac{2c^4k^4}{\alpha^2} + ck\alpha\left[e^{\frac{kc^2}{\alpha}(\xi + C)} - \frac{\alpha}{kc}\right], \quad \xi = kx + \frac{k(\alpha^4 - 2c^4k^4)}{\alpha^2}t. \tag{16}$$

4. Conclusion

In this article, we investigated the modified Fornberg–Whitham equation. The $\tan(\phi/2)$-expansion method (TEM) was a useful method for finding travelling wave solutions of nonlinear evolution equations. This method has been successfully applied to obtain some new generalized solitonary solutions to the modified Fornberg–Whitham equation. The TEM was more powerful in searching for exact solutions of nonlinear partial differential equations. It can be concluded that this method is a very powerful and efficient technique in finding exact solutions for wide classes of problems.

References

1. Fornberg, B. Whitham, G.B. (1978) *A numerical and theoretical study of certain nonlinear wave phenomena*, Phil. Trans. R. Soc. Lond., **289**, 373–404.
2. Shirvani, V., Nadjafikhah, M. (2014) *Conservation laws and exact solutions of the Whitham-type equations*, Commun. Nonlinear Sci. Numer. Simulat. **19**, 2212–2219.
3. Dehghan, M., Heris, J.M. (2012) *Study of the wave–breaking's qualitative behavior of the Fornberg–Whitham equation via quasi-numeric approaches*, Int. J. Num. Meth. Heat Fluid Flow **22**, 537–53.
4. Wazwaz, A.M. (2006) *Travelling wave solutions for combined and double combined sine-cosine-Gordon equations by the variable separated ODE method*, Appl. Math. Comput, **177**, 755-760.
5. Dehghan, M., J. Manafian, J. (2009) *The solution of the variable coefficients fourth–order parabolic partial differential equations by homotopy perturbation method*, Z. Naturforsch, *64a*, 1-11.
6. Dehghan, M., Manafian, J., Saadatmandi, A. (2010) *Solving nonlinear fractional partial differential equations using the homotopy analysis method*, Num. Meth. Partial Differential Eq. J. **26**, 448-479.
7. Manafian, J., Lakestani, M. (2016) *Abundant soliton solutions for the Kundu-Eckhaus equation via tan(ϕ/2)-expansion method*, Optik **127**, 5543-5551.
8. Zhang, S. (2008) *Application of Exp-function method to high-dimensional nonlinear evolution equation*, Chaos Solitons Fract. **38**, 270-276.
9. Manafian, J., Lakestani, M. (2015) *Solitary wave and periodic wave solutions for Burgers, Fisher, Huxley and combined forms of these equations by the G'/G-expansion method*, Pramana **130**, 31-52.
10. Shen, X., Manafian, J., M. Jiang, M., Ilhan, O.A., Shafikk, S.S. Zaidi, M. (2022) *Abundant wave solutions for generalized Hietarinta equation with Hirota's bilinear operator*, Modern Phys. Lett. B **36**(10), 2250032.
11. Wu, W., Manafian, J., Ali, K.K., Karakoc, S.B.G., Taqik, A.H., Mahmoud, M.A. (2022) *Numerical and analytical results of the 1D BBM equation and 2D coupled BBM-system by finite element method*, Int. J. Modern Phys. B **36**(28), 2250201.

Analytical treatment of the Black-Scholes equation for European option pricing by Saul'yev finite difference scheme

Mehdi Fazli Aghdaei[1,*] and Jalil Manafian[2,3]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m.fazliaghdaei@gmail.com
[2]Department of Applied Mathematics, Faculty of Mathematical Science, University of Tabriz, Tabriz, Iran.
Email: manafeian2@gmail.com
[3]Natural Sciences Faculty, Lankaran State University, 50, H. Aslanov str., Lankaran, Azerbaijan.

Abstract. The analytical solution of the Black-Scholes equation can lead to the attainment of the price of an option in an idealized fiscal market. However, this is not practically beneficial enough. This happens due to the constricting assumptions based on which the Black-Scholes model is derived. In the real financial market, one can question the constant nature of the coefficients of the Black-Scholes equation. In this paper, the solution of the Black-Scholes equation with constant parameters is reviewed.

Keywords: Linear and nonlinear Black-Scholes equations; Barles' and Soner's model; Saul'yev scheme

AMS Mathematics Subject Classification [2020]: 58E11, 53B30, 53C50

1. Introduction

The Black-Scholes model (BSM) [1] indeed was accorded as some sort of victory for mathematical modeling in finance in a way that it has been relied on as an inherent tool in options trading as well as financial derivatives. As stated earlier in [2], the interest in pricing financial derivatives, including pricing choices is likely to be resulted from the fact that the minimization of losses develops out of fluctuations in prices regarding the underlying assets. We recall that a Brownian motion is in parallel with a process, whose increments are independent stationary normal random variables [4]. Since the stock price cannot be negative, Samuelson [9] offered the exploitation of this process for the illustration of the return of the stock price. Authors of [7] discovered solutions for the inhomogeneous Black-Scholes equations with time dependent coefficients. They also studied min-max estimates, gradient estimates, monotonicity and convexity of the solutions with respect to the stock price variable. Farnoosh et al. [5] explored the problem of discrete double barrier option pricing considering the time dependent models. In these models, the parameters risk free rate, dividend and volatility have been deduced to be deterministic functions of

*Speaker.

time. They proposed a numerical method and calculated the Greeks of contract. Barls and Soner [3] provided a model based on the assumption that an exponential utility function can characterize the investor's preferences. After utilizing the exponential utility function, they proved the implementation of the theory of stochastic optimal control, when V is the unique viscosity solution of the Black-Scholes equation

$$\frac{\partial u}{\partial t}+\frac{\sigma^2}{2}S^2\frac{\partial^2 u}{\partial S^2}+rS\frac{\partial u}{\partial S}-ru=0, \tag{1}$$

with modified volatility function

$$\sigma^2=\sigma_0^2\left(1+\Psi\left[\exp(r(T-t)a^2S^2\frac{\partial^2 V}{\partial S^2})\right]\right), \tag{2}$$

and with the following non-differentiable terminal and time-dependent boundary conditions

$$V(S,T)=f(S),\qquad V(0,t)=0,\qquad \lim_{S\to\infty}V(S,t)=S, \tag{3}$$

where S is the price of the underlying asset, T is the maturity date, r is the risk-free interest rate, σ_0 is the asset volatility, and a is transaction cost. Function $\Psi(A)$ is the solution of the following nonlinear ordinary differential equation

$$\Psi'(A)=\frac{\Psi(A)+1}{2\sqrt{A\Psi(A)}-A},\quad A\neq 0,\quad \Psi(0)=0. \tag{4}$$

In this work, first we present the solution of the Black-Scholes equation with constant parameters following [10].

2. Linear Black-Scholes equation with constant coefficients

The Black-Scholes equation for a European call option with value $u(S,t)$ is

$$\frac{\partial u}{\partial t}+\frac{\sigma_c^2}{2}S^2\frac{\partial^2 u}{\partial S^2}+r_cS\frac{\partial u}{\partial S}-r_cu=0, \tag{5}$$

which σ_c and r_c are constant. Its condition is in backward form, at $t=T$ $u(S,T)=\max(S-E,0)$, which E is the strike price and boundary conditions $u(0,t)=0,\quad u(S,t)\sim S,\quad S\to\infty$. We follow [10] to put $S=Ee^x,\ t=T-\frac{2\tau}{\sigma_c^2},\ u=E\nu(x,\tau)$. This results in the equation

$$\frac{\partial \nu}{\partial \tau}=\frac{\partial^2 \nu}{\partial x^2}+(k-1)\frac{\partial \nu}{\partial x}-k\nu, \tag{6}$$

where $k=\frac{2r_c}{\sigma_c^2}$ and $\nu(x,0)=\max(e^x-1,0)$. By setting $\nu=e^{\alpha x+\beta\tau}U(x,\tau)$, which α and β should be found, one gets

$$\beta U+\frac{\partial U}{\partial \tau}=\alpha^2U+2\alpha\frac{\partial U}{\partial x}+\frac{\partial^2 U}{\partial x^2}+(k-1)\Big(\alpha U+\frac{\partial U}{\partial x}\Big)-kU. \tag{7}$$

By considering $\beta=\alpha^2+(k-1)\alpha-k$ and $2\alpha+(k-1)=0$, we will have an equation without U and $\frac{\partial U}{\partial x}$. Therefore, α and β are $\alpha=\frac{1-k}{2},\quad \beta=\frac{-(k+1)^2}{4}$. Then

$$\nu=e^{\frac{1-k}{2}x-\frac{(k+1)^2}{4}\tau}U(x,\tau), \tag{8}$$

and

$$\frac{\partial U}{\partial \tau}=\frac{\partial^2 U}{\partial x^2}, \tag{9}$$

with

$$U(x,0) = U_0 = \max(e^{\frac{(k+1)x}{2}} - e^{\frac{(k-1)x}{2}}, 0). \tag{10}$$

The solution of the diffusion equation (9) with initial condition (10) is

$$U(x,\tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} U_0(s) e^{-\frac{(x-s)^2}{4\tau}} ds. \tag{11}$$

By change of variable $-\frac{x-s}{\sqrt{2\tau}} = X$

$$U(x,\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U_0(X\sqrt{2\tau} + x) e^{-\frac{X^2}{2}} dX = I_1 - I_2, \tag{12}$$

which

$$I_1 = \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{(k+1)(x+X\sqrt{2\tau})}{2}} e^{-\frac{X^2}{2}} dX, \tag{13}$$

and

$$I_2 = \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{(k-1)(x+X\sqrt{2\tau})}{2}} e^{-\frac{X^2}{2}} dX. \tag{14}$$

The cumulative distribution function for the normal distribution is

$$N(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{z^2}{2}} dz. \tag{15}$$

We consider

$$d_1 = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(k+1)\sqrt{2\tau}, \quad d_2 = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(k-1)\sqrt{2\tau}, \tag{16}$$

then

$$I_1 = e^{\frac{(k+1)x}{2} + \frac{(k+1)^2\tau}{4}} N(d_1), \quad I_2 = e^{\frac{(k-1)x}{2} + \frac{(k-1)^2\tau}{4}} N(d_2). \tag{17}$$

Variables of (??) can be written as

$$x = \ln(\frac{S}{E}), \quad \tau = \frac{\sigma_c^2(T-t)}{2}, \quad u = E\nu(x,\tau). \tag{18}$$

Therefore by (12), (17) and (18) we obtain the solution of (5) as

$$u(S,t) = SN(d_1) - Ee^{-r_c(T-t)} N(d_2), \tag{19}$$

which

$$d_1 = \frac{\ln(\frac{S}{E}) + (r_c + \frac{\sigma_c^2}{2})(T-t)}{\sigma_c\sqrt{T-t}}, \quad d_2 = \frac{\ln(\frac{S}{E}) + (r_c - \frac{\sigma_c^2}{2})(T-t)}{\sigma_c\sqrt{T-t}}. \tag{20}$$

We compute the Black-Scholes price for several volatilities in $t = 0$. As we can see in Figure 1 (left), the value of the call option increases as σ_0 increases. Moreover, the Black-Scholes price is computed for several interest rates and maturity times. In Figures 1 (middle) and (right) we can see that the value of the call option increases as r and T increase.

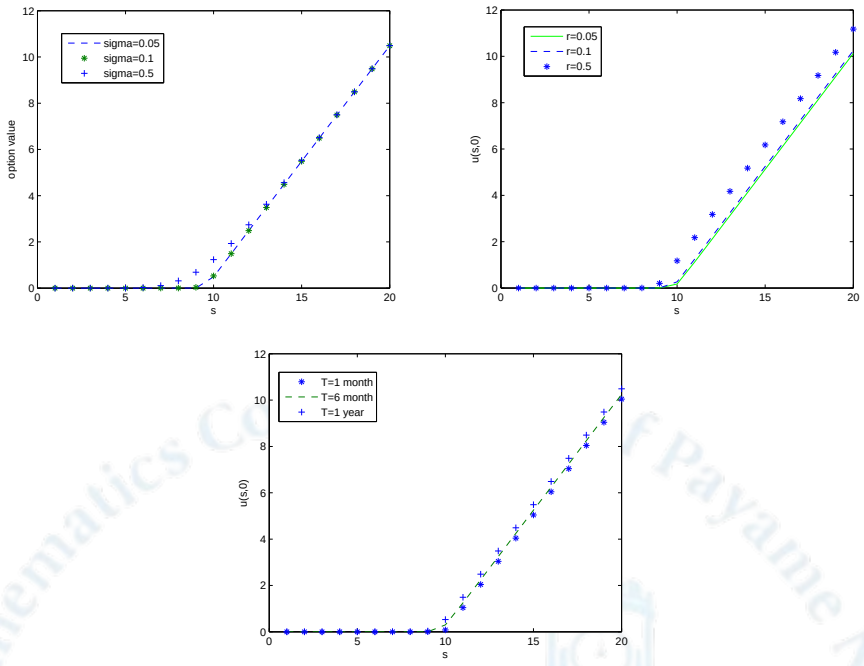


FIGURE 1. Price of the European call option with (left) E=10, r=0.2, T=0.25 year, (middle) E=10, T=0.25 year, $\sigma = 0.05$ and (right) E=10, r=0.05, $\sigma = 0.05$.

3. Conclusion

We appraise the Black-Scholes equation solution with constant parameters in this article. In the real financial market, the volatility is more complicated which leads to the fully nonlinear Black-Scholes equation. Because this equation does not have an exact solution, we solve this nonlinear partial differential equation numerically.

References

1. Black, F., Scholes, M. (1973) *The pricing of options and corporate liabilities*, J. Polit. Econ. **81**, 637-59.
2. J. Ankudinova, and M. Ehrhardt, *On the numerical solution of nonlinear Black-Scholes equations*, Comput. Math. Appl. *56* (2008), 799-812.
3. G. Barles and H. M. Soner, *Option pricing with transaction costs and a nonlinear Black-Scholes equation*, Finance and Stochastics. *2* (1998) 369-397.
4. L. Bachelier, *Thorie de la Spculation, Annales Scientifiques de lEcole Normale Suprieure*, Troisime Srie, *2*, 1900.
5. R. Farnoosh, A. Sobhani, H. Rezazadeh, and M. Beheshti, *Numerical method for discrete double barrier option pricing with time-dependent parameters*, Compu. Math. Appl. *70* (2015), 2006-2013.
6. C. F. Lo, H. C. Lee, and C. H. Hui, *A simple approach for pricing barrier options with time-dependent parameters*, Quantitive Finance. *3* (2003), 98-107.
7. H. C. O, J. J. Jo, and J. S. Kim, *General properties of solutions to inhomogeneous BlackScholes equations with discontinuous maturity payoffs* , J. Differential Equations, *260* (2016) 3151-3172.
8. J. Perello, J. M. Porra, M. Montero and J. Masoliver, *Black Scholes option pricing within Ito and Stratonovich conventions*, Physica A. *278* (2000) 260-274.
9. P. A. Samuelson, *Rational Theory of Warrant Pricing Industrial Management Review*, 6, 1965.
10. P. Wilmott, S. Howison, J. Dewynne, *The Mathematics of Financial Derivatives*, Cambridge University Press, New York, 1995.

Construction of Operational Matrix for Solving Non-linear Fractional Differential Equations Via Genocchi Polynomials and Collocation Method

Azar Sadat Shabani[1], Academic member, Department of Mathematics,
Payame Noor University, P.O. Box, 19395-3697, Tehran, Iran
Shabani_azarsadat@pnu.ac.ir

Abstract: In this paper, operational method based on Genocchi polynomials for numerical solutions of non-linear fractional differential equations (NFDEs) is proposed. The Genocchi operational matrix of fractional derivative is first constructed by using some important properties of Genocchi polynomials. These operational matrices together with the collocation method are used to reduce the NFDEs into a system of non-linear algebraic equations. The error bound for this proposed method is shown.

Keywords: Genocchi polynomials, operational matrix of fractional derivatives, fractional differential equations, collocation points.

1. Introduction

In this article, we consider NFDEs of the form:

$$D^{\alpha_i} y_n(x) = f_n(x, y_1, y_2, \dots, y_n) \quad (1)$$

where, D^{α_i} is the fractional derivative of order α_i in Caputo sense and α_i is an arbitrary order, subject to initial conditions $y_i(0) = d\ , i = 1,2, \dots, n.$

2. Fractional derivative and integration

Definition 1. The Riemann-Liouville fractional integral of order α of $f(t)$ is given by

$$I^{\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau) d\tau \ \ , \ t > 0, \alpha \in \mathbb{R}^+ \quad (2)$$

Where $\Gamma(.)$ is the well known gamma function. The Riemann-Liouville fractional derivative of order $\alpha > 0$ is also defined by $(D_i^{\alpha} f)(t) = \left(\frac{d}{dt}\right)^m (I^{m-\alpha} f)(t), (\alpha > 0, m-1 < \alpha < m)$ Some properties of I^{α} are as follows:

$$I^{\alpha} I^{\beta} f(t) = I^{\alpha+\beta}\ , \alpha > 0, \beta > 0 \quad , I^{\alpha} t^{\beta} = \frac{\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)} t^{\alpha+\beta} \quad (3)$$

Definition 2. The Caputo fractional derivative D^{α} of a function $f(t)$ is defined as**:**

$$D^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau \ \ , n-1 < \alpha \leq n, n \in \mathbb{N} \quad (4)$$

3. Genocchi polynomials and some properties

Genocchi polynomials can be used for numerical integration and function approximation. The

[1] . Corresponding Author

Genocchi polynomials $G_n(x)$and numbers $G_n$are usually expressed utilizing the exponential generating functions $Q(t,x)$and$Q(t)$, respectively, as follows [1,3]:

$$Q(t)=\frac{2t}{e^t+1}=\sum_{n=0}^{\infty}G_n\frac{t^n}{n!}\ ,\quad(|t|<\pi) \tag{5}$$

$$Q(t,x)=\frac{2t}{e^t+1}=\sum_{n=0}^{\infty}G_n(x)\frac{t^n}{n!}\ ,\quad(|t|<\pi) \tag{6}$$

$G_n(x)$ is the Genocchi polynomial of order n, and is defned as follows:

$$G_n(x)=\sum_{k=0}^{n}\binom{n}{k}G_{n-k}x^k \tag{7}$$

where G_{n-k} is the Genocchi number, which can be calculated from

$$G_n=2(1-2^n)B_n \tag{8}$$

is Bernoulli numbers.

Some of the important properties of Genocchi polynomials are:

$$\int_0^1 G_nG_m dx=\frac{2(-1)^n n!\,m!}{(m+n)!}G_{m+n},\qquad n,m\geq 1$$

$$\frac{dG_n(x)}{dx}=nG_{n-1}(x),\ \ n\geq 1$$

$$G_n(1)+G_n(0)=0,\ \ n>1 \tag{9}$$

$$G_n(t)=\int_0^t nG_{n-1}(x)dx+G_n\ ,\ \ n\geq 1$$

3.1 Function approximation

Suppose that $G(t)=[G_1(t),G_2(t),\dots,G_N(t)]\subset L^2[0,1]$ is the set of Genocchi polynomials and $Y=span[G_1(t),G_2(t),\dots,G_N(t)]$.Let $f(t)$is an arbitrary function belonging to $L^2[0,1]$, since Y is a finite dimensional subspace of $L^2[0,1]$ space, then $f(t)$ has a unique best approximation in Y, say $f^*(t)$ such that

$$\forall\, y(t)\in Y, \|f(t)-f^*(t)\|_2\leq\ \|f(t)-y(t))\|_2 \tag{10}$$

This implies that $\forall\, y(t)\in Y$

$$\langle f(t)-f^*(t),y(t)\rangle=0 \tag{11}$$

Where $\langle.\rangle$ denotes inner product. Since $f^*(t)\in Y$, then there exist the unique coefficients $C=[c_1,c_2,\dots,c_N]^T$ and $G(t)=[G_1(t),G_2(t),\dots,G_N(t)]$ such that

$$f(t)\approx f^*(t)=\sum_{i=1}^{N}c_iG_i(t)=C^T\,G(t)=C^TG\chi_i \tag{12}$$

In which $C=[c_1,c_2,\dots,c_N]^T$ is unknown vectors; $\chi_i=[1,t,t^2,\dots,t^N]^T$.

Using (18), we have

$$\langle f(t)-C^TG(t),G_i(t))\rangle=0\qquad i=1,2,\dots,N \tag{13}$$

For simplicity we write

$$C^T\langle G(t),G(t)\rangle=\langle f(t),G(t)\rangle \tag{14}$$

Where $\langle G(t),G(t)\rangle$ is an $N\times N$ matrix. Let $D=\langle G(t),G(t)\rangle=\int_0^1 G(t)G^T(t)dt$ the entries of the matrix D can be calculated from (9). Therefore, any function $f(t)\in L^2[0,1]$ can be expanded by Genocchi polynomials as $f(t)=C^TG(t)$, where

$$C=D^{-1}\langle f(t),G(t)\rangle \tag{15}$$

3.2. Error Bound

In this section we provide the error bound for the approximated function $f(t)$. Therefore, we suppose that $f(t) \in C^{n+1}[0,1]$ then

$$\|f(t) - C^T G(t)\| \leq \frac{h^{\frac{2n+3}{2}} R}{(n+1)!\sqrt{2n+3}} \quad (16)$$

Where $R = \max_{t\in[t_i,t_{i+1}]} \left|f^{(n+1)}(t)\right|$ and $h = t_{i+1} - t_i$. Hence we conclude that at each sub interval $[t_i, t_{i+!}]$, $i = 1,2, \dots, n$. f(t) has a local error bound of O $\left(h^{\frac{2n+3}{2}}\right)$. Thus, $f(t)$has a global error of O $\left(h^{\frac{2n+1}{2}}\right)$ on the whole interval $[0,1]$.

Lemma 1. Let $G_i(t)$be the Genocchi then, $D^\alpha G_i(t) = 0, for \;\; i = 1,2, \dots, [\alpha] - 1, \alpha > 0$.

4. Genocchi operational matrix of fractional derivative

If we consider the Genocchi vector $G(t)$ given by $G(t) = [G_1(t), G_2(t), \dots, G_N(t)]$, then the derivative of $G(t)$ with the aid of (9) can be expressed in the matrix form by $\frac{dG(t)^T}{dt} = MG(t)^T$ where

$$M = \begin{bmatrix} 000\cdots & 0 & 00 \\ 200\cdots & 0 & 00 \\ 000\cdots & 0 & 00 \\ 034\cdots & 0 & 00 \\ \vdots\vdots\vdots\cdots & \vdots & \vdots\vdots \\ 000\cdots N-1 & 0 & 0 \\ 000\cdots & 0 & N0 \end{bmatrix} \quad (17)$$

Thus, $M$is $N \times N$operational matrix of derivative. It is not difficult to show inductively that, the k^{th}derivative of $G(t)$ is given by

$$\frac{d^k G(t)^T}{dt^k} = G(t)(M^T)^k \quad (18)$$

Theorem 1. Suppose $G(t)$is the Genocchi vector given in (12) and let$\alpha > 0$. Then,

$$D^\alpha G(t)^T = P^\alpha G(t)^T \quad (19)$$

where $P^\alpha$is $N \times N$ operational matrix of fractional derivative of order α in Caputo sense and is defined as follows:

$$P^{(\alpha)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \\ \sum_{k=[\alpha]}^{[\alpha]} \rho_{[\alpha],k,1} & \sum_{k=[\alpha]}^{[\alpha]} \rho_{[\alpha],k,2} & \dots & \sum_{k=[\alpha]}^{[\alpha]} \rho_{[\alpha],k,N} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{k=[\alpha]}^{i} \rho_{i,k,1} & \sum_{k=[\alpha]}^{i} \rho_{i,k,2} & \cdots & \sum_{k=[\alpha]}^{i} \rho_{i,k,N} \\ \vdots & \vdots & \dots & \vdots \\ \sum_{k=[\alpha]}^{N} \rho_{N,k,1} & \sum_{k=[\alpha]}^{N} \rho_{N,k,2} & \cdots & \sum_{k=[\alpha]}^{N} \rho_{N,k,N} \end{bmatrix} \quad (20)$$

where $\rho_{i,j,k}$ is given by:

$$\rho_{i,j,k} = \frac{i! G_{i-k}}{(i-k)!\Gamma(k+1-\alpha)} c_j \quad (21)$$

G_{i-k} is the Genocchi number and c_j can be obtained from (15).

5. Collocation method based on Genocchi operational matrix of fractional derivative

In this section, we use the collocation method based on the Genocchi operational matrix of fractional derivatives to solve the NFDEs (1) numerically. To do this, we first approximate $y_j(t), j = 1, 2 \dots, n$ by Genocchi polynomials as follows:

$$y_j(t) = \sum_{k=1}^{N} c_{j,k} G_k(t) = C_j G(t)^T \qquad j = 1,2, \dots, n \tag{22}$$

Where $C_j = [c_{j,1}, c_{j,2}, \dots, c_{j,N}]$ is an unknown vector. Now employing (16) in (22), we have

$$D^{\alpha} y_j(t) \cong C_j P^{\alpha} G(t)^T, \ j = 1,2, \dots, n \tag{23}$$

Therefore, substituting (28) and (29) in (1), we have

$$C_j P^{(\alpha)} G(t)^T = f_j(t, C_1 G(t)^T, C_2 G(t)^T, \dots, C_n G(t)^T) \ \ j = 1,2, \dots, n \tag{24}$$

From the initial conditions we have

$$C_j G(0)^T = d_j \quad j = 1,2, \dots, n \tag{25}$$

To find the solution of (1), we collocate (30) at the collocation points $t_i = \frac{i}{N-i}, \ i = 1,2, \dots, N-1$ to obtain

$$C_j P^{(\alpha)} G(t_i)^T = f_j(t_i, C_1 G(t_i)^T, C_2 G(t_i)^T, \dots, C_n G(t_i)^T) \quad i = 1,2, \dots, N-1, j = 1,2, \dots, n \tag{26}$$

Thus, (26) contains $n(N-1)$ algebraic equations. These equations together with (25) make $n(N)$ algebraic equations which can be solved through Newton's iterative method. Thus, $y_j(t)$ given in (22) can be calculated. The procedure can be easily extend to solve the nonlinear system of fractional differential equations (NSFDEs).

6. Numerical Examples

Consider the following NSFDE**:**

$$\begin{cases} D^{\alpha} y_1(t) = \dfrac{y_1(t)}{2} \\ D^{\alpha} y_2(t) = \left(y_1(t)\right)^2 + y_2(t) \\ y_1(0) = 1 \, , y_2(0) = 0 \end{cases}$$

The exact solution of this system when $\alpha = 1$is known to be$y_1(t) = e^{\frac{t}{2}}$ and $y_2(t) = te^t$. We consider this example when $\alpha = 0.5, 0.7$. We reported the numerical results for $y_1(t)$ and $y_2(t)$ in Table 1.

Table 1: Numerical solutions $y_1(t)$ and $y_2(t)$, when $\alpha = 0.5, 0.7$ obtained by the present method.

	$\alpha = 0.5$		$\alpha = 0.7$	
t	$y_1(t)$	$y_2(t)$	$y_1(t)$	$y_2(t)$
0.2	1.2931031230	1.0835866399	1.1892580591	0.5361674631
0.4	1.4695250791	2.3646303626	1.3428046690	1.2320138557
0.6	1.6293827449	4.0691446057	1.4944201345	2.1872479471
0.8	1.7841485799	6.3194120920	1.6499360492	3.4822704268

7. Conclusion

In this paper, a new operational matrix based on the Genocchi polynomials is derived and applied together with the collocation method to numerically solve the NFDEs. The present method is a simple and good mathematical tool for finding the numerical solutions of NFDEs.

References

1. Isah, A., Phang, C.: On Genocchi operational matrix of fractional integration for solving fractional diferential equations. AIP Conf. Proc. 1795, 020015 (2017).
2. Loh, J.R., Phang, C., Isah, A.: New operational matrix via Genocchi polynomials for solving Fredholm-Volterra fractional integro-diferential equations. Adv. Math. Phys. 2017, 112 (2017).
3. Hashemizadeh, E., Ebadi, M.A., Noeiaghdam, S.: Matrix method by Genocchi polynomials for solving non-linear Volterra integral equations with weakly singular kernels. Symmetry 12, 2105 (2020).

He's semi-inverse variational method for the fifth-order integrable equations

Mehdi Fazli Aghdaei[1,*] and Jalil Manafian[2,3]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m.fazliaghdaei@gmail.com
[2]Department of Applied Mathematics, Faculty of Mathematical Science, University of Tabriz, Tabriz, Iran.
Email: manafeian2@gmail.com
[3]Natural Sciences Faculty, Lankaran State University, 50, H. Aslanov str., Lankaran, Azerbaijan.

Abstract. In this paper, the nonlinear partial differential equations including the $(1+1)$-dimensional and $(2+1)$-dimensional fifth-order integrable equations are studied by He's semi-inverse variational method based upon the integration tool. The merits of the presented method is finding the further solutions of the considering problems including soliton, periodic, kink, kink-singular wave solutions. Finally, these solutions might play important role in engineering, physics and applied mathematics fields.

Keywords: He's semi-inverse variational method, Fifth-order integrable equations, Soliton wave solutions

AMS Mathematics Subject Classification [2020]: 65D19, 65H10, 35A20

1. Introduction

The $(1+1)$- and $(2+1)$-dimensional fifth-order integrable equations are given as

$$u_{ttt} + u_{txxxx} - \alpha(u_x u_t)_{xx} - \beta(u_x u_{xt})_x = 0, \tag{1}$$

$$u_{ttt} + u_{tyyyy} - u_{txx} - \alpha(u_y u_{yt})_y = 0, \tag{2}$$

where were established by Wazwaz [1] and give multiple kink solutions. Kink solutions for three new fifth order nonlinear equations was investigated by Wazwaz [2]. Also the Hirota's direct method is used to derive multiple kink solutions for Eq. (1) for the case $\alpha = \beta = 4$, and only two soliton solutions for Eq. (2) for $\alpha = 4$, have been investigated by Wazwaz [2]. In [3], the Bäcklund transformation and the simplified Hirota's method were be used to study the derived couplings by Wazwaz. Some new fifth-order nonlinear equations for obtaining the exact solutions have used the G'/G expansion and rational sine-cosine methods by Qawasmeh and Alquran [4]. Authors of [5,6] applied new efficient methods for solving some nonlinear partial differential equations and obtained new exact solutions for the related equations. First, we introduce a general form of the ITEM [7–9],

*Speaker.

which is a new method. Second, we use the He's semi-inverse variational principle method to the Eqs. (1) and (2) for obtaining the dark and bright soliton wave solutions.

2. The He's semi-inverse variational principle method

Step 1. Suppose the following nonlinear partial differential equation as

$$\mathcal{N}(u, u_x, u_t, u_{xx}, u_{tt}, ...) = 0, \tag{3}$$

and can be converted to an ODE as:

$$\mathcal{Q}(u, ku', wu', k^2u'', w^2u'', ...) = 0, \tag{4}$$

by the transformation $\xi = kx + wt$ as the wave variable.
Step 2. According to He's semi-inverse method, we construct the following trial-functional

$$J(U) = \int Ld\xi, \tag{5}$$

where L is an unknown function of U and its derivatives.
Step 3. By the Ritz method, we search the solitary wave solutions in the form

$$U(\xi) = Asech(B\xi), \qquad U(\xi) = A\tanh(B\xi), \tag{6}$$

where A and B are constants. Putting (6) into (5), we get

$$\frac{\partial J}{\partial A} = 0, \qquad \frac{\partial J}{\partial B} = 0, \tag{7}$$

Solving Eqs. (7), A and B and the solitary wave solutions (6) are well determined.
Example 1. We consider the following

$$(\mu^2 - \lambda^2)u' + \mu^2 u''' + \frac{\mu}{2}(\alpha + \beta)(u')^2 = 0, \tag{8}$$

by setting $w = u'$, then Eq. (8) will be as

$$(\mu^2 - \lambda^2)w + \mu^2 w'' + \frac{\mu}{2}(\alpha + \beta)(w)^2 = 0. \tag{9}$$

By He's semi-inverse principle [10–12], the following formulation for (9) is obtained

$$J = \int_0^\infty \left[(\mu^2 - \lambda^2)\frac{w^2}{2} + \mu^2\frac{(w')^2}{2} + \frac{\mu}{6}(\alpha + \beta)w^3\right] d\xi. \tag{10}$$

By a Ritz-like method, we search a solitary wave solution $w(\xi) = Asech(B\xi)$ to find A and B as constants. Thus, we have

$$\begin{aligned} J &= \int_0^\infty \tfrac{1}{2}A^2\left[(\mu^2 - \lambda^2)sech(B\eta)^2 + \mu^2 sech(B\eta)^2\tanh(B\eta)^2B^2 + \tfrac{1}{3}\mu(\alpha + \beta)Asech(B\eta)^3\right] d\eta \\ &= \tfrac{1}{2B}(\mu^2 - \lambda^2)A^2 + \tfrac{1}{6}\mu^2A^2B + \tfrac{1}{24B}\mu A^3(\alpha + \beta)\pi. \end{aligned} \tag{11}$$

Making J stationary with A and B yields

$$\begin{aligned} \tfrac{\partial J(A,B)}{\partial A} &= \tfrac{1}{B}(\mu^2 - \lambda^2)A + \tfrac{1}{3}\mu^2AB + \tfrac{1}{8B}\mu A^2(\alpha + \beta)\pi = 0, \\ \tfrac{\partial J(A,B)}{\partial B} &= -\tfrac{1}{2B^2}(\mu^2 - \lambda^2)A^2 + \tfrac{1}{6}\mu^2A^2 - \tfrac{1}{24B^2}\mu A^3(\alpha + \beta)\pi = 0. \end{aligned} \tag{12}$$

Solving Eqs. (12), we obtain $A = -\frac{48(\mu^2-\lambda^2)}{5\mu\pi(\alpha+\beta)}$, $B = \sqrt{\frac{3}{5}\left(1 - \frac{\lambda^2}{\mu^2}\right)}$. By using the transformation $u = \int w(\xi)d\xi$, we will have

$$u(x,t) = -\frac{48}{5\pi(\alpha + \beta)}\sqrt{\frac{5}{3}(\mu^2 - \lambda^2)}\arctan\left[\sinh\left(\sqrt{\frac{3}{5}(\mu^2 - \lambda^2)}(x - \lambda t)\right)\right]. \tag{13}$$

Also, we search a solitary wave solution in the form

$$w(\xi) = A\tanh(B\xi), \tag{14}$$

where A and B are unknown constants. Putting (14) into (10), we have

$$\begin{aligned} J &= \int_0^\infty \tfrac{1}{2}A^2\left[(\mu^2-\lambda^2)\tanh(B\eta)^2 + \mu^2(1-\tanh(B\eta)^2)^2B^2 + \tfrac{1}{3}\mu(\alpha+\beta)A\tanh(B\eta)^3\right]d\eta \\ &= -\tfrac{1}{2B}(\mu^2-\lambda^2)A^2 + \tfrac{1}{3}\mu^2A^2B - \tfrac{1}{12B}\mu A^3(\alpha+\beta). \end{aligned} \tag{15}$$

Making J stationary with A and B yields

$$\begin{aligned} \tfrac{\partial J(A,B)}{\partial A} &= -\tfrac{1}{B}(\mu^2-\lambda^2)A + \tfrac{2}{3}\mu^2AB - \tfrac{1}{4B}\mu A^2(\alpha+\beta) = 0, \\ \tfrac{\partial J(A,B)}{\partial B} &= \tfrac{1}{2B^2}(\mu^2-\lambda^2)A^2 + \tfrac{1}{3}\mu^2A^2 + \tfrac{1}{12B^2}\mu A^3(\alpha+\beta) = 0. \end{aligned} \tag{16}$$

Solving Eqs. (16), we obtain $A = -\frac{24(\mu^2-\lambda^2)}{5\mu(\alpha+\beta)}$, $B = \sqrt{\frac{3}{10}\left(\frac{\lambda^2}{\mu^2}-1\right)}$. By using the transformation $u = \int w(\xi)d\xi$, we get

$$u(x,t) = -\frac{12}{5(\alpha+\beta)}\sqrt{\frac{10}{3}(\lambda^2-\mu^2)}\ln\left[-sech^2\left(\sqrt{\frac{3}{10}(\lambda^2-\mu^2)}(x-\lambda t)\right)\right]. \tag{17}$$

Example 2. We consider the following

$$(1-\lambda^2)u' + \mu^2u''' + \frac{\alpha\mu}{2}(u')^2 = 0, \tag{18}$$

by setting $w = u'$, then Eq. (18) will be as

$$(1-\lambda^2)w + \mu^2w'' + \frac{\alpha\mu}{2}(w)^2 = 0. \tag{19}$$

By He's semi-inverse principle [10–12], we obtain the following formulation for (19)

$$J = \int_0^\infty \left[(1-\lambda^2)\frac{w^2}{2} + \mu^2\frac{(w')^2}{2} + \frac{\alpha\mu}{6}w^3\right]d\xi. \tag{20}$$

Substituting $w(\xi) = Asech(B\xi)$ into (20), we have

$$\begin{aligned} J &= \int_0^\infty \tfrac{1}{2}A^2\left[(1-\lambda^2)sech(B\eta)^2 + \mu^2sech(B\eta)^2\tanh(B\eta)^2B^2 + \tfrac{1}{3}\mu\alpha Asech(B\eta)^3\right]d\eta \\ &= \tfrac{1}{2B}(1-\lambda^2)A^2 + \tfrac{1}{6}\mu^2A^2B + \tfrac{1}{24B}\mu A^3\alpha\pi. \end{aligned} \tag{21}$$

$$\begin{aligned} \tfrac{\partial J(A,B)}{\partial A} &= \tfrac{1}{B}(1-\lambda^2)A + \tfrac{1}{3}\mu^2AB + \tfrac{1}{8B}\mu A^2\alpha\pi = 0, \\ \tfrac{\partial J(A,B)}{\partial B} &= -\tfrac{1}{2B^2}(1-\lambda^2)A^2 + \tfrac{1}{6}\mu^2A^2 - \tfrac{1}{24B^2}\mu A^3\alpha\pi = 0. \end{aligned} \tag{22}$$

Solving Eqs. (22), we obtain $A = -\frac{48(1-\lambda^2)}{5\mu\pi\alpha}$, $B = \sqrt{\frac{3}{5}\left(\frac{1-\lambda^2}{\mu^2}\right)}$. By using the transformation $u = \int w(\xi)d\xi$, we will have

$$u(x,y,t) = -\frac{48}{5\pi\alpha}\sqrt{\frac{5}{3}(1-\lambda^2)}\arctan\left[\sinh\left(\sqrt{\frac{3}{5}(1-\lambda^2)}(x+y-\lambda t)\right)\right]. \tag{23}$$

Substituting $w(\xi) = A\tanh(B\xi)$ into (20), we have

$$\begin{aligned} J &= \int_0^\infty \tfrac{1}{2}A^2\left[(1-\lambda^2)\tanh(B\eta)^2 + \mu^2(1-\tanh(B\eta)^2)^2B^2 + \tfrac{1}{3}\mu\alpha A\tanh(B\eta)^3\right]d\eta \\ &= -\tfrac{1}{2B}(1-\lambda^2)A^2 + \tfrac{1}{3}\mu^2A^2B - \tfrac{1}{12B}\mu A^3\alpha, \end{aligned} \tag{24}$$

$$\begin{array}{l} \frac{\partial J(A,B)}{\partial A} = -\frac{1}{B}(1-\lambda^2)A + \frac{2}{3}\mu^2 AB - \frac{1}{4B}\mu A^2\alpha = 0, \\ \frac{\partial J(A,B)}{\partial B} = \frac{1}{2B^2}(1-\lambda^2)A^2 + \frac{1}{3}\mu^2 A^2 + \frac{1}{12B^2}\mu A^3\alpha = 0. \end{array} \tag{25}$$

Solving Eqs. (25), we obtain $A = -\frac{24(1-\lambda^2)}{5\mu\alpha}$, $B = \sqrt{\frac{3}{10}\left(\frac{\lambda^2-1}{\mu^2}\right)}$. By using the transformation $u = \int w(\xi)d\xi$, we will get

$$u(x,y,t) = -\frac{12}{5\alpha}\sqrt{\frac{10}{3}(\lambda^2-1)}\ln\left[-sech^2\left(\sqrt{\frac{3}{10}(\lambda^2-1)}(x+y-\lambda t)\right)\right]. \tag{26}$$

3. Conclusion

In this paper, the He's semi-inverse variational principle method was considered. The exact solutions were presented in terms of the hyperbolic, the trigonometric functions, the polynomial functions and the rational functions. By using He's semi-inverse variational principle method dark and bright soliton wave solutions have been obtained. Finally, these solutions might play important role in engineering, physics and applied mathematics fields.

References

1. Wazwaz, A.M. (2011) *A new fifth order nonlinear integrable equation: multiple soliton solutions*, Phys. Scr. **83**, 015012.
2. A. M. Wazwaz, Kink solutions for three new fifth order nonlinear equations, Appl. Math. Model., **38** (2014) 110-118.
3. A. M. Wazwaz, Couplings of a fifth order nonlinear integrable equation: Multiple kink solutions, Computers Fluids, **84** (2013) 97-99.
4. A. Qawasmeh, M. Alquran, Reliable study of some new fifth-order nonlinear equations by means of G'/G expansion method and rational sine-cosine method, Appl. Math. Sci., **8** (2014) 5985-5994.
5. Q. Zhou, M. Mirzazadeh, E. Zerrad, A. Biswas, M. Belic, Bright, dark, and singular solitons in optical fibers with spatio-temporal dispersion and spatially dependent coefficients, J. Modern Opt., **63** (2016) 950-954.
6. Q. Zhou, S. Liu, Dark optical solitons in quadratic nonlinear media with spatio-temporal dispersion, Nonlinear Dyn., **81** (2015) 733-738.
7. J. Manafian, M. Lakestani, Application of $tan(\phi/2)$-expansion method for solving the Biswas-Milovic equation for Kerr law nonlinearity, Optik-Int. J. Elec. Opt., **127** (2016) 2040-2054.
8. J. Manafian, M. Lakestani, Dispersive dark optical soliton with Tzitzéica type nonlinear evolution equations arising in nonlinear optics, Opt. Quant. Electron, **48** (2016) 1-32.
9. J. Manafian, Optical soliton solutions for Schrödinger type nonlinear evolutionequations by the $tan(\phi/2)$-expansion method, Optik-Int. J. Elec. Opt., **127** (2016) 4222-4245.
10. J.H. He, Some asymptotic methods for strongly nonlinear equations, Int. J. Modern Phys. B., **20** (2006) 11411199.
11. R . Kohl, D . Milovic, E. Zerrad, A. Biswas, Optical solitons by He's variational principle in a non-Kerr law media, J. Infrared Milli. Terahertz Waves, **30** (5) (2009) 526-537.
12. J. Zhang, Variational approach to solitary wave solution of the generalized Zakharov equation, Comput. Math. Appl., **54** (2007) 1043-1046.

An iterative method for computing eigenpairs of symmetric matrices

Hadi Azizi[1,*], Hamid Reza Navabpour [2] and Behzad Kafash[3]

[1]Department of Mathematics, Taft. C., Islamic Azad University, Taft, Iran.
Email: Ha.Azizi@iau.ac.ir
[2]Department of Mathematics, Faculty of Mathematics, Yazd University, Yazd, Iran.
Email: navabpour@yazd.ac.ir
[3]Faculty of Engineering, Ardakan University, Ardakan, Iran.
Email: bkafash@ardakan.ac.ir

Abstract. An iterative algorithm is proposed in this paper, expanding the classical power method by incorporating the Gram-Schmidt orthogonalization process. With this improvement, it becomes possible to compute all eigenvalues and eigenvectors of a given symmetric matrix simultaneously. The algorithm has been tested on a variety of benchmark matrices to determine the robustness of the full eigen decomposition, and the results are reported to be accurate.

Keywords: Eigenpairs, Iterative method, Power method, Eigenvalues, Eigenvectors.
AMS Mathematics Subject Classification [2020]: 65F15, 15A18

1. Introduction

Computing the eigenpairs of large matrices is a fundamental problem in computational physics [2]. This task has broad applications across engineering and science, including structural analysis, quantum mechanics, and signal and image processing. Several methods exist for determining the eigenvalues and eigenvectors of symmetric matrices, such as the power method, QR algorithm, and various specialized techniques [1,3,4]. However, many of these approaches are limited to finding only a subset of the eigenpairs. In this work, we present an extension of the classical power method by incorporating the Gram-Schmidt orthogonalization process, enabling the complete determination of all eigenpairs for any symmetric matrix A.

2. Preliminaries

The theoretical framework of our method rests upon several key mathematical theorems and defenitions. In this section, we now introduce these essential results, which will be invoked to prove the convergence properties of our algorithm.

*Speaker.

DEFINITION 2.1. Let $A \in \mathbb{R}^{n\times n}$. A nonzero vector $x \in \mathbb{R}^n$ is called an eigenvector of A with corresponding eigenvalue $\lambda \in \mathbb{C}$ if $Ax = \lambda x$.

THEOREM 2.2. *All of the eigenvalues of any symmetric real matrix A, are real.*

PROOF. See [1]. □

THEOREM 2.3. *For every symmetric real matrix, $A \in \mathbb{R}^{n\times n}$, there exists a set of eigenvectors of A as an orthonormal base of the vector space $\mathbb{R}^n$ on the real field $\mathbb{R}$.*

PROOF. See [1]. □

THEOREM 2.4. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $\lambda_1, \lambda_2, ..., \lambda_n$ are eigenvalues of A. Let $|\lambda_1| > |\lambda_2| \geqslant ... \geqslant |\lambda_n|$ and v_1 is an eigenvector correspond with λ_1 such that $\|v_1\|_2 = 1$, also let $x \in \mathbb{R}^n$ such that $x^T v_1 \neq 0$ then:*

$$\lim_{k\to\infty} \frac{A^k x}{\|A^k x\|_2} = \pm v_1, \quad \lim_{k\to\infty} \frac{x^T A^k x}{x^T A^{k-1} x} = \lambda_1.$$

PROOF. See [1]. □

DEFINITION 2.5. The dominant eigenvalue of any matrix is the eigenvalue with the largest magnitude [1].
For any positive integer i, the i^{th} dominant eigenvalue of any matrix A is an eigenvalue μ of A if there exist exactly $(i-1)$ distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_{i-1}$ of A with the property $\mu < |\lambda_k|, k = 1, 2, ..., i-1$, also the dominant eigenvalue of A is called the first dominant eigenvalue of A.

THEOREM 2.6. *Let $A \in \mathbb{R}^{n\times n}$ and $\lambda_1, \lambda_2, ..., \lambda_n$ are all eigenvalues of A and $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n|$. Let $v_1, v_2, ..., v_n$ are correspond eigenvectors. Let $|\lambda_1| = |\lambda_2| = ... = |\lambda_l|, 1 \geq l \geq n$. If $x \in \mathbb{R}^n$ is any non-zero vector and not orthogonal to at least one of $v_1, v_2, ..., v_l$, then:*

$$\lim_{k\to\infty} \frac{A^k x}{\|A^k x\|_2} = u, \quad \lim_{k\to\infty} \frac{x^T A^k x}{x^T A^{k-1} x} = \lambda_1.$$

Also u is unit and belong to span$\{v_1, v_2, ..., v_l\}$.

LEMMA 2.7. *Let $B = \{v_1, v_2, ..., v_n\}$ be an orthonormal base of $\mathbb{R}^n$ from eigenvectors of A. Let x be any non zero vector in $\mathbb{R}^n$ such that for a vector $v_i \in B$ we have $\langle x, v_i\rangle = x^T v_i = 0, i = 1, 2, ..., n$ then for any positive integer k we have, $\langle A^k x, v_i\rangle = (A^k x)^T v_i = 0$.*

COROLLARY 2.8. *Let $B = \{v_1, v_2, ..., v_n\}$ be an eigenvectors of matrix A from an orthonormal base of $\mathbb{R}^n$. Let x be any vector and orthogonal whit m vectors $v_{i_1}, v_{i_2}, ..., v_{i_m}, 1 \leq m < n$ then so is $A^k x, k = 1, 2, ...$*

3. Algoritm of the proposed method

Now we can explain the proposed method. Let us find the first dominant eigenvalue and its corresponding eigenvector v_1 of matrix A by the power method. Theoretically, the approach for finding the second eigenpair of A from corollary 2.8, it is enough to choose a vector x such that $\langle x, v_1\rangle = 0$ and apply the power method. But the power method gives an approximation of the eigenvector, so we have some error by using x. This error expands by iterations of the power method and may never give us a true approximation of the second eigenvector. Because of this, we added the Gram-Schmidt process to our algorithm as follows.

In this section, we propose a new algorithm for approximating all non-zero eigenvalues and eigenvectors of any symmetric matrix $A \in \mathbb{R}^{n\times n}$. It should be noted that in this algorithm $\lambda_1, \lambda_2, ..., \lambda_n$ are eigenvalues and $v_1, v_2, ..., vn$ are corresponding eigenvectors of A.

INPUT
Entries of matrix $A = (u_1, u_2, ..., u_n)$ where u_i is the i^{th} column of A, $i = 1, 2..., n$, maximum number of iterations N and tolerance ε, δ.

OUTPUT
All eigenpairs of A.

Step 1. Choose u_1 as initial guess of v_1 and apply power method for finding λ_1 and v_1.

Step 2. For $j = 1, 2, ..., n-1$ do step 3 to step 12.

Step 3. Use Gram-Schmidt process for u_{j+1} on $v_1, ..., v_j$. So u_{j+1} is orthogonal with $v_1, ..., v_j$.

Step 4. For $i = 1, 2, ..., N$ do step 5 to step 12.

Step 5. $u_{j+1} = \dfrac{u_{j+1}}{\|u_{j+1}\|_2}$.

Step 6. $y = Au_{j+1}$.

Step 7. $\mu_{j+1} = u^T y$.

Step 8. $y_1 = \dfrac{y}{\|y\|_2}$, $d = \|y_1 - u_{j+1}\|_2$.

Step 9. If $\mid \mu_{j+1} \mid < \delta$ then print $v_1, v_2, ..., v_j, \lambda_1, ..., \lambda_j$ and other eigenvalues are zero. **stop**

Step 10. If $d < \varepsilon$ then $v_{j+1} = y_1, \lambda_{j+1} = \mu_{j+1}, j = j + 1$ and go to step 3.

Step 11. If $i = N$ then print the number of iterations was exceeded and $(\lambda_1, v_1), ..., (\lambda_j, v_j)$, **stop**.

Step 12. Apply Gram-Schmidt process for y and $v_1, ..., v_j$. (so y is orthogonal with $v_1, v_2, ..., v_j$), $u_{j+1} = y, i = i + 1$ and go to step 5.

Step 13.
Output
Print $(v_1, \lambda_1), (v_2, \lambda_2), ..., (v_n, \lambda_n)$ and print "The procedure was successful." **stop**.

4. Numerical results

In this section, we utilize the proposed algorithm for finding eigenvalues and eigenvectors of matrices. In these examples the positive tolerance ε is an upper bound for $\|u_j^{(k)} \pm u_j^{(k-1)}\|_2$ in iteration k for approximating of eigenvector v_j. Also the positive real

number δ shows the maximum of the absolute value of non-zero eigenvalues, such that if $|\lambda| < \delta$ then we put $\lambda = 0$ (step 9).
In tables v_a and λ_a illustrate the approximation of eigenvector and eigenvalue respectively, and λ_e is the exact eigenvalue and S is $\|Av_a - \lambda_a v_a\|_2$.

EXAMPLE 4.1. Consider the matrix

$$A = \begin{pmatrix} 4 & -1 & 1 \\ -1 & 3 & -2 \\ 1 & -2 & 3 \end{pmatrix},$$

with eigenpairs

$$(6, \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}), \quad (3, \frac{1}{\sqrt{6}}\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}), \quad (1, \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}).$$

We see the results in table 1 by using the present algorithm with $\varepsilon = 10^{-8}$ and $\delta = 0.01$.

TABLE 1. Matrix with dimension 3 and its approximation of eigenpairs

initial vector	iterations	v_a	λ_a	λ_e
$\begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix}$	27	$\begin{bmatrix} -.577350 \\ 0.577350 \\ -.577350 \end{bmatrix}$	6.0	6.0
$\begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$	1	$\begin{bmatrix} 0.816496 \\ 0.408248 \\ -.408248 \end{bmatrix}$	3.0	3.0
$\begin{bmatrix} 1 \\ -2 \end{bmatrix}$	2	$\begin{bmatrix} 0.707106 \\ -.707106 \end{bmatrix}$	1.0	1.0

5. Conclusion

This paper presented an enhanced iterative algorithm for computing all eigenpairs of symmetric matrices by integrating the Gram-Schmidt orthogonalization process with the classical power method. This modification enables the simultaneous and robust determination of all eigenvalues and eigenvectors. The numerical results demonstrate the algorithm's effectiveness and accuracy in obtaining the full eigen decomposition for benchmark matrices. Therefore, the proposed method provides a reliable and comprehensive solution for the complete eigen analysis of symmetric matrices, addressing a key limitation of the standard power method.

References

1. Burden, R. L., Douglas J. (2005), Faires Numerical Analysis, 8nd ed., THOMSON.
2. Dayanada, M. A. (2017) Determination of eigenvalues, eigenvectors and interdiffusion coeffcients in ternary diffision from diffusional constraints at the Matano plane, Acta Materialia, **129**, 474-481.
3. Gubernatis, J. E., Booth, T. E. (2008) Multiple external eigenpairs by the power method, Journal of Computational Physics, **227**, 8508-8522.
4. Navarro-Gonález, F. J., Compañ, P., Satorre, R., Villacampa, Y., (2016) Numerical determination for solving the symmetric eigenvector problem using genetic algorithm,Applied Mathematical Modelling, **40**, 4935-4947.

A Boundary Element Method for the Numerical Solution of the Variable-Order Time-Fractional Diffusion Equation Based on the Caputo Derivative

Leila Hasani[1,*], Allahbakhsh Yazdani Charati[2]

[1]Department of Mathematics, Mazandaran University, Babolsar, Iran.
Email: leilahasani92@gmail.com
[2]Department of Mathematics, Mazandaran University, Babolsar, Iran.
Email: Yazdani@umz.ac.ir

Abstract. In this study, a boundary element method is employed for the numerical solution of the variable-order time-fractional diffusion equation based on the Caputo fractional derivative. Variable-order fractional equations are particularly effective in modeling physical and engineering processes with time-dependent memory effects. In the proposed approach, the fractional diffusion equation is reformulated in an integral form, and time discretization based on the Caputo definition is applied to accurately account for the variations in the fractional order over time. The numerical algorithm is implemented in MATLAB, and results obtained from a numerical example demonstrate that the proposed method provides high accuracy, good stability, and fast convergence. These features make the boundary element method an efficient and reliable tool for solving variable-order time-fractional diffusion problems.

Keywords: Variable-order, fractional diffusion, Caputo derivative, Boundary Element Method

AMS Mathematics Subject Classification [2020]: 35R11, 65N38

1. Introduction

Fractional differential equations (FDEs) describe memory and nonlocal phenomena in diffusion, viscoelasticity, and relaxation. Among available definitions, the Caputo form is preferred because it accommodates standard initial and boundary conditions. Variable-order fractional differential equations (VO-FDEs), introduced by Samko and Ross [1] and formalized by Lorenzo and Hartley [2], allow the differentiation order to vary with time or space, improving modeling fidelity for heterogeneous systems. Recent developments by Garrappa et al. [3] refined their theoretical framework, while Patnaik et al. [4] surveyed broad engineering applications. However, existing numerical schemes remain costly due to full-domain discretization. To overcome this, the present work proposes a Variable-Order Caputo Boundary Element Method (VO-CBEM) that embeds variable

*Speaker.

fractional order within a boundary-only formulation, achieving efficient and stable solutions for time-fractional diffusion problems.

2. Main results

2.1. Mathematical Model. In a bounded domain $\Omega \subset \mathbb{R}^2$ with boundary $\Gamma = \Gamma_D \cup \Gamma_N$, the variable-order time-fractional diffusion equation is formulated as

$$ {}^{C}D_t^{\alpha(t)} u(x,y,t) = D_x \frac{\partial^2 u}{\partial x^2} + D_y \frac{\partial^2 u}{\partial y^2} + F(x,y,t), \quad (x,y) \in \Omega,\ t > 0, \tag{1}$$

where u denotes temperature or concentration, D_x, D_y are diffusion coefficients, and ${}^{C}D_t^{\alpha(t)}$ is the Caputo derivative of order $\alpha(t)$ (see Definition 2.1).

The boundary-initial conditions are prescribed as

$$ u(x,y,0) = u_0(x,y), \tag{2}$$

$$ u(x,y,t) = f(x,y,t) \quad \text{on } \Gamma_D, \tag{3}$$

$$ \frac{\partial u(x,y,t)}{\partial n} = g(x,y,t) \quad \text{on } \Gamma_N, \tag{4}$$

with $\mathbf{n}$ denoting the outward normal. This variable-order formulation enables modelling of non-stationary diffusion with mixed boundaries, suitable for BEM discretization.

Definition 2.1 (Variable-order Caputo derivative). For a function $u(t) \in C^1[0,T]$ and a variable order $\alpha : [0,T] \to (0,1]$, the Caputo derivative of order $\alpha(t)$ is defined as

$$ {}^{C}D_t^{\alpha(t)} u(t) = \frac{1}{\Gamma(1-\alpha(t))} \int_0^t (t-\tau)^{-\alpha(t)} \frac{du(\tau)}{d\tau}\, d\tau, \tag{5}$$

where $\Gamma(\cdot)$ is the Gamma function. This nonlocal operator accounts for time-dependent memory intensity governed by $\alpha(t)$, providing a continuous transition between classical ($\alpha = 1$) and fractional ($0 < \alpha < 1$) dynamics.

2.2. Formulation and Implementation of the BEM. According to the standard CD–BEM scheme [5], the variable-order time-fractional diffusion equation is converted into the boundary–domain integral form:

$$ \begin{aligned} c(\xi)u(\xi,t) = &\int_\Gamma q(X,t)\, w(\xi,X)\, d\Gamma - \int_\Gamma u(X,t)\, Q(\xi,X)\, d\Gamma \\ &- \int_\Omega {}^{C}D_t^{\alpha(t)} u(X,t)\, w\, d\Omega + \int_\Omega F(X,t)\, w\, d\Omega, \end{aligned} \tag{6}$$

where $w(\xi,X) = \frac{1}{2\pi\sqrt{D_x D_y}} \ln(1/r)$ is the anisotropic fundamental solution, $Q = \partial w/\partial n$ and $c(\xi)$ denotes the geometric coefficient, equal to 1 for interior points and $1/2$ for smooth boundary nodes.

The time-fractional term is discretized by

$$ {}^{C}D_t^{\alpha(t_{n+1})} u = \frac{1}{\Gamma(2-\alpha_{n+1})\Delta t^{\alpha_{n+1}}} \sum_{k=0}^{n} \omega_{n+1,k+1}^{(\alpha_{n+1})} (u_{k+1} - u_k), \tag{7}$$

with memory weights $\omega_{n+1,k+1}^{(\alpha_{n+1})} = (n+1-k)^{1-\alpha_{n+1}} - (n-k)^{1-\alpha_{n+1}}$.

Spatial discretization over Γ and Ω leads to the algebraic VO–CBEM system

$$\mathbf{H}u_{n+1} = \mathbf{G}q^b_{n+1} - \frac{1}{\Gamma(2-\alpha_{n+1})\Delta t^{\alpha_{n+1}}}\mathbf{M}\sum_{k=0}^{n}\omega^{(\alpha_{n+1})}_{n+1,k+1}(u_{k+1}-u_k) + \mathbf{M}F_{n+1}, \tag{8}$$

where $\mathbf{H}$ and $\mathbf{G}$ are the standard BEM influence matrices for potential and flux, and $\mathbf{M}$ corresponds to the domain integration matrix associated with the fractional-time term.

2.3. Results. To evaluate the accuracy and stability of the proposed VO-CBEM scheme, a two-dimensional transient problem [6] is solved on $\Omega = (0,\pi)^2$ with $\Delta t = 0.005$, $n_\Gamma = 64$, and $N_{\text{ele}} = 512$. The governing equation employs $\alpha(t) = 2+\sin(t)/4$, and diffusion coefficients $D_x = D_y = 1$. Mixed boundary conditions are specified: Dirichlet on $x = 0, \pi$ and Neumann on $y = 0, \pi$. The analytical solution $u(x,y,t) = (t^3+3t^2+1)\sin(x)\sin(y)$ ensures full consistency with the VO-fractional model.

The corresponding source term is

$$f(x,y,t) = \left[\frac{6t^{3-\alpha(t)}}{\Gamma(4-\alpha(t))} + \frac{6t^{2-\alpha(t)}}{\Gamma(3-\alpha(t))} + 2(t^3+3t^2+1)\right]\sin(x)\sin(y). \tag{9}$$

Table 1 compares analytical and VO-CBEM results along $y = \pi/2$ at $t = 1$. The numerical data follow the analytical profile closely, confirming high precision and smooth convergence. The maximum relative error is below 0.006949, and the error distribution remains smooth and symmetric across the domain, indicating temporal stability of VO-CBEM. All computations were performed using a MATLAB R2023b code developed for the VO-CBEM formulation.

TABLE 1. Spatial distribution along $y = \pi/2$ at $t = 1$: analytical vs. VO-CBEM numerical results.

No.	x	Analytical u	Numerical u	$\lvert e\rvert$	Rel. error ε
1	0.0000	0.000000	0.000000	0.000000	0.000000
2	0.1963	0.975452	0.968884	0.006567	0.001313
3	0.3927	1.913417	1.900304	0.013114	0.002623
4	0.5890	2.777851	2.758753	0.019098	0.003820
5	0.7854	3.535534	3.511205	0.024329	0.004866
6	0.9817	4.157348	4.128659	0.028689	0.005738
7	1.1781	4.619398	4.587388	0.032010	0.006402
8	1.3744	4.903926	4.869839	0.034087	0.006817
9	1.5708	5.000000	4.965256	0.034744	0.006949
10	1.7671	4.903926	4.869911	0.034015	0.006803
11	1.9635	4.619398	4.587424	0.031974	0.006395
12	2.1598	4.157348	4.128609	0.028739	0.005748
13	2.3562	3.535534	3.511125	0.024409	0.004882
14	2.5525	2.777851	2.758702	0.019149	0.003830
15	2.7489	1.913417	1.900288	0.013129	0.002626
16	2.9452	0.975452	0.968885	0.006567	0.001313
17	3.1416	0.000000	0.000000	0.000000	0.000000

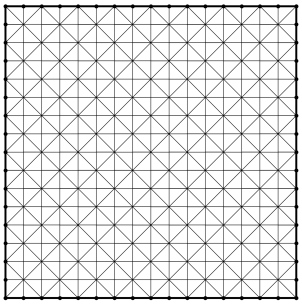


FIGURE 1. Example1: setup and mesh discretization for $\Omega = (0, \pi)^2$.

2.4. Discussion. Results confirm that a variable order $\alpha(t)$ enriches diffusion dynamics by introducing time-dependent memory. Early deviations stem from strong transient effects, while stabilization of $\alpha(t)$ yields rapid convergence. The VO-CBEM framework remains stable under mixed boundary conditions, effectively capturing non-uniform flux and preserving numerical symmetry.

3. Conclusion

The proposed VO-CBEM provides an accurate and stable tool for modeling variable-order fractional diffusion. It efficiently represents memory-dependent processes and maintains convergence across complex boundaries. The formulation offers a reliable basis for future three-dimensional and distributed-order extensions.

Acknowledgement

The author gratefully acknowledges the support and facilities provided by the Department of Mathematics, University of Mazandaran, Iran. Special thanks are extended to colleagues for valuable discussions and technical assistance during the preparation of this work.

References

1. Samko, S. G. & Ross, B. (1993) *Integration and differentiation to a variable fractional order*, Integral Transforms Spec. Funct., **1**, 277–300.
2. Lorenzo, C. F. & Hartley, T. T. (2002) *Variable order and distributed order fractional operators*, Nonlinear Dyn., **29**(1), 57–98.
3. Garrappa, R., Giusti, A. & Mainardi, F. (2021) *Variable-order fractional calculus: A change of perspective*, Commun. Nonlinear Sci. Numer. Simul., **102**, 105904.
4. Patnaik, S., Hollkamp, J. P. & Semperlotti, F. (2020) *Applications of variable-order fractional operators: A review*, Proc. R. Soc. A, **476**(2234), 20190498.
5. Carrer, J. A. M., Solheid, B. S., Trevelyan, J. & Seaid, M. (2021) *A boundary element method formulation based on the Caputo derivative for the solution of the anomalous diffusion problem*, Eng. Anal. Bound. Elem., **122**, 132–144.
6. Zhang, J.-L., Fang, Z.-W. & Sun, H.-W. (2021) *Fast second-order evaluation for variable-order Caputo fractional derivative with applications to fractional sub-diffusion equations*, arXiv preprint arXiv:2102.02960.

A Boundary Element Method Approach to Fractional Diffusion-Reaction Equations with Caputo Derivative: Applications to Mathematical Biology

Leila Hasani[1,*], Allahbakhsh Yazdani Cherati[2]

[1]Department of Mathematics, Mazandaran University, Babolsar, Iran.
Email: leilahasani92@gmail.com
[2]Department of Mathematics, Mazandaran University, Babolsar, Iran.
Email: Yazdani@umz.ac.ir

Abstract. In this study, a Boundary Element Method (BEM) is developed for solving the time-fractional diffusion-reaction equation. The model describes anomalous transport behavior accompanied by a linear decay reaction. The numerical algorithm is implemented in MATLAB, providing an efficient and accessible framework for fractional analysis. Numerical results for a two-dimensional strip domain show excellent agreement with the analytical solution, with a maximum relative error below 0.2%. The proposed method demonstrates high accuracy and stability for reaction parameter values in the range of 0 to approximately 0.1, where the internal and boundary errors remain negligible. However, for larger reaction parameters, the internal errors increase noticeably, indicating the sensitivity of the method to stronger reaction effects. The presented approach offers a reliable computational tool for modeling fractional diffusion-reaction phenomena and can be applied to biological processes such as drug absorption, molecular transport, and chemical reactions in complex tissues.

Keywords: Fractional diffusion, Caputo derivative, Boundary Element Method, Reaction-diffusion, Mathematical biology

AMS Mathematics Subject Classification [2020]: 35R11, 65N38, 92C45

1. Introduction

Fractional differential equations (FDEs) are powerful models for phenomena exhibiting memory and nonlocal behavior [1,2,6]. Among fractional operators, the Caputo derivative is preferred for its compatibility with standard boundary and initial conditions. The Boundary Element Method (BEM) offers a dimensionally reduced framework for fractional diffusion models, requiring boundary-only discretization [3].

Although existing formulations are accurate and stable, most have neglected reaction mechanisms that control growth and decay in biological contexts [4]. Extending the time-fractional diffusion equation with a reaction term λu enables realistic modeling of transformation and biochemical kinetics.

*Speaker.

Fractional reaction-diffusion equations, introduced to capture anomalous transport and memory-dependent behavior [5], now form a robust mathematical base for complex systems.

This study extends the Caputo-based domain BEM (CD-BEM) to include a constant reaction term, derives analytical solutions using the Mittag–Leffler function, and validates numerical accuracy. The proposed formulation integrates memory, reaction dynamics, and nonlocal diffusion, establishing an efficient framework for fractional biological modeling. All numerical implementations and visualizations were carried out using MATLAB, ensuring reproducibility and computational efficiency.

2. Main results

2.1. Mathematical Model. The anisotropic two-dimensional time-fractional reaction-diffusion equation is

$$\frac{\partial^\alpha u}{\partial t^\alpha} = D_x \frac{\partial^2 u}{\partial x^2} + D_y \frac{\partial^2 u}{\partial y^2} - \lambda u, \quad (x,y)\in\Omega\subset\mathbb{R}^2,\ t>0, \tag{1}$$

where $u(x,y,t)$ is the field variable (e.g., concentration or cell density); $0<\alpha\le 1$ is the Caputo fractional order representing memory; $D_x, D_y > 0$ are diffusion coefficients; and $\lambda \ge 0$ defines a linear reaction rate. This equation describes diffusion–reaction processes with temporal nonlocality such as molecular degradation or cell mortality.

Boundary and Initial Conditions.

$$u = f(x,y,t), \qquad (x,y)\in\Gamma_D \text{ (Dirichlet)}, \tag{2}$$

$$\frac{\partial u}{\partial n} = g(x,y,t), \qquad (x,y)\in\Gamma_N \text{ (Neumann)}, \tag{3}$$

$$u(x,y,0) = u_0(x,y), \qquad (x,y)\in\Omega, \tag{4}$$

with $\Gamma_D \cup \Gamma_N = \partial\Omega$ and $\partial u/\partial n = \nabla u\cdot\mathbf{n}$ denoting the outward normal flux.

Definition 2.1 (Caputo fractional derivative). For $u(t)\in C^1[0,t]$ and $0<\alpha<1$,

$${}^C D_t^\alpha u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} \frac{du(\tau)}{d\tau}\, d\tau,$$

where $\Gamma(\cdot)$ is the Gamma function. It quantifies the memory effect inherent in anomalous diffusion.

2.2. Boundary Element Formulation. Starting from the residual form of the governing PDE,

$$R(X,t) = \frac{\partial^\alpha u}{\partial t^\alpha} - D_x \frac{\partial^2 u}{\partial x^2} - D_y \frac{\partial^2 u}{\partial y^2} + \lambda u, \tag{5}$$

the weighted residual method is applied to enforce $\int_\Omega R(X,t)\, w(\xi,X)\, d\Omega = 0$, where $w(\xi,X)$ is the fundamental solution of the anisotropic steady-state operator

$$D_x \frac{\partial^2 w}{\partial x^2} + D_y \frac{\partial^2 w}{\partial y^2} = -\delta(\xi,X), \qquad w(\xi,X) = \frac{1}{2\pi\sqrt{D_x D_y}} \ln\!\left(\frac{1}{r}\right), \tag{6}$$

with $r = \sqrt{(x-\xi_x)^2 + \frac{D_x}{D_y}(y-\xi_y)^2}$. Integration by parts and use of the delta property lead to the classical BEM boundary integral equation

$$c(\xi)u(\xi,t) = \int_\Gamma q(X,t) w\, d\Gamma - \int_\Gamma u(X,t) Q\, d\Gamma - \int_\Omega \left(\frac{\partial^\alpha u}{\partial t^\alpha} + \lambda u\right) w\, d\Omega, \tag{7}$$

where $Q = \partial w/\partial n$ and $c(\xi)$ is the geometric coefficient,

$$c(\xi) = \tfrac{1}{2\pi}\left[\tan^{-1}\!\left(\frac{D_x}{D_y}\tan\theta_2\right) - \tan^{-1}\!\left(\frac{D_x}{D_y}\tan\theta_1\right)\right], \tag{8}$$

reducing to $c(\xi) = \frac{\theta_2-\theta_1}{2\pi}$ for isotropic media, with $c(\xi) = 1/2$ at a smooth boundary, $> 1/2$ at internal corners, and $< 1/2$ at external corners.

Temporal Discretization. Time is divided in steps Δt, and the Caputo derivative at t_{n+1} is approximated by the convolution quadrature scheme

$$\frac{\partial^\alpha u(X,t_{n+1})}{\partial t^\alpha} = \frac{1}{\Gamma(2-\alpha)\Delta t^\alpha}\left[u_{n+1} - u_n + \sum_{j=0}^{n-1} B_{(n+1)(j+1)}(u_{j+1}-u_j)\right], \tag{9}$$

with $B_{(n+1)(j+1)} = (n+1-j)^{1-\alpha} - (n-j)^{1-\alpha}$ being memory weights. Substituting Eq. (9) into (7) yields a time-discrete integral system.

Spatial Discretization and Matrix Form. Discretization of Γ into N_Γ boundary elements and Ω into N_Ω triangular cells transforms the integral equation into a matrix system

$$\mathbf{H}u_{n+1} = \mathbf{G}q^b_{n+1} - \frac{\mathbf{M}}{\Gamma(2-\alpha)\Delta t^\alpha}\left[(u_{n+1}-u_n) + \sum_{j=0}^{n-1} B_{(n+1)(j+1)}(u_{j+1}-u_j)\right] - \lambda\mathbf{M}u_{n+1}, \tag{10}$$

where $\mathbf{H}$ and $\mathbf{G}$ arise from boundary integrals, and $\mathbf{M}$ from domain integrals. Collecting terms in u_{n+1} gives a symmetric linear system that is solved iteratively over time.

2.3. Results and Validation. To assess CD-BEM performance, a rectangular domain $\Omega = [0, L_x]\times[0, L_y]$ with isotropic diffusion ($D_x = D_y = D$) and reaction coefficient λ is considered. Dirichlet conditions: $u(0, y, t) = 10$, $u(L_x, y, t) = 0$; Neumann zero-flux on $y = 0, L_y$; initial $u_0(x, y) = 0$. For $L_x = 2$, $L_y = 1$, $\Delta t < 0.1$, and $\lambda < 0.01$, simulations (see Fig. 1) show excellent agreement with analytical solutions derived from the Mittag-Leffler function.

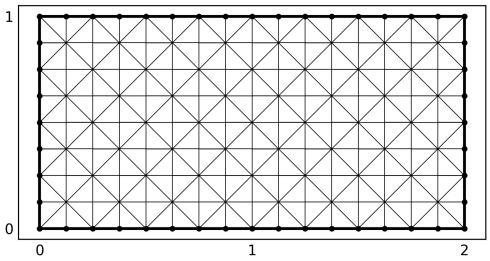
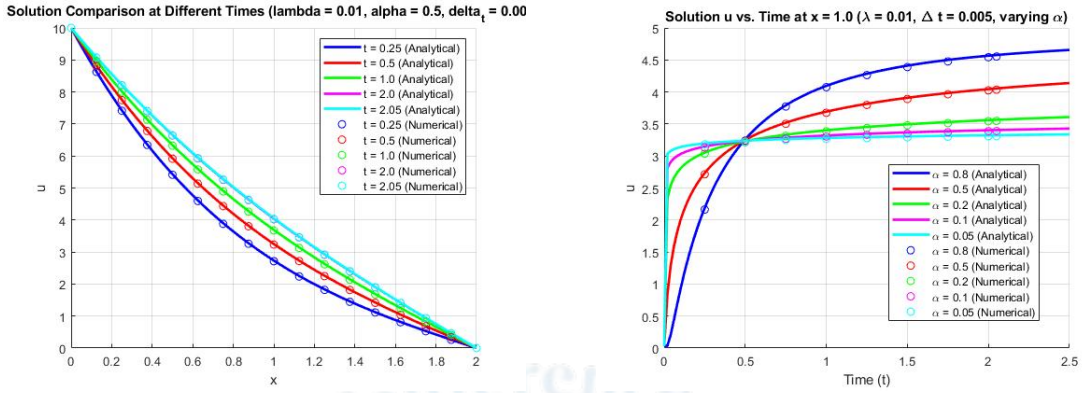


FIGURE 1. Rectangular mesh with 52 boundary and 105 interior nodes.

The reduced one-dimensional analytical form can be found in [3], providing the exact solution for fractional diffusion–reaction systems.

Results (see Fig. 2 and Table 1) confirm accurate and stable CDBEM performance with relative errors below 0.2%, smoother decay for smaller α, faster attenuation for larger λ, and low sensitivity to Δt.



(A) $u(x)$ for $\alpha = 0.5$, $\lambda = 0.01$, $\Delta t = 0.005$

(B) $u(x)$ for various α at $t = 2.05$

FIGURE 2. Analytical and CD-BEM results for spatial distribution.

TABLE 1. Relative error vs. λ at $t = 2.0$.

λ	E_{rel}	Stability
0.5	8.2×10^{-2}	stable
0.05	3.2×10^{-3}	stable
0.01	4.8×10^{-4}	highly stable
0.005	2.9×10^{-4}	highly stable
0.001	1.0×10^{-4}	highly stable

3. Conclusion

An extended time-fractional diffusion-reaction model based on the Caputo derivative was solved via the CD-BEM approach, yielding highly accurate and stable results across various $(\alpha, \lambda, \Delta t)$. The method remained robust for small fractional orders and weak reactions, supporting reliable simulation of transport-decay phenomena in complex media. This framework can be further adapted for multidimensional or nonlinear fractional systems.

References

1. I. Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. Elsevier, 1998.
2. R. Gorenflo and F. Mainardi, "Integral and differential equations of fractional order," in *Fractals and Fractional Calculus in Continuum Mechanics*, 1977, pp. 223–276.
3. J. A. M. Carrer, B. S. Solheid, J. Trevelyan, and M. Seaid, "A boundary element method formulation based on the Caputo derivative for the solution of the anomalous diffusion problem," *Engineering Analysis with Boundary Elements*, vol. 122, pp. 132–144, 2021.
4. J. D. Murray, *Mathematical Biology: I. An Introduction*. Springer, 2007.
5. K. Seki, M. Wojcik, and M. Tachiya, "Fractional reaction–diffusion equation," *The Journal of Chemical Physics*, vol. 119, no. 4, pp. 2165–2170, 2003.
6. C. A. Brebbia, J. C. F. Telles, and L. C. Wrobel, *Boundary Element Techniques: Theory and Applications in Engineering*. Springer, 2012.

A local RBF-FD scheme for the two-dimensional fractional integro-differential equation with weakly singular kernel

Samira Eslami[1,*], Mohammad Ilati[2]

[1]Department of Applied Mathematics, Faculty of Basic Sciences, Sahand University of Technology, Tabriz, Iran.
Email: s_eslami98@sut.ac.ir
Email: ilati@sut.ac.ir

Abstract. This paper presents a numerical solution for a two-dimensional fractional integro-differential equation featuring a weakly singular kernel. The temporal discretization is achieved through a finite difference technique combined with a second-order approximation for the integral, while spatial discretization relies on the local radial basis function-finite difference (RBF-FD) approach. The efficiency and convergence of the method are confirmed by a numerical example.

Keywords: fractional integro-differential equations, weakly singular kernel, RBF-FD method, finite difference method

AMS Mathematics Subject Classification [2020]: 65M22, 65R20, 35R11

1. Introduction

In recent decades, extensive research has focused on understanding complex phenomena. Fractional differential equations offer a powerful framework for modeling such processes due to their simplicity and ability to capture intricate dynamics with high accuracy. As a result, they have been widely applied in various fields, including thermal systems, image denoising, finance, and many others. In this study, the two-dimensional fractional integro-differential equation featuring a weakly singular kernel is investigated in the following form [**?**]

$$v_t(\mathbf{x},t) - \gamma\Delta v(\mathbf{x},t) = I^{\alpha}\Delta v(\mathbf{x},t) + g(\mathbf{x},t), \qquad (\mathbf{x},t) \in \Omega_T = \Omega \times (0,T] \tag{1}$$

with the following initial and boundary conditions

$v(\mathbf{x},0) = \varphi(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad v(\mathbf{x},t) = 0, \quad (\mathbf{x},t) \in \partial\Omega \times (0,T]$, where γ is a positive constant and I^{α} denotes the Riemann–Liouville fractional integral operator, which is defined as follows: $I^{\alpha}v(t) = \int_0^t \eta(t-s)\, v(s)\, ds$, where $\eta(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$, $0 < \alpha < 1$. Throughout the paper, we assume that the exact solution v of problem (1.1) in Ω_T satisfies the following regularity conditions: $\|v_t(\cdot,t)\| \le Ct^{\alpha}$, $\|v_{tt}(\cdot,t)\| \le Ct^{\alpha-1}$, $\|v_{xxyy}(x,y,\cdot)\| \le C$, $0 < C < \infty$. The regularity condition describes how

*Speaker.

the second derivative of the solution with respect to time behaves near the initial moment. It indicates that this derivative becomes weakly singular as time approaches zero, meaning that the solution does not possess full smoothness at the beginning of the time interval. In other words, the solution is not twice continuously differentiable in time, which represents a more realistic and less restrictive assumption compared with those typically adopted in earlier studies.Equation of the form (1) can be regarded as prototype problem that arise in areas such as heat conduction in materials with memory, population dynamics, and viscoelasticity, among others.

2. RBF-FD Method

In the global radial basis function (RBF) approach, the coefficient matrix becomes large, dense, and ill-conditioned as the number of interpolation points increases. To alleviate this issue, local RBF-based methods have been developed. These local approaches extend the classical finite difference (FD) method to scattered node distributions. Since the FD weights are obtained using RBF interpolation, the technique is known as the RBF-FD method [3]. In this method, the local approximation of a function $u(\mathbf{x})$ is expressed as $u(\mathbf{x}) = \sum_{j=1}^{s} \alpha_j \, \phi(\|\mathbf{x} - \mathbf{x}_j\|)$, where $\phi(\cdot)$ is a chosen radial basis function and $\{\mathbf{x}_j\}_{j=1}^{s}$ represents the local stencil of s neighboring nodes. The coefficients α_j are determined such that the interpolation conditions are satisfied at the stencil nodes.

The action of a linear differential operator L at the center $\mathbf{x}_c$ can be approximated as a weighted linear combination of the function values at these nodes: $Lu(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_c} = \sum_{j=1}^{s} \varpi_j \, u(\mathbf{x}_j)$, where ϖ_j are the unknown RBF-FD weights.

To compute the weights, one solves the following local linear system:

$$\underbrace{\begin{bmatrix} \phi(\|\mathbf{x}_1 - \mathbf{x}_1\|) & \phi(\|\mathbf{x}_1 - \mathbf{x}_2\|) & \dots & \phi(\|\mathbf{x}_1 - \mathbf{x}_s\|) \\ \phi(\|\mathbf{x}_2 - \mathbf{x}_1\|) & \phi(\|\mathbf{x}_2 - \mathbf{x}_2\|) & \dots & \phi(\|\mathbf{x}_2 - \mathbf{x}_s\|) \\ \vdots & \vdots & & \vdots \\ \phi(\|\mathbf{x}_s - \mathbf{x}_1\|) & \phi(\|\mathbf{x}_s - \mathbf{x}_2\|) & \dots & \phi(\|\mathbf{x}_s - \mathbf{x}_s\|) \end{bmatrix}}_{A} \begin{bmatrix} \varpi_1 \\ \varpi_2 \\ \vdots \\ \varpi_s \end{bmatrix} = \begin{bmatrix} L\phi(\|\mathbf{x} - \mathbf{x}_1\|)\big|_{\mathbf{x}=\mathbf{x}_c} \\ L\phi(\|\mathbf{x} - \mathbf{x}_2\|)\big|_{\mathbf{x}=\mathbf{x}_c} \\ \vdots \\ L\phi(\|\mathbf{x} - \mathbf{x}_s\|)\big|_{\mathbf{x}=\mathbf{x}_c} \end{bmatrix}.$$

To enhance the accuracy and ensure polynomial reproduction, the RBF interpolation can be augmented by a polynomial of degree q. In this case, the RBF-FD approximation becomes:

$$Lu(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_c} = \sum_{j=1}^{s} \varpi_j \, u(\mathbf{x}_j) + \sum_{i=1}^{(q+1)(q+2)/2} c_i P_i(\mathbf{x}),$$

where $P_i(\mathbf{x})$ are the polynomial basis functions and c_i are their coefficients.

For example, when $q = 1$, the augmented system that provides the weights $\varpi_1, \dots, \varpi_s$ and coefficients c_1, c_2, c_3 takes the block form:

$$\begin{bmatrix} & & & \begin{bmatrix} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_s & y_s \end{bmatrix} \\ & A & & \\ & & & \\ \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_s \\ y_1 & \cdots & y_s \end{bmatrix} & & & 0 \end{bmatrix} \begin{bmatrix} \varpi_1 \\ \vdots \\ \varpi_s \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} L\phi(\|\mathbf{x} - \mathbf{x}_1\|)\big|_{\mathbf{x}=\mathbf{x}_c} \\ \vdots \\ L\phi(\|\mathbf{x} - \mathbf{x}_s\|)\big|_{\mathbf{x}=\mathbf{x}_c} \\ L1\big|_{\mathbf{x}=\mathbf{x}_c} \\ Lx\big|_{\mathbf{x}=\mathbf{x}_c} \\ Ly\big|_{\mathbf{x}=\mathbf{x}_c} \end{bmatrix}.$$

This augmented formulation improves the stability and accuracy of the RBF-FD approximation, especially for scattered and irregular node distributions.

3. Temporal discretization

In this section, we introduce some notations such as $t_k = k\tau, \quad k = 1, \ldots, N, \tau = \frac{T}{N}$, and $\delta_t v^k = \frac{v^k - v^{k-1}}{\tau}, \quad v^k = v(x, y, t_k)$ to discretize the time variable. Moreover, the following lemma is employed to approximate the integral term.

LEMMA 1. *[1] Let $v(t) \in C^2[0, T]$. So, there is a positive constant $\mathcal{C}$ depends only on $0 < \eta < 1$ such that*

$$\left| \int_0^{t_k} (t_k - \varsigma)^{\eta-1} \, v(\varsigma) d\varsigma - \sum_{r=0}^{k} p_{k-r,k}^k \, v(t_{k-r}) \right| \leqslant \mathcal{C} \max_{0 \leqslant t \leqslant T} \left| v''(t) \right| \; t_k^{\eta} \; \tau^2, \quad 1 \leqslant k \leqslant N,$$

where

$$(2) \quad p_{r,k}^k = \frac{\tau^\eta}{\eta(\eta+1)} \times \begin{cases} (k-1)^{\eta+1} - (k-1-\eta) \; k^\eta, & r = 0, \\ (k-r+1)^{\eta+1} - 2 \; (k-r)^{\eta+1} + (k-r-1)^{\eta+1}, & 1 \leqslant r \leqslant k-1, \\ 1, & r = k. \end{cases}$$

Using the above notations and lemma, the time-discrete scheme for Equation (1) is derived as follows:

$$(3) \quad \frac{v^k - v^{k-1}}{\tau} - \gamma \Delta v^k - \frac{1}{\Gamma(\eta)} \sum_{r=0}^{k} p_{k-r,k}^k \; \Delta v^{k-r} = g^k + R^k,$$

then (3) can be written as:

$$(4) \quad v^k - v^{k-1} - \tau\gamma \Delta v^k - \frac{\tau}{\Gamma(\eta)} \left(p_{k,k}^k \; \Delta v^k + \sum_{r=1}^{k} p_{k-r,k}^k \; \Delta v^{k-r} \right) = \tau g^k + R^k,$$

where $\left| R^k \right| \leqslant C\tau^2$, and $C > 0$ is a constant independent of τ. By neglecting the small term R_k, the following result is obtained.

$$\hat{v}^k - \tau\gamma \Delta \hat{v}^k - \frac{\tau}{\Gamma(\eta)} p_{k,k}^k \Delta \hat{v}^k = \hat{v}^{k-1} + \frac{\tau}{\Gamma(\eta)} \sum_{r=1}^{k} p_{k-r,k}^k \Delta \hat{v}^{k-r} + \tau g^k.$$

Now, the RBF-FD method is applied to discretize the above time-stepping equation. To this end, the computational domain Ω is discretized using m scattered nodes $\{\mathbf{x}_d\}_{d=1}^m = \{\mathbf{x}_d\}_{d=1}^{m_1} \cup \{\mathbf{x}_d\}_{d=m_1+1}^m$. Let $\vartheta = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_s}\}$ denote a local stencil consisting of the center point $\mathbf{x}_i$ and its s nearest neighboring nodes. At each node $\mathbf{x}_i \in \Omega$, the approximate RBF-FD representations of $\hat{v}(\mathbf{x}_i, t)$ and its Laplacian $\Delta \hat{v}(\mathbf{x}_i, t)$ are given by

$$\hat{v}_i^k \approx \sum_{j \in \vartheta} \varpi_j \hat{v}_j^k, \quad \Delta \hat{v}_i^k \approx \sum_{j \in \vartheta} (\varpi_{xx,j} + \varpi_{yy,j}) \, \hat{v}_j^k.$$

Substituting these approximations into the discrete equation yields the following RBF-FD formulation:

$$\sum_{j \in \vartheta} \varpi_j \hat{v}_j^k - \tau\gamma \sum_{j \in \vartheta} (\varpi_{xx,j} + \varpi_{yy,j}) \hat{v}_j^k - \frac{\tau}{\Gamma(\eta)} p_{k,k}^k \sum_{j \in \vartheta} (\varpi_{xx,j} + \varpi_{yy,j}) \hat{v}_j^k = \sum_{j \in \vartheta} \varpi_j \hat{v}_j^{k-1}$$

$$+ \frac{\tau}{\Gamma(\eta)} \sum_{r=1}^{k} p_{k-r,k}^k \sum_{j \in \vartheta} (\varpi_{xx,j} + \varpi_{yy,j}) \hat{v}_j^{k-r} + \tau g_i^k.$$

This relation provides the RBF-FD-based discrete formulation of the given equation at each node $\mathbf{x}_i \in \Omega$.

4. Numerical results

To verify the accuracy and convergence of the proposed method, we consider the following exact solution: $u(x, y, t) = \frac{4t^{\alpha+1}}{3\sqrt{\pi}} \sin(2\pi x) \sin(2\pi y)$. The source term $g(\mathbf{x}, t)$, as well as the initial and boundary conditions, are obtained by substituting this exact solution into the governing equation. In this test, the parameters are chosen as $\gamma = 1$ and $\alpha = 0.75$, with final time $T = 1$. The spatial convergence order is evaluated by refining the number of spatial nodes m with a fixed time step. The results and CPU times are reported in Table 1. The computational orders are calculated by $Rate = \frac{\log_{10}(E_1/E_2)}{\log_{10}(h_1/h_2)}$, where E_1 and E_2 are errors for h_1 and h_2 , respectively.

TABLE 1. Errors and convergence orders with $\tau = \frac{1}{100}$.

m	$\|E_u\|_\infty$	Rate	CPU-time (s)
8	1.1441×10^{-3}	–	0.0714
16	2.7045×10^{-4}	2.0808	0.3300
32	6.0935×10^{-5}	2.1500	3.9354
64	1.6152×10^{-5}	1.9156	132.70

5. Conclusion

An efficient numerical scheme was presented for solving a two-dimensional fractional integro-differential equation with a weakly singular kernel. The method combined a finite difference technique in time with a second-order approximation for the singular integral term. For spatial discretization, the local radial basis function-finite difference (RBF-FD) approach was employed. A numerical example verified the accuracy and convergence of the proposed method, demonstrating its effectiveness as a reliable approach for this class of problems.

References

1. Diethelm, K., Ford, N. J., & Freed, A. D. (2004) *Detailed error analysis for a fractional Adams method*, Numer. Algorithms, **36**, 31–52.
2. Qiao, L. and Xu, D. (2021) *A fast ADI orthogonal spline collocation method with graded meshes for the two-dimensional fractional integro-differential equation*, Adv. Comput. Math., **47**, 1–22.
3. Sarra, S. A. (2012) *A local radial basis function method for advection–diffusion–reaction equations on complexly shaped domains*, Appl. Math. Comput., **218**(19), 9853–9865.

APPLICATION OF LOCAL MESHLESS MOVING KRIGING METHOD FOR SOLVING 2D STOCHASTIC ADVECTION–DIFFUSION EQUATIONS

Zahra Jeihouni[1,*] and Mohammad Ilati[1]

[1]Department of Applied Mathematics, Faculty of Basic Sciences, Sahand University of Technology, Tabriz, Iran.
Email: z_jeihounikozekanan402@sut.ac.ir

[1]Department of Applied Mathematics, Faculty of Basic Sciences, Sahand University of Technology, Tabriz, Iran.
Email: ilati@sut.ac.ir

Abstract. In this article, a local meshless technique is investigated for solving the two-dimensional stochastic advection–diffusion equation. The time discretization of the equation is done by using the Crank-Nicholson method, and then a local meshless moving Kriging method is applied in the space direction. At the end, a numerical example is presented to show the accuracy and efficiency of the method.

Keywords: Stochastic advection–diffusion equation, Moving Kriging interpolation, Meshless method, Crank-Nicolson scheme, Brownian motion

AMS Mathematics Subject Classification [2020]: 60H15,65M99

1. Introduction and Preliminaries

Partial differential equations (PDEs) have been widely used to model many problems in applied sciences and engineering. For example, this occurs in advection–diffusion models arising in ground water flows where exact knowledge of the permeability of the soil, magnitude of source terms, inflow or outflow conditions are exactly not known. The existence of uncertainties in such problems can be described by random fields. This requires to include, in the PDEs modeling, a rational assessment of uncertainty. Consequently, this leads to the notion of stochastic PDEs. In this work, we numerically investigate the linear stochastic advection–diffusion equation which can be formulated as follows [1]:

$$du + (\nu \cdot \nabla u - \kappa \Delta u - f)dt = \sigma dW(t), \tag{1}$$

where κ and ν are positive constants. Here, $W(t)$ is a Wiener process (Brownian motion) with

$$E[W(t)] = 0, \qquad E[W(t)W(s)] = \min(t, s).$$

*Speaker.

Let $\delta W_k = W(t_k) - W(t_{k-1})$, then

$$E[\delta W_k] = 0, \qquad E[(\delta W_k)^2] = \tau,$$

introducing a random perturbation with zero mean and variance at each time step.

In the spatial domain, the stochastic term is modeled as a Gaussian random field:

$$\xi(x) = \sigma \delta W_k(x),$$

with

$$E[\xi(x), \xi(y)] = \sigma^2 \tau q(x, y).$$

In this work, we employ

$$q(x, y) = \frac{1}{(|x - y|^2 + 1)^2}.$$

The stochastic forcing can be represented by the random vector

$$\xi_x = (\xi_1, \xi_2, \ldots, \xi_N) \sim \mathcal{N}(0, Q), \quad Q = (\sigma^2 \tau q(x_i, x_j))_{i,j=1}^N.$$

2. Moving Kriging (MK) Interpolation

In this section, the construction of meshless MK interpolation is described. The MK approximation of $u(x)$ can be written as [2, 3]

$$u_h(x) = \sum_{j=1}^{m} p_j(x) a_j + z(x) = \mathbf{p}^T(x)\mathbf{a} + z(x), \tag{2}$$

where $p_j(x)$ are the monomial basis functions, a_j are their corresponding coefficients, and $z(x)$ is a realization of a stochastic process with zero mean, variance σ^2, and nonzero covariance. The covariance matrix of $z(x)$ is defined by

$$\text{cov}\{z(x_i), z(x_j)\} = \sigma^2 R[\eta(x_i, x_j)]. \tag{3}$$

A Gaussian function is chosen as the correlation function, defined by

$$\eta(x_i, x_j) = e^{-\theta r_{ij}^2}, \quad r_{ij} = |x_i - x_j|, \quad \theta = \frac{\omega}{h^2}, \tag{4}$$

where ω is a constant.

Equation (2) can then be rewritten in matrix form as

$$u_h(x) = \mathbf{p}(x)\Psi\mathbf{u} + \mathbf{r}(x)\Gamma\mathbf{u}, \tag{5}$$

with

$$\Psi = (P^T R^{-1} P)^{-1} P^T R^{-1}, \qquad \Gamma = R^{-1}(I - P\Psi), \tag{6}$$

$$\mathbf{r}(x) = [\eta(x_1, x), \eta(x_2, x), \ldots, \eta(x_n, x)], \quad \mathbf{p}(x) = [p_1(x), p_2(x), \ldots, p_m(x)], \tag{7}$$

$$P = \begin{bmatrix} \mathbf{p}(x_1) \\ \vdots \\ \mathbf{p}(x_n) \end{bmatrix}, \qquad R = \begin{bmatrix} \mathbf{r}(x_1) \\ \vdots \\ \mathbf{r}(x_n) \end{bmatrix}, \tag{8}$$

and I is the identity matrix.
The moving Kriging shape functions are then obtained as

$$\Phi(x) = [\Phi_1(x), \ldots, \Phi_n(x)] = \mathbf{p}(x)\Psi + \mathbf{r}(x)\Gamma. \tag{9}$$

Finally, the approximation is given by

$$u_h(x) = \Phi(x)\mathbf{u} = \sum_{j=1}^{n} \phi_j(x) u_j. \tag{10}$$

3. Discretization Process

To discretize Eq. (1) in the temporal direction, let $t_k = k\tau$, $k = 0, 1, \ldots, N$, where $\tau = T/N$ is the time step size. The Crank–Nicolson method is used to obtain:

$$\frac{u^k - u^{k-1}}{\tau} = -\nu \cdot \left(\frac{\nabla u^k + \nabla u^{k-1}}{2}\right) + \kappa \left(\frac{\Delta u^k + \Delta u^{k-1}}{2}\right) + f^{k-\frac{1}{2}} + \sigma \delta W_k. \tag{11}$$

After simplification, we have

$$u^k + \frac{\tau\nu}{2}\left(\nabla u^k\right) - \frac{\tau\kappa}{2}\left(\Delta u^k\right) = u^{k-1} - \frac{\tau\nu}{2}\left(\nabla u^{k-1}\right) + \frac{\tau\kappa}{2}\left(\Delta u^{k-1}\right) + \tau f^{k-\frac{1}{2}} + \tau\sigma\delta W_k.$$

For spatial discretization, consider that the boundary points and interior points span the entire computational domain. By inserting the approximation

$$u^k(\mathbf{x}) = \sum_{j=1}^{n} \phi_j(\mathbf{x}) u_j^k, \tag{12}$$

into Eq. (11) and applying the collocation procedure at the interior points, the following discrete equations are obtained:

$$\begin{aligned} &\delta_\Xi\left[u^k\right] + \frac{\tau\nu}{2}\lambda_\Xi\left[u^k\right] - \frac{\tau\kappa}{2}\gamma_\Xi\left[u^k\right] = \delta_\Xi\left[u^{k-1}\right] - \frac{\tau\nu}{2}\lambda_\Xi\left[u^{k-1}\right] \\ &+\frac{\tau\kappa}{2}\gamma_\Xi\left[u^{k-1}\right] + \tau\delta_\Xi\left[f^{k-\frac{1}{2}}\right] + \tau\sigma\delta W_k, \end{aligned} \tag{13}$$

where δ_Ξ represents the point evaluation functional at $\mathbf{x}_\Xi$, and the functional γ_Ξ and λ_Ξ are defined by

$$\gamma_\Xi[u^k] := \gamma_\Xi[u^k(\mathbf{x})] = \sum_{j=1}^{n} \Delta\phi_j(\mathbf{x}_\Xi) u_j^k, \tag{14}$$

$$\lambda_\Xi[u^k] := \lambda_\Xi[u^k(\mathbf{x})] = \sum_{j=1}^{n} \nabla\phi_j(\mathbf{x}_\Xi) u_j^k. \tag{15}$$

4. Numerical Results

Consider Eq. (1) with the exact solution $u(x, y, t) = \exp(t)\cos(xy)(x+y)^2$. The initial condition, Dirichlet boundary conditions, and the source term f are derived from this exact solution. We solve this problem using the parameters $\nu = 1$, $\kappa = 1$, and $\sigma = 1$ and the *RMS-error* is computed using the statistical mean obtained from multiple realizations as follows:

$$\text{RMS-error} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left| E(u_j^k) - u(\mathbf{x}_j, t_k)\right|^2}$$

where $E(u_j^k)$ denotes the statistical mean obtained from multiple realizations at point $\mathbf{x}_j$ and time t_k, and $u(\mathbf{x}_j, t_k)$ represents the analytical expected solution.

The obtained results are reported in Table 1. It comes from this table, the present method are accurate than the method of [1].

TABLE 1. L_∞ and RMS errors for Test problem 1.

Realization	h	Method of [1]		Present Method	
		$\|E\|_\infty$	RMS	$\|E\|_\infty$	RMS
25	$\frac{1}{2}$	3.9033×10^{-1}	3.9033×10^{-1}	8.4071×10^{-2}	2.8024×10^{-2}
100	$\frac{1}{4}$	2.4696×10^{-1}	1.5403×10^{-1}	1.9618×10^{-2}	7.3031×10^{-3}
400	$\frac{1}{8}$	4.6090×10^{-2}	3.0809×10^{-2}	8.5318×10^{-3}	3.4759×10^{-3}
1600	$\frac{1}{16}$	2.3211×10^{-2}	8.9816×10^{-2}	3.3147×10^{-3}	1.6347×10^{-3}

5. Conclusions

In this paper, a local meshless method based on the Moving Kriging interpolation was successfully applied to solve two-dimensional stochastic advection–diffusion equation. The Crank–Nicolson scheme was used for time discretization, while the spatial discretization was performed using the Moving Kriging method. The numerical results are compared with those reported in [1]. A comparison shows that the results obtained by the proposed method exhibit higher accuracy than those presented in [1].

References

[1] Dehghan, M. and Shirzadi, M. (2015) *Meshless simulation of stochastic advection–diffusion equations based on radial basis functions*, Eng. Anal. Bound. Elem.,**53**, 18–26.

[2] Hidayat, M. I. P. (2023) *A meshfree approach based on moving Kriging interpolation for numerical solution of coupled reaction-diffusion problems*, Int. J. Comput. Methods, **20**(5), 2350002

[3] Ilati, M. (2020) *A meshless local moving Kriging method for solving Ginzburg–Landau equation on irregular domains*, Eur. Phys. J. Plus,**135**(11), 1–18.

Rational Radial Basis Functions for Solving Ordinary Differential Equations

Mansour Shiralizadeh*

Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m.shiralizadeh@pnu.ac.ir

Abstract. This paper presents an effective meshfree method for solving ordinary differential equations (ODEs) using rational radial basis functions (RRBFs). Traditional radial basis function (RBF) methods suffer from ill-conditioning and difficulties in handling steep gradients or singularities. To address these issues, we explore the use of rational forms of RBFs, which exhibit improved numerical stability and higher accuracy near singularities or boundary layers. The method is tested on one ODE with steep boundary layer. Numerical results demonstrate that RRBFs provide superior accuracy, particularly in regions with steep fronts or sharp gradients.

Keywords: Rational radial basis functions, Ordinary differential equations, Mesh-free methods, Numerical approximation

AMS Mathematics Subject Classification [2020]: 65D15, 65N35, 41A30,65L05

1. Introduction

Ordinary differential equations (ODEs) are fundamental in modeling physical, biological, and engineering systems. While analytical solutions exist for some ODEs, most require numerical techniques such as finite differences, finite elements, or spectral methods. Recently, mesh-free methods, particularly those based on radial basis functions (RBFs), have gained attention as powerful meshfree techniques for solving such problems due to their flexibility in handling scattered data and high-order smoothness. However, conventional RBFs (such as Gaussian or multiquadric bases) may suffer from ill-conditioning and loss of accuracy in regions with sharp variations, also when the shape parameter is small or when the solution has a steep gradient. To overcome these issues, Rational Radial Basis Functions (RRBFs) have been proposed, where the approximant is expressed as a ratio of two RBF expansions. This rational form yields better stability and local adaptability, especially in boundary-layer or singular perturbation problems [3]. In this paper, we develop and apply an RRBF collocation method for solving ODEs. A classic problem with a steep boundary layer is considered to test the method's accuracy.

*Speaker.

2. RBF interpolation and Rational RBF interpolation

2.1. RBF interpolation. Let $X = \{\boldsymbol{x}_1^c, ..., \boldsymbol{x}_N^c\}$ be a set of N distinct points, hereinafter referred to as centers, and $F = \{f(\boldsymbol{x}_1^c), ..., f(\boldsymbol{x}_N^c)\}$ a set of function values. A RBF $\phi(\boldsymbol{x}) = \phi\left(\|\boldsymbol{x} - \boldsymbol{x}^c\|_2\,,\,\epsilon\right)$ is a function of one variable $r = \|\boldsymbol{x} - \boldsymbol{x}^c\|_2$ that is centered at $\boldsymbol{x}^c$, which ϵ is a free parameter and it is known as the shape parameter [1, 3]. The Inverse quadratic RBF $\phi(r) = 1/1 + (\epsilon r)^2$ is a strictly positive definite RBF that we use it in numerical example. A RBF interpolant takes the form

$$s(\boldsymbol{x}) = \sum_{j=1}^{N} a_j\ \phi(\|\boldsymbol{x} - \boldsymbol{x}_j^c\|_2\,,\,\epsilon)$$

where the coefficients a_j are obtained by solving the linear system $B\mathbf{a} = \mathbf{f}$, based on the interpolation conditions $s\left(\boldsymbol{x}_i^c\right) = f_i$ where $\mathbf{f} = [f(\boldsymbol{x}_1^c), ..., f(\boldsymbol{x}_N^c)]^T$. The entries of the matrix B are of the form

$$b_{ij} = \phi(\|\boldsymbol{x}_i^c - \boldsymbol{x}_j^c\|_2\,,\,\epsilon),\ \ i, j = 1, ..., N.$$

B is a symmetric positive definite matrix and thus invertible. The evaluation of the interpolant at M points $\boldsymbol{x}_j$ is done by multiplying $\mathbf{a}$ by H where the entries of the evaluation matrix H are of the form

$$h_{ij} = \phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j^c\|_2\,,\,\epsilon),\ \ i = 1, ..., M,\ \ j = 1, 2, ..., N.$$

The first and second derivatives of RBF interpolant are of the form

$$D(s(\boldsymbol{x})) = \sum_{j=1}^{N} a_j\ D(\phi(\|\boldsymbol{x} - \boldsymbol{x}_j^c\|_2\,,\,\epsilon)),$$

thus $D(s(\boldsymbol{x}_i^c)) = \sum_{j=1}^{N} a_j\ D\phi(\|\boldsymbol{x}_i^c - \boldsymbol{x}_j^c\|_2\,,\,\epsilon)$, i.e. $D\mathbf{f} \simeq H_D\mathbf{a}$, where the entries of H_D are of the form $(H_D)_{ij} = D\phi(\|\boldsymbol{x}_i^c - \boldsymbol{x}_j^c\|_2\,,\,\epsilon),\ \ i, j = 1, ..., N.$ and $D(D(s(\boldsymbol{x}))) = \sum_{j=1}^{N} a_j\ D(D\phi(\|\boldsymbol{x} - \boldsymbol{x}_j^c\|_2\,,\,\epsilon))$, thus $D(D(s(\boldsymbol{x}_i^c))) = \sum_{j=1}^{N} a_j\ D(D\phi(\|\boldsymbol{x}_i^c - \boldsymbol{x}_j^c\|_2\,,\,\epsilon))$, i.e. $D(D\mathbf{f}\,) \simeq H_{DD}\mathbf{a}$, where the entries of H_{DD} are of the form $(H_{DD})_{ij} = D(D\phi(\|\boldsymbol{x}_i^c - \boldsymbol{x}_j^c\|_2\,,\,\epsilon)),\ \ i, j = 1, ..., N.$

2.2. Rational RBF interpolation. The RRBF interpolant of function f is of the form $\mathcal{R}(\boldsymbol{x}) = p(\boldsymbol{x})/q(\boldsymbol{x})$, which satisfies in the interpolation conditions $\mathcal{R}(\boldsymbol{x}_k^c) = f(\boldsymbol{x}_k^c),\ \ k = 1, 2, ..., N$ and $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ are the RBF interpolants $p(\boldsymbol{x}) = \sum_{j=1}^{N} a_j^p\ \phi(\|\boldsymbol{x} - \boldsymbol{x}_j^c\|_2\,,\,\epsilon)$, $q(\boldsymbol{x}) = \sum_{j=1}^{N} a_j^q\ \phi(\|\boldsymbol{x} - \boldsymbol{x}_j^c\|_2\,,\,\epsilon)$ By applying the interpolation conditions we have a system of equations that is underdetermined, thus in order for the rational interpolant to be uniquely defined, we add an additional condition (for more descriptions see [2, 3]), which leads the native space semi-norms [1] of the RBF interpolants $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ to be minimized. By adding the condition we will have a minimization problem with the solution $\mathbf{q}$ that is the eigenvector corresponding to the smallest eigenvalue problem $S\mathbf{q} = \lambda\mathbf{q}$ where

$$S = \operatorname{diag}\left(1/\left(\frac{\mathbf{f}^{\,2}}{\|\mathbf{f}\|_2^2} + 1\right)\right)\left(\frac{DB^{-1}D}{\|\mathbf{f}\|_2^2} + B^{-1}\right), \tag{1}$$

and B is the RBF system matrix, $D = \operatorname{diag}\left(f(\boldsymbol{x}_1^c), ..., f(\boldsymbol{x}_N^c)\right)$. Moreover, $\mathbf{f}^{\,2}$ is an elementwise squaring of the elements of the vector $\mathbf{f} = [f(\boldsymbol{x}_1^c), ..., f(\boldsymbol{x}_N^c)]^T$ and division is elementwise. When $\mathbf{q}$ is found, then the vector $\mathbf{p}$ is obtained by $\mathbf{p} = D\mathbf{q}$. When $\mathbf{p}$ and $\mathbf{q}$ are found, the expansion coefficients of the RBF interpolants are found by solving two

linear systems, $Ba^p = \mathbf{p}$, and $Ba^q = \mathbf{q}$. Now the rational interpolant at M points $\boldsymbol{x}_j$ is evaluated by $\mathbf{R} = \frac{Ha^p}{Ha^q}$ where $\mathbf{R} = [\mathcal{R}(\boldsymbol{x}_1), ..., \mathcal{R}(\boldsymbol{x}_M)]^T$, H is the RBF evaluation matrix, and division is elementwise.

Now, we calculate the first and second derivatives of the rational interpolant at N centers $\boldsymbol{x}_i^c$ by applying quotient rule as below

$$\mathbf{R}_1^{'} = \frac{(Ba^q).(H_D a^p) - (Ba^p).(H_D a^q)}{(Ba^q)^2}, \tag{2}$$

$$\mathbf{R}_1^{''} = \frac{2(Ba^p).(H_D a^q)^2 + (Ba^q)^2.(H_{DD} a^p)}{(Ba^q)^3} - \frac{(Ba^q).\Big(2\,(H_D a^p).(H_D a^q) + (Ba^p).(H_{DD} a^q)\Big)}{(Ba^q)^3}, \tag{3}$$

where $\mathbf{R}_1^{'} = [\mathcal{R}^{'}(\boldsymbol{x}_1^c), ..., \mathcal{R}^{'}(\boldsymbol{x}_N^c)]^T$, $\mathbf{R}_1^{''} = [\mathcal{R}^{''}(\boldsymbol{x}_1^c), ..., \mathcal{R}^{''}(\boldsymbol{x}_N^c)]^T$, B is the RBF system matrix, and H_D and H_{DD} are the first and second derivatives of evaluation matrix at N centers $\boldsymbol{x}_i^c$.

3. Main results

Now, we use the RRBF method to find the numerical solution of the ODEs. In fact, we consider a numerical example of the ODEs to validate the presented scheme.

Example 3.1. Consider the following problem

$$\varepsilon u''(x) + u'(x) = 0, \quad 0 \le x \le 1, \tag{4}$$

with $\varepsilon = 0.01$ and boundary conditions $u(0) = 0$, $u(1) = 1$. The exact solution is

$$u(x) = \frac{1 - e^{-x/\varepsilon}}{1 - e^{-1/\varepsilon}}. \tag{5}$$

We solve this problem with the RRBF method and inverse quadratic kernel with $N = 100$ uniformly spaced centers and use a shape parameter $\epsilon = 8$. The graph of approximate and exact solution are shown in Figure 1. It can be seen that the RRBF method resolves the problem accurately, also the results obtained by this method are in good agreement with exact solutions.

4. Conclusion

This paper demonstrated the effectiveness of rational radial basis functions for solving ordinary differential equations especially for ODEs with steep boundary layer also in cases that the solution of equation is a function with steep front or sharp gradients. The rational radial basis function collocation approach provided enhanced accuracy, particularly for boundary-layer problems. For the tested equation $\varepsilon u''(x) + u'(x) = 0$ with $\varepsilon = 0.01$, the RRBF solution exhibited excellent agreement with the exact analytical result.

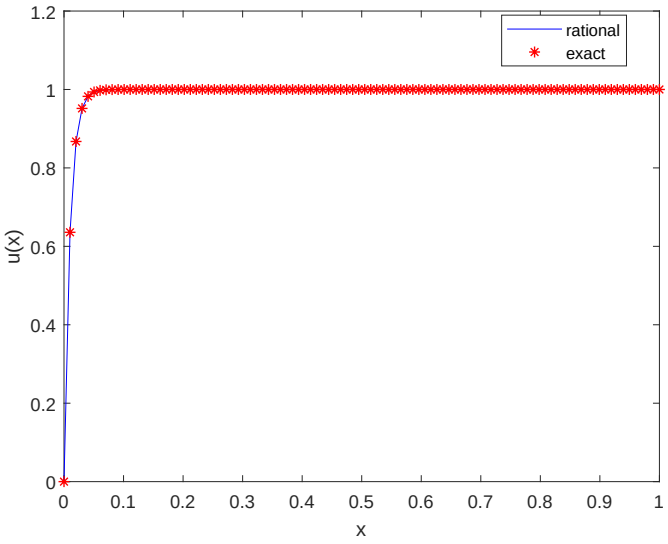


FIGURE 1. Exact solution and RRBF approximation for $\varepsilon = 0.01$.

References

1. Fasshauer. G.E, *Meshfree Approximation Methods with Matlab*, World Scientific, 2007.
2. Jakobsson. S, Andersson. B, Edelvik. F, *Rational radial basis function interpolation with applications to antenna design*, J. Comput. Appl. Math. **233** (2009), 889-904.
3. Sarra. S.A, Bai. Y, *A rational radial basis function method for accuratly resolving discontinuities and steep gradients*, Appl. Numer. Math. **130** (2018), 131-142.

A review of the various fields of wavelet applications

Sarkout Abdi[1],Aram Azizi[2*],Mahmoud Shafiee[3] and Jamshid Saeidian[4]

[1]Department of Mathematics, Ra.C., Islamic Azad University, Rasht, Iran.
Email: sarkout.abdi@iau.ac.ir

[2]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: a.azizi@pnu.ac.ir

[3]Department of Mathematics, Ra.C., Islamic Azad University, Rasht, Iran.
Email: Mahmoud.shafiee@iau.ac.ir

[4]Faculty of Mathematical Sciences and Computer , Kharazmi University , Tehran , Iran ,
Email: j.saeidian@khu.ac.ir

Abstract. over the last decade or so, wavelets have had a growing impact on signal processing theory and practice, both because of their unifying role and their successes in applications. Filter banks, which lie at the heart of wavelet-based algorithms, have become standard signal processing operators, used routinely in applications ranging from compression to modems. The contributions of wavelets have often been in the subtle interplay between discrete-time and continuous-time signal processing. The purpose of this article is to look at recent wavelet advances from a signal processing perspective. In particular, approximation results are reviewed, and the implication on compression algorithms is discussed. New constructions and open problems are also addressed.

Keywords: Filter Bank, Orthogonal, Compression, Random Wavelet.

AMS Mathematics Subject Classification [2020]: 58E11, 53B30, 53C50

1. Orthogonal Filter Banks

When thinking of filtering, one usually thinks about frequency selectivity. For example, an ideal discrete-time lowpass filter with cut-off frequency$\omega_c < \pi$takes any input signal and projects it onto the subspace of signals bandlimited to$[-\omega_c, \omega_c]$. Orthogonal discrete-time filter banks perform a similar projection which we now review. Assume a discrete-time filter with finite impulse response$g_0[n] = \{g_0[1], g_0[1], g_0[L-1]\}$,$L$ even, and the property

$$\langle g_0[n], g_0[n-2k]\rangle = \delta_k, \tag{1}$$

that is, the impulse response is orthogonal to its even shifts, and $\|g_0\|_2 = 1$.Denote by$G_0(z)$ the z-transform of the impulse response $g_0[n]$

$$G_0(z) = \sum_{n=0}^{L-1} g_0[n] z^{-n}, \tag{2}$$

*Speaker.

2. Discrete-Time Polynomials and Filter Banks

Signal processing specialists intuitively think of problems in terms of sinusoidal bases. Approximation theory specialists think often in terms of other series, like the Taylor series, and thus, of polynomials as basic building blocks. We now look at how polynomials are processed by filter banks. A discrete-time polynomial signal of degree Mis composed of a linear combination of monomial signals

$$\rho^{(m)}[n] = n^m, \qquad 0 \leq m \leq M. \tag{3}$$

3. Continuous-Time Polynomials and Wavelets

As is well known, a strong link exists between iterated filter banks and wavelets. For example, filter banks can be used to generate wavelet bases [1], and filter banks can be used to calculate wavelet series [2]. It comes thus as no surprise that the properties seen in discrete time regarding polynomial representation carry over to continuous time. While these properties are directly related to moment properties of wavelets and thus hold in general, we review them in the context of wavelets generated from orthogonal finite impulse response (FIR) filter banks. Assume again that the lowpass filter has$N$ zeros at $\omega = \pi$, and thus, the highpass has N zeros at$\omega = 0$. From the two scale relation of scaling function and wavelet, we get that the Fourier transform of the wavelet can be factored as

$$\Psi(w) = \frac{1}{\sqrt{2}} G_1(l^{jw/2}).\, \phi\left(\frac{w}{2}\right). \tag{4}$$

4. Discontinuities in Filter Bank and Wavelet Representations

What happens if a signal is discontinuous at some point t_0 ? We know that Fourier series do not like discontinuities, since they destroy uniform convergence. Wavelets have two desirable properties as far as discontinuities are concerned. First, they focus locally on the discontinuity as we go to finer and finer scales. That is because of the scaling relation of wavelets.

5. Compression of Piecewise Polynomial Signals

Let us return to one-dimensional piecewise smooth signals. Wavelets are well suited to approximate such signals when nonlinear approximation is allowed. To study compression behavior, consider the simpler case of piecewise polynomials, with discontinuities. To make matters easy, let us look again at the signal we used earlier to study nonlinear approximation, but this time include quantization and bit allocation. A simple analysis of the approximate rate distortion behavior of a step function goes as follows. Coefficients decay as$2^{m/2}$, so the number of scales J involved, if a quantizer of size Δ is used, is of the order of $\log_2(1/\Delta)$. The number of bits per coefficient is also of the order of $\log_2(1/\Delta)$, so the rate R is of the order

$$R \sim \log_2^2 (1/\Delta) \sim J^2 \tag{5}$$

6. Wavelet-like Transforms that Map Integers to Integers

Integer transforms are especially looks at the difference between the ‘‘true’’ odd$s_{j,2l+1}$ and the ‘‘predicted’’ odd $s_{j,2l+1}$ —this difference is the detail information d_j ; finally, one has to adjust the even $s_{j,2l}$ to correct for aliasing, leading to the $s_{j-1,l}$ for more details!.

A few examples are the Haar transform:
classically:

$$s_{j-1,l} = \frac{1}{\sqrt{2}}, (s_{j,2l} + s_{j,2l+1}), \qquad d_{j-1,l} = \frac{1}{\sqrt{2}}, (-s_{j,2l} + s_{j,2l+1}); \tag{6}$$

lifting:

$$\{s_{j,l}\} \;\rightarrow\; e^0_{j-1,l} = s_{j,2l}, \qquad o^0_{j-1,l} = s_{j,2l+1}$$

$$e^1_{j-1,l} = e^0_{j-1,l}, \qquad o^1_{j-1,l} = o^0_{j-1,l} - e^0_{j-1,l}$$

$$e^2_{j-1,l} = e^1_{j-1,l} + \tfrac{1}{2} o^1_{j-1,l}, \qquad o^2_{j-1,l} = o^1_{j-1,l}$$

$$s_{j-1,l} = \sqrt{2}\, e^2_{j-1,l}, \qquad d_{j-1,l} = \frac{1}{\sqrt{2}}\, o^2_{j-1,l}.$$

7. Wavelet Compression ≠ Karhunen-Loe‘ve Approximation

In theoretical models for the mathematical study of compression, signals and particularly images are often viewed as realizations of an unknown! stochastic process. The corresponding Karhunen- Loe‘ve KL! basis , as the orthonormal basis that optimally decorrelates this process. The basis $(\varphi_n)_{n\in N}$ that minimizes $\mathbb{E}\left(\left\|s - \sum_{n=1}^{N} \langle s, \varphi_n\rangle \varphi_n\right\|^2\right)$ for every N, is then viewed as the best possible basis on which to compress the signals or images. In practice, determining this KL basis exactly may be cumbersome and computationally intensive, suggesting the use of a basis that is easier to work with and that is still ‘‘close’’ to the KL basis, in the sense that it also decorrelates well although not optimally!. This has been argued as a justification both for direct current transform DCT! methods and for wavelet transforms.
Although the usefulness of KL bases is well documented and beyond dispute in many applications, there has been a growing realization that optimizing decorrelation for the stochastic process may not be the final or even the most important point in signal compression. In the terms of mathematical approximation theory, this corresponds to a shift from linear approximation to nonlinear approximation.

8. Wavelets for Nonuniformly Sampled Data

When wavelet bases are constructed via a lifting scheme, as described above, the computation of the wavelet coefficients consists of a prediction step for the ‘‘odds’’ from the ‘‘evens,’’ and a comparison of the true ‘‘odds’’ with these predictions. If the wavelet coefficients are zero, i.e., if we are considering a scaling function, then the predictions are exact at all levels: to build a scaling function for this scheme one thus needs only to iterate the prediction scheme level after level, generating an increasingly finer sampling of the scaling function through a subdivision scheme. This approach used, in fact, to plot all compactly supported wavelets and scaling functions in, e.g., is not limited to the case where the sampling points are uniformly distributed. Two types of nonuniform cases can be considered. In the semi-regular case, the original samples at level 0! are not equally spaced, but the subdivision scheme still introduces new grid points midway between old ones. This scheme is used in computer graphics applications, where subdivision is applied to generate smooth curves or surfaces. In the irregular case, new grid points need not be in the middle between old points, even at infinity.This irregular setting comes up

naturally in the case of compression of, or multiresolution analysis for, irregular samples. The user provides data, sampled on a closely spaced but irregular grid, which one can think of as the ''finest'' level grid. Resampling onto a regular grid is typically costly and may generate unwanted artifacts. One can then build a multiresolution analysis and an associated wavelet transform for the irregular grid, using the lifting scheme, leading to spatially variant filters.

9. Random Wavelets

***Theorem*1**.Suppose that $\varphi(x)$ is a bounded function with supp $\varphi(x) \subseteq [-a, a]$, $0 < a < +\infty$ and satisfies the following conditions:
(i) $\sum_{j=-\infty}^{\infty} \varphi(x-j) \equiv 1 \quad$ on $\mathbb{R}$;
(ii) there is a number b such that $\varphi(x)$ is non-decreasing if $x \leq b$ and is non-increasing if $x \geq b$.
Then, for $f \in C(\mathbb{R})$, if f is a non-decreasing function, the linear wavelet operators $A_k(f)$ defined by (1) are also non-decreasing functions on $\mathbb{R}$ and satisfy

$$\big|A_k(f)(x) - f(x)\big| \leq \omega\big(f, 2^{-k+1}\alpha\big), \qquad x \in \mathbb{R},\ k \in \mathbb{Z}, \tag{7}$$

where $\omega(f, h)$ is the modulus of continuity of f. Moreover, the inequalities (7) are sharp.
***Theorem*2**.Suppose that $\varphi(x)$ is a bounded right-continuous function with supp $\varphi(x) \subseteq [-a, a]$, $0 < a < +\infty$ Let $F(X)$ be a continuous disiribuiion function on R. Then the linear wavelet opemtors$A_k(F)$ defined by frmoula are distribution functions and satisfy

$$|A_k(F)(x) - F(x)| \leq \omega(F, 2^{-k+1}a), x \in R, k \in Z$$

Proof. From the assumption on φ and Lemma , it follows that $A_t(F)$ are right-continuous on R. Since $F(z)$ is non-decreasing and $\lim_{x\to+\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$ from Theorem and Lemma we know that $A_k(F)(r)$ are non-decreasing and $\lim_{x\to+\infty} A_k F(x) = 1$ and $\lim_{x\to-\infty} A_k F(x) = 0$ Hence At(F) are distribution functions on R. Theorem gives the desired estimates of $|A_k(F)(x) - F(x)|$. The example of $g(z)$ used for the sharpness of formoula in Theorem is a distribution function. Hence these estimates are still sharp for the distribution functions.

References

1. I. Daubechies(1988), Orthonormal bases of compactly supported wavelets, *Commun. Pure Appl. Math.*, **41** 909–996.
2. S. Mallat(1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Recognition Machine Intell.*, **11** 674–693.

Fibonacci-Based Spectral Method for Caputo Fractional PDEs

Shahed Mashhoodi[1,*], Esmail Babolian[2,†] and Mahmoud Shafiee[3]

[1]Department of Mathematics, Ra.C., Islamic Azad University, Rasht, Iran.
Email: shahed.mashhoodi@iau.ac.ir

[2]Department of Computer Science, Faculty of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran.
Email: babolian@khu.ac.ir

[3]Department of Mathematics, Ra.C., Islamic Azad University, Rasht, Iran.
Email: shafiee@iau.ac.ir

Abstract. This study introduces a spectral numerical method using the Fibonacci truncated series expansion to solve fractional-order differential equations involving Caputo derivatives. The unknown functions are expressed in finite series of Fibonacci polynomials with unknown coefficients, and operational matrices are applied to transform the problem into linear algebraic equations. In last section, the results of the numerical example demonstrate the accuracy and effectiveness of the applied method compared to exact solutions.

Keywords: Spectral method, Fibonacci polynomials, Fractional partial differential equations (FPDEs) system, Caputo derivative.

AMS Mathematics Subject Classification [2020]: 65M22, 35R11, 11B39.

1. Introduction

Fractional calculus (FC) generalizes differentiation and integration to non-integer orders and provides powerful tools for modeling complex physical and engineering phenomena [5, 6]. Since the introduction of the Caputo derivative [4], many numerical techniques have been developed for solving fractional equations [1, 2]. In this work, a Fibonacci collocation method is proposed to solve a two-dimensional system of fractional partial differential equations of the general form

$$(1)\qquad \begin{cases} {}_cD_x^{\alpha}u + u_x + h_1(u,v) = g_1(x,t), \\ {}_cD_t^{\beta}v + v_t + h_2(u,v) = g_2(x,t), \end{cases} \qquad (x,t)\in[0,1]\times[0,1], \quad 1<\alpha,\beta\le 2,$$

*Speaker.
†Corresponding.

where ${}_cD_x^\alpha$ and ${}_cD_t^\beta$ denote Caputo derivatives of order α and β, respectively, and h_1, h_2 are given linear functions. The initial conditions are

$$u(0,t) = u(x,0) = v(0,t) = v(x,0) = 0. \tag{2}$$

The proposed approach based on Fibonacci polynomials [3], offers accurate and efficient numerical approximations and can be extended to higher-dimensional systems.

2. Spectral Method

In this section, a numerical technique based on the Fibonacci collocation method is presented for solving a two-dimensional system of fractional partial differential equations (FPDEs) involving Caputo derivatives.

2.1. Fibonacci Polynomial Approximation. Let $\{F_i(x)\}_{i=0}^N$ be the Fibonacci polynomials defined recursively as

$$F_0(x) = 0, \quad F_1(x) = 1, \quad F_{n+1}(x) = xF_n(x) + F_{n-1}(x), \quad n \geq 1. \tag{3}$$

These polynomials form a suitable basis for function approximation on the interval $[0,1]$. The unknown functions $u(x,t)$ and $v(x,t)$ are approximated by truncated series:

$$u(x,t) \approx \sum_{i=0}^{N} a_i(t)F_i(x), \qquad v(x,t) \approx \sum_{i=0}^{N} b_i(t)F_i(x), \tag{4}$$

where $a_i(t)$ and $b_i(t)$ are unknown time-dependent coefficients.

Substituting these series into the FPDE system and applying the Caputo derivatives yield

$${}_cD_x^\alpha u(x,t) \approx \sum_{i=0}^{N} a_i(t)\,{}_cD_x^\alpha F_i(x), \qquad {}_cD_t^\beta v(x,t) \approx \sum_{i=0}^{N} \dot{b}_i(t)\,{}_cD_t^\beta F_i(x). \tag{5}$$

To simplify the computation, the derivatives of Fibonacci polynomials are expressed in matrix form as

$$\mathbf{D}_x^{(\alpha)}\mathbf{F}(x) = \begin{bmatrix} {}_cD_x^\alpha F_0(x) \\ {}_cD_x^\alpha F_1(x) \\ \vdots \\ {}_cD_x^\alpha F_N(x) \end{bmatrix} = \mathbf{M}_\alpha \mathbf{F}(x), \tag{6}$$

where $\mathbf{M}_\alpha$ is the operational matrix of fractional derivatives with respect to x. Similarly, for the time derivative, we have

$$\mathbf{D}_t^{(\beta)}\mathbf{F}(t) = \mathbf{M}_\beta \mathbf{F}(t). \tag{7}$$

2.2. Formation of the Algebraic System. By substituting the approximate expansions and the operational matrices into the FPDE system, we obtain

$$\mathbf{M}_\alpha \mathbf{A} + \mathbf{A}_x + \mathbf{H}_1(\mathbf{A},\mathbf{B}) = \mathbf{G}_1, \qquad \mathbf{M}_\beta \mathbf{B} + \mathbf{B}_t + \mathbf{H}_2(\mathbf{A},\mathbf{B}) = \mathbf{G}_2, \tag{8}$$

where $\mathbf{A} = [a_0, a_1, \ldots, a_N]^T$ and $\mathbf{B} = [b_0, b_1, \ldots, b_N]^T$ are the unknown coefficient vectors.

Next, we apply the collocation method by enforcing the above equations at discrete collocation points

$$x_i = \frac{i}{N}, \quad t_j = \frac{j}{N}, \qquad i,j = 0,1,\ldots,N. \tag{9}$$

At each point (x_i, t_j), the residuals of the FPDEs are set to zero, resulting in a system of algebraic equations that can be written as

$$\mathbf{T}\begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix} = \mathbf{G}, \tag{10}$$

where $\mathbf{T}$ is the global coefficient matrix obtained from the Fibonacci operational matrices and collocation conditions, and $\mathbf{G}$ is the known vector formed from $g_1(x,t)$ and $g_2(x,t)$.

2.3. Numerical Implementation. The algebraic system is solved using standard linear solvers (e.g., LU decomposition). The approximate solutions are reconstructed as

$$u_N(x,t) = \mathbf{A}^T\mathbf{F}(x), \qquad v_N(x,t) = \mathbf{B}^T\mathbf{F}(x). \tag{11}$$

The absolute error functions are defined as

$$E_u(x,t) = |u(x,t) - u_N(x,t)|, \qquad E_v(x,t) = |v(x,t) - v_N(x,t)|. \tag{12}$$

Numerical experiments indicate that the proposed Fibonacci collocation approach provides high accuracy with a small number of basis functions, demonstrating efficient convergence and computational stability for solving fractional PDE systems.

3. Numerical simulation

In this part, we apply the proposed technique to obtain approximate results for certain case involving fractional-order partial differential systems, as illustrated in the forthcoming example.

Example 3.1. Consider the system of fractional PDEs as follows [7]

$$\begin{cases} D_x^{\alpha}u(x,t) - u = (x - \frac{1}{6}x^3)\sin(2\pi t) \\ D_t^{\alpha}v(x,t) - v = (t - \frac{1}{6}t^3)\sin(2\pi x) \end{cases}, \tag{13}$$

taking $u(0,t) = u(x,0) = v(0,t) = v(x,0) = 0$ as the initial condition. The analytical solutions of the system for $\alpha = 2$ are $u(x,t) = \frac{1}{6}x^3\sin(2\pi t), v(x,t) = \frac{1}{6}t^3\sin(2\pi x)$.

The absolute errors of the system for $m = n = 4$ and $m = n = 6$ at $t = 0.6$ are listed in Table 1. In Figure 1, the green/yellow areas correspond to higher errors, since they appear at the top of the surface, while the red areas correspond to lower errors near the base. The analytical solutions are closely matched by the approximate solutions of the system, as evidenced by their strong convergence.

Table 1. Absolute errors at $t = 0.6$ for exapmle 2

t	x	error$(u(x,t))$		error$(v(x,t))$	
		$m = n = 4$	$m = n = 6$	$m = n = 4$	$m = n = 6$
0.6	0.1	1.27×10^{-5}	5.13×10^{-8}	6.49×10^{-3}	2.51×10^{-5}
	0.2	1.01×10^{-4}	8.53×10^{-8}	2.63×10^{-3}	8.17×10^{-6}
	0.3	3.41×10^{-4}	8.47×10^{-8}	1.98×10^{-3}	7.39×10^{-7}
	0.4	8.08×10^{-4}	3.23×10^{-8}	2.73×10^{-3}	2.81×10^{-7}
	0.5	1.58×10^{-3}	8.93×10^{-7}	4.26×10^{-14}	2.83×10^{-7}
	0.6	2.73×10^{-3}	2.97×10^{-7}	2.27×10^{-3}	2.05×10^{-7}
	0.7	4.33×10^{-3}	6.09×10^{-6}	1.98×10^{-3}	1.98×10^{-7}
	0.8	6.47×10^{-3}	1.04×10^{-6}	2.63×10^{-3}	9.40×10^{-6}
	0.9	9.21×10^{-3}	1.61×10^{-6}	6.49×10^{-3}	2.31×10^{-5}

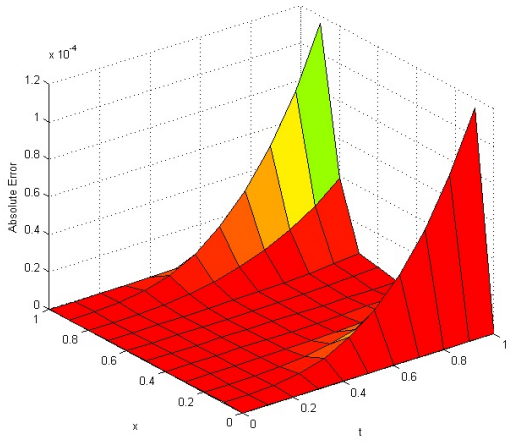


FIGURE 1. Absolute errors of the system for t=0.6

4. conclusion

A Fibonacci-based spectral collocation method was presented for solving fractional PDEs with Caputo derivatives. The approach transforms the problem into simple algebraic equations using Fibonacci operational matrices. Numerical results confirm its high accuracy, fast convergence, and computational efficiency. This method can be easily extended to more complex fractional systems.

5. Acknowledgment

This manuscript is prepared based on PhD thesis of first author at Rasht Branch, Islamic Azad University, Rasht, Iran.

References

1. Abdulazeez, S.T. and Modanli, M. (2022), *Solutions of fractional order pseudo-hyperbolic telegraph partial differential equations using finite difference method.* Alexandria Engineering Journal **61**, **12** 12443–12451.
2. Abd-El hamid, W.M.; Machado, J.A.T. and Youssri, H.Y.(2022), *Hyper geometric fractional derivatives formula of shifted Chebyshev polynomials: tau Algorithm for a type of fractional delay differential equations.* Walter de Gruyter GmbH, **23**, **7-8**.
3. Bednarz, U. and Musiał, M.W. (2020), *Distance Fibonacci Polynomials.* Symmetry **12**, **9**, 1540.
4. Caputo, M. (1967) *Linear model of dissipation whose Q is almost frequency independent-II*, Geophysical Journal of the Royal Astronomical Society, **13**, 529–539.
5. He, J.H. (1999), *Some applications of nonlinear fractional differential equations and their applications*, Bull. Sci. Technol. **15**, **2**, 86–90 .
6. He, J. (1998), *Nonlinear oscillation with fractional derivative and its applications*, Int. Conf. Vibr. Eng. **98**, 288–291 .
7. Zhao, F.; Huang, Q.; Xie, J.; Ma, Y.Li.L. and Wang, J. (2017), *Chebyshev polynomials approach for numerically solving system of two-dimensional fractional PDEs and convergence analysis*, Appl. Math. Comput. **313**, 321–330.

A spectral framework for numerical solution of the time-fractional heat conduction problems

Younes Talaei[1,*], Golnoosh Azizipour[2]

[1] Department of Engineering Sciences, Faculty of Advanced Technologies, University of Mohaghegh Ardabili, Namin, Iran.
Email: Talaei@uma.ac.ir

[2]Department of Mathematics, University of Mohaghegh Ardabili, Ardabil, Iran.
Email: g_azizipour@yahoo.com

Abstract. This paper introduces novel numerical method to solve time-fractional diffusion equations based on Lanczos spectral method. The method accounts for the non-smooth behavior of solutions over time by employing fractional-order canonical basis polynomials, thereby enhancing the convergence rate of the method.

Keywords: Spectral method, Fractional diffusion problems, Fractional canonical polynomials

1. Introduction

In this work, we consider the time-fractional diffusion equation

$$\partial_t^\alpha u(x,t) = \kappa u_{xx}(x,t) + g(x,t), \ (x,t) \in \Lambda, \tag{1}$$

with the initial condition $u(x,0) = f(x)$ and boundary condition $u(0,t) = u(l,t) = 0$, in which $\partial_t^\alpha u$ is Caputo type time-fractional derivative of order $\alpha \in (0,1)$ defined by $\partial_t^\alpha u(x,t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} u_s(x,s) ds$, and $\Lambda = [0,l] \times (0,T]$, κ is the diffusion coefficient and $g(x,t)$ is the source function. Here, $\Gamma(\cdot)$ is the gamma function.

2. Main Results

Given a set of $n+1$ distinct interpolation nodes $x_0, x_1, ..., x_n$ in $[0,l]$ and corresponding function values $y_0, y_1, ..., y_n$, the Barycentric Lagrange polynomials [3] are expressed as

$$\varphi_i(x) = \frac{\dfrac{w_i}{(x-x_i)}}{\displaystyle\sum_{k=0}^{n} \frac{w_k}{(x-x_k)}}; \quad w_i = \left(\prod_{\substack{k \neq i \\ k=0}}^{n} (x_i - x_k) \right)^{-1}, \quad i = 0, ..., n, \tag{2}$$

*Speaker.

in which $x_i = \frac{l}{2}\big(1 - cos(\frac{i\pi}{n})\big)$ are Chebyshev-Gauss-Lobatto (CGL) nodes on $[0, l]$ and w_i are the barycentric weights. The Barycentric Lagrange interpolation polynomial is expressed as

$$\mathcal{P}_n(x) = \sum_{i=0}^{n} y_i \varphi_i(x) = \mathbf{\Phi}_n^T(x)\mathbf{Y}, \quad \mathbf{Y} = [y_0, ..., y_n]^T, \tag{3}$$

where $\mathbf{\Phi}_n(x) = [\varphi_0(x), ..., \varphi_n(x)]^T$. The main advantages of Barycentric Lagrange interpolation include high accuracy, excellent stability, and the ability to circumvent the Runge phenomenon. The ℓ-th derivative of $\mathcal{P}_n(x)$ can be expressed in terms of its samples in the following form

$$y_n^{(\ell)}(x) = \mathbf{\Phi}_n^T(x)\mathbf{D}_n^{(\ell)}\mathbf{Y}, \tag{4}$$

where $\mathbf{D}_n^{(\ell)} = [d_{ij}^{(\ell)}]$ is the differentiation matrix with $d_{ij}^{(\ell)} := \varphi_j^{(\ell)}(x_i)$ for $i, j = 0, ..., n$. In [4] derived a very useful recursive formula

$$d_{ij}^{(\ell)} = \begin{cases} \dfrac{\ell}{x_i - x_j}\left(\dfrac{\lambda_j}{\lambda_i} d_{ii}^{(\ell-1)} - d_{ij}^{(\ell-1)}\right), & i \neq j \\ -\displaystyle\sum_{j=0, j\neq i}^{n} d_{ij}^{(\ell)}, & i = j \end{cases} \tag{5}$$

Now, consider to the Barycenteric Lagrange interpolation of the solution of the problem (1)

$$\widetilde{u}(x,t) = \sum_{i=1}^{M-1} \psi_i(t)\varphi_i(x) \tag{6}$$

with unknown time-dependent coefficients functions $\psi_i(t) = u(x_i, t)$. It is clear that $\widetilde{u}(0,t) = \widetilde{u}(l,t) = 0$. Define $U_M(t) := [\psi_1(t), ..., \psi_{M-1}(t)]^T$, we can rewrite Eq. (6) in the form

$$\widetilde{u}(x,t) = \widehat{\mathbf{\Phi}}_M^T(x)U_M(t). \tag{7}$$

Therefore, from (4) and (7), we get

$$\widetilde{u}_{xx}(x,t) \simeq \widehat{\mathbf{\Phi}}_M^T(x)\widehat{\mathbf{D}}_M^{(2)}U_M(t), \tag{8}$$

where $\widehat{\mathbf{\Phi}}_M(x) = [\varphi_1(x), ..., \varphi_{M-1}(x)]^T$ and $\widehat{\mathbf{D}}_n^{(2)}$ is derived by differentiation matrix $\mathbf{D}_n^{(2)}$ by removing the first and last rows and columns. Substituting Eqs. (7) and (8) into (1) leads to

$$\widehat{\mathbf{\Phi}}_M^T(x)\partial_t^\alpha U_M(t) \simeq \kappa^2 \widehat{\mathbf{\Phi}}_M^T(x)\widehat{\mathbf{D}}_M^{(2)}U_M(t) + g(x,t). \tag{9}$$

By evaluating Eq. (9) at the interior points x_k for $k = 1, ..., M-1$, we obtain the following system of differential equations

$$\partial_t^\alpha U_M(t) = \kappa^2 A U_M(t) + G(t); \quad U_M(0) = U_0, \tag{10}$$

where $A := \widehat{\mathbf{D}}_M^{(2)}$, $U_0 = [f(x_1), ..., f(x_{M-1})]^T$ and $G(t) = [g(x_1, t), ..., g(x_{M-1}, t)]^T$. By taking the fractional integrating of order α on both sides of (10) with respect to time from 0 to t, we obtain an equivalent system of linear Volterra integral equations

$$U_M(t) = U_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1} G(s)ds + \frac{\kappa^2}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1} A U_M(s)ds. \tag{11}$$

Rewriting (11), we obtain a system of Volterra integral equations in the form

$$U_M(t) = \mathbf{G}(t) + \frac{\kappa^2}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} A U_M(s) ds, \quad \mathbf{G}(t) = [\mathbf{G}_1(t), ..., \mathbf{G}_{M-1}(t)]^T, \tag{12}$$

with

$$\mathbf{G}_i(t) = f(x_i) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} g(x_i, s) ds, \quad i = 1, ...M-1. \tag{13}$$

The Lanczos Tau method aims to find an exact polynomial solution to the perturbed problem by introducing a perturbation term to the right-hand side of the original problem [1]. Here, in order to cover the low order of convergence of the spectral Tau method for problem (12), we investigate the approximate solution $U_{M,N}(t)$ is the fractional order vector space as

$$U_{M,N}(t) = [\psi_{1,N}(t), ..., \psi_{M-1,N}(t)]^T$$

where $\psi_{i,N}(t) \in \mathcal{M}_{N,\alpha} := span\{1, t^\alpha, t^{2\alpha}, ..., t^{N\alpha}\}$ which represents the exact solution to the perturbed problem

$$\mathbf{L}U_{M,N}(t) = \mathbf{G}_N(t) + \mathcal{T}(t). \tag{14}$$

Here,

$$\mathbf{L}U_M(t) = U_M(t) - \frac{\kappa^2}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} A U_M(s) ds, \tag{15}$$

$$\mathbf{G}_N(t) = [\mathbf{G}_{1,N}(t), ..., \mathbf{G}_{M-1,N}(t)]^T, \quad \mathcal{T}_N(t) = [\mathcal{T}_{1,N}(t), ..., \mathcal{T}_{M-1,N}(t)]^T.$$

$\mathbf{G}_{i,N}(t)$ are fractional orthogonal projections of $\mathbf{G}_i$ using fractional Legendre polynomials

$$\mathbf{G}_{i,N}(t) := \mathbf{\Pi}_N^\alpha \mathbf{G}_i(t) = \sum_{j=0}^{N} \widehat{g}_{i,j} P_j(2(\frac{t}{T})^\alpha - 1) = \sum_{j=0}^{N} g_{i,j} t^{j\alpha} \in \mathcal{M}_{N,\alpha}. \tag{16}$$

$P_i(t)$ denotes the standard Legendre polynomials on $[-1, 1]$, orthogonal with respect to the weight function $w(t) = \alpha t^{\alpha-1}$. The perturbation terms are in the from $\mathcal{T}_{i,N}(t) = \tau_{i,N} p_{N+1}(t^\alpha)$, where $p_i(t^\alpha) = \sqrt{\frac{(2i+1)\alpha}{T^\alpha}} P_i(t^\alpha)$ is orthonormal fractional Legendre polynomial on $[0, 1]$.

Definition 2.1. The vector canonical polynomial $\mathbf{Q}_i^j(t)$ is called the i-th vector canonical polynomial of degree $r\alpha$ associated with the linear operator $\mathbf{L}$ defined in (15), if

$$\mathbf{L}(\mathbf{Q}_i^j(t)) = t^{j\alpha}\mathbf{e}_i, \quad j \in \{0, 1, ..., \}, \quad i = 1, ..., M-1, \tag{17}$$

where $\mathbf{e}_i$ is the i-th column of the identity matrix $\mathbf{I}$ of dimension $(M-1)$.

Now, we aim to derive a recurrence relation to generate the canonical polynomials $\mathbf{Q}_i^j(t)$ related to linear operator (15). Given that

$$\mathbf{L}(t^{j\alpha}\mathbf{e}_i) = t^{j\alpha}\mathbf{e}_i - \frac{\kappa^2\Gamma(j\alpha+1)}{\Gamma((j+1)\alpha+1)} A t^{(j+1)\alpha}\mathbf{e}_i = \mathbf{L}\left(\mathbf{Q}_i^j(t) - \frac{\kappa^2\Gamma(j\alpha+1)}{\Gamma((j+1)\alpha+1)} A\mathbf{Q}_i^{j+1}(t)\right) \tag{18}$$

we obtain

$$A\mathbf{Q}_i^{j+1}(t) = \frac{\Gamma((j+1)\alpha+1)}{\kappa^2\Gamma(j\alpha+1)} \left(\mathbf{Q}_i^j(t) - t^{j\alpha}\mathbf{e}_i\right). \tag{19}$$

Define

$$\mathbf{Q}_*^j(t) := [\mathbf{Q}_1^j(t), \mathbf{Q}_2^j(t), ..., \mathbf{Q}_{M-1}^j(t)]^T. \tag{20}$$

It follows straightforwardly that $\mathbf{L}(\mathbf{Q}_*^j(t)) = t^{j\alpha}\mathbf{I}$ and $\mathbf{Q}_*^{j+1}(t) = \frac{\Gamma((j+1)\alpha+1)}{\kappa^2\Gamma(j\alpha+1)}\left(\mathbf{Q}_*^j(t) - t^{j\alpha}\mathbf{I}\right)A^*$ in which $A^* = (A^{-1})^T$. Since $\mathbf{Q}_*^0(t)$ cannot be generated from relation (19), we formulate this recurrence relation as

$$\mathbf{Q}_*^j(t) = \mathcal{Q}_*^j(t) + \mathcal{E}_j\mathbf{Q}_*^0(t),$$

where $\mathcal{Q}^0(t) = [\mathbf{0}, \mathbf{0}, ..., \mathbf{0}]$ and $\mathcal{E}_0 = \mathbf{I}$. Therefore, we obtain

$$\mathcal{Q}_*^{j+1}(t) = \frac{\Gamma((j+1)\alpha+1)}{\kappa^2\Gamma(j\alpha+1)}\left(\mathcal{Q}_*^j(t) - t^{j\alpha}\mathbf{I}\right)A^*; \quad \mathcal{E}_*^{j+1} = -\frac{\Gamma((j+1)\alpha+1)}{\kappa^2\Gamma(j\alpha+1)}\mathcal{E}_*^j A^*.$$

Let $p_{N+1}(t^\alpha) = \sum_{\ell=0}^{N+1} c_\ell t^{\ell\alpha}$. By reformulating the right hand side of (14), we obtain

$$\mathbf{L}U_{M,N}(t) = \sum_{i=1}^{M-1}\sum_{j=0}^{N} g_{i,j}t^{j\alpha}\mathbf{e}_i + \sum_{i=1}^{M-1}\tau_{i,N}p_{N+1}(t^\alpha)\mathbf{e}_i = \mathbf{L}\left(\sum_{j=0}^{\sigma}\mathcal{G}_{j,N}^T\mathbf{Q}_*^j(t) + \tau_N^T\Big(\sum_{\ell=0}^{N+1}c_\ell\mathbf{Q}_*^\ell(t)\Big)\right)$$

where $\mathcal{G}_{j,N} := [f_{1,j}, ..., f_{M-1,j}]^T$ and $\tau_N := [\tau_{1,N}, ..., \tau_{M-1,N}]^T$. Therefore,

$$\begin{aligned} U_{M,N}^T(t) &= \sum_{j=0}^{\sigma}\mathcal{G}_{j,N}^T\mathbf{Q}_*^j(t) + \tau_N^T\Big(\sum_{\ell=0}^{N+1}c_\ell\mathbf{Q}_*^\ell(t)\Big) \\ &= \sum_{j=0}^{\sigma}\mathcal{G}_{j,N}^T\left[\mathcal{Q}_*^j(t) + \mathcal{E}_j\mathbf{Q}_*^0(t)\right] + \tau_N^T\left[\sum_{\ell=0}^{N+1}c_\ell\mathcal{Q}_*^\ell(t) + \sum_{\ell=0}^{n+1}c_\ell\mathcal{E}_\ell\mathbf{Q}_*^0(t)\right] \\ &= \sum_{j=0}^{\sigma}\mathcal{G}_{j,N}^T\mathcal{Q}_*^j(t) + \tau_N^T\Big(\sum_{\ell=0}^{N+1}c_\ell\mathcal{Q}_*^\ell(t)\Big) + \left[\sum_{j=0}^{\sigma}\mathcal{G}_{j,N}^T\mathcal{E}_j + \tau_N^T\Big(\sum_{\ell=0}^{N+1}c_\ell\mathcal{E}_\ell\Big)\right]\mathbf{Q}_*^0(t) \end{aligned} \tag{21}$$

The parameters τ_i can be found by setting the coefficients of $\mathbf{Q}_i^0(t)$ for $i = 1, ..., M-1$ to zero

$$\Big(\sum_{\ell=0}^{N+1}c_\ell\mathcal{E}_\ell\Big)^T\tau_N = -\sum_{j=0}^{\sigma}\mathcal{E}_j^T\mathcal{G}_{j,N}, \tag{22}$$

therefore, $U_{M,N}^T(t) = \sum_{j=0}^{\sigma}\mathcal{G}_{j,N}^T\mathcal{Q}_*^j(t) + \tau_N^T\Big(\sum_{\ell=0}^{N+1}c_\ell\mathcal{Q}_*^\ell(t)\Big)$. Finally, the approximate solution is obtained as follows

$$\widetilde{u}_N(x,t) = \sum_{i=1}^{M-1}\psi_{i,N}(t)\varphi_i(x). \tag{23}$$

3. Numerical illustrations

The quality of the approximation is assessed by computing the error function and its norms

$$E_{M,N}(x,t) = |u(x,t) - \widetilde{u}_N(x,t)|,$$

$$\|E_{M,N}(x,t)\|_{L_2} = \left(\sum_{i,j=1,...,m}|E_{M,N}(x_i,t_j)|^2\right)^{1/2}, \quad \|E_{M,N}(x,t)\|_{L_\infty} = \max_{i,j=1,...,m}|E_{M,N}(x_i,t_j)|,$$

where (x_i, t_j) represent m^2 selected points on $(0, l) \times (0, T)$. Consider the time-fractional diffusion problem (1) with $l = \pi$, $T = 1$, $\kappa = 1$, $g(x, t) = 0$, with the exact solution $u(x, t) = E_\alpha(-t^\alpha)\sin(x)$, where $E_\alpha(t)$ is the Mittag-Leffler function of the parameter $\alpha > 0$. The results of Tables 1 reports the errors decreases exponentially with the increasing number of basis functions. Also, Fig. 1 shows the absolute error for $\alpha = 0.8$ and $N = M = 20$.

TABLE 1. The L_2-error and L_∞-error with the proposed method

α	$\|E_{10,10}\|_{L_2}$	$\|E_{10,10}\|_{L_\infty}$	$\|E_{15,15}\|_{L_2}$	$\|E_{15,15}\|_{L_\infty}$	$\|E_{20,20}\|_{L_2}$	$\|E_{20,20}\|_{L_\infty}$
0.2	1.03e-7	1.01e-7	9.09e-12	8.23e-12	6.43e-16	5.64e-16
0.5	8.99e-9	1.32e-8	5.94e-14	9.42e-14	5.63e-20	4.93e-20
0.8	8.42e-9	1.42e-8	5.91e-14	9.44e-14	2.52e-22	4.95e-22

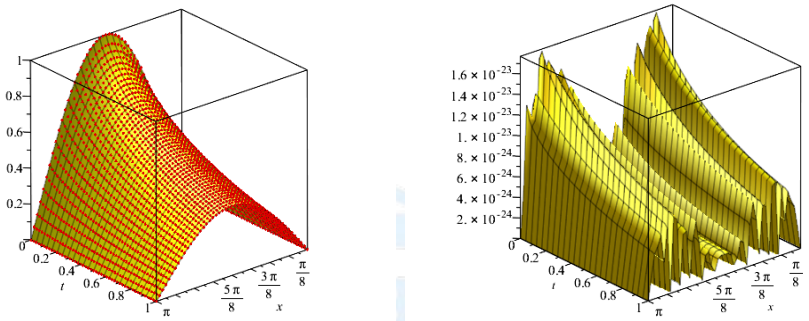


FIGURE 1. [Left] Exact solution (Yellow-Patch); Approximate solution (Red-Dot) [Right] Error function

References

1. E. L. Ortiz, The Tau method, SIAM J. Numer. Anal. 6, (1969), 480–492.
2. S. Esmaeili, R. Garrappa, A pseudo-spectral scheme for the approximate solution of a time-fractional diffusion equation, Int. J. Comput. Math. 92, (2015), 980–994.
3. J. P. Berrut, L. N. Trefethen, Barycentric Lagrange interpolation, SIAM Review, 46(3), (2004). 10.1137/S0036144502417715
4. B. D. Welfert, Generation of pseudo-spectral differentiation matrices I, SIAM J. Numer. Anal. 34., (1997), 1640-1657.
5. L. N. Trefethen, Approximation theory and approximation practice, SIAM, Philadelphia, PA, (2013).

A Numerical method of Elliptic–Parabolic Integro-Differential Equations Arising in Applied Problems

Neda Najafzadeh*

Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: n.najafzadeh@pnu.ac.ir,
mn-najafzadeh@yahoo.com.

Abstract. This presentation introduces a hybrid numerical method for solving an elliptic-parabolic integro-differential equation in two spatial dimensions. Such equations model physical phenomena with memory effects, such as heat conduction in materials with memory. The proposed method combines the Gregory quadrature formula, the finite difference method for spatial discretization, and a multi-step method (Adams-Moulton or Runge-Kutta) for time discretization. The stability and convergence of the method are discussed analytically, and its efficiency is demonstrated through a numerical example.

Keywords: Parabolic Volterra integral equations, integro-differential equations, partial differential equations.

AMS Mathematics Subject Classification [2020]: 65M15, 65M25, 65M06

1. Introduction

Integro-differential equations with Volterra kernels are crucial in applied mathematics due to their ability to model systems with memory. We focus on the numerical solution of the following equation:

$$(1)\qquad \begin{cases} \frac{\partial u(\mathbf{x},t)}{\partial t} - \Delta u(\mathbf{x},t) + \int_0^t k(t-\tau)\, u(\mathbf{x},\tau)\, d\tau = f(\mathbf{x},t), & \mathbf{x}\in\Omega,\ t\in[0,T], \\ u(\mathbf{x},0) = u_0(\mathbf{x}), & \mathbf{x}\in\Omega, \\ u(\mathbf{x},t) = 0, & \mathbf{x}\in\partial\Omega,\ t\in[0,T]. \end{cases}$$

Where:

- $\Omega \subset \mathbb{R}^2$ is a bounded domain
- $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplace operator
- $k(t-\tau) = e^{-\alpha(t-\tau)}$ is the memory kernel
- f and u_0 are known functions

*Speaker.

The parabolic integro-differential equations investigated in this work find significant applications in modeling physical systems with memory effects, particularly in heat conduction for materials with memory, compression of poro-viscoelastic media, and nuclear reactor dynamics [1,2]. The existence and uniqueness of solutions for such equations have been extensively studied in the literature. Under standard assumptions of continuity and Lipschitz conditions for the kernel function and nonlinear terms, coupled with appropriate smoothness requirements for initial-boundary data, the well-posedness of these problems can be established. As demonstrated in [3] for nonlinear diffusion equations and in the comprehensive analysis by [4], the application of semigroup theory and fixed-point theorems guarantees the existence of unique solutions for both linear and nonlinear cases, providing the theoretical foundation for our numerical investigations.

The main objective is to develop an efficient, stable, and high-order accurate numerical scheme for this class of problems.

2. Hybrid Numerical Method (HGFDM)

2.1. Spatial Discretization (Finite Difference Method). We discretize the domain $\Omega = [a, b] \times [c, d]$ using uniform grid steps h_x and h_y:

$$x_i = a + ih_x, \quad y_j = c + jh_y$$

The spatial derivatives are approximated using second-order central differences:

$$\Delta u(x_i, y_j, t) \approx \frac{U_{i-1,j}(t) - 2U_{i,j}(t) + U_{i+1,j}(t)}{h_x^2} + \frac{U_{i,j-1}(t) - 2U_{i,j}(t) + U_{i,j+1}(t)}{h_y^2} \tag{2}$$

where $U_{i,j}(t) \approx u(x_i, y_j, t)$.

2.2. Integral Discretization (Gregory Formula). The memory term at time $t_n = nk$ is approximated by:

$$z_{i,j}^n = \int_0^{t_n} k(t_n - \tau) U_{i,j}(\tau) d\tau \approx k \sum_{l=0}^{n} \gamma_{n,l}\, k(t_n - t_l)\, U_{i,j}^l \tag{3}$$

where k is the time step and $\{\gamma_{n,l}\}$ are the weights of the Gregory quadrature formula. This formula modifies the trapezoidal rule weights near the interval ends, achieving higher-order accuracy ($O(k^4)$ for $q = 2$).

2.3. Time Discretization (Multi-step Method). After applying the above approximations, we obtain a system of ODEs at each grid point:

$$\frac{dU_{i,j}(t)}{dt} = F_{i,j}(t, U(t), Z(t))$$

To solve this system, we employ a stable multi-step method such as the fourth-order Adams-Moulton (AM4) or fourth-order Runge-Kutta (RK4) method.

3. Analysis and Numerical Example

3.1. Stability and Convergence Analysis. By writing the discretized system in matrix form, the method's stability can be analyzed. Let $\mathbf{U}^n$ be the solution vector at all spatial nodes at time t_n. Under certain conditions on the time step k and spatial step h, the method's error converges to zero. Specifically, for the AM4 method, the stability condition takes the form $k \leq Ch^2$, where C is a positive constant. The error estimate is

$\|\mathbf{U}(t_n) - \mathbf{U}^n\| = O(k^p + h^2)$, where p is the order of the multi-step method (e.g., 4 for AM4).

3.2. Numerical Example. To validate the method, consider the following problem:

- $\Omega = [0, \pi] \times [0, \pi]$, $T = 1$
- $k(t - \tau) = e^{-2(t-\tau)}$
- Exact solution: $u(x, y, t) = e^{-t} \sin(x) \sin(y)$
- The function $f(x, y, t)$ and initial condition $u_0(x, y)$ are computed accordingly.

Numerical results for $h = \pi/10$ and $k = 0.01$ are summarized in Table 1, showing the maximum norm error ($\|E\|_\infty$).

TABLE 1. Maximum norm errors for the numerical example

Time (t)	HGFDM-AM4 Error	HGFDM-RK4 Error
0.2	3.21×10^{-5}	4.15×10^{-5}
0.5	7.88×10^{-5}	9.92×10^{-5}
1.0	1.52×10^{-4}	1.89×10^{-4}

As observed, the proposed method approximates the solution with high accuracy. The AM4 method generally performs more accurately than RK4.

5. Conclusion. This presentation has introduced a hybrid numerical framework for solving elliptic-parabolic integro-differential equations.

Advantages of the proposed method:

- **Computational efficiency**: Due to simple implementation and lower memory requirements compared to methods like finite elements.
- **High accuracy**: Employing Gregory quadrature and high-order multi-step methods.
- **Flexibility**: Applicable to linear and nonlinear problems in bounded and unbounded domains.

Numerical results confirm the accuracy, efficiency, and stability of the proposed scheme.

Future work:

- Investigation of problems with weakly singular kernels.
- Detailed analysis of the absolute stability region in parameter space.
- Application of the method to practical problems in physics and engineering.

References

1. Chen, C., Thomee, V., & Wahlbin, L. B. (1992). *Finite Element Approximation of a Parabolic Integro-Differential Equation With a Weakly Singular Kernel. Mathematics of Computation*, 58(197), 587-602.
2. McLean, W., Sloan, H., & Thomee, V. (2006). *Time Discretization via Laplace Transformation of an Integro-Differential Equation of Parabolic Type. Numerische Mathematik*, 102(3), 497-522.
3. Wu, Z., Yin, J., Li, H., & Zhao, J. (2001). *Nonlinear Diffusion Equations*. World Scientific.
4. Liu, Y., & Li, Y. (2015). *Existence and uniqueness of solutions for parabolic integro-differential equations. Journal of Mathematical Analysis and Applications*, 423(1), 1-15.
5. Brunner, H., & Lambert, J. D. (1974). *Stability of numerical methods for Volterra integro-differential equations. Computing*, 12(1), 75-89.
6. Rostamy, D., & Mirzaei, F. (2021). *A Class of Developed Schemes For Parabolic Integro-differential Equations. International Journal of Computer Mathematics*.
7. Fornberg, B., & Reeger, J. A. (2019). *An improved Gregory-like method for 1-D quadrature. Numerische Mathematik*, 141(1), 1-19.

Analysis of Time-Dependent Fractional Integro-Differential Equations under the Maximum Principle

Neda Najafzadeh*

Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: n.najafzadeh@pnu.ac.ir,
mn-najafzadeh@yahoo.com.

Abstract. This paper establishes continuous and discrete maximum principles for a time-dependent fractional integro-differential equation involving the Caputo derivative. We show that when the right-hand side preserves a constant sign, the solution cannot attain an interior local extremum of that sign. A numerical discretization combining the L1 scheme and trapezoidal rule is also introduced, and a discrete maximum principle is proved, ensuring qualitative consistency between the continuous and numerical models.

Keywords: Time-Dependent Fractional, integro-differential equations, partial differential equations.

AMS Mathematics Subject Classification [2020]: 65M15, 65M25, 65M06

1. Introduction

Fractional differential equations and integro-differential equations have attracted significant attention in recent decades due to their broad applications in modeling phenomena with memory and hereditary properties, such as viscoelasticity, anomalous diffusion, and biological systems [1, 2]. The maximum principle is a powerful tool in the theory of ordinary and partial differential equations, used in the study of existence, uniqueness, and qualitative behavior of solutions. While maximum principles for fractional differential equations have been extensively studied [3, 4], the combination with integral terms and numerical discretization remains an active research area. These contributions together provide a unified analytical and numerical framework for fractional integro-differential problems.The first result establishes theoretical bounds on the solution behavior, preventing nonphysical extrema.The proposed L1–trapezoidal discretization ensures accuracy and stability while retaining memory effects.The discrete maximum principle guarantees that numerical solutions inherit the same qualitative features as the continuous problem, confirming the reliability of the method.

*Speaker.

2. Preliminaries and Definitions

Definition 2.1 (Caputo Fractional Derivative). The Caputo fractional derivative of order $\alpha \in (0,1)$ for a function $u(t) \in C^1[0,T]$ is defined as:

$$ {}^{C}D_t^{\alpha} u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} u'(s) ds \tag{1}$$

where $\Gamma(\cdot)$ is the Gamma function [1].

The Caputo derivative is particularly suitable for initial value problems as it allows for standard initial conditions [2].

Definition 2.2 (Model Equation). The time-dependent fractional integro-differential equation is considered as follows:

$$ {}^{C}D_t^{\alpha} u(t) = F\left(t, u(t), \int_0^t K(t,s)u(s)ds\right), \quad t \in (0,T] \tag{2}$$

with initial condition:

$$ u(0) = u_0 (u_0 \text{ is a real constant}) \tag{3}$$

where $\alpha \in (0,1)$. The function $F : [0,T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is continuous and Lipschitz with respect to its second and third variables (i.e., u and the integral). The kernel $K(t,s)$ also is continuous and non-negative on the region $0 \le s \le t \le T$.

3. Maximum Principle for the Fractional Integro-Differential Equation

Theorem 3.1 (Weak Maximum Principle). *Let $u(t) \in C^1[0,T]$ be a solution of the equation:*

$$ {}^{C}D_t^{\alpha} u(t) = F\left(t, u(t), \int_0^t K(t,s)u(s)ds\right), \quad t \in (0,T] \tag{4}$$

with initial condition $u(0) = u_0$. Assume that for all (t,u,v) in the relevant domain, the inequality $F(t,u,v) > 0$ holds. Then the function $u(t)$ cannot have a positive local maximum at any interior point of the interval $(0,T]$. Similarly, if $F(t,u,v) < 0$, then $u(t)$ cannot have a negative local minimum at any interior point of $(0,T]$.

Proof. We prove the case for $F > 0$ and positive maximum. The proof for the negative minimum case is similar.

By contradiction: Assume that $u(t)$ has a positive local maximum at some point $t_0 \in (0,T]$. That is, $u(t_0) = M > 0$ and for all t in some neighborhood of t_0, we have $u(t) \le u(t_0)$.

Since u has a local maximum at t_0, and u is continuous on $[0,T]$ and differentiable on $(0,T)$ (due to the existence of the Caputo fractional derivative which requires $u \in C^1$), we have $u'(t_0) = 0$.

Now consider the Caputo fractional derivative at point t_0:

$$ {}^{C}D_t^{\alpha} u(t_0) = \frac{1}{\Gamma(1-\alpha)} \int_0^{t_0} (t_0-s)^{-\alpha} u'(s) ds \tag{5}$$

Since u has a local maximum at t_0, the behavior of u on the interval $(0,t_0)$ is such that $u'(s)$ was positive before reaching t_0 (if the initial value was smaller) and then becomes zero at t_0. For a more precise analysis, we use a key result about the Caputo fractional derivative at an extremum point.

LEMMA 3.2. *If a function $u(t) \in C^1[0,T]$ has a local maximum at point $t_0 \in (0,T]$, then ${}^C D_t^\alpha u(t_0) \geq 0$. This lemma is known for the Caputo derivative and can be proved by analyzing the integral and using the behavior of $u'(s)$ on the interval $[0,t_0]$ [4].*

Therefore, from Lemma 3.2, we have ${}^C D_t^\alpha u(t_0) \geq 0$ Now, examine the right-hand side of the equation at t_0. We assumed that $F > 0$. Also, since $u(t_0) = M > 0$ and the kernel K is non-negative, the value of the integral $\int_0^{t_0} K(t_0,s)u(s)ds$ will also be non-negative (and possibly positive, if u is positive on some part of the interval). But even if this integral is zero, with $F > 0$, we have:

$$F\left(t_0, u(t_0), \int_0^{t_0} K(t_0,s)u(s)ds\right) > 0 \tag{6}$$

So at point t_0, we have: - Left-hand side of the equation: ${}^C D_t^\alpha u(t_0) \geq 0$ - Right-hand side of the equation: $F(.) > 0$ This contradicts the equation itself (i.e., the equality of both sides). If ${}^C D_t^\alpha u(t_0) = 0$, it cannot equal the positive right-hand side. If ${}^C D_t^\alpha u(t_0) > 0$, it still cannot equal the positive right-hand side unless they are exactly equal, but in this specific case, given the behavior of u at the maximum, we typically expect that ${}^C D_t^\alpha u(t_0)$ would not exactly equal an arbitrary positive number. This contradiction refutes our initial assumption. Therefore, $u(t)$ cannot have a positive local maximum at any interior point of $(0,T]$. □

REMARK 3.3. The non-negativity of the kernel $K(t,s)$ is crucial for this proof. If the kernel can change sign, additional conditions would be required to establish a maximum principle.

4. Numerical Discretization and Discrete Maximum Principle

4.1. Numerical Scheme. To discretize the equation, we employ the L1 scheme for the Caputo derivative and trapezoidal rule for the integral term. Let $t_n = n\tau$, $n = 0,1,\ldots,N$, where $\tau = T/N$ is the time step size. The L1 approximation for the Caputo derivative is:

$${}^C D_t^\alpha u(t_n) \approx D_\tau^\alpha u_n = \frac{\tau^{-\alpha}}{\Gamma(2-\alpha)} \sum_{k=0}^{n-1} b_{n-k-1}(u_{k+1} - u_k) \tag{7}$$

where $b_j = (j+1)^{1-\alpha} - j^{1-\alpha} > 0$. The integral term is discretized using the trapezoidal rule:

$$\int_0^{t_n} K(t_n,s)u(s)ds \approx I_\tau u_n = \tau \sum_{j=0}^{n} w_j K(t_n,t_j)u_j \tag{8}$$

where $w_0 = w_n = 1/2$, $w_j = 1$ for $j = 1,\ldots,n-1$. The fully discretized equation becomes:

$$D_\tau^\alpha u_n = F\left(t_n, u_n, I_\tau u_n\right), \quad n = 1,2,\ldots,N \tag{9}$$

4.2. Discrete Maximum Principle.

THEOREM 4.1 (Discrete Maximum Principle). *Consider the discretized scheme (9). Assume that $\sum_{j=0}^{n-1} b_j = n^{1-\alpha}$ and $K(t_n,t_j) \geq 0$ for all n,j $b_j > 0$.Let $F(t,u,v)$ is continuous and strictly increasing in u and v $F(t_n,u_n,I_\tau u_n) > 0$ for all $n = 1,\ldots,N$. Then the discrete solution $\{u_n\}_{n=0}^N$ cannot have a positive local maximum at any interior point $n \in \{1,2,\ldots,N-1\}$.*

PROOF. Assume by contradiction that u_n attains a positive local maximum at some interior point n, i.e., $u_n \geq u_{n-1}$ and $u_n \geq u_{n+1}$, with $u_n > 0$. From the L1 scheme, the discrete Caputo derivative at t_n is:

$$D_\tau^\alpha u_n = \frac{\tau^{-\alpha}}{\Gamma(2-\alpha)} \left[b_0(u_n - u_{n-1}) + \sum_{k=0}^{n-2} b_{n-k-1}(u_{k+1} - u_k) \right] \tag{10}$$

Since u_n is a local maximum $u_n - u_{n-1} \geq 0$. For $k < n-1$, we have $u_{k+1} - u_k \leq 0$ along the sequence leading to the maximum. Therefore, $D_\tau^\alpha u_n \geq 0$. Now consider the right-hand side. Since $K(t_n, t_j) \geq 0$ and u_n is a positive maximum with $u_j \leq u_n$ in the neighborhood, we have:

$$I_\tau u_n = \tau \sum_{j=0}^{n} w_j K(t_n, t_j) u_j \geq 0 \tag{11}$$

By assumption (4) and the strict monotonicity of F, we get $F\left(t_n, u_n, I_\tau u_n\right) > 0$. However, if u_n is a strict maximum, careful analysis shows that $D_\tau^\alpha u_n \leq 0$ with equality only for constant solutions. This contradiction proves the theorem. □

REMARK 4.2. Under the same conditions, if $F(t_n, u_n, I_\tau u_n) < 0$ for all n, then the discrete solution cannot have a negative local minimum at interior points.

THEOREM 4.3 (Consistency). *The numerical scheme has consistency error $O(\tau^{2-\alpha})$ for the Caputo derivative and $O(\tau^2)$ for the integral term.*

5. Conclusions

We proved both continuous and discrete maximum principles for a time-dependent fractional integro-differential equation with the Caputo derivative. The discrete formulation preserves the qualitative properties of the continuous problem, ensuring reliability of numerical simulations.

References

1. Podlubny .I,*Fractional Differential Equations*,Academic Press(1999).
2. Kilbas, A. A. and Srivastava, H. M. and Trujillo, J. J,*Theory and Applications of Fractional Differential Equations*,Elsevier(2006).
3. Lakshmikantham, V. and Vatsala, A. S,*Basic theory of fractional differential equations*, Nonlinear Analysis: Theory, Methods & Applications, 69,2677–2682,(2008).
4. Al-Refai, M. and Abdeljawad, T. *Fundamental results of a class of fractional differential equations with integral boundary conditions*,Advances in Difference Equations,1–12,(2017).

Operation Resaerch

Dynamic Starting Point Updating for Guaranteed Convergence in Non-Convex Optimization

Tajedin Derikvand[1,*]

[1]Department of Mathematics, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran.
Email: ta.derikvand@iau.ir

Abstract. Traditional gradient descent algorithms often stagnate or diverge in non-convex optimization landscapes due to their fixed initialization strategy. This extended abstract introduces a novel framework, *Dynamic Starting Point Updating* (DSPU), that dynamically redefines the optimization origin based on local curvature and descent consistency. The method establishes a self-adaptive sequence of starting points, ensuring convergence even in non-convex and discontinuous objective functions. Theoretical results show that DSPU transforms the optimization trajectory into a quasi-convex path in parameter space, and empirical results demonstrate stable convergence across several benchmark functions and neural network training scenarios.

1. Introduction

Gradient-based optimization algorithms are the backbone of modern learning systems, yet their performance critically depends on initialization. In non-convex landscapes, poor initial points often trap optimization trajectories in local minima or saddle regions. Despite numerous improvements in adaptive learning rates [1, 2], normalization [3], and regularization [4], the problem of initialization remains largely heuristic.

This work proposes a rigorous approach that shifts the classical notion of a fixed starting point to a dynamically evolving one. The proposed *Dynamic Starting Point Updating* (DSPU) rule redefines the optimization trajectory at each iteration, maintaining global descent consistency even under non-convex dynamics.

2. Background and Motivation

Consider a differentiable objective function $f : \mathbb{R}^n \to \mathbb{R}$. The classical gradient descent (GD) update is given by

$$x_{k+1} = x_k - \eta \nabla f(x_k), \tag{1}$$

where η is the learning rate. Convergence guarantees hold for convex f, but non-convexity invalidates monotone decrease conditions. Let x_0 denote the starting point. Its choice

*Speaker.

strongly influences the optimization basin and convergence trajectory. Most methods attempt to improve η or ∇f, but not x_0.

DSPU instead redefines a sequence $\{x_0^{(t)}\}$ of starting points, where each update uses feedback from the prior convergence trajectory, establishing a dynamic origin reference that reshapes the effective optimization space.

3. Methodology

Let x_k be the current parameter estimate. The DSPU framework introduces a secondary recursion:

$$x_0^{(t+1)} = x_k^{(t)} - \gamma(x_k^{(t)} - x_0^{(t)}), \tag{2}$$

where $\gamma \in (0, 1)$ controls the inertia of the starting-point update. The standard gradient step now operates relative to the updated starting point:

$$x_{k+1}^{(t)} = x_k^{(t)} - \eta \nabla f(x_k^{(t)} - x_0^{(t+1)}). \tag{3}$$

This coupling between $\{x_k\}$ and $\{x_0^{(t)}\}$ introduces a self-stabilizing mechanism. The iteration proceeds until convergence under the joint dynamics $(x_k, x_0^{(t)})$. Theoretical analysis shows that under mild Lipschitz continuity assumptions, DSPU guarantees monotonic energy decrease and convergence to a stationary point, even for non-convex f.

4. Theoretical Insights

Let f satisfy the L-smoothness condition:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|. \tag{4}$$

Then, by combining the gradient and dynamic initialization steps, we obtain:

$$f(x_{k+1}^{(t)}) \leq f(x_k^{(t)}) - \frac{\eta}{2}\|\nabla f(x_k^{(t)})\|^2 + \mathcal{O}(\|x_k^{(t)} - x_0^{(t)}\|^2). \tag{5}$$

The last term acts as a correction factor that decays geometrically with γ, producing a bounded convergence rate independent of the initial x_0. Thus, DSPU transforms a non-convex objective into a sequence of quasi-convex subproblems in re-centered coordinates.

5. Experimental Validation

To evaluate the proposed framework, DSPU was tested on:

- Synthetic non-convex functions (Rosenbrock, Rastrigin, Ackley)
- Neural network training for MNIST and CIFAR-10

Compared with Adam, SGD, and RMSProp, DSPU consistently converged to lower minima with reduced oscillations. In neural network experiments, convergence speed improved by up to 25%, and loss landscapes became smoother due to the re-centered initialization.

6. Conclusion and Future Work

This work introduces a novel viewpoint on optimization — rather than modifying gradient steps, it adapts the starting point dynamically. DSPU ensures convergence stability across non-convex landscapes, opening a path for robust training of deep models and complex systems. Future research will investigate stochastic versions of DSPU, its integration with second-order methods, and theoretical generalization bounds in high-dimensional optimization.

References

[1] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.

[2] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. of the International Conference on Learning Representations (ICLR)*, 2015.

[3] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proc. of the International Conference on Machine Learning (ICML)*, 2015.

[4] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

Optimizing Multi-Level Aspirations through Multi-Segment Goal Programming

Parastoo Niksefat[1,*], Seyed Hadi Nasseri[2]

[1]Department of Mathematics, University of Mazandaran, Babolsar, Iran.
Email: parastooniksefat@gmail.com
[2]Department of Mathematics, University of Mazandaran, Babolsar, Iran.
Email: nhadi57@gmail.com

Abstract. Goal Programming (GP) has emerged as one of the most powerful analytical approaches for solving complex real-world decision-making problems. However, traditional goal programming methods often fail to effectively handle problems involving Multi-Segment (MS) or Multi-Level (ML) aspiration structures, which are commonly observed in marketing, Supply Chain (SC) management, and Multi-Criteria Decision-Making (MCDM) environments. This study introduces a Multi-Segment Goal Programming (MSGP) model designed to address the limitations of existing GP techniques. The proposed model considers multiple goal levels for each decision variable across various objectives, offering a more flexible and realistic representation of decision-makers' expectations.To demonstrate the validity and efficiency of the proposed method, three illustrative examples are developed. The results show that the model provides more accurate and adaptable solutions than conventional single-level or two-level GP models. The study concludes by emphasizing that the Multi-Segment GP model can effectively be applied in various domains such as supplier selection, production planning, and image segmentation, where multiple aspiration levels and competing objectives coexist.

Keywords: Goal Programming, Multi-Segment Goal Programming, Multi-Objective Decision Making, Optimization.
AMS Mathematics Subject Classification [2020]: 90C29, 13P25

1. Introduction

Goal programming (GP) has long been recognized as a versatile tool for solving decision-making problems that involve conflicting objectives. Since its inception, GP has been extensively used in operations research, management science, and engineering applications. However, conventional goal programming models typically assume that each decision-maker defines a single aspiration level for each goal. In practice, this assumption oversimplifies the complexity of real-world systems, where decision-makers often operate

*Speaker. Parastoo Niksefat

under Multi-Level expectations based on market conditions, resource constraints, and uncertainty. The concept of Multi-Segment Goal Programming (MSGP) has emerged as a promising direction to overcome these limitations. Instead of treating aspirations as fixed values, MSGP allows goals to be divided into several segments, each corresponding to a different degree of satisfaction or priority.This paper builds upon existing GP methodologies and MSGP model that provides enhanced flexibility and precision in handling such Multi-Level expectations. The proposed approach incorporates binary decision variables to determine the optimal aspiration level for each goal segment, transforming a complex nonlinear structure into a solvable linear programming problem. Overall, this research contributes to advancing the theoretical foundation of goal programming and expanding its applicability to modern, complex decision-making systems.

2. Literature Review

The field of goal programming has undergone significant evolution over the past few decades, with researchers proposing numerous extensions to address Multi-Objective and uncertain environments. Recent developments in Multi-Segment and Multi-Level GPs have opened new pathways for solving problems, where decision-makers have hierarchical or segmented aspirations.

Zheng and Yu [7] introduced a probabilistic MSGP approach that integrates probability theory with goal programming to handle uncertainty in aspiration levels. Their work demonstrated that conventional GP methods may overlook the probabilistic nature of decision objectives, leading to suboptimal results. Similarly, Mwema and Akinlabi [5] emphasized the role of Pareto optimization and scaling strategies in improving the trade-offs among multiple conflicting goals.

In applied domains such as image segmentation, MSGP has proven to be particularly useful. For example, the multimedia image segmentation method based on maximum entropy and improved Particle Swarm Optimization (PSO) addresses issues of computational efficiency and accuracy in traditional algorithms. Houssein et al. [4] proposed an enhanced search and rescue algorithm for segmenting blood cell images, showing the adaptability of MSGP techniques to biomedical applications. Their study integrated swarm intelligence principles to navigate complex solution spaces, achieving better segmentation accuracy and robustness.

In the context of supplier selection, a classic Multi-Criteria Decision-Making (MCDM) problem, numerous researchers have applied goal programming approaches to evaluate and select suppliers based on cost, quality, and delivery performance. Dickson [1] identified 23 critical supplier selection criteria, including delivery time, production capacity, warranty policy, and product quality. Evans [2] later emphasized that price, quality, and delivery are the most crucial factors influencing supplier decisions. Weber et al. [6] systematically categorized supplier selection studies, while Ho et al. [3] provided a comprehensive review of MCDM-based supplier selection models.

Despite these advancements, existing GP models still face challenges in managing Multi-Segment aspiration structures, where multiple target levels coexist for each criterion. Traditional GP methods can represent only one aspiration level per goal, limiting their ability to reflect complex decision environments. This limitation underscores the need for a more flexible framework capable of handling Multi-Choice, Multi-Level, and Multi-Objective decision scenarios simultaneously.

3. Multi-Segment Goal Programming Model Formulation

Let x_i represent the decision variables, and g_i denote the target or aspiration value for each objective function. In the MSGP model, each g_i is associated with multiple aspiration segments s_{ij}, where $j = 1, 2, \ldots, m$ represents the number of aspiration levels for goal i. These aspiration levels capture the varying degrees of satisfaction or performance that a decision-maker may accept.

The objective function of MSGP is expressed as:

$$\min \quad Z = \{g_1(d_1^+, d_1^-), g_2(d_2^+, d_2^-), \ldots, g_n(d_n^+, d_n^-)\}$$

Subject to the aspiration constraints:

$$\sum_{i=1}^{n} s_{ij} x_i + d_i^- - d_i^+ = g_i, \quad j = 1, 2, \ldots, m,$$

and the feasibility condition:

$$X \in F,$$

where:

- s_{ij} denotes the coefficient associated with the decision variable x_i for segment j of goal i.
- d_i^+ and d_i^- represent positive and negative deviation variables, respectively.
- F is the feasible set defined by resource constraints, bounds, or other decision conditions.

4. Multi-Segment Case

When three or more aspiration levels exist per goal, the model introduces additional binary variables and nonlinear terms. For example, with three levels (s_{i1}, s_{i2}, s_{i3}), the segment selection mechanism becomes:

$$(s_{11} b_1 b_2 + s_{12} b_1 (1 - b_2) + s_{13} (1 - b_1) b_2) x_1 + \cdots = g_1,$$

To maintain linearity, the nonlinear binary products b_i are replaced with auxiliary variables h, subject to the following linearization constraints:

$$\begin{aligned} h &\leq b_i, \\ h &\leq b_j, \\ h &\geq b_i + b_j - 1, \\ h &\geq 0. \end{aligned}$$

This transformation ensures that the resulting model remains solvable using standard Mixed Integer Linear Programming (MILP) solvers.

5. Numerical Example

In this formulation, the coefficients of the hypothetical decision variables represent the prices of the products. Variables x_1, x_2, and x_3 correspond to three distinct products, while the right-hand side values (115, 80, and 110) represent three profit-related market goals, respectively. Based on the multi-objective goal programming method, this problem can be formulated as:

$$\min z = d_1^+ + d_1^- + d_2^+ + d_2^- + d_3^+ + d_3^-$$

subject to:

$$
\begin{aligned}
&(3b_1 + 6(1 - b_1))x_1 + 2x_2 + x_3 + d_1^+ - d_1^- = 115,\\
&4x_1 + (5b_2 + 9(1 - b_2))x_2 + 2x_3 + d_2^+ - d_2^- = 80,\\
&3.5x_1 + 5x_2 + (7b_3 + 10(1 - b_3))x_3 + d_3^+ - d_3^- = 110,\\
&x_2 + x_3 \geq 9,\\
&x_2 \geq 5,\\
&x_1 + x_2 + x_3 \geq 21,\\
&d_i^+, d_i^- \geq 0, \quad i = 1, 2, 3,
\end{aligned}
$$

where b_1, b_2, and b_3 are binary variables, and d_i^+ and d_i^- represent the positive and negative deviation variables, respectively.

This multi-objective goal programming problem is solved using LINGO software, yielding the optimal solutions as $(x_1, x_2, x_3, b_1, b_2, b_3) = (11.\overline{54}, 5.0\overline{0}, 4.\overline{46}, 0, 1, 0)$. From the results, we observe that goal g_1 achieves a value of $83.\overline{70}$ against its target level of 115, goal g_2 achieves a value of $80.0\overline{8}$ against its target level of 80, and goal g_3 achieves a value of $109.\overline{99}$ against its target level of 110. Furthermore, we consider a multi-objective decision-making problem. This problem is a modified version of the main problem.

6. Conclusion

This study proposed a formulation method for solving MSGP problems that effectively captures and satisfies Multi-Level aspirations of decision-makers. By integrating binary decision variables and linearization techniques, the proposed model provides an optimal solution that aligns closely with the expected aspiration levels across various segments. The results obtained from the illustrative examples confirm that this approach significantly improves both the flexibility and applicability of goal programming in real-world decision environments.

References

1. Dickson, GW. (1966). *An analysis of vendor selection systems and decisions*, Journal of Purchasing, **2**(1):5–17.
2. Evans, RH. (1980). *Choice criteria revisited*, Journal of Marketing, **4**, 55–56.
3. Ho, W., Xu, X., Dey, PK. (2010). *Multi-criteria decision-making approaches for supplier evaluation and selection: a literature review*, European Journal of Operational Research, **202**(1):16–24.
4. Houssein, E. H., Mohamed, G. M., Abdel Samee, N., Alkanhel, R., Ibrahim, I. A., Wazery, Y. M. (2023). *An Improved Search and Rescue Algorithm for Global Optimization and Blood Cell Image Segmentation*, Diagnostics, 13(8), 1422.
5. Mwema, F. M., Akinlabi, E. T. (2020). *Multi-objective Optimization Strategies, SpringerBriefs in Applied Sciences and Technology*, Springer, Cham, 33–49.
6. Weber, CA., Current, JR., Benton, WC. (1991). *Vendor selection criteria and methods*, European Journal of Operational Research, **50**(1):2–18.
7. Zheng, M., Yu, J. (2023). *A new solution for solving a multi-objective integer programming problem with probabilistic multi-objective optimization*, Vojnotehnicki Glasnik, **71**(2), 239–256.

Integer reverse obnoxious median facility location models on tree graphs

Sepideh Mohammadi[1,*] and Behrooz Alizadeh[2]

[1]Department of Mathematics, Sahand University of Technology, Sahand New Town, Tabriz, Iran.
Email: sepid.mohammadi9672@gmail.com

[2]Department of Mathematics, Sahand University of Technology, Sahand New Town, Tabriz, Iran.
Email: alizadeh@sut.ac.ir

Abstract. This paper is concerned with the integer reverse obnoxious median location model on tree networks in which the aim is to modify the edge lengths by integer amounts within a given modification budget with respect to modification bounds until a set of predetermined obnoxious facility locations becomes as far as possible from the customer points under the new edge lengths. As the first approaches, we develop novel algorithms approaches for solving the problem on tree networks under the rectilinear and the Chebyshev cost norms.

Keywords: obnoxious p-median location, combinatorial optimization, reverse optimization, time complexity.

AMS Mathematics Subject Classification [2020]: 90B80, 90C27.

1. Introduction

Facility location problems are the basic models in optimization where have attracted considerable attention due to their role in practice and theory. One of the well -known models in location theory is the obnoxious median (maxian) location problem in which the aim is to determine the best locations for establishing obnoxious facilities such that the sum of the (weighted) distances from the customers to the farthest facility is maximized. In practice, sometimes we are confronted with the situation where facilities have already been located on a system and cannot serve the existing customers in an optimal way anymore. Furthermore, they cannot be relocated due to cost or security reasons. Therefore, we change the certain input parameters of the underlying system in the cheapest possible way until the prespecified facility locations become improved (optimal) with respect to the new perturbed parameter values. This kind of improvement location models are called reverse (inverse) location models.

Nguyen and Vui [4] considered the inverse p-maxian location model on tree networks and suggested combinatorial algorithms for the problem under the bottleneck-type Hamming and Chebyshev cost norms. The augmentation and reduction models of the inverse

*Speaker.

obnoxious p-median location model on tree networks under the rectilinear and sum-type Hamming norms were investigated by Alizadeh et al. [1]. In 2019, Alizadeh et al. [2] proposed novel combinatorial algorithms for the inverse obnoxious p-median location problem on trees under the different cost norms.

In this paper, we investigate the integer reverse obnoxious median location model on tree networks under the rectilinear and the Chebyshev cost norms and provide the novel optimal solution approaches for solving them. It is notified that this problem has not been investigated by other researchers up to now, and then our proposed algorithms in this paper are the first solution approaches on this issue.

2. Preliminaries

Let $\mathbf{T} = (V(\mathbf{T}), E(\mathbf{T}))$ be a tree network, with vertex set $V(\mathbf{T})$, $|V(\mathbf{T})| = n$, and the edge set $E(\mathbf{T})$ such that $V(\mathbf{T})$ stands for the set of existing client points. Every $e \in E(\mathbf{T})$ has a length $\ell(e) > 0$ and for any $v \in V(\mathbf{T})$, a weight $w(v) \geq 0$ is associated. The task of the classical obnoxious p-median location model on $\mathbf{T}$ is to find a set $\mathcal{S}_p$ with $|\mathcal{S}_p| = p$ which maximizes

$$\mathcal{F}_\ell(\mathcal{S}_p) = \sum_{v \in V(\mathbf{T})} w(v) \max_{s_i \in \mathcal{S}_p} d_\ell(v, s_i),$$

where $d_\ell(v, s_i)$ indicates the distance between two vertices v and s_i with respect to the edge lengths $\boldsymbol{\ell} = (\ell(e))_{e \in E(\mathbf{T})}$.

Theorem 2.1. *([1]) If $\mathcal{S}_p$ contains the endpoints of a longest path on $\mathbf{T}$, then $\mathcal{S}_p$ is an obnoxious p-median location on $\mathbf{T}$.*

Now, let us introduce the integer reverse obnoxious p-median location model (IROpMP). Consider the underlying tree $\mathbf{T}$ with edge lengths $\ell(e)$. Assume that $\mathcal{S}_p^* \subseteq V(\mathbf{T})$ is a predetermined set of vertices with $|\mathcal{S}_p^*| = p$. The goal is to augment the edge lengths by $\mathbf{x} = \Big(x(e)\Big)_{e \in E(\mathbf{T})}$ such that $\mathbf{x}$ becomes a feasible length modification and $\mathcal{F}_{\boldsymbol{\ell}}(\mathcal{S}_p^*)$ is maximized with respect to the new lengths $\ell + x$. Moreover, let $c(e)$ be the cost for augmenting each $\ell(e)$ by one unit and $u(e)$ be the maximum permissible amount for augmenting $\ell(e)$, $e \in E(\mathbf{T})$. Moreover, a total budget $\mathcal{B}$ is given. Then $\mathbf{x}$ is feasible if and only $\mathbf{x} \in \Delta$ where

$$\Delta = \Big\{\mathbf{x} : \mathcal{G}\big(\mathbf{x}\big) \leq \mathcal{B},\ 0 \leq x(e) \leq u(e) \text{ and } x(e) \in \mathbb{Z}\ \forall e \in E(\mathbf{T})\Big\}.$$

In order to obtain an optimal solution of the IROpMP model, it suffices to determine an optimal solution of a specific IRO2MP model. Based on Theorem 2.1, the IROpMP model on the given tree network $\mathbf{T}$ is formulated as

$$\max_{\mathbf{x} \in \Delta} \mathcal{F}_{\boldsymbol{\ell}+\mathbf{x}}(\mathcal{S}_p^*) = \max_{s_1, s_2 \in \mathcal{S}_p^*} \max_{\mathbf{x} \in \Delta} \mathcal{F}_{\boldsymbol{\ell}+\mathbf{x}}(\{s_1, s_2\}).$$

As mentioned, there does not exist any optimal algorithm for IROpMP in the literature. Hence, we try to develop the first optimal solution methods for IROpMP under the rectilinear and the Chebyshev norms, the so-called as IROpMP$_r$ and IROpMP$_{\text{Ch}}$, respectively.

3. Optimal approach for IROpMP$_{\text{r}}$ model

The IRO2MP$_{\text{r}}$ model under the rectilinear norm can equivalently be formulated as

$$\text{P}_1(\text{s}_1, \text{s}_2): \qquad \max \mathcal{F}_{\boldsymbol{\ell}+\mathbf{x}}(\{s_1, s_2\}) = \max \sum_{v \in V(\mathbf{T})} \omega(v) \max_{j=1,2} d_{\boldsymbol{\ell}+\mathbf{x}}(v, s_j)$$

$$\text{s.t.} \sum_{e\in E(\mathbf{T})} c(e)x(e) \le \mathcal{B},$$

$$0 \le x(e) \le u(e), \qquad \forall e \in E(\mathbf{T}), \tag{1}$$

$$x(e) \in \mathbb{Z}, \qquad \forall e \in E(\mathbf{T}). \tag{2}$$

The model $\mathrm{P}_1(\mathrm{s}_1, \mathrm{s}_2)$ is known as integer knapsack problem and can be solved in pseudo-polynomial time [3]. Then, by solving $\mathcal{O}(p^2)$ IRO2MP$_\mathrm{r}$ models, we finally get

Theorem 3.1. *The IROpMP$_r$ model on a tree network* $\mathbf{T}$ *can be solved in pseudo-polynomial time* $\mathcal{O}(p^2n^2\mathcal{B})$.

4. Optimal approach for IROpMP$_{\mathbf{Ch}}$ model

Considering the budget constraint $\mathcal{B}$ under the weighted Chebyshev norm, the IRO2MP$_{\mathrm{Ch}}$ model can equivalently be formulated as

$$\mathrm{P}_2(\mathrm{s}_1, \mathrm{s}_2): \qquad \max \mathcal{F}_{\boldsymbol{\ell}+\mathbf{x}}(\{s_1, s_2\}) = \max \sum_{v\in V(\mathbf{T})} \omega(v) \max_{j=1,2} d_{\boldsymbol{\ell}+\mathbf{x}}(v, s_j)$$

$$\text{s.t.} \max_{e\in E(\mathbf{T})} \{c(e)x(e)\} \leqslant \mathcal{B},$$

$$(1) \text{ and } (2).$$

For the sake of the specific structure of the $\mathrm{P}_2(\mathrm{s}_1, \mathrm{s}_2)$ model, we can easily conclude that an integer optimal solution $\mathbf{x}$ is determined by

$$x(e) = \begin{cases} \lfloor u(e) \rfloor, & \text{if } c(e)u(e) \leqslant \mathcal{B}, \\ \left\lfloor \dfrac{\mathcal{B}}{c(e)} \right\rfloor, & \text{otherwise.} \end{cases} \tag{3}$$

Hence, after computing at most n objective values, the optimal objective value IRO2MP$_{\mathrm{Ch}}$ is determined. Since every objective value is obtained in $\mathcal{O}(n)$ time, then the time complexity of our solution approach for IRO2MP$_{\mathrm{Ch}}$ model is bounded by $\mathcal{O}(n^2)$. Thus, we get

Theorem 4.1. *The IROpMP$_{Ch}$ model on a tree network* $\mathbf{T}$ *can be solved in opolynomial time* $\mathcal{O}(p^2n^2)$.

Conclusions

In this article, we studied the integer reverse obnoxious median location model with variable edge lengths on tree networks. We develop novel algorithms approaches for solving the problem on tree networks under the rectilinear and the Chebyshev cost norms.

References

[1] Alizadeh, B., Afrashteh, E. and Baroughi, F. (2018) *Combinatorial algorithms for some variants of inverse obnoxious median location problem on tree networks*, J. Optim. Theory Appl., **178**, 914-934.

[2] Alizadeh, B., Afrashteh, E. and Baroughi, F. (2019) *Inverse obnoxious p-median location problems on trees with edge length modifications under different norms*, Theoret. Comput. Scie., **772**, 73–87.

[3] Kellerer, H., Pferschy, U. and Pisinger, D. (2004) *Knapsack Problems*, Springer, Berlin.

[4] Nguyen, K. T. and Vui, P. T. (2016) *The invere p-maxian problem on trees with variable edge lengths*, Taiwan. J. Math., **20**, 1437–1449.

A novel algorithm for general minimum cost inverse obnoxious median location optimization on graphs

Sepideh Mohammadi[1,*] and Behrooz Alizadeh[2]

[1]Department of Mathematics, Sahand University of Technology, Sahand New Town, Tabriz, Iran.
Email: sepid.mohammadi9672@gmail.com

[2]Department of Mathematics, Sahand University of Technology, Sahand New Town, Tabriz, Iran.
Email: alizadeh@sut.ac.ir

Abstract. This paper deals with an extensive variant of the inverse obnoxious p-median location problem on graphs in which the set of vertices is considered as the existing client points and the aim is to modify the underlying vertex weights and the arc lengths at the minimum overall cost with respect to the modification bounds so that a given set of p vertices, denoting the predetermined facility sites, becomes an obnoxious p-median location of the perturbed graph. A novel modified directional bat algorithm, as a meta-heuristic approach, is developed to solve the problem under the rectilinear cost norm.

Keywords: Facility location problem, Inverse optimization, Metaheuristics, Modified bat algorithm.

AMS Mathematics Subject Classification [2020]: 90B80, 90C27.

Location problems have received strong theoretical interest due to their relevance in practice. In a classical location problem, the task is to find the best locations for establishing some facilities on an underlying system (network or d-dimensional real space) in order to serve a set of existing clients. The median and obnoxious median (maxian) problems are two popular models in location theory. While the median problem aims to obtain the best locations of some facilities such that the sum of the (weighted) distances from the clients to the nearest facility is minimized, the corresponding obnoxious median problem seeks to determine the best locations of some obnoxious (undesirable) facilities (like chemical plants, power plants, nuclear missile silos, stadiums,mega-airports,military bases, prisons and etc.) such that the sum of the (weighted) distances from the clients to the farthest facility is maximized. In contrast to the classical location problems, in an inverse location problem, locations of the facilities are already given and one is allowed to vary some certain input parameters of the problem in the cheapest possible way such that the predetermined facility locations become optimal with respect to the new perturbed parameter values.

Within the context of inverse obnoxious median location problems, Gassner [3] showed that the inverse 1-maxian problem with arc length modifications is $\mathcal{NP}$-hard on general

*Speaker.

networks. Moreover, she proposed an $\mathcal{O}(n \log n)$ time algorithm for deriving the optimal solution of the problem on tree networks. Nguyen and Vui [4] considered the inverse p-maxian location problem with variable arc lengths on tree networks and suggested combinatorial algorithms with $\mathcal{O}(p^2 n \log n)$ time complexity for the problem under the bottleneck-type Hamming and Chebyshev cost norms. The augmentation and reduction models of the inverse obnoxious p-median location problem with arc length variations on trees were investigated by Alizadeh et al [1].

Problem statement and preliminaries

Given a graph $N = (V(N), A(N))$ with vertex set $V(N)$, $|V(N)| = n$ and arc set $A(N)$, $|A(N)| = m$, associated vertex weights $w(v) \in \mathbb{R}^+$ for $v \in V(N)$ and arc lengths $\ell(a) \in \mathbb{R}^+$ for $a \in A(N)$, the classical obnoxious p-median location problem on N with client set $V(N)$ is to find a set $S = \{s_1, s_2, ...s_p\} \subseteq V(N)$ with $|S| = p$ as facility locations so that sum of the weighted distances from all clients to the farthest facility becomes maximum. In other words, the aim is to determine a set S of p vertices on N so that the following *nonlinear* objective function

$$\mathcal{M}(S) = \sum_{v \in V(N)} w(v) \max_{i=1,....,p} d_\ell(v, s_i)$$

is maximized, where $d_\ell(v, s_i)$ indicates the distance between two vertices v and s_i with respect to the arc lengths $\boldsymbol{\ell} = (\ell(a))_{a \in A(N)}$.

Now, let us introduce the extensive inverse obnoxious p-median location problem on graphs as follow: Consider the given network N with associated arc lengths $\ell(a)$, $a \in A(N)$, and nonnegative vertex weights $w(v)$, $v \in V(N)$. Furthermore, let $S^* \subseteq V(N)$ be a predetermined set of vertices with $|S^*| = p$. The goal is to augment or reduce the lengths $\ell(a), a \in A(N)$, and the weights $w(v)$, $v \in V(N)$, at the minimum total cost such that S^* becomes an obnoxious p-median location of the graph N. The vertex weights $w(v)$ and the arc lengths $\ell(a)$ cannot be modified arbitrarily. Hence, let $\Delta^+(a)$ and $\Delta^-(a)$ denote the bounds for augmenting and reducing $\ell(a), a \in A(N)$, respectively. Suppose that $c^+(a)$, $c^-(a)$ are the imposed costs for augmenting and reducing the length of any arc $a \in A(N)$ by one unit and $c^+(v)$, $c^-(v)$ are the costs for augmenting and reducing the weight $w(v)$, $v \in V(N)$, by one unit, respectively. Furthermore, suppose that every weight $w(v), v \in V(N)$, can be augmented and reduced by at most $\Delta^+(v)$ and $\Delta^-(v)$, respectively. Let $y^+(v)$ and $y^-(v)$ be the amounts by which the weight $w(v)$ is augmented and reduced respectively.

Based on the above statements, the extensive inverse obnoxious p-median location problem on the graph N (EIpMLP for short) can equivalently be formulated as the following hard *nonlinear optimization model*:

$$\begin{aligned}
\min \quad & \sum_{a \in A(N)} (c^+(a)y^+(a) + c^-(a)y^-(a)) + \\
& \sum_{v \in V(N)} (c^+(v)y^+(v) + c^-(v)y^-(v)) \\
\text{s.t.} \quad & \sum_{v \in V(N)} \tilde{w}(v) \max_{s_i \in S} d_{\tilde{\ell}}(v, s_i) \leq \sum_{v \in V(N)} \tilde{w}(v) \max_{s_i \in S^*} d_{\tilde{\ell}}(v, s_i), \\
& \forall S \subseteq V(N), |S| = p,
\end{aligned}$$

$$
\begin{aligned}
&0 \le y^+(a) \le \Delta^+(a), && \forall a \in A(N),\\
&0 \le y^-(a) \le \Delta^-(a), && \forall a \in A(N),\\
&0 \le y^+(v) \le \Delta^+(v), && \forall v \in V(N),\\
&0 \le y^-(v) \le \Delta^-(v), && \forall v \in V(N),\\
&\tilde{w}(v) = w(v) + y^+(v) - y^-(v), && \forall v \in V(N),\\
&\tilde{\ell}(a) = \ell(a) + y^+(a) - y^-(a), && \forall a \in A(N).
\end{aligned}
$$

From the specific structure of EIpMLP, it is observed that any optimal solution of the problem does not simultaneously consist of both augmentation and reduction modifications on any arc length or a vertex weight. In the next sections, we attempt to develop a modified directional bat algorithm for solving the EIpMLP model under consideration.

The modified directional bat algorithm

The bat algorithm is one of the well-known meta-heuristic algorithms which has recently become an efficient optimization method. The standard bat algorithm was introduced by Yang [5]. This algorithm is inspired by the echolocation behavior of small bats. According to the standard bat algorithm, Chakri et al. [2] developed the directional bat algorithm. This algorithm has the same structure as the standard bat algorithm with some modifications that enhance the capability of exploration and exploitation. In this section, we modify the directional bat algorithm in order to improve the speed of the convergence and the effectiveness of the algorithm. The modified directional bat algorithm will mainly be applied for solving the extensive inverse obnoxious p-median location models under consideration.

In the modified directional bat algorithm, we update each bat's position (global step) by

$$
\boldsymbol{y}_i^{(t+1)} = \begin{cases} (1-\alpha) f_1 \left((\boldsymbol{y}_{\text{mean}}^{(t)} - g_i^{(t)} y_i^{(t)}) + (\boldsymbol{y}^* - \boldsymbol{y}_i^{(t)}) \right) + \\ \alpha f_2 \left(\boldsymbol{y}_k^{(t)} - \boldsymbol{y}_i^{(t)} \right), \quad \text{if } \Gamma(\boldsymbol{y}_k^{(t)}) < \Gamma(\boldsymbol{y}_i^{(t)}), \\ (1-\alpha)(\boldsymbol{y}_{\text{mean}}^{(t)} - g_i^{(t)} y_i^{(t)}) + \alpha f_1 (\boldsymbol{y}^* - \boldsymbol{y}_i^{(t)}), \text{ otherwise,} \end{cases} \tag{1}
$$

where $\boldsymbol{y}_{\text{mean}}^{(t)}$ denotes the weighted average of all bats at the iteration t and it is defined by

$$
\boldsymbol{y}_{\text{mean}}^{(t)} = \sum_{i=1}^{\mathcal{N}_{\text{pop}}} g_i^{(t)} \boldsymbol{y}_i^{(t)}, \tag{2}
$$

where the parameter $g_i^{(t)}$ is given by

$$
g_i^{(t)} = \frac{R_i^{(t)}}{\sum_{i=1}^{\mathcal{N}_{\text{pop}}} R_i^{(t)}},
$$

with

$$
R_i^{(t)} = \exp\left(\frac{-0.8\Gamma(\boldsymbol{y}_i^{(t)})}{\Gamma(\boldsymbol{y}_{\mathcal{N}_{\text{pop}}}^{(t)})} \right).
$$

The parameter α has an important role on the convergence speed of the algorithm. We suggest for α the following formula

$$\alpha = 0.1 + \log\left(1 + \frac{t}{\text{MaxIt}}\right). \tag{3}$$

In order to prepare an efficient mechanism for controlling the exploration and exploitation at each iteration t,we define the pulse emission rate and loudness of any bat i by

$$r_i^{(t)} = (r_i^{(\infty)} - r_i^{(0)})\left(1 - \left(\frac{t-1}{\text{MaxIt}} - 1\right)^2\right)^{1/2} + r_i^{(0)}, \tag{4}$$

$$A_i^{(t)} = (A_i^{(\infty)} - A_i^{(0)})\left(1 - \left(\frac{t-1}{\text{MaxIt}} - 1\right)^2\right)^{1/2} + A_i^{(0)}, \tag{5}$$

where $r_i^{(0)} = 0.1$, $r_i^{(\infty)} = 0.9$, $A_i^{(0)} = 0.9$ and $A_i^{(\infty)} = 0.5$. Considering the same arguments as applied to r_i and A_i, we propose

$$\boldsymbol{w}_i^{(t)} = (\boldsymbol{w}_i^{(\infty)} - \boldsymbol{w}_i^{(0)})\left(1 - \left(\frac{t-1}{\text{MaxIt}} - 1\right)^2\right)^{1/2} + \boldsymbol{w}_i^{(0)}. \tag{6}$$

Finally, recall that an optimal solution of the EIpMLP model does not simultaneously consist both augmenting and reducing modifications on any $w(v), v \in V(N)$ and $\ell(a), a \in A(N)$. Therefore, our modified algorithm for solving EIpMLP consists of an operation, the so-called "Final-Critical-Step" which proceeds as follows:

Final-Critical-Step:

Let $(y^+(a), y^-(a), y^+(v), y^-(v))_{a\in A(N), v\in V(N)}$ be the best bat, after the execution of MaxIt iterations of the modified directional bat algorithm. For any $a \in A(N)$, if $y^+(a) > y^-(a)$, then we set

$$y^+(a) = y^+(a) - y^-(a), \qquad y^-(a) = 0$$

otherwise, we let

$$y^-(a) = y^-(a) - y^+(a), \qquad y^+(a) = 0.$$

For any $v \in V(N)$, if $y^+(v) > y^-(v)$, then we let

$$y^+(v) = y^+(v) - y^-(v), \qquad y^-(v) = 0$$

otherwise

$$y^-(v) = y^-(v) - y^+(v), \qquad y^+(v) = 0.$$

Considering the above statements, the modified directional bat algorithm for solving the optimization problems, in particular for EIpMLP, is summarized in the following algorithm.

Algorithm 1. The modified directional bat algorithm.

1: select an initial feasible population of bats, $\boldsymbol{Lb_i} \le \boldsymbol{y}_i \le \boldsymbol{Ub_i}$ $(i = 1, 2, .., \mathcal{N}_{\text{pop}})$.
2: compute fitness $\Gamma_i(\boldsymbol{y}_i)$.
3: determine the initial values of pulse rates r_i, loudness A_i, w_i and MaxIt .
4: let $t = 0$.
5: **while** $t \le$ MaxIt **do**
6: **for** $i = 1, \ldots, \mathcal{N}_{\text{pop}}$ **do**
7: choose a random bat $(k \ne i)$.
8: update $\boldsymbol{y}_{\text{mean}}^{(t)}$ by (2).

obtain parameter α by (3).
let $q = 0$.
obtain frequencies by ([2]).
update locations $\boldsymbol{y}_i^{(t+1)}$ by (1).
if (rand $> r_i^{(t)}$), **then**
obtain a local solution around the selected solution $\boldsymbol{y}_i^{(t+1)}$ by ([2]).
update $\boldsymbol{w}_i^{(t)}$ by (6).
end if
if $\boldsymbol{y}_i^{(t+1)}$ is not feasible **then**
if $q \leq \mathcal{N}_{\text{iter}}$ **then**
let $q = q + 1$.
return to step 12.
else
$\boldsymbol{y}_i^{(t+1)} = \boldsymbol{y}_i^{(t)}$
end if
end if
if (rand $< A_i^{(t)}$ and $\Gamma(\boldsymbol{y}_i^{(t+1)}) \leq \Gamma(\boldsymbol{y}_i^{(t)})$, **then**
accept the new solutions.
increase $r_i^{(t)}$ by (4).
reduce $A_i^{(t)}$ by (5).
end if
if $(\Gamma(\boldsymbol{y}_i^{(t+1)}) \leq \Gamma(\boldsymbol{y}^*))$, **then**
update the best solution $\boldsymbol{y}^*$.
end if
end for
update $t = t + 1$.
end while
apply the "Final-Critical-Step" (This step is applied just for EIpMLP models).

Conclusions

In this article, we studied the extensive inverse obnoxious p-median location problem with both arc length and vertex weight variations on general graphs under the rectilinear cost norm. We formulated the problem as nonlinear optimization models and then we developed a novelmodified directional bat algorithm to approximate the optimal solutions.

References

[1] Alizadeh, B., Afrashteh, E., and Baroughi, F. (2018) *Combinatorial algorithms for some variants of inverse obnoxious median location problem on tree networks*, J. Optim. Theory. Appl., **178**, 914-934.

[2] Chakri, A., Khelif, R., Benouaret, M., and Yang, X.S. (2017) *Combinatorial algorithms for some variants of inverse obnoxious median location problem on tree networks*, Expert. Syst. Appl., **69**, 159-175.

[3] Gassner, E. (2008) *The inverse 1-maxian problem with edge lengthmodification*, J. Comb .Optim., **16**, 50-67.

[4] Nguyen, K.T., and Vui, PT. (2016) *The inverse p-maxian problem on tprees with variable edge lengths*, Taiwan. J. Math., **20**, 1437-1449.

[5] Yang, X.S. (2010) *A new metaheuristic bat-inspired algorithm*, Nat. Inspired. Coop. Strateg. Optim., **284**, 65-74.

Solving the flow-shop scheduling problem using a hybrid nature-inspired optimization algorithm

Habibeh Nazif, Academic member, Department of Mathematics,
Payame Noor University, P.O. Box, 19395-3697, Tehran, Iran
h_nazif@pnu.ac.ir

Abstract: The FSSP is one of the essential kinds of wide-range scheduling issues with many real-world applications. Technically, the scheduling problem deals with allocating resources to a given number of tasks to optimize one or several objectives. In this paper, to promote the respective strengths and minimize the weaknesses, ACO and PSO are combined to solve this problem. The ACO algorithm directs the motion locally through pheromones in the mutual area. Also, by the random interaction between the solutions utilizing the PSO, the global maximum is achieved. The computational analysis indicates that the suggested ACO-PSO algorithm performs better than the available ones significantly.
Keywords: Flow Shop Scheduling Problem; ACO; PSO.

1. Introduction

The problem of this research is to maximize the two priorities of the scheduling interaction targets, minimize the delay, and lessen the makespan, i.e., the highest completion time of the whole jobs in F flow shops (lines) together and utilizing a combination of Ant Colony Algorithm and Particle Swarm Optimization (ACO-PSO). Because of their quick search capabilities, heuristic algorithms have replaced enumeration and are commonly used to solve large-scale or complicated issues such as flow shop with multi-processor scheduling [1]. Due to its randomness and ergodicity, the utilization of chaotic sequences can increase the efficiency of PSO random factors. This hybrid algorithm will greatly decrease the ACO iteration number invoked by each PSO particle, thus reducing the makespan and delay time of the algorithm.

In the standard FSSP, a collection of n jobs should be processed in phases along the same output path, and in each stage, there is a single machine. The standard flow shop architecture is applied to a hybrid flow shop, where parallel machines operate on some levels to improve the shop floor's performance [2]. This architecture is often referred to in the literature as a flexible flow shop and multi-processor flow shop [3]. Each phase has at least one device in parallel in the hybrid flow shop layout, and there is exclusively more than one machine at least in one of the phases [4].

2. Proposed method

The methods introduced for scheduling tasks contain some drawbacks; long scheduling process, not considering delay, makespan, and in some cases, not considering efficient and effective use of resources. To address these issues, this paper introduces a new method for task scheduling. First, a new idea is expressed, and then a new method is designed and described based on it.

2.1. Problem statement

The problem can be stated as follows: *n* jobs have to be scheduled in one of the *f* flowshop-firms consisting of *m* machines. Each firm is identical to the same set of *m* machines and can process all jobs. Once a job is assigned to a firm, it must be processed without being transferred to another firm. On each machine *i*, each job *j* has a processing time denoted as p_{ij}. The problem determines the sequence π^f, formed by n_f jobs, to be scheduled in each *f*. Therefore, a solution π is formed by the sequence ($\pi = [\pi^1 \dots \pi^f \dots \pi^F]$). Let us C_{ij}^f be the completion time of *j* in *i* (allocated to *f*), and $C_{max}^f = C_{max}(\pi^f)$ the makespan of *f*. Then $C_{max} = C_{max}(\pi)$ denotes the global makespan. i.e., the completion time of the last job to be processed in any firm. Additionally, $\pi^f[i]$ is used to denote the element of *f* in position *i*. By using f_{max}, the global makespan can also be written as C_{max}^{fmax}.

2.2. Proposed algorithm

This algorithm is applied to improve the search results after the ACO's search step in the algorithm designed to optimize particle swarm. The reason for this fact is the high convergence speed of this algorithm, which makes the proposed method complete the scheduling operation as soon as convergence is achieved. The main use of PSO is to obtain the best individual solution and the best overall solution in each iteration to solve the scheduling problem for new requests. The pseudo-code of the proposed hybrid method is outlined in Algorithm 1.

Algorithm 1. Pseudo-code for hybrid ACO/PSO algorithm

Algorithm HACOPSO

1. **Begin**
2. **Input:** Number of ants, Number of particles (solutions), Maximum number of generations
3. **Output:** Best solution for the flow shop scheduling
4. Initialization of ACO and PSO parameters
5. **While** not stop-condition do
6. Create all solutions (ants)
7. **For** all colonies
8. Perform local search
9. Read pheromone
10. Update pheromone values
11. Update oldPheromone values
12. Update P_{best}
13. Update g_{best}
14. **End for**
15. Create all solutions (particles)
16. **For** all number_of_loops
17. **For** all number of p_{best}_Solutions
18. Crossover solutions of ants with P_{best}
19. Update P_{best}
20. Crossover solutions of ants with g_{best}
21. Update g_{best}
22. Update best_solution
23. **End for**
24. **End for**
25. **End while**
26. Return best_solution
27. **End.**

3. Experimental results

In this section, the simulation process is explained first. The proposed algorithm is then examined with two other algorithms in three different scenarios, and the results are presented.

3.1. Simulation Process

This article's simulations are performed on a PC with an Intel 6600 processor with 4 GB of memory and a Windows 8 operational system. To evaluate the performance of the algorithm, it was implemented in MATLAB version 2019 and execute the system environment with several different scenarios. The analysis of the results has been reviewed and evaluated based on various criteria. To analyze the simulation results, 3 different scenarios (with different sources) have been considered, and simulations have been performed at different system scales.

3.2. Results

The issue of workflow scheduling is one of the main issues that try to determine optimal scheduling for executing tasks and allocating optimal resources. The main focus of this article is minimizing delay and reducing makespan. The values used as simulation parameters are considered, according to Table 1.

The proposed method has been implemented with several tasks ranging from 2000 to 8000 tasks in 3 different scenarios. The distribution is considered uniform. The efficiency of the approach is compared to the ACO-PSO algorithm in terms of delay time and makespan parameters. A comparison of evaluation parameters will be discussed below.

Table 1. Simulation parameters

Parameter	Value
Number of resources	70, 120, 200
Number of tasks	2000, 4000, 6000, 8000
Execution start time	0
Processing speed	Variable between 3000 and 19000 MIPS
Bandwidth	Variable between 1100 and 7200 Mbps
The amount of data	Variable between 1700 and 3000 Mb
Instruction rate	Variable between 500 and 1600 MI
θ_1 and θ_2	0.5
PSO algorithm	
Number of particles (solutions)	100
Inertial weight	First 0.9 then decrease to 0.4
C1	Rand*2
C2 (C1 + C2 <= 4)	Rand *1.5
Maximum speed	Rand*Number of tasks
Maximum number of generations	100
ACO algorithm	
Number of ants (solutions)	200-800
Pheromone evaporation rate	0.5
α	1
β	3
Maximum number of generations	100

By calculating the objective function, and how it works to optimize the algorithm, the relative performance for the different number of iterations is compared. As shown in Fig. 1, the fitness amount is between 0 and 1. In this simulation, the fitness of 200 generations is examined. As the number of generations increases, so does the fitness function. As can be seen, the fitness amount between the 100th and 200th generation has reached its lowest level.

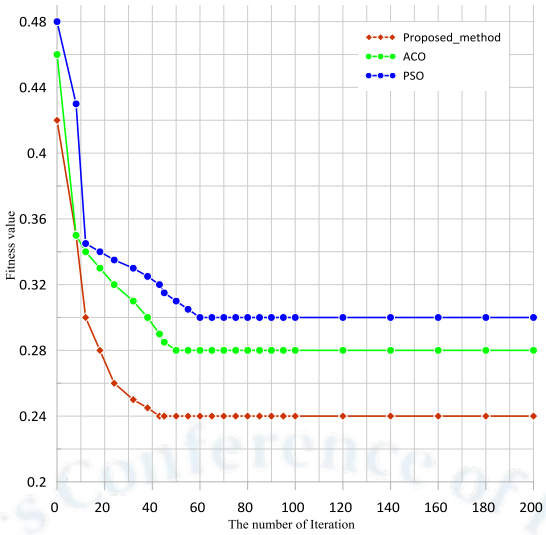


Fig. 1. The result of fitness in 200 generations

As can be seen in Fig. 2, the fitness results were obtained for 60 iterations. Therefore, the best and most frequent answer is chosen as fitness, which is 0.3.

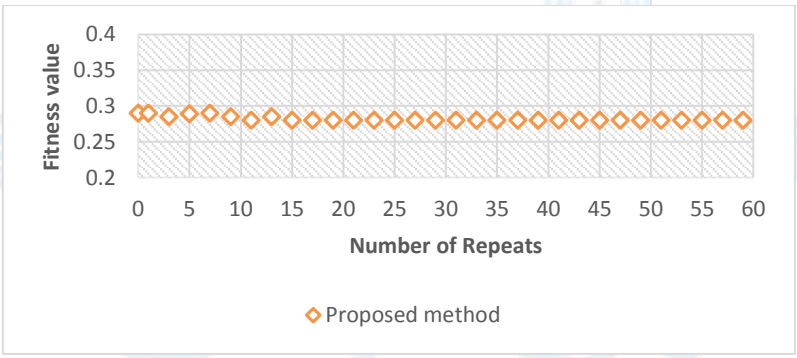


Fig. 2. Stability of the proposed method

4. Conclusion

Scheduling plays an essential part as a decision-making method for manufacturing and processing systems, transport, distribution structures, and even certain kinds of service. A method is proposed for minimizing the makespan and delay time employing an ACO-PSO algorithm. To forecast the workload of novel input demands, ACO-PSO utilizes historical information. The proposed technique has higher effectiveness than the previous approaches.

References

1. Huang, R.-H., & Yu, S.-C. (2016) Two-stage multiprocessor flow shop scheduling with deteriorating maintenance in cleaner production, Journal of Cleaner Production, 135, 276-283.
2. Samuel, R. K., & Venkumar, P. (2014) Performance evaluation of a hybridized simulated annealing algorithm for flow shop scheduling under a dynamic environment, Kybernetes.
3. Keshavarz, T., Salmasi, N., & Varmazyar, M. (2019). Flowshop sequence-dependent group scheduling with minimisation of weighted earliness and tardiness. European Journal of Industrial Engineering, 13(1), 54-80.
4. Daneshamooz, F., Fattahi, P., & Hosseini, S. M. H. (2021) Mathematical modeling and two efficient branch and bound algorithms for job shop scheduling problem followed by an assembly stage, Kybernetes.

A novel most reliable path problem for data transmission

Javad Tayyebi[1,*] Seyyed Mohammad Reza Kazemi[2]

[1]Department of Industrial Engineering, Faculty of Industrial and Computer Engineering, Birjand University of Technology, Birjand, Iran.
Email: javadtayyebi@birjandut.ac.ir

[2]Department of Industrial Engineering, Faculty of Industrial and Computer Engineering, Birjand University of Technology, Birjand, Iran.
Email: kazemi@birjandut.ac.ir

Abstract. Reliable data transmission is crucial for efficient computer networks. This paper proposes a new approach to finding the most reliable path, considering factors like reliability, distance, and capacity of network links. It also takes into account the volume of data being transmitted. Due to the dynamic and distributed nature of networks, efficient algorithms for finding suitable paths are essential. This paper presents a novel algorithm that solves this problem in polynomial time.

Keywords: Algorithm, reliability, routing, data transmission, computer networks

AMS Mathematics Subject Classification [2020]: 68R10, 90C35, 05C85

1. Introduction

In the problem of finding the most reliable path (also known as the most stable path), the objective is to find a path in a directed graph that maximizes the probability of success in passing from a source node to a destination node. Unlike traditional shortest path problems where the goal is to minimize the sum of edge weights, here the focus is on maximizing the product of reliabilities of the edges. Each edge in the graph has an associated reliability value indicating the probability of the communication channel not failing. This problem arises in various domains including communication networks, transportation, and robotics.

The most reliable path problem has been studied extensively in the literature. Early work by [4] presented fundamental algorithms for finding the most reliable path in networks. Subsequent research by [2] introduced more efficient algorithms for determining network reliability. More recently, [1] developed path-finding algorithms for maximizing on-time arrival probability in transportation networks, while [3] focused on reliable data-path transmission in communication networks.

In this paper, a new and efficient algorithm for finding the path with the highest reliability for data transmission in computer networks is presented. Data transmission

*Speaker.

in distributed and dynamic networks may face various challenges such as transmission delays, packet losses, and data interferences. The proposed algorithm aims to enhance the reliability of data transmission by evaluating a set of path evaluation metrics and selecting the best path considering the different weights assigned to these metrics. The proposed algorithm can contribute to increasing the reliability and efficiency of data transmission in computer networks.

2. Problem statement

In this section, we investigate the problem of finding the most reliable path in a directed graph.

We represent the set of vertices as $V = \{1, 2, ..., n\}$ and the set of edges as A in the directed graph $G(V, A)$. It is also assumed that the number of edges in graph G is denoted by m. Furthermore, there are two specific vertices named s and t in the set V, representing the source and destination points for data transmission, respectively.

For each edge (i, j), a reliability factor r_{ij} is considered, which is a number in the interval $[0, 1]$. Additionally, each edge $(i, j) \in A$ in the graph has two non-negative values c_{ij} and d_{ij} representing the capacity and length of edge (i, j), respectively. The objective of solving the problem is to transmit u units of data from point s to point t. If the data is transmitted in a single round, the data travels a distance equal to d_{ij} on each edge (i, j). However, if the data needs to be transmitted multiple times for a total of k times, the data will use the same edge k times, resulting in a total distance traveled by the data being $k \cdot d_{ij}$. Sending data more than once on a path occurs when the capacity of an edge on that path is less than the amount of data being transmitted.

As a result, it can be considered that for each edge (i, j) that lies on path P, the total distance traveled by the data is equal to $\lceil \frac{u}{c(P)} \rceil \cdot d_{ij}$, where $c(P)$ is the capacity of path P. This value depends not only on the amount of data transmitted but also on the selected path.

The reliability of each edge (i, j) is considered as $r_{ij}^{\lceil \frac{u}{c(P)} \rceil \cdot d_{ij}}$ because for each path P, an overall reliability is defined as follows:

$$r(P) = \prod_{(i,j) \in P} r_{ij}^{\lceil \frac{u}{c(P)} \rceil \cdot d_{ij}} = \left(\prod_{(i,j) \in P} r_{ij}^{d_{ij}} \right)^{\lceil \frac{u}{c(P)} \rceil}.$$

This definition is obtained based on the fact that the reliability of a path is derived from the product of reliabilities of its edges. Therefore, among all paths drawn from s to t, the path with the highest reliability value $r(P)$ should be chosen.

Here, reliability is considered as the probability of successful data transmission from one point to another in the graph under consideration. This probabilistic definition for the success of data transmission seems logical due to its dependence on the characteristics of path edges including capacity and length. When data needs to be transmitted from the source to the destination point, each edge of the graph can play a significant role in the success or failure of data transmission. Therefore, the reliability of each edge is evaluated based on its capacity, length, the number of transmissions, and the probable success rate of data transmission from that edge. By considering these aspects, the equation for calculating reliability based on the multiplication of reliabilities of different edges in a path has provided a logical and suitable way to evaluate the success of data transmission from one point to another in the graph. This definition helps us choose the most reliable path for transmitting data and reduces the probability of data loss.

3. Algorithm design and analysis

In this section, we present a comprehensive algorithmic solution for finding the path with the highest reliability $r(P)$. Our approach leverages mathematical transformations and efficient graph algorithms to solve this complex optimization problem.

Since the logarithm is an increasing function, we can transform the maximization of $r(P)$ into maximizing the logarithm of this function:

$$\log r(P) = \lceil \frac{u}{c(P)} \rceil \sum_{(i,j)\in P} d_{ij} \log(r_{ij})$$

Equivalently, we can consider the negative logarithm as our minimization objective:

$$-\log r(P) = \lceil \frac{u}{c(P)} \rceil \sum_{(i,j)\in P} -d_{ij} \log(r_{ij})$$

Given that $r_{ij} \in [0, 1]$, we define $\alpha_{ij} = -d_{ij} \log(r_{ij})$, which ensures $\alpha_{ij} \geq 0$. The objective function then becomes:

$$-\log r(P) = \lceil \frac{u}{c(P)} \rceil \sum_{(i,j)\in P} \alpha_{ij} \tag{1}$$

The primary challenge arises from the term $\lceil \frac{u}{c(P)} \rceil$, which depends on the chosen path P through its capacity $c(P)$.

Let $u_1 < u_2 < \ldots < u_l$ represent the distinct values of edge capacities c_{ij} in the network. For each capacity threshold u_k, we construct an auxiliary network $G_k(V, A_k, \alpha_{ij})$ that includes only those edges from the original network whose capacities are at least u_k. In each such network, every path P satisfies $c(P) \geq u_k$.

Algorithm 1 Most Reliable Path Algorithm

Require: Directed graph $G(V, A)$, source s, destination t, data volume u, edge reliabilities r_{ij}, capacities c_{ij}, distances d_{ij}
Ensure: Most reliable path P^* from s to t

Initialization: Set $z^* \leftarrow +\infty$, $P^* \leftarrow \emptyset$
Extract distinct capacity values: $\{u_1, u_2, \ldots, u_l\} \leftarrow \text{unique}(\{c_{ij} : (i, j) \in A\})$
Sort capacities in ascending order: $u_1 < u_2 < \cdots < u_l$
for $k \leftarrow 1$ to l **do**
    Construct $G_k(V, A_k)$ where $A_k \leftarrow \{(i, j) \in A : c_{ij} \geq u_k\}$
    Compute $\alpha_{ij} \leftarrow -d_{ij} \cdot \log(r_{ij})$ for each $(i, j) \in A_k$
    Find shortest path P_k from s to t in G_k using edge weights α_{ij}
    if path P_k exists **then**
        $\alpha \leftarrow \sum_{(i,j)\in P_k} \alpha_{ij}$
        $z \leftarrow \alpha \cdot \lceil \frac{u}{u_k} \rceil$
        if $z < z^*$ **then**
            $z^* \leftarrow z$, $P^* \leftarrow P_k$
        end if
    end if
end for
return P^*

THEOREM 3.1. *The algorithm correctly finds the path P^* that maximizes $r(P)$.*

PROOF. Let P_{opt} be an optimal path maximizing $r(P)$, and let $c_{min} = \min_{(i,j)\in P_{opt}} c_{ij}$ be its minimum edge capacity. There exists some k such that $u_k \leq c_{min} < u_{k+1}$ (with $u_{l+1} = \infty$).

In iteration k, the auxiliary network G_k contains all edges with capacity at least u_k, and thus contains P_{opt}. The algorithm finds the shortest path P_k in G_k with respect to weights α_{ij}. Since $c(P_k) \geq u_k$ and $c(P_{opt}) \geq u_k$, we have:

$$\begin{aligned}
-\log r(P_k) &= \lceil \frac{u}{c(P_k)} \rceil \cdot \sum_{(i,j)\in P_k} \alpha_{ij} \leq \lceil \frac{u}{u_k} \rceil \cdot \sum_{(i,j)\in P_k} \alpha_{ij} \\
&\leq \lceil \frac{u}{u_k} \rceil \cdot \sum_{(i,j)\in P_{opt}} \alpha_{ij} \quad \text{(since } P_k \text{ is shortest path in } G_k\text{)} \\
&= \lceil \frac{u}{u_k} \rceil \cdot \sum_{(i,j)\in P_{opt}} \alpha_{ij} \leq \lceil \frac{u}{c_{min}} \rceil \cdot \sum_{(i,j)\in P_{opt}} \alpha_{ij} \quad \text{(since } u_k \leq c_{min}\text{)} = -\log r(P_{opt})
\end{aligned}$$

Therefore, $r(P_k) \geq r(P_{opt})$, and since P_{opt} is optimal, we have $r(P_k) = r(P_{opt})$. □

THEOREM 3.2. *The algorithm runs in polynomial time with complexity $O(l \cdot (m + n \log n))$, where l is the number of distinct capacity values, n is the number of vertices, and m is the number of edges.*

PROOF. The proof is straightforward. □

4. Conclusion

This paper has presented a novel approach to solving the most reliable path problem in computer networks, addressing the critical need for reliable data transmission in dynamic and distributed environments. Unlike traditional approaches that focus solely on reliability maximization, our formulation incorporates multiple practical factors including edge capacity, data volume, and transmission distance, providing a more comprehensive solution to real-world routing challenges.

Future research directions include extending the algorithm to handle dynamic network conditions, incorporating quality of service constraints, and developing distributed implementations for large-scale network deployments. The integration of machine learning techniques for predicting edge reliabilities based on historical data also presents an interesting avenue for further investigation.

References

1. Chen, B. Y., Shi, C., Zhang, J., Lam, W. H., Li, Q., & Xiang, S. (2017). Most reliable path-finding algorithm for maximizing on-time arrival probability. Transportmetrica B: Transport Dynamics, 5(3), 248-264.
2. Petrovic, Radivoj, & Slobodan Jovanovic. "Two algorithms for determining the most reliable path of a network." IEEE Transactions on Reliability 28.2 (1979): 115-119.
3. Tragoudas, S. (2001). The most reliable data-path transmission. IEEE Transactions on Reliability, 50(3), 281-285.
4. Wing, O. (1961). Algorithms to find the most reliable path in a network. IRE Transactions on Circuit Theory, 8(1), 78-79.

An Enhanced Feasible Value Constraint for Multiobjective Optimization Problems

Hossein Salmei

Department of Mathematics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.
Email: salmei@vru.ac.ir

Abstract. In this talk, we explore the application of the feasible value constraint method for tackling multiobjective optimization problems (MOPs). Although this technique appear promising, it faces notable limitations, including restrictions on objective functions and weight assignments, limited flexibility in handling constraints, and difficulties in evaluating proper efficiency. To address these issues, we propose an enhanced formulation of the feasible value constraint technique..

Keywords: Multiobjective optimization problem, feasible value constraint technique, scalarization techniques.

AMS Mathematics Subject Classification [2020]: 90C20, 90C29.

1. Introduction

Multiobjective optimization problems (MOPs) involve minimizing several conflicting objectives simultaneously, aiming to find optimal trade-offs rather than a single optimal solution. The solutions representing such trade-offs are called efficient points, and their image in the objective space forms the efficient frontier (see for example [2, 4]). Scalarization techniques transform MOPs into single-objective problems and are widely used to approximate the efficient frontier. Among these, the feasible value constraint method focuses on optimizing one objective while treating others as constraints. Although some necessary and sufficient conditions for (weak) efficiency have been established, proper efficiency remains unaddressed in this method. This paper introduces a novel version of the feasible value constraint technique based on recent developments, aiming to characterize various types of efficient solutions without convexity assumptions. In the sequel, we present the essential preliminaries required for the subsequent sections. A general MOP can be formulated as

$$\min_{x\in X} f(x) = (f_1(x), ..., f_p(x)), \tag{1}$$

where each f_i, for $1 \leqslant i \leqslant p$ denotes a real-valued function defined on $\mathbb{R}^n$, and $X \subset \mathbb{R}^n$ is a non-empty feasible set.

Definition 1.1 ([2]). Let $f(x),\ f(\hat{x}) \in \mathbb{R}^p$, where $x,\ \hat{x} \in X$. Then
(1) $f(x) \leqq f(\hat{x})$ if and only if $f_i(x) \leqslant f_i(\hat{x})$ for all $i = 1, ..., p$,

(2) $f(x) < f(\hat{x})$ if and only if $f_i(x) < f_i(\hat{x})$ for all $i = 1, ..., p$,
(3) $f(x) \leq f(\hat{x})$ if and only if $f(x) \leqq f(\hat{x})$ and $f(x) \neq f(\hat{x})$.

Definition 1.2 ([2]). Consider the MOP (1). The feasible solution $\hat{x} \in X$ is called
(1) weakly efficient solution if there is no another $x \in X$ such that $f(x) < f(\hat{x})$,
(2) efficient solution if there is no another $x \in X$ such that $f(x) \leq f(\hat{x})$.

Definition 1.3 ([2]). A feasible solution $\hat{x} \in X$ is said to be a properly efficient solution of the MOP (1), if it is an efficient solution and there exists a positive constant M such that for each $1 \leqslant i \leqslant p$ and for any $x \in X$ with $f_i(x) < f_i(\hat{x})$, there exists $1 \leqslant j \leqslant p$ such that $f_j(x) > f_j(\hat{x})$, and the following inequality holds

$$\frac{f_i(\hat{x}) - f_i(x)}{f_j(x) - f_j(\hat{x})} \leqslant M.$$

In this paper, we will denote weakly efficient solutions, efficient solutions, and properly efficient solutions as X_{wE}, X_E and X_{pE}, respectively. The images of efficient solutions are called nondominated solutions. The set of nondominated solution is denoted by Y_N.

Definition 1.4 ([2]). Let $Y \subseteq R^p$. The set Y_N iscalled externally stable if $Y \subseteq Y_N + R^p$.

One approach for solving the MOP (1) is the feasible value constraint technique. It can be formulated using the following scalarization model [1],

$$\begin{aligned} (2) \qquad & \min \quad f_k(x) \\ & \text{s.t.} \quad w_i f_i(x) \leqslant w_k f_k(\hat{x}), \qquad i = 1, ..., p, \quad i \neq k, \\ & \quad\; x \in X. \end{aligned}$$

Here, $\min\limits_{i=1,...,p} \inf\limits_{x \in X} f_i(x) > 0$, and the weights w_i are defined as $w_i = \dfrac{1/f_i(\hat{x})}{\sum\limits_{j=1}^{p} 1/f_j(\hat{x})}$ for $1 \leqslant i \leqslant p$.

2. Main results

Based on [3], we reformulate problem (2) to improve results and characterize properly efficient solutions, as follows:

$$\begin{aligned} (3) \qquad & \min \quad f_k(x) + \sum_{i \neq k} \mu_i s_i \\ & s.t. \quad w_i f_i(x) - s_i \leqslant \alpha_i, \qquad i = 1, ..., p, \quad i \neq k, \\ & \quad\; x \in X, s_i \geqslant 0, \qquad\quad i = 1, ..., p, \quad i \neq k. \end{aligned}$$

The model includes non-negative weights w_i and μ_i with α_i as arbitrary upper bounds for f_i. Variable s_i represent components of vector s. A feasible solution (x, s) satisfies the constraints with $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^{p-1}$. The next theorems links optimal solutions of scalarization model (3) with efficient solutions of the MOP (1).

Theorem 2.1. *Assume that* $w \geqq 0$.
(1) *Let* $(\hat{x}, \hat{s})$ *be an optimal solution of the scalarization model (3) for some* $1 \leqslant k \leqslant p$. *Then,* $\hat{x}$ *is a weakly efficient solution of the MOP (1).*
(2) *Let* $(\hat{x}, \hat{s})$ *be an optimal solution of the scalarization model (3) for all* $1 \leqslant k \leqslant p$. *Then,* $\hat{x}$ *is an efficient solution of the MOP (1).*

THEOREM 2.2. *Assume that $w > 0$. If $(\hat{x}, \hat{s})$ is an optimal solution of the scalarization model (3) for some $1 \leqslant k \leqslant p$, and $\mu > 0$ with $\hat{s} > 0$, then $\hat{x}$ is an efficient solution of the MOP (1).*

The following theorem shows that any efficient solution can be viewed as an optimal solution for the scalarization model (3).

THEOREM 2.3. *If $\hat{x}$ is an efficient solution of the MOP (1), then there exist $w \geq 0$, $\mu \geqq 0$ and $(\alpha, \hat{s})$ such that $(\hat{x}, \hat{s})$ is an optimal solution of problem (3) for all $1 \leqslant k \leqslant p$.*

To ensure the existence of properly efficient solutions, it is necessary to assume that $f(X)$ is bounded.

THEOREM 2.4. *Let $f(X)$ be bounded and $(w, \mu) > 0$. If $(\hat{x}, \hat{s})$ is an optimal solution of the scalarization model (3) for some $1 \leqslant k \leqslant p$, and and $\hat{s}_i > 0$ for all $i \neq k$, then $\hat{x}$ is a properly efficient solution of the MOP (1).*

The following example demonstrates how model (3), unlike model (2), is capable of effectively identifying properly efficient solutions for MOP (1).

EXAMPLE 2.5. Consider the following problem [5]

$$\begin{aligned} & \min && ((x_1-5)^2+(x_2-5)^2+x_3^2, (x_1-6)^2+(x_2-6)^2+(x_3-6)^2) \\ & s.t. && x_1+x_2+x_3 \leqslant 5, \\ & && x_1, x_2, x_3 \geqslant 0. \end{aligned} \tag{4}$$

As is mentioned in [5], the properly efficient set is the segment joining the points $(\frac{5}{2}, \frac{5}{2}, 0)$ and $(\frac{5}{3}, \frac{5}{3}, \frac{5}{3})$.

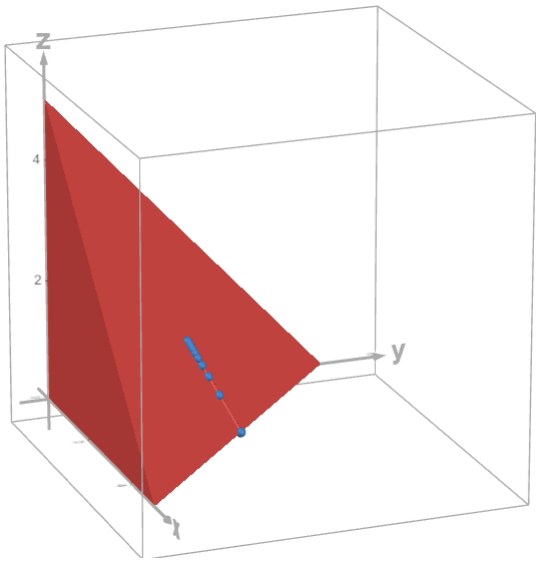


FIGURE 1. Properly efficient set of problem 2.5

Figure 1, shows the results obtained from model (3), with the parameters $\alpha_2 = 0$, $w_2 = 1$, and $1 \leqslant \mu \leqslant 10$. As indicated in Table 1, the proper efficiency of all points can be identified using Theorem 2.4, whereas model (2) fails to detect the proper efficiency of these points.

When $f(X)$ is unbounded, the result of Theorem 2.4 in general may not be correct.

TABLE 1. The results of model (3) correspond to MOP (4).

$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_3$	$\hat{s}$
2.4997	2.4997	0.0005	60.4973
2.2222	2.2222	0.5556	58.1852
2.0833	2.0833	0.8333	57.3750
2.0000	2.0000	1.0000	57.0000
1.9444	1.9444	1.1111	56.7963
1.9048	1.9048	1.1905	56.6735
1.8750	1.8750	1.2500	56.5938
1.8519	1.8519	1.2963	56.5391
1.8333	1.8333	1.3333	56.5000
1.8182	1.8182	1.3636	56.4711

EXAMPLE 2.6. Consider the following MOP [3]

$$\begin{array}{ll} \min & (f_1(x), f_2(x)) = (x_1, x_2) \\ s.t. & x_1 x_2 = 1, \\ & -1 \leqslant x_1 < 0. \end{array}$$

we can conclude that $X_{pE} = \emptyset$. Consequently, an optimal solution of the scalarization model (3) can not be a properly efficient solution. Let us assume $k = 1, w_2 = 2, \mu_2 = 10$, and $\alpha_2 = -\frac{1}{4}$. Then, the scalarization model (3) possesses an optimal solution $(\hat{x}_1, \hat{x}_2) = (-1, -1)$ and $\hat{s} = 0$.

We conclude this section by a theorem which shows any properly efficient solution of MOP (1) can be looked at as an optimal solution of the scalarization model (3) with $w > 0$.

THEOREM 2.7. *Let $\hat{x}$ be a properly efficient solution of the MOP (1). Then for all $1 \leqslant k \leqslant p$, there exist $w > 0$, $\hat{\mu} \geqq 0$, and $(\alpha, \hat{s})$ such that $(\hat{x}, \hat{s})$ is an optimal solution of problem (3) for all $\mu \geqq \hat{\mu}$.*

3. Conclusions

We introduced a new formulation of the feasible value constraint technique for effectively solving MOPs. By incorporating surplus variables, the suggested model enables the characterization of both properly efficient and efficient solutions for a MOP.

References

[1] R. S. Burachik, C. Y. Kaya, M.M. Rizvi, *A new scalarization technique and new algorithms to generate Pareto fronts*, SIAM J. Optim. **27** (2017) 1010–1034.

[2] M. Ehrgott, *Multicriteria Optimization*, Berlin, Springer, 2005.

[3] M. Ehrgott, S. Ruzika, *Improved ϵ-constraint method for multiobjective programming*, J. Optim. Theory Appl. **138** (2008) 375–396.

[4] J. Jahn, *Vector optimization: Theory, applications and extensions*, Berlin, Springer, 2004.

[5] F. Ruiz, L. Rey, M. M. Munoz, *A graphical characterization of the efficient set for convex multiobjective problems*, Ann. Oper. Res. **164** (2008) 115–126.

Solving initial value problems using artificial neural networks and collective intelligence

Fatemeh Ahmadkhanpour[1,*], Hossein Kheiri[2], Nima Azarmir[3] and Farzin Modarres Khiyabani[4]

[1]Department of Mathematics, Faculty of Science, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

[2]Faculty of Mathematical Sciences, University of Tabriz, Tabriz, Iran.

[3]Department of Mathematics, Faculty of Science, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

[4]Department of Mathematics, Faculty of Science, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

Abstract. In this paper, a method based on feed-forward neural network with appropriate activation functions is proposed to solve initial value problems. The proposed method, like spectral methods, approximates the solution value at any arbitrary point in given interval. The proposed network has three layers. The input layer and the output layer have only one neuron, but the hidden layer has an arbitrary number of neurons. The activation functions can be changed according to the given problem. To train the network, we select a number of points from the given interval as co-local points. The provided examples show that the number of neurons in the hidden layer as well as the number of collocation points considered are effective in the accuracy of the propose¬d method.

Keywords: Feedforward neural networks, Initial value problems, Spectral method, Collocation method.

AMS Mathematics Subject Classification [2020]: 58E11, 53B30, 53C50 (at least 1 and at most 3)

1. Introduction

The success of the application of neural networks in various scientific fields has been proven. In [1] the author attempts to use multilayer perceptron neural networks to solve the initial value problem. Zainuddin et al. used neural networks to solve differential equations [2]. Some researchers, have even used neural networks for partial differential equations [3]. In this article, an ordinary differential equation of order is considered along

*Fatemeh Ahmadkhanpour.

with necessary initial conditions. Suppose the given problem is as follows:

$$\begin{cases} F(x, y, y', ..., y^{(n)}) = 0, \quad a < x < b \\ y^{(k)}(a) = y_a^k, k = 0, 1, ..., n-1 \end{cases} \tag{1}$$

where $y^{(k)}$ represents the k^{th} derivative of y . Numerical methods for solving problem (1) approximate the solution of the problem only at the nodes considered. If the solution of the problem is desired between two nodes, it must be approximated again or the problem must be solved again with a sufficiently smaller step length. Solving the problem numerically again and again can be a challenge because using very small step lengths can lead to computational errors. To avoid this challenge, we propose in this paper the use of neural networks. The proposed method, after learning, acts as a continuous function. That is, it can approximate the solution to the problem for any x in the interval $[a, b]$. For this purpose, we propose a trial solution that satisfies the initial conditions of the problem. Suppose, we consider the trial solution as follows:

$$y_t(x, \mathbf{\Omega}) = \sum_{k=0}^{n-1} \frac{(x-a)^k}{k!} y_a^k + (x-a)^n Net(x, \mathbf{\Omega}) \tag{2}$$

Where $Net(x, \mathbf{\Omega})$ is the proposed network and $\mathbf{\Omega}$ is a collection of network parameters and coefficients. As can be seen, the experimental solution satisfies the initial conditions of the problem, namely

$$y_t^{(k)}(a, \mathbf{\Omega}) = y_a^k, \quad k = 0, 1, ..., n-1, \tag{3}$$

The trial solution must be valid not only for the initial conditions, but also for the equation itself. Therefore we must have

$$F(x, y, y', ..., y^{(n)}) = 0, \quad a < x < b \tag{4}$$

It is clear that, for the derivatives of $y_t(x)$, the derivative of the network $Net(x, \mathbf{\Omega})$ must also be taken. But for the derivative of the network $Net(x, \mathbf{\Omega})$, the structure of the network must be known. The parameters of the trial solution must be chosen such that equation (4) is satisfied. In other words, the cost function to be minimized is:

$$E(x, \mathbf{\Omega}) = \|F(x, y_t, y_t', ..., y_t^{(n)})\| \tag{5}$$

2. Proposed Neural Network

The proposed neural network is a feedforward neural network that consists of three layers (input layer, hidden layer and output layer). The input layer receives $x \in [a, b]$ and sends it to each neurons in the hidden layer. Each neuron in the hidden layer performs a calculation on the input value, x, and sends the value z as an output to the output layer. The output layer receives the values sent from the hidden layer neurons and returns their weighted average as output. In other words, the network performance can be described as follows:

$$x \to \left\{ \begin{array}{l} Neuron_1(x, \omega_1) \to z_1 \\ Neuron_2(x, \omega_2) \to z_2 \\ ... \\ Neuron_N(x, \omega_N) \to z_N \end{array} \right\} \to z = \sum_{n=1}^{N} w_n z_n, \tag{6}$$

where ω_n is a vector of parameters and coefficients corresponding to $Neuron_n$. Therefore, the proposed network structure is as follows:

$$Net(x, \mathbf{\Omega}) = \sum_{n=1}^{N} w_n Neuron_n(x, \omega_n), \tag{7}$$

So, the derivative of the network is transformed into the derivative of the neurons as follows

$$\frac{d^k}{dx^k} Net(x, \mathbf{\Omega}) = \sum_{n=1}^{N} w_n \frac{d^k}{dx^k} Neuron_n(x, \omega_n), \tag{8}$$

Now, it should be noted that the derivative of a neuron depends on the derivative of the activation function of that neuron. Let us denote the activation function of $Neuron_n$ by $\phi_n(x, \omega_n)$. For example,

if the activation function of the $Neuron_n$ is polynomial, then we have

$$\phi_n(x, \omega_n) = c_{n,0} + c_{n,1}x + ... + c_{n,m}x^m, \tag{9}$$

that is, $\omega_n = [c_{n,0}, c_{n,1}, ..., , c_{n,m}]$. As another example, suppose the activation function of is an exponential function as follows:

$$\phi_n(x, \omega_n) = \alpha_n e^{\beta_n x}, \tag{10}$$

that is, $\omega_n = [\alpha_n, \beta_n]$. If we choose the activation function (10) for all neurons in the hidden layer, we will have

$$\frac{d^k}{dx^k} Net(x, \mathbf{\Omega}) = \sum_{n=1}^{N} w_n \frac{d^k}{dx^k} (\alpha_n e^{\beta_n x}) = \sum_{n=1}^{N} w_n (\alpha_n \beta_n^k e^{\beta_n x}), \tag{11}$$

3. Proposed learning algorithm

Since the experimental solution is satisfy in the initial conditions, we must choose the network parameters such that the cost function in (5) is minimized.For this purpose, we propose evolutionary methods. In the following, we use genetic algorithm and particle swarm algorithm for this purpose and compare their results. We evaluate the cost function at a number of collocation points $\{x_i \in (a, b)\}$ and take their maximum absolute value as the cost value corresponding to the selected parameters. In other words:

$$E_\infty = maxE(x_i, \mathbf{\Omega}) = \|F(x, y_t, y_t', ..., y_t^{(n)})\| \tag{12}$$

4. Numerical resoults

Consider the following initial value problem

$$\begin{cases} y' + y = 1, 0 < x < 1 \\ y(0) = 2, \end{cases} \tag{13}$$

The exact solution is $y_e(x) = 1 + e^{-x}$ and the trial solution is $y_t(x) = 2 + xNet(x, \mathbf{\Omega})$ where $Net(x, \mathbf{\Omega}) = \sum_{m=1}^{5} \alpha_m e^{-\beta x}$. Therefore, $\frac{d}{dx} Net(x, \mathbf{\Omega}) = -\sum_{m=1}^{5} \alpha_m \beta_m e^{-\beta x}$ and

$$\frac{d}{dx} y_t(x) = Net(x, \mathbf{\Omega}) + x \frac{d}{dx} Net(x, \mathbf{\Omega}) \tag{14}$$

$$E(x,\mathbf{\Omega}) = (Net(x,\mathbf{\Omega}) + x\frac{d}{dx}Net(x,\mathbf{\Omega})) + (2 + xNet(x,\mathbf{\Omega})) - 1$$
$$= 1 + (1+x)Net(x,\mathbf{\Omega}) + x\frac{d}{dx}Net(x,\mathbf{\Omega}) \tag{15}$$

we use the collocation point $x_i = 0.1i, i = 1, 2, ..., 10$ and setting $-2 < \alpha_i < 2$ and $-2 < \beta_i < 2$. Finally, for minimizing the following problem, we use the PSO and GA.

$$\min_{\mathbf{\Omega}}\{\max\{E(x_i,\mathbf{\Omega}), i = 1, 2, ..., 10\}\} \tag{16}$$

We put the optimal parameters obtained from Algorithms PSO and GA into the $Network$ and obtain trial solutions y_t^{PSO} and y_t^{GA}. Figures 1 compare the trial solutions with the exact solution.

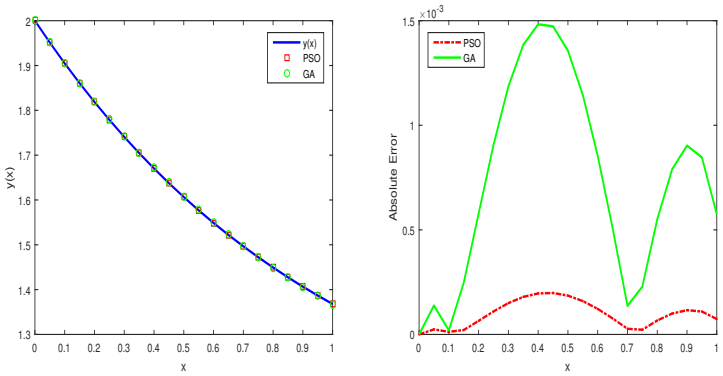


FIGURE 1. Comparison of trial solutions obtained from the PSO and GA with exact solution.

5. Conclusion

In this paper, a three-layer feedforward neural network was proposed to solve initial value problems. A number of collocation points were used to determine the parameters and coefficients of the proposed network. The maximum absolute value of the error at the collocation points was considered as cost function. The particle swarm optimization algorithm PSO and genetic algorithm GA were used to minimize the cost function. The results obtained show the success of the method. Figure 1, compar the exact solution by trial solutions obtained by proposed network where training by PSO and GA. Figure 2, compare absolute error these trial solutions. As can be seen, PSO has very beter performance then GA.

References

1. F. Ahmadkhanpour1, H. Kheiri, N. Azarmir, and F. M. Khiyabani, (2023), *Solving initial value problems using multilayer perceptron artificial neural networks*, Computational Methods for Differential Equations, DOI:10.22034/cmde.2024.58774.2486
2. L. S. Tan, Z. Zainuddin, and P. Ong, (2018) *Solving ordinary differential equations using neural networks*, In AIP Conference Proceedings, AIP Publishing LLC., 1974(1) , 020070.
3. A. Sacchetti, B. Bachmann, K. Loffel, U. M. Kunzi, and B. Paoli, (2022) *Neural networks to solve partial differential equations: a comparison with fnite elements*, IEEE Access, 10 , 32271-32279.

Optimization And Control

Theoretical Perspectives and Open Problems in Generalized Semi-Infinite (Multiobjective) Nonsmooth Optimization

Nader Kanzi*

Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: N.Kanzi@pnu.ac.ir

Abstract. This paper investigates the theoretical aspects of generalized semi-infinite optimization problems (GSIP) in both single- and multiobjective nonsmooth settings. A concise overview of existing results on optimality conditions, subdifferential analysis, and constraint qualifications is presented. The discussion highlights unresolved theoretical issues, including extensions of Karush–Kuhn–Tucker-type conditions, stability analysis, duality, and characterization of efficient solutions under nonsmoothness. Finally, several open problems are identified to motivate further research in (multiobjective) GSIP.

Keywords: Generalized Semi-Infinite Optimization, Nonsmooth Optimization, Multiobjective Programming, Optimality Conditions, Open Problems

AMS Mathematics Subject Classification [2020]: 90C34, 90C29, 49J52

1. Introduction

Generalized Semi-Infinite Programming (GSIP) is a versatile framework for modeling optimization problems that involve infinitely many constraints. In recent years, GSIP has attracted considerable attention due to its applications in optimal control, robust optimization, and hierarchical decision-making; see, e. g., [5]. This paper is connecting three active research fields: nonsmooth analysis, multiobjective optimization, and semi-infinite programming. All these three issues have strong intersections with global optimization because of the possibility of the presence of nonsmooth nonconvex objective/constraint functions.

A multiobjective GSIP (MGSIP) problem can be formulated as

$$(\mathcal{P}): \quad \min_{x\in\mathbb{R}^n} \quad \mathbf{f}(x)=(f_1(x),\ldots,f_p(x)) \quad \text{subject to} \quad g(x,y)\geq 0, \quad \forall y\in Y(x),$$

where $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^p$ represents a vector-valued objective function, $g:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is the upper-level constraint function, and $Y(x)\subseteq\mathbb{R}^m$ is a parameter-dependent index set, defined as

$$Y(x):=\{y\in\mathbb{R}^m \mid h_t(x,y)\leq 0,\ t\in T\},$$

where T is an index set, and $h_t:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, for $t\in T$, are the lower-level constraint functions.

*Speaker.

In practice, many GSIP problems involve nonsmooth functions, for which classical derivatives are inadequate [5]. To analyze optimality in such settings, several notions of subdifferentials are employed. The Clarke subdifferential is suitable for locally Lipschitz functions and local analysis [1], the Mordukhovich subdifferential provides a more general tool for necessary optimality conditions [3], the quasiconvex subdifferentials are useful for exploiting the structure of quasiconvex objectives [4], and the convex subdifferential is helpful for analyze of D.C. (difference of convex) structures [2]. These tools allow rigorous derivation of optimality conditions in nonsmooth and multiobjective settings.

The the following assumptions are *standing* throughout the whole paper:

- The appearing functions f_i, g, and h_t as $(i,t) \in \{1,\dots,p\}\times T$ are locally Lipschitz.
- The index set T is finite.
- The set-valued mapping $x \mapsto Y(x)$ is uniformly bounded; i.e., for each $x_0 \in S$ there exists a neighborhood U of x_0 such that the set $\bigcup_{x\in U} Y(x)$ is bounded.

GSIP problems can often be interpreted as comprising upper- and lower-level structures. Upper-level problems represent the main decision-making process, while lower-level problems appear as parametric constraints or subproblems. This structure is closely related to bilevel optimization, where the feasible set of the upper-level problem depends on the solution of a lower-level problem. Understanding these hierarchical interactions is crucial for both theoretical analysis and applications (see, [5, 6]).

The aim of this paper is to provide a comprehensive theoretical overview of open problems in GSIP with a focus on nonsmooth and multiobjective contexts. We review existing results on subdifferential-based optimality conditions, discuss challenges in extending these conditions to more general nonsmooth structures, and highlight open issues related to upper- and lower-level problem formulations, thereby motivating future research in generalized semi-infinite (multiobjective) optimization.

2. Main results

The feasible set of problem $(\mathcal{P})$ is denoted by S, and the index set of active constraints at each $x_0 \in S$ is defined by $Y_0(x_0) := \{y \in Y(x_0) \mid g(x_0,y) = 0\}$. The lower-level problem at $x_0 \in S$ is

$$LP(x_0): \qquad \min\ g(x_0,y), \quad \text{s.t.} \quad y \in Y(x_0),$$

and the set (probably empty) of active inequalities of $LP(x_0)$ at each $y_0 \in Y(x_0)$ is defined by $T_0(x_0,y_0) := \{t \in T \mid h_t(x_0,y_0) = 0\}$. Let $y_0 \in Y_0(x_0)$. Then, y_0 is a minimizer of the lower-level problem $LP(x_0)$, and by Fritz-John first-order optimality condition [1, Theorem 6.1.1], we can find some multipliers $\alpha \geq 0$ and $\beta := \big(\beta_t,\ t \in T_0(x_0,y_0)\big)$ satisfying $\beta_t \geq 0$ for each $t \in T_0(x_0,y_0)$ and

$$\alpha + \sum_{t\in T_0(x_0,y_0)} \beta_t = 1, \qquad 0 \in \partial_y^c \mathcal{L}_{y_0}^{x_0}(x_0,y_0,\alpha,\beta), \tag{1}$$

where $\partial_y^c \mathcal{L}(x_0,y_0,\alpha,\beta)$ denotes the Clarke subdifferential of $\mathcal{L}(x_0,\cdot,\alpha,\beta)$ at y_0, and $\mathcal{L}_{y_0}^{x_0}$ refers to Lagrangian function, defined as $\mathcal{L}_{y_0}^{x_0}(x,y,\alpha,\beta) := \alpha g(x,y) + \sum_{t\in T_0(x_0,y_0)} \beta_t h_t(x,y)$. Let $F(x_0,y_0) := \big\{(\alpha,\beta) \mid (\alpha,\beta) \text{ fulfills } (1)\big\}$ be the Fritz-John (FJ) multipliers set of $LP(x_0)$ at $y_0 \in Y_0(x_0)$. For each $x_0 \in S$ put

$$\mathfrak{D}(x_0) := \bigcup_{y\in Y_0(x_0)} \Big(\bigcup_{(\alpha,\beta)\in F(x_0,y)} \partial_x^c \mathcal{L}_y^{x_0}(x_0,y,\alpha,\beta) \Big).$$

The following assumption is totally characterized in the complete paper:

Assumption A: The set-valued mapping $y \mapsto \partial_x^c \mathcal{L}_y^{x_0}(x_0, y, \alpha, \beta)$ is upper semi-continuous and

$$\partial^c\Big(\inf_{y \in Y(\cdot)} g(\cdot, y)\Big)(x_0) \subseteq conv\big(\mathfrak{D}(x_0)\big),$$

for each $x_0 \in S$, where $conv(A)$ denotes the convex hull of $A \subseteq \mathbb{R}^n$.

LEMMA 2.1. *Suppose that Assumption A is satisfied. Then, $\mathfrak{D}(\hat{x})$ is a compact set for all $\hat{x} \in S$.*

THEOREM 2.2. **(FJ Necessary Condition):** *Suppose that $\hat{x}$ is a weakly efficient solution for $(\mathcal{P})$ and that Assumption A is satisfied.*

(i):: *If $Y_0(\hat{x}) \neq \emptyset$, then there exist finitely many indices $y^1, \ldots, y^q \in Y_0(\hat{x})$, Fritz John multipliers $(\alpha^\nu, \beta^\nu) \in F(\hat{x}, y^\nu)$ as $\nu = 1, \ldots, q$, as well as non-negative scalars $\lambda_i \geq 0$ as $i = 1, \ldots, p$ and $\mu_\nu \geq 0$ as $\nu = 1, \ldots, q$, satisfying*

$$0 \in \sum_{i=1}^{p} \lambda_i \partial^c f_i(\hat{x}) - \sum_{\nu=1}^{q} \mu_\nu \partial_x^c \mathcal{L}_{y^\nu}^{\hat{x}}(\hat{x}, y^\nu, \alpha^\nu, \beta^\nu), \tag{2}$$

$$\sum_{i=1}^{p} \lambda_i + \sum_{\nu=1}^{q} \mu_\nu = 1.$$

(ii):: *If $Y_0(\hat{x}) = \emptyset$, then there exist non-negative scalars $\lambda_i \geq 0$ as $i = 1, \ldots, p$, such that*

$$0 \in \sum_{i=1}^{p} \lambda_i \partial^c f_i(\hat{x}), \qquad \text{and} \qquad \sum_{i=1}^{p} \lambda_i = 1.$$

PROOF. Because the proof relies on preliminaries that are beyond the scope of this extended abstract, we refer the reader to the full paper for its detailed presentation. □

As we know, Theorem 2.2 can be true even when $\lambda_i = 0$, for all $i = 1, \ldots, p$, and the vector valued objective function disappears from the main inclusion (2). For this reason, we state a Karush-Kuhn-Tucker (KKT) type necessary optimality condition as follows. We know from classical optimization that a KKT type necessary condition requires a suitable constraint qualification. As generalization of [6, Theorem 3.1], we set the following theorem under a Mangasarian-Fromovitz type constraint qualification.

THEOREM 2.3. **(KKT Necessary Condition):** *Suppose that $\hat{x}$ is a weakly efficient solution for $(\mathcal{P})$, that Assumption A is satisfied, and that the following constraint qualification holds:*

$$\Big\{z \in \mathbb{R}^n \mid \langle z, d\rangle \geq 0, \quad \forall d \in \mathfrak{D}(\hat{x})\Big\} \neq \emptyset.$$

Then, there exist some $y^\nu \in Y_0(\hat{x})$ and $(\alpha^\nu, \beta^\nu) \in F(\hat{x}, y^\nu)$ as $\nu = 1, \ldots, q$, and some non-negative scalars $\lambda_i \geq 0$ as $i = 1, \ldots, p$ and $\mu_\nu \geq 0$ as $\nu = 1, \ldots, q$, satisfying $\sum_{i=1}^{p} \lambda_i = 1$ and

$$0 \in \sum_{i=1}^{p} \lambda_i \partial^c f_i(\hat{x}) - \sum_{\nu=1}^{q} \mu_\nu \partial_x^c \mathcal{L}_{y^\nu}^{\hat{x}}(\hat{x}, y^\nu, \alpha^\nu, \beta^\nu).$$

PROOF. The proof depends on certain preliminaries that cannot be fully covered in this extended abstract; we therefore present it in the full paper. □

In almost all examples, some of the multipliers λ_i as $i = 1, \ldots, p$, named λ_k, may be equal to zero in Theorem 2.3, and the $k-$component of the vector-valued objective function does not play the role in the necessary condition. We assert that strong KKT condition holds for a multiobjective optimization problem, when the KKT multipliers are positive for all components of the objective function. In the below theorem, we establish the strong KKT necessary conditions for weakly efficient solution of $(\mathcal{P})$ under a Cottle type constraint qualification (that is stronger than Mabgasarian-Fromovitz constraint qualification).

THEOREM 2.4. **(Strong KKT Necessary Condition):** *Suppose that $\hat{x}$ is a weakly efficient solution for $(\mathcal{P})$, that Assumption A is satisfied, and that the following constraint qualification holds:*

$$\Big\{z \in \mathbb{R}^n \mid \langle z, d\rangle \geq 0, \quad \forall d \in \mathfrak{D}(\hat{x}) \cup \bigcup_{i=1}^{p} \partial^c f_i(\hat{x})\Big\} \neq \emptyset.$$

Then, there exist some $y^\nu \in Y_0(\hat{x})$ and $(\alpha^\nu, \beta^\nu) \in F(\hat{x}, y^\nu)$ as $\nu = 1, \ldots, q$, and some positive scalars $\lambda_i > 0$ as $i = 1, \ldots, p$ and $\mu_\nu \geq 0$ as $\nu = 1, \ldots, q$, satisfying $\sum_{i=1}^{p} \lambda_i = 1$ and

$$0 \in \sum_{i=1}^{p} \lambda_i \partial^c f_i(\hat{x}) - \sum_{\nu=1}^{q} \mu_\nu \partial_x^c \mathcal{L}_{y^\nu}^{\hat{x}}(\hat{x}, y^\nu, \alpha^\nu, \beta^\nu).$$

PROOF. Due to the need for foundational material, the proof is omitted here and is provided in detail in the full paper. □

3. Conclusion

In this paper, we presented three essential optimality conditions for nonsmooth multiobjective semi-infinite problems. The significance and novelty of these results were highlighted. In an upcoming presentation, these conditions will be compared with previously established results, and several open problems in the field will be discussed. These insights aim to guide and inspire future research in this area.

Acknowledgement

We warmly thank the conference organizers and reviewers for their guidance and thoughtful feedback.

References

1. Clarke, F.H. (1983), *Optimization and Nonsmooth Analysis*, Wiley-Interscience.
2. Kanzi, N. (2013) *Lagrange multiplier rules for non-differentiable DC generalized semi-infinite programming problems*, J. Glob. Optim., **56**(2), 417–430.
3. Kanzi, N., & Nobakhtian, S. (2010) *Necessary optimality conditions for nonsmooth generalized semi-infinite programming problems*, Eur. J. Oper. Res., **205**(2), 253–261.
4. Kanzi, N., & Soleimani-damaneh, M. (2020) *Characterization of the weakly efficient solutions in nonsmooth quasiconvex multiobjective optimization*, J. Glob. Optim., **77**(3), 627–641.
5. Stein, O. (2003), *Bi-level strategies in semi-infinite programming*, Kluwer, Boston.
6. Vazquez, F.G., Rebollar, L.A.H., & Ruckmann, J.J. (2018) *On Vector Generalizrd Semi-infinite Programming*, Rev. Investig. Oper., **39**(2), 341–352.

Numerical method for time-invariant delay optimal control problems

Seyed Mehdi Mirhosseini-Alizamini[1,*]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: m_mirhosseini@pnu.ac.ir

Abstract. This paper, the Adomian decomposition method and variational iteration method are applied to obtain suboptimal control for linear time-varying systems with multiple state and control delays and with quadratic cost functional. The optimal control law obtained consists of an accurate linear feedback term and a nonlinear compensation term which is the limit of an adjoint vector sequence. The feedback term is determined by solving Riccati matrix differential equation. Through the finite iterations of algorithm, a suboptimal control law is obtained for the nonlinear optimal control problem.

Keywords: Time-delay systems; Pontryagin's maximum principle; Adomian decomposition method; Variational iteration method.

AMS Mathematics Subject Classification [2020]: 58E11, 53B30, 53C50

1. Introduction

The dynamics of many control systems may be expressed by time-delay equations. The delays may appear in the system state, control input and/or output. Delays occur frequently in incubation periods, mechanics, viscoelasticity, physics, physiology, population dynamics, communication, information technologies, stability of networked control systems, maturation times, age structure, blood transfusions, biological, chemical, electronic and transportation systems [1]- [4]. Therefore the control of time-delay systems has been interested by many engineers and scientists, due to its variety presence in realistic models of phenomena. On the other hand, in the context of numerical analysis, the Adomian Decomposition Method (ADM) which was proposed originally by Adomian, has been proved by many authors to be a powerful mathematical tool for various kinds of linear and nonlinear ODE's or PDE's. Unlike the traditional numerical methods, ADM needs no discretization, linearization, transformation or perturbation. The method, has been widely applied to solve nonlinear problems, and different modifications are suggested to overcome the demerits arising in the solution procedure. The aim of present paper to introduce a numerical method to solve the quadratic optimal control problem with delay systems. The optimal control law obtained consists of an accurate linear feedback term and a nonlinear compensation term which is the limit of an adjoint vector sequence.

*Speaker.

2. Main Results

Consider a time-varying delays system described by:

$$(1)\quad \begin{cases} \dot{x}(t) = A(t)x(t) + A_1(t)x(t-\tau_x) + B(t)u(t) + B_1(t)u(t-\tau_u), & t \geq t_0, \\ x(t) = \phi(t), & t_0 - \tau_x \leqslant t \leqslant t_0, \\ u(t) = \psi(t), & t_0 - \tau_u \leqslant t \leqslant t_0, \end{cases}$$

where $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$, are the state and control vectors respectively; $A(t)$, $A_1(t)$, $B(t)$ and $B_1(t)$ are real, piecewise continuous matrices of appropriate dimensions defined on the appropriate intervals; $\phi(t)$ and $\psi(t)$ are specified initial functions; τ_x and τ_u are constant positive scalars. Here, it is assumed that the system (1) is controllable and assume that $\tau_u < \tau_x$. The objective is to find the optimal control law $u^*(t)$ over $t \in [t_0, t_f]$, which minimizes the following quadratic cost functional subject to the system (1)

$$(2)\quad J = \frac{1}{2}x^T(t_f)Q_f x(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)\right) dt,$$

where, the matrix $Q_f \in \mathbb{R}^{n\times n}$ is symmetric positive semi-definite, the matrix $Q(t) \in \mathbb{R}^{n\times n}$ is symmetric positive semi-definite and piecewise continuous for $t_0 \leq t \leq t_f$, and the matrix $R(t) \in \mathbb{R}^{m\times m}$ is symmetric positive definite and piecewise continuous for $t_0 - \tau_u \leq t \leq t_f$.

The Hamiltonian function for the problem is

$$(3)\quad \begin{aligned} \mathcal{H}(x,u,\lambda,t) &= \frac{1}{2}x^T(t)Q(t)x(t) + \frac{1}{2}u^T(t)R(t)u(t) \\ &+ \lambda^T(t)[A(t)x(t) + A_1(t)x(t-\tau_x) + B(t)u(t) + B_1(t)u(t-\tau_u)], \end{aligned}$$

where $\lambda(t) \in \mathbb{R}^n$ is the Lagrange multiplier vector corresponding to dynamic equality constraint (1). According to the necessary conditions for optimality, we can obtain the following nonlinear TPBVP [4]:

$$(4)\ \dot{x}(t) = \begin{cases} A(t)x(t) + A_1(t)x(t-\tau_x) - (S_1(t) + S_2(t))\lambda(t) - S_3(t)\lambda(t+\tau_u) \\ -S_4(t)\lambda(t-\tau_u), \quad t_0 \leqslant t < t_f - \tau_u, \\ A(t)x(t) + A_1(t)x(t-\tau_x) - S_1(t)\lambda(t) - S_4(t)\lambda(t-\tau_u), \quad t_f - \tau_u \leqslant t \leqslant t_f, \end{cases}$$

and

$$(5)\quad \dot{\lambda}(t) = \begin{cases} -Q(t)x(t) - A^T(t)\lambda(t) - A_1^T(t+\tau_x)\lambda(t+\tau_x), & t_0 \leqslant t < t_f - \tau_x, \\ -Q(t)x(t) - A^T(t)\lambda(t), & t_f - \tau_x \leqslant t \leqslant t_f, \end{cases}$$

with initial conditions

$$(6)\quad \begin{cases} x(t) = \phi(t), & t_0 - \tau_x \leq t \leqslant t_0, \\ u(t) = \psi(t), & t_0 - \tau_u \leqslant t \leqslant t_0, \\ \lambda(t_f) = Q_f x(t_f), \end{cases}$$

where

$$\begin{aligned} S_1(t) &= B(t)R^{-1}(t)B^T(t) \\ S_2(t) &= B_1(t)R^{-1}(t-\tau_u)B_1^T(t) \\ S_3(t) &= B(t)R^{-1}(t)B_1^T(t+\tau_u) \\ S_4(t) &= B_1(t)R^{-1}(t-\tau_u)B^T(t-\tau_u), \end{aligned}$$

$x(t-\tau)$ is time-delay term and $\lambda(t+\tau)$ is time-advance term, furthermore $\lambda(t) \in \mathbb{R}^n$ is the co-state vector. Also, the optimal control law is given by:

$$(7)\quad u^*(t) = \begin{cases} -R^{-1}(t)B^T(t)\lambda(t) - R^{-1}(t)B_1^T(t+\tau_u)\lambda(t+\tau_u), & t_0 \leqslant t < t_f - \tau_u \\ -R^{-1}(t)B^T(t)\lambda(t), & t_f - \tau_u \leqslant t \leqslant t_f. \end{cases}$$

Note that, Eqs. (2)-(4) form a linear TPBVP with time-varying coefficient involving both delay and advance terms. The exact solution of this problem is, in general, extremely difficult, if not impossible. To overcome this difficulty, an iterative approach, based on the ADM and VIM, will be introduced in the next section.

We may find the suboptimal control law in practical applications by replacing infinite with a finite positive integer N. Thus, the Nth order suboptimal control law is obtained as follows:

$$(8)\quad u_N(t) = \begin{cases} -R^{-1}(t)B^T(t)[P(t)x_N(t) + g_N(t)] - R^{-1}(t)B_1^T(t+\tau_u)[P(t+\tau_u)x_N(t+\tau_u) \\ +g_N(t+\tau_u)], \quad t_0 \leqslant t < t_f - \tau_u \\ -R^{-1}(t)B^T(t)[P(t)x_N(t) + g_N(t)], \quad t_f - \tau_u \leqslant t \leqslant t_f. \end{cases}$$

In (8) $x_N(t)$ is an approximation of x_∞ with finite-step iteration, and g_N is an approximation by substituting a finite-step iteration of g_N for g_∞. We let $N = k$ and the Nth order suboptimal control law from (8). Then, the following quadratic performance index can be calculated:

$$(9)\quad J_N = \frac{1}{2}x_N^T(t_f)Q_f x_N(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x_N^T(t)Q(t)x_N(t) + u_N^T(t)R(t)u_N(t)\right) dt,$$

where $u_N(t)$ has been obtained from (8) and $x_N(t)$ is the corresponding state trajectory obtained from applying $u_N(t)$ to the original TDOCP in (1).
For the accuracy analysis, we consider the following criterion. The suboptimal control law has the desirable accuracy, if for given positive constants $\epsilon > 0$, the following condition hold jointly:

$$(10)\quad \left|\frac{J_N - J_{N-1}}{J_N}\right| < \epsilon,$$

If the tolerance error bound be chosen small enough, the Nth order suboptimal control law will be very close to the optimal control law $u^*(t)$, and thus, the value of quadratic performance index in (9) and its optimal value J^* will be almost identical.

Algorithm: Suboptimal control law of system (1)

Step 1: Obtain the positive-semidefinite matrix $P(t)$ from Riccati matrix differential equation. Let $x_0(t) = x(t_0) = \phi(t), g_0(t) = g(t_0)$ and $k = 1$.

Step 2: Compute $x_k(t)$ and $g_k(t)$ using ADM or VIM method. Store these values.

Step 3: Letting $N = k$, calculate $u_N(t)$ from Eq. (8).

Step 4: Calculate J_N according to (9). If $\left|\frac{J_N - J_{N-1}}{J_N}\right| < \epsilon$, then stop and output $u_N(t)$, go to step 5; else, replace k by $k+1$ and go to step 2.

TABLE 1. Simulation results of Example 4.4 at different iteration times.

k (iteration time)	J_k	$\left\lvert\frac{J_k - J_{k-1}}{J_k}\right\rvert$
1	22.12432	-
2	22.02267	0.04215
3	22.02254	0.06390
4	22.02231	0.00232

Step 5: Stop the algorithm; set $u_N(t)$ is the desirable close-loop suboptimal control law.

3. Numerical Example

Consider the following time-varying system with both state and control delays:

$$\begin{cases} \dot{x}_1(t) = x_2(t) + x_1(t-1), & t \geq 0, \\ \dot{x}_2(t) = tx_1(t) + 2x_1(t-1) + x_2(t-1) + u(t) - u(t-0.5), & t \geq 0 \\ x_1(t) = x_2(t) = 1, & -1 \leqslant t \leqslant 0, \\ u(t) = 5(t+1), & -0.5 \leqslant t \leqslant 0, \end{cases}$$

with the cost functional

$$J = \frac{1}{2}x_1^2(3) + x_2^2(3) + \frac{1}{2}\int_0^3 [2x_1^2(t) + 2x_1(t)x_2(t) + x_2^2(t) + \frac{u^2(t)}{(t+2)}]dt,$$

Here we solve this problem by using the suggested algorithm with the tolerance error bounds $\epsilon = 4 \times 10^{-4}$. From Table (1) it is observed that, the convergence is achieved after only four iterations, i.e. $\left|\frac{J_4 - J_3}{J_4}\right| = 1.125 \times 10^{-4} < 4 \times 10^{-4}$, and a minimum value of $J_4 = 22.02231$ is obtained.

4. Conclusion

In this paper, the optimal control obtained consists of both feedback and forward portions. The procedure converts the solution of a coupled TPBVP with advance and delay terms into the solution of a single Riccati differential equation and a sequence of ODEs. The feedback term is determined by solving Riccati matrix differential equation. By using the ADM and VIM with the finite-step iteration of a nonlinear compensation sequence. We used four examples to demonstrate the validity and applicability of the method.

References

1. Dadebo, S. and Luus, R. (1992), *Optimal control of time-delay systems by dynamic programming*, Optim. Control. Appl. Methods., **13** (1), 29–41.
2. Jabbari-Khanbehbin, T. (2022) *Shooting continuous Runge–Kutta method for delay optimal control problems*, Iranian Journal of Numerical Analysis and Optimization, **12** (3), 680–703.
3. Khaledi, G. (2025) *LQR technique based SMC design for a class of uncertain time-delay Conic nonlinear systems*, Computational Methods for Differential Equations, **13** (2), 505–523.
4. Golman, G. (2021) *Optimal control of a delayed alcoholism model with saturated treatment*, Differ. Equ. Dyn. Syst., **14**, 1–16.

A Genetic Optimization Approach to Brachytherapy Dose Rate Planning with Fuzzy Constraints

Mohammad Mohammadi Najafabadi[*1], Department of Mathematics,
Payame Noor University, Tehran, Iran. mm.najafabadi@pnu.ac.ir
Davood Darvishi, Department of Mathematics, Payame Noor University,
Tehran, Iran. d.darvishi@pnu.ac.ir

Abstract: This study addresses the complexity of high-dose-rate brachytherapy (HDR-BT) planning by proposing a fuzzy-constrained integer programming (FIP) model. The model integrates fuzzy set theory with clinical dosimetric indices to handle uncertainty and enhance flexibility in treatment design. To solve the optimization problem, three evolutionary algorithms—Genetic Algorithm (GA), Evolutionary Programming (EP), and Genetic Programming (GP)—were applied to prostate cancer cases. Results show that the FIP framework produces clinically acceptable and patient-specific treatment plans. GA proved the fastest in achieving acceptable dose coverage, while GP delivered the most effective tumor overage overall.

Keywords: Brachytherapy, Optimization, Genetic Algorithm, Fuzzy Logic, Cancer.

1. Introduction

HDR brachytherapy irradiates the prostate through 14–20 catheters by modulating dwell times at fixed positions, demanding inverse optimization that maximizes CTV dose while satisfying OAR dose–volume constraints[8]. The optimization problem is constrained by clinical thresholds (e.g., V100 ≥ 96 % for CTV, D2cc < 74 % for rectum) and must be solved rapidly (< 1 h) for theatre workflow. Although the study focuses on prostate HDR, the dwell-time optimization framework is equally applicable to LDR or other anatomical sites[3].

2. Definition of model

In HDR-BT, the goal is to maximize dose coverage of the CTV while adhering to DVH constraints that keep OAR doses below clinical thresholds[5].

- V_d^o Criteria specify how large the cumulative volume of an organ o receiving at least the radiation dose level d (relative to the planning-aim dose) should be.
- D_v^o Criteria specify how high the radiation dose level that covers the most-radiated cumulative volume v of an organ o should be.

Table 1 lists the required DV indices; the planner must therefore tune dwell times to maximize target coverage while minimizing dose to healthy tissues[1].

Table 1: Dosimetric criteria for treatment with high dose brachytherapy

Prostate	***Bladder***	***Rectum***	***Urethra***	***Vesicles***
$V_{100} > 96\%$	$D_{1cc} < 86\%$	$D_{1cc} < 78\%$	$D_{0.1cc} < 110\%$	$\boldsymbol{V_{80} > 95\%}$
$V_{150} < 50\%$	$D_{2cc} < 74\%$	$D_{2cc} < 74\%$		
$V_{200} < 20\%$				

3. Model Formulation

In this section, a complex integer program (IP) model with fuzzy constraints is presented for programming on high-dose brachytherapy. Table 2 describes the infrastructure, parameters, and variables in this model[2].

Table 2: Definitions parameters and variables of the model IP.

parameters and variables	Definitions
S	set for organs
I	Set for dose points
J	Set for dwell positions
G_s	Set of dose values in organ $\boldsymbol{S}$
P_{si}	Three-dimensional coordinates of dose point in $\boldsymbol{G_s}$
N_s	The number of dose points in $\boldsymbol{G_s}$
T_j	Three-dimensional coordinates of position $\boldsymbol{J}$
N_T	The number of dwell positions for the patient
D_{sij}	The dose rate transferred from $\boldsymbol{T_j}$ to $\boldsymbol{P_{si}}$
t_j	Dwell time at $\boldsymbol{T_j}$
D_{si}	Dose rate in $\boldsymbol{P_{si}}$
R_s	The tolerance threshold value for $\boldsymbol{G_s}$
M_s	Maximum dose for $\boldsymbol{G_s}$
X_{si}	Index variable for $\boldsymbol{P_{si}}$
V_S	Dosimetric index for $\boldsymbol{G_s}$
L_s	Lower bound for $\boldsymbol{V_S}$
U_s	Upper bound for $\boldsymbol{V_S}$

Stop times are continuous variables that represent the source stop time in T_j. The total dose received in $\boldsymbol{P_{si}}$ is equal to the average dose received from dwell position.

$$D_{si} = \frac{1}{N_T}\sum_{j=1}^{N_T} D_{sij}t_j$$

Indicator variables are binary variables that should behave as follows:

$$X_{si} = \begin{cases} 1 & D_{si} \geq R_s \\ 0 & otherwise \end{cases}$$

Finally, the dosimeter index for $\boldsymbol{G_s}$ is equal to the sum of all index variables, such that

$$V_S = \frac{1}{N_s}\sum_{i=1}^{N_s} X_{si}$$

Given the nature of such problems, two objectives are considered here. The first goal is to maximize the target coverage. Which is as follows:

$$f_o(t_j) = \sum_{j=1}^{N_T} D_{oij}t_j \tilde{\gg} R_o + \delta \qquad \forall\, i \in G_o \qquad (7)$$

In the above equation, inequality $\tilde{\gg}$ is fuzzy. The second goal is to reduce the exposure of the endangered organs to the problem as a limitation.

$$f_s(t_j) = \sum_{j=1}^{N_T} D_{sij}t_j \tilde{\ll} M_s - \varepsilon \qquad \forall\, s,i \in G_s \qquad (8)$$

Again, in the above limit, inequality $\tilde{\ll}$ is fuzzy. Because organs at risk need not be exposed to radiation as much as possible. In both of these limitations, t_j is a dwell time and should not exceed one hour, causing damage to healthy cells[7]. In such cases, the optimization is performed on the dwell time variables. And we have

$$0 \leq t_j \leq 3600 \quad second \qquad \forall j \qquad (9)$$

The proposed heuristic converts the fuzzy-integer program into a fast-solvable linear formulation while retaining clinical dose–volume limits.

$$\begin{aligned} \max \quad & f(t_j) = \sum_{i=1}^{N_0} X_{0i} + \sum_{s=0}^{T} w_s\, \mu_s(t_j) \\ s.t. \quad & \sum_{j=1}^{N_T} D_{oij} t_j - \delta \times \mu_o(t_j) \geq R_o \\ & (\mu_s(t_j) - 1) \times \varepsilon + f_s(t_j) \leq M_s - \varepsilon \\ & \sum_{s=0}^{T} w_s = 1 \\ & 0 \leq X_{si} \leq 1 \qquad \forall\, i \in G_0 \\ & 0 \leq t_j \leq 3600 \quad second \quad \forall j \end{aligned}$$

4. Evolutionary algorithm

These three evolutionary algorithms—Genetic Algorithm, Evolutionary Programming, and Genetic Programming—each evolve a population of candidate solutions through distinct mechanisms such as crossover, mutation, and selection to identify optimal solutions. [9].

5. Implementation

Evolutionary algorithms (GA, EP, GP) were applied to optimize dwell times across 20 patients, each executed 20 times within one hour. The results provide comparative performance insights, visualized through solution sets plotted for clinical evaluation.

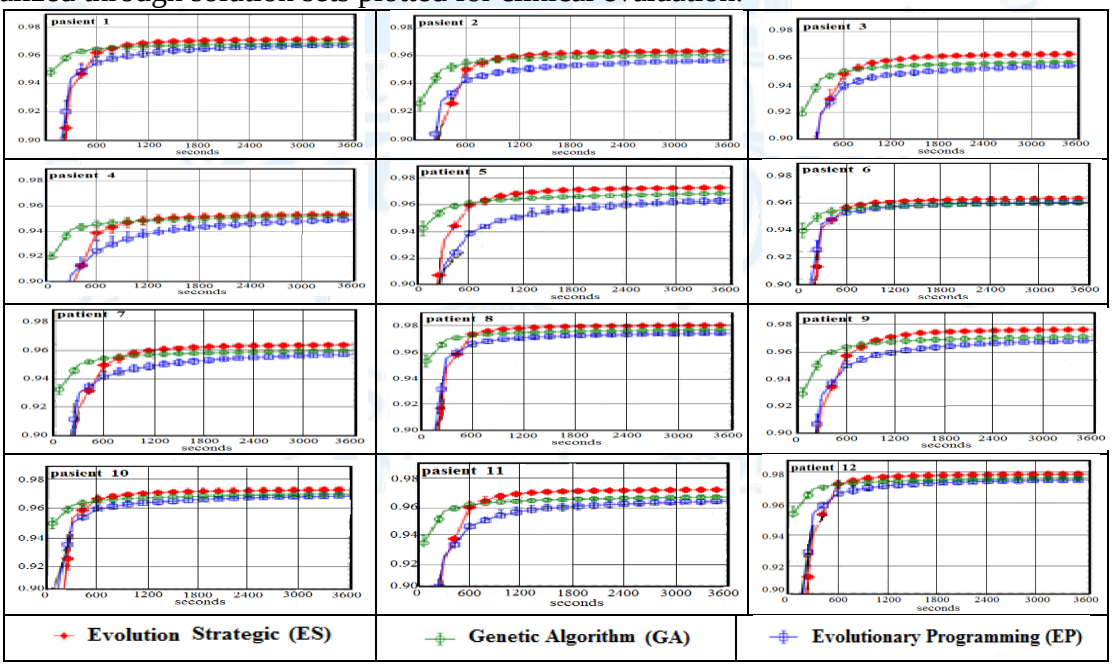


Figure 4: 20 runs/patient (<1 h) quantify each algorithm's performance.

6. Results

BT planning demands high-dimensional, one-hour optimisation that simultaneously covers the target yet spares OARs under strict DV limits, exceeding the capability of conventional simplified models.

Table 2: All three evolutionary algorithms covered >96 % of the tumor volume in minimal time.

Patient ID	GA	EP	GP
1	0.955372 (506 *second*)	0.960003 (853 *second*)	**0.960089 (600 *second*)**

2	0.960007 (1973 *second*)	*	**0.960033 (1297 *second*)**
3	0.960001 (600 *second*)	0.960000 (2400 *second*)	**0.960000 (600 *second*)**
4	0.960000 (2400 *second*)	0.960000 (2400 *second*)	**0.960000 (853 *second*)**
5	0.967947 (214 *second*)	0.96 (403 *second*)	**0.960000 (403 *second*)**
6	0.96 (411 *second*)	0.96 (1183 *second*)	**0.965743 (772 *second*)**
7	0.96 (297 *second*)	0.96 (600 *second*)	**0.968237 (600 *second*)**
8	0.96 (448 *second*)	0.96 (1500 *second*)	**0.960000 (600 *second*)**
9	0.963984 (228 *second*)	0.962718 (600 *second*)	**0.974823 (600 *second*)**
10	0.96 (1493 *second*)	*	**0.96000 0(1352 *second*)**
11	0.960015 (287 *second*)	0.960108 (600 *second*)	**0.970000 (600 *second*)**
12	0.96 (442 *second*)	0.96 (1135 *second*)	**0.962548 (772 *second*)**

The results indicate that all three evolutionary algorithms (GA, EP, GP) can achieve target coverage above 96% for prostate cancer HDR-BT planning, though with varying efficiency. Instances marked with * denote cases where no algorithm achieved the required coverage[4].
Comparative results show that all three evolutionary algorithms (GA, EP, GP) achieved high tumor coverage, with GP providing the highest value (0.974823) among tested cases. Overall, GP is identified as the most effective method for maximizing tumor volume coverage in HDR-BT planning.

7. Conclusion

Three mainstream evolutionary algorithms were benchmarked; GP consistently achieved >96 % target coverage in the shortest runtime and enclosed the largest tumor volume. Fuzzy set theory effectively reconciles the inherent trade-offs and enables integration of additional patient factors, yielding clinically selectable plans. Consequently, GP is recommended for rapid, high-coverage BT planning, and fuzzy multi-objective extensions are encouraged for future protocols.

References

[1] Bao, S., Wang, Y., Hu, Y., Lin, S., Wang, Y., Yuan, K., & Zuo, Z. (2025). Research trends in cervical cancer brachytherapy: a bibliometric analysis. *Discover Oncology*, *16*(1), 1512.
[2] Cao, R., Si, L., Li, X., Guang, Y., Wang, C., Tian, Y., & Zhang, X. (2022). A conjugate gradient-assisted multi-objective evolutionary algorithm for fluence map optimization in radiotherapy treatment. *Complex & Intelligent Systems*, *8*(5), 4051-4077.
[3] Gao, Y., Shen, C., Jia, X., & Park, Y. K. (2023). Implementation and evaluation of an intelligent automatic treatment planning robot for prostate cancer stereotactic body radiation therapy. *Radiotherapy and Oncology*, *184*, 109685.
[4] Kim, H., Lee, M., & Choi, H. J. (2025). Improving treatment quality of gynecologic cancers with online adaptive brachytherapy using deep learning-based CT imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *1072*, 170149.
[5] Mohammadi Najafabadi, M., Nazif, H., & Soltanian, F. (2023). Dose Optimization in a Fuzzy Model of High-Dose Rate Brachytherapy Problem. *Control and Optimization in Applied Mathematics*, *8*(2), 33-47.
[6] Mohammadi, N. M., Nazif, H., & Soltanian, F. (2022). Optimization of Fuzzy Model of High Dose Brachytherapy Problem for the Treatment of Prostate Cancer Using Evolutionary Algorithms. *Razi Journal of Medical Sciences* 29 (1), 84-95.
[7] Wang, T., Feng, Y., Beaudry, J., Nunez, D. A., Gorovets, D., Kollmeier, M., & Damato, A. L. (2025). Instant plan quality prediction on transrectal ultrasound for high-dose-rate prostate brachytherapy. *Brachytherapy*, *24*(1), 171-176.
[8] Xun, S., Li, Q., Liu, X., Huang, P., Zhai, G., Sun, Y., & Tan, T. (2025). Charting the path forward: CT image quality assessment-an in-depth review. *Journal of King Saud University Computer and Information Sciences*, *37*(5), 1-24.
[9] Yang, Y., Ma, Y., Zhao, Y., Zhang, W., & Wang, Y. (2024). A dynamic multi-objective evolutionary algorithm based on genetic engineering and improved particle swarm prediction strategy. *Information Sciences*, *660*, 120125.

Piecewise orthogonal function neural network: A general framework for function approximation

Ghasem Ahmadi*

Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: g.ahmadi@pnu.ac.ir

Abstract. Accurate approximation and modeling of nonlinear dynamic systems remain a central challenge in computational intelligence and control theory. This paper introduces a general form of neural network architectures termed the Piecewise Orthogonal Function Neural Networks (POFNNs), which integrate the concept of orthogonal functional bases with localized piecewise representation. In the proposed framework, the input domain is partitioned into several subregions, each associated with a set of orthogonal basis functions that form locally independent subspaces. This structure enables the network to capture distinct nonlinear behaviors in different regions while preserving numerical stability and interpretability. The results confirm that the proposed framework provides a flexible, stable, and mathematically interpretable foundation for advanced neural modeling of nonlinear processes.

Keywords: neural network, piecewise orthogonal functions, function approximation
AMS Mathematics Subject Classification [2020]: 68T05, 62M10

1. Introduction

Neural networks have emerged as powerful tools for modeling complex nonlinear systems, performing function approximation, and forecasting time series data. Despite their remarkable success, conventional neural architectures such as multilayer perceptrons (MLPs) and radial basis function (RBF) networks often face challenges related to convergence speed, overfitting, and the requirement for extensive training data. To address these limitations, researchers have explored the integration of orthogonal basis functions within neural network frameworks, leading to the development of neural networks based on piecewise orthogonal functions (POFNNs) [1–3].

The use of piecewise orthogonal functions introduces a mathematically rigorous representation that enhances the approximation capability and numerical stability of the network. By decomposing the input domain into subregions and employing orthogonal basis functions within each segment, POFNNs achieve local adaptivity and global smoothness simultaneously. This structure enables the network to efficiently capture abrupt changes and discontinuities in nonlinear mappings.

*Speaker.

Moreover, the orthogonality property of the basis functions reduces redundancy in the representation and simplifies the learning process by decoupling the network parameters. This not only accelerates convergence but also improves generalization performance. Consequently, POFNNs have found applications in diverse fields such as system identification, signal processing, function approximation, and time series prediction. Therefore, The integration of piecewise orthogonal functions into neural network design represents a promising direction in computational intelligence.

2. Piecewise Orthogonal Functions: Theoretical Foundations

2.1. Definition and Motivation. Orthogonal functions play a crucial role in numerical analysis, signal representation, and approximation theory due to their ability to form complete and nonredundant bases in function spaces. However, when modeling nonlinear systems or functions with local discontinuities, global orthogonal functions (such as Legendre, Chebyshev, or Fourier bases) often fail to provide efficient local representation. To overcome this limitation, piecewise orthogonal functions (POFs) are introduced as locally defined, mutually orthogonal basis functions over subdivided intervals of the input domain.

Consider a bounded domain $D = [a, b] \subset \mathbb{R}$. Let this domain be partitioned into M nonoverlapping subintervals:

$$D = \bigcup_{m=1}^{M} D_m, \qquad D_m = [x_{m-1}, x_m], \qquad a = x_0 < x_1 < \cdots < x_M = b. \tag{1}$$

Within each subdomain D_m, a set of locally orthogonal basis functions $\{\phi_{m,k}(x)\}_{k=1}^{K_m}$ is defined, satisfying the orthogonality condition

$$\int_{D_m} \phi_{m,i}(x)\, \phi_{m,j}(x)\, w_m(x)\, dx = \begin{cases} 0, & i \neq j, \\ \alpha_{m,i}, & i = j, \end{cases} \tag{2}$$

where $w_m(x)$ is a positive weighting function and $\alpha_{m,i} > 0$ are normalization constants.

2.2. Construction of Piecewise Orthogonal Bases. The general form of a POF over the entire domain D can be written as

$$\Phi_{m,k}(x) = \begin{cases} \phi_{m,k}(x), & x \in D_m, \\ 0, & x \notin D_m. \end{cases} \tag{3}$$

This construction ensures local support (each function is nonzero only in its subdomain) and global orthogonality, since for any two functions $\Phi_{m,i}$ and $\Phi_{n,j}$,

$$\int_a^b \Phi_{m,i}(x)\, \Phi_{n,j}(x)\, w(x)\, dx = \begin{cases} \alpha_{m,i}, & m = n,\ i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Hence, the complete set $\{\Phi_{m,k}(x)\}$ forms an orthogonal basis over D.

2.3. Examples of Orthogonal Function Families. Various classical orthogonal functions can be adapted into the piecewise framework. For instance:

- Piecewise Legendre functions [4]:

$$\phi_{m,k}(x) = P_k\left(\frac{2(x - x_{m-1})}{x_m - x_{m-1}} - 1\right), \tag{5}$$

where $P_k(\cdot)$ denotes the Legendre polynomial of degree k.

- Piecewise hyperbolic functions [2]:

$$\phi_{m,k}(x) = \sinh\left(\frac{k(x - x_{m-1})}{x_m - x_{m-1}}\right), \qquad \psi_{m,k}(x) = \cosh\left(\frac{k(x - x_{m-1})}{x_m - x_{m-1}}\right). \tag{6}$$

2.4. Approximation Property. A square-integrable function $f(x) \in L^2([a, b])$ can be expanded in terms of the piecewise orthogonal basis as

$$f(x) \approx \sum_{m=1}^{M} \sum_{k=1}^{K_m} c_{m,k}\, \Phi_{m,k}(x), \tag{7}$$

where the coefficients are determined using the projection formula

$$c_{m,k} = \frac{1}{\alpha_{m,k}} \int_{D_m} f(x)\, \phi_{m,k}(x)\, w_m(x)\, dx. \tag{8}$$

Because the bases are orthogonal, the approximation minimizes the mean square error within each subdomain, and convergence in L^2 is guaranteed as $M, K_m \to \infty$.

3. Neural Networks Based on Piecewise Orthogonal Functions

3.1. Network Architecture. POFNN is a class of function approximation networks that integrate the analytical properties of orthogonal functions with the adaptive learning capability of neural networks. Unlike conventional feedforward neural networks, where activation functions are typically sigmoidal or radial, the POFNN employs POFs as neuron activation or basis functions. This design allows the network to achieve local adaptivity, rapid convergence, and high numerical stability.

The structure of the POFNN consists of three main layers: an input layer, a hidden layer composed of locally supported orthogonal neurons, and an output layer that linearly combines the hidden outputs to generate the final prediction. The output of the POFNN with one-dimensional input x can be represented as

$$y(x) = \sum_{m=1}^{M} \sum_{k=1}^{K_m} w_{m,k}\, \Phi_{m,k}(x), \tag{9}$$

where $\Phi_{m,k}(x)$ are the piecewise orthogonal basis functions defined over the subdomains D_m, and $w_{m,k}$ are the corresponding synaptic weights learned from data.

Each neuron in the hidden layer is therefore associated with a particular subinterval D_m and a local orthogonal function $\phi_{m,k}(x)$, giving the network a highly interpretable and modular structure.

3.2. Convergence and Approximation Property. Let $f(x)$ be a target function in $L^2([a, b])$. According to the orthogonal expansion theory, as the number of subdomains M and the local basis functions K_m increase, the POFNN approximation converges to the true function:

$$\lim_{M,\, K_m \to \infty} \|f(x) - y(x)\|_{L^2([a,b])} = 0. \tag{10}$$

This property confirms that POFNNs possess the *universal approximation capability*, ensuring that any continuous nonlinear mapping can be approximated with arbitrary precision.

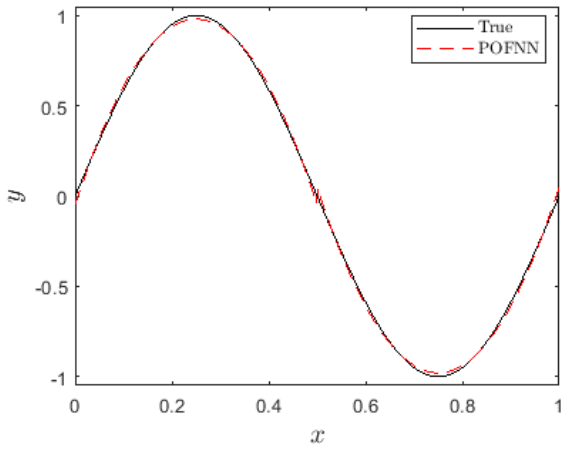


FIGURE 1. Approximation of $f(x) = \sin(2\pi x)$ using a POFNN with two subintervals and local Legendre bases of degree 0,1, 2.

4. Illustrative Example

To demonstrate the effectiveness of POFNN, we consider a simple nonlinear function approximation problem. The target function is defined as

$$f(x) = \sin(2\pi x), \qquad x \in [0, 1]. \tag{11}$$

The objective is to approximate this function using a POFNN composed of two subintervals and locally defined Legendre orthogonal basis functions. A set of $N = 500$ uniformly spaced samples over $[0, 1]$ is used for training.

Figure 1 shows the comparison between the true function $f(x)$ and the POFNN approximation. The model successfully captures the nonlinear shape of the sine function across both subintervals. The mean squared error (MSE) between the true and approximated values is 3.02×10^{-4}.

5. Conclusion

This study introduced the Piecewise Orthogonal Function Neural Network (POFNN) as a general and efficient framework for nonlinear function approximation. The model leverages piecewise orthogonal functions to provide local adaptability and analytical clarity. The localized representation of POFNN makes it well-suited for modeling nonstationary and piecewise nonlinear behaviors. The framework can be effectively applied to system identification, signal reconstruction, and time series forecasting.

References

1. Li, S., Jia, K., Wen, Y., Liu, T., and Tao, D. (2021). *Orthogonal Deep Neural Networks*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(4), pp. 1352-1368
2. Ahmadi, G. (2022) *Stochastic gradient-based hyperbolic orthogonal neural networks for nonlinear dynamic systems identification*, J. Math. Model. **10**, 529–547.
3. Pan, Y., Yu, H., Li, S., and Huang, R. (2024). *Orthogonal Neural Network: An Analytical Model for Deep Learning*, Appl. Sci., 14, 1532.
4. Zhagharian, S., Heydari, M.H., and Razzaghi, M. (2024).*Piecewise fractional Legendre functions for nonlinear fractional optimal control problems with ABC fractional derivative and non-smooth solutions*, Asian Journal of Control, 26(1), pp. 490-503.

A Mathematical Model for cancer by using immunotherapy

Maryam Nikbakht[1],

Department of Mathematics, Payame Noor University, P.O. Box, 19395-3697, Tehran, Iran
m_nikbakht@pnu.ac.ir

.

Abstract: In this paper, an effective treatment strategy for optimal control of cancer cells in a limited time is investigated. In order to improve the treatment of cancer cells through tumor-immune interactions, a nonlinear mathematical model using immunotherapy is analyzed. The main goal of this research is to reach the conditions under which cancer cells can be effectively controlled. For this purpose, by designing a quadratic control function, optimal treatment strategies are created that maximize the number of immune effector cells and minimize the number of cancer cells.

Keywords: Optimal control, mathematical model, immunotherapy, differential equations.

1. Introduction

Cancer, as one of the major challenges in global health, remains a leading cause of death worldwide. Remarkable advances in cellular research have shown that tumors are composed of heterogeneous and highly differentiated cell populations with diverse genetic differences [1].

In recent decades, cellular research has allowed us to better understand the mechanisms of cell progression, tumor prevention, and tumor cell destruction during infancy. However, current therapies against cancer cells still face major challenges.

In the last two decades, advances in the understanding of cancer immunogenesis have opened the door to cancer immunotherapy approaches [2]. This approach offers a promising new therapeutic strategy and is used to destroy cancer cells. In immunotherapy, genetically modified immune control cells recognize tumor-associated antigens and deliver a specific cytotoxic agent to the tumor cells. Immunotherapy works by slowing or stopping the spread of cancer cells to other parts of the body and also helps the immune system to increase its efficiency by eliminating cancer cells.

[1] . Corresponding Author

Immunotherapy has emerged as a promising new approach in this area. This approach uses genetically modified immune cells to recognize tumor-associated antigens and then targets the tumor cells with a specific cytotoxic agent.
The main goal of this research is to minimize the tumor toxicity burden and thus maintain a sufficient amount of host cells. In this regard, we are trying to provide optimized strategies to achieve the desired goals in immunotherapy. This research uses mathematical models and optimal control techniques to help increase the understanding of the dynamics of tumor-immune interactions and find better solutions to therapeutic problems. For this purpose, the relevant mathematical model is presented below.

2. Mathematical model

As we know, a better understanding of the tumor and its correct prediction can be very useful for improving cancer and its better treatment, hence mathematical modeling, dynamical systems and differential equations can be of great help in this field. In the last few decades, significant advances have been made in theoretical, experimental and clinical approaches to understand the dynamics of cancer cells and their interactions with the immune system. In addition, advances have also been made in analytical and computational models to help provide insight into clinical observations. The presented mathematical model is based on the model proposed in [3]. Considering the control variable u which represents the drug dose and σ which refers to the adaptive cellular immunotherapy treatment; the control model is presented as follows:

$$\begin{aligned}
\frac{dT}{dt} &= a_1 T(1 - b_1 T) - m_1 TH - \frac{n_1 Tl}{a_1 + T}, \\
\frac{dH}{dt} &= a_2 H(1 - b_2 H) - m_2 TH, \\
\frac{dl}{dt} &= \frac{n_2 Tl}{\alpha_2 + T} - \rho Tl - \delta l + u\sigma, \qquad (1)
\end{aligned}$$

Which apply under the following initial conditions.

$$\mathrm{T}\left(\cdot\right) = T_0 \ \geq 0.\ \mathrm{H}\left(\cdot\right) = H_0 \ \geq 0.\ \mathrm{I}\left(\cdot\right) = \mathrm{I}_0 \ \geq 0.$$

I, H, T represent tumor cell, host cell and immune cell, respectively. Tumor cells grow logistic with proliferation rate and maximum tumor burden $a_1$and b_1, respectively. Tumor cells, based on Michel-Menten motion, are killed by immune effector cells at a rate $\mathrm{n}_1$due to the limited immune cell effect against the tumor immunosuppressive activity and the stiffness factor a_1. The introduction of parameters is continued in [3]. It should be noted that σrefers to the treatment of Adoptive cellular immunotherapy which is administered as an external injection of immune-effector cells and is controlled by u (amount of dose).

Since the goal is to maintain the patient's health during the treatment period with a state constraint for the host cell or healthy cell. Therefore, we must minimize the growth of tumor cells along with the cost of treatment control or the harmful effects of the drug and maximize the growth of immune factor cells. Therefore, the cost function is defined as:

$$\Im(u,w) = \int_0^{t_f} [T - I + \epsilon_u u^2] \qquad (2)$$

Theorem 2-1: For sufficiently small time (t_f), the bounded solutions of the optimal system are unique.

Proof of the theorem: Due to the limitation in the pages of the article, its proof is not stated in this section.

3. Numerical results

In this section, we discuss the optimal control problem numerically to investigate the effects of immunotherapy on the proposed model. The model parameters are given in [3]. The simulation is performed with initial conditions and the back-and-forth sweep method, and finally, the optimal system is solved using an iterative process based on the fourth-order Runge-Kutta approach. In the case where immunotherapy is considered for disease improvement, the control u^* is obtained as follows:

$$u^* = \begin{cases} 1 & 0 \le t \le t_1 \\ 0 & t_1 < t \le t_f \end{cases} \qquad (3)$$

Which is true in $t_1 = 103.2$.

The numerical results of the optimal system are:

$$J(u^*) = 335129 \times 10^5, T(t_f) = 1.001 \times 10^3, I(t_f) = 10.081 \times 10^6, H(t_f) = 8.0201 \times 10^3$$

We observe that the tumor burden size decreases after immunotherapy. If we look closely, the cancer cell is eliminated after about 100 days in Figure (1). Here, the weights are $\epsilon_u = 100,, \sigma = 60$.

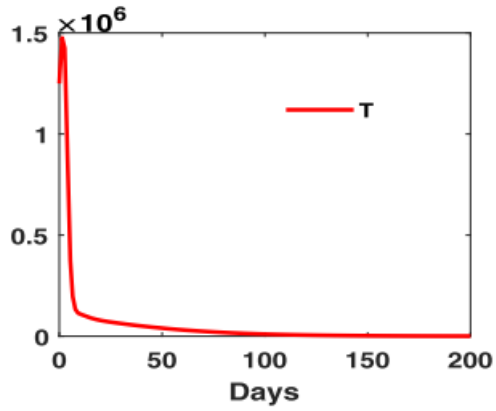


Figure 1: Tumor cell changes with immunotherapy

4. Conclusion

Immunotherapy is a promising new approach to cancer treatment. It uses genetically modified immune cells to recognize tumor-associated antigens and target tumor cells with specific cytotoxic agents. In this paper, we investigated the effects of immunotherapy on a model of the tumor immune system. The results showed that cancer cells were significantly reduced in the presented model.

References

1. Sweilam N.H., Al-Mekhlafi S.M., Assiri T. and AtanganaA. (2020) Optimal control for cancer treatment mathematical model using Atangana–Baleanu–Caputo fractional derivative, Advances in Difference Equations, https://doi.org/10.1186/s13662-020-02793-9

2. Castiglione F., Piccoli B., Cancer immunotherapy, mathematical modeling and optimal control, Journal of Theoretical Biology, Vol. 247(4), 723-73. DOI: 10.1016/j.jtbi.2007.04.003.

3. **3.** Das P., Mukherjee S., Das P., (2019) An investigation on Michaelis - Menten kinetics based complex dynamics of tumor - immune interaction, https://doi.org/10.1016/j.chaos.2019.08.006

A Mathematical Model for cancer by using chemo-therapy

Maryam Nikbakht[1],

Department of Mathematics, Payame Noor University, P.O. Box, 19395-3697, Tehran, Iran
m_nikbakht@pnu.ac.ir

Abstract: In this paper, a nonlinear mathematical model with drug therapy and treatment control is presented and analyzed. To understand under what conditions cancer cells can be destroyed, an optimal control problem is formulated with treatment as the control parameter. Next, a quadratic control function is designed to create optimal treatment strategies that minimize the number of cancer cells and reduce the harmful effects of the drug dose. In the proposed model, the uniqueness of the optimal solution is investigated, and then numerical simulations are performed and the graph related to the simulation of the paper is given. The results show that cancer cells are eradicated with little harmful effects for cancer patients.
Keywords: Optimal control, mathematical model, chemotherapy, differential equations.

1. Introduction

According to research, one of the leading causes of death worldwide is cancer. Cancer progression is associated with the emergence of highly heterogeneous populations of tumor cells, which is a defining feature of most advanced tumors[1]. Human knowledge and understanding of the mechanisms of cellular phenomena, prevention and destruction of tumor cells is still in its infancy. In the human body, tumor-specific antigens are recognized by the adaptive immune system to stimulate antitumor immune responses, and the immune system can only eliminate tumor cells at an early stage and before clinical interventions. Although new effective medical treatments have been dedicated to combating cancer, cancer treatments are still a challenging problem in medicine. In fact, host cells or normal cells must be maintained above a minimum level during the recovery of cancer cells in the whole body. Hence, modern techniques, for example, surgery, chemotherapy, radiotherapy, are performed as a strategy to eliminate cancer cells.

So far, various mathematical models have been studied with chemotherapy for cancer treatment [2]. In the following, this article first presents a mathematical model related to tumor stability

[1] . Corresponding Author

analysis in the form of differential equations; then, changes and examination of chemotherapy applications on it are discussed.

2. Mathematical model

In today's information world, one of the promising approaches involves mathematical modeling. The proposed model includes the identification of cells that play a role in cancer spread, interactions between cells, and a description of the dynamics of these interactions, which has helped to estimate the effective parameters in the analysis of tumor stability. It should be noted that its basic model has been previously described [3]. Next, by applying the control variable of chemotherapy that affects tumor, healthy, and immune cells, the following mathematical model based on a system of differential equations has been proposed:

$$\begin{aligned}
\frac{dT}{dt} &= a_1 T(1-b_1 T) - m_1 TH - \frac{n_1 Tl}{a_1+T} - r_1(1-e^{-V})T,\\
\frac{dH}{dt} &= a_2 H(1-b_2 H) - m_2 TH - r_2(1-e^{-V})H,\\
\frac{dl}{dt} &= \frac{n_2 Tl}{\alpha_2+T} - \rho Tl - \delta l - r_3(1-e^{-V})I, \qquad (1)\\
\frac{dV}{dt} &= \mathrm{w} - d_1 \mathrm{V}.
\end{aligned}$$

Which apply under the following initial conditions:

$$\mathrm{T}\left(\cdot\right) = T_0 \geq 0.\ \mathrm{H}\left(\cdot\right) = H_0 \geq 0.\ \mathrm{I}\left(\cdot\right) = \mathrm{I}_0 \geq 0.\ \mathrm{V}\left(\cdot\right) = V_0 \geq 0 \qquad (2)$$

I, H, T represent tumor cells, host cells, and immune cells, respectively. The amount of chemotherapy administered is represented by the variable V. Tumor cells have a logistic growth with $a_1$and $\mathrm{b}_1$proliferation rate and maximum tumor burden, respectively. Tumor cells, based on Michel-Menten motion, are destroyed by immune factor cells at a rate $\mathrm{n}_1$due to the limited immune cell effect against the tumor immunosuppressive activity and the stiffness coefficient a_1. The introduction of parameters is continued in [3]. r_1, r_2, and r_3 represent the rate of destruction of tumor cells, host cells, and immune cells, respectively, and are considered with values of 0.8, 0.6, and 0.6.

Since the goal is to maintain the patient's health during the treatment period with a state limitation for the host cell, that is, healthy cells. Therefore, we should minimize the tumor cell growth along with the cost of treatment control or the harmful effects of the drug and maximize the growth of immune factor cells. So the cost function is defined as:

$$\Im(u,w) = \int_0^{t_f} \left[T - I + \frac{1}{2}(\epsilon_w w^2)\right] \qquad (3)$$

where *w* refers to the amount of dose of chemo which is injected into the system and d_1 represents decay rate during drug administration. The following theorem is presented to demonstrate the existence of an optimal solution.

Theorem 2.1. An optimal solution

$$(X^*, u^*, w^*) \in S^{1,\infty}\left([0, t_f], R_+^4\right) \times L^{\infty}\left([0, t_f], R_+^2\right)$$

For (1) and (2) exist optimal control such that

$$\mathcal{J}(u^*, w^*) = min\{\mathcal{J}(u, w): u, w \in S\} \qquad (4)$$

in which

$$X^* = [T^*, H^*, I^*, V^*]^T$$

and S is an acceptable control set in $[0, t_f]$ with initial conditions

$$T(0) = T_0, H(0) = H_0, I(0) = I_0. V(0) = V_0$$

is defined.

Proof: Due to the limitation of the article pages, its proof is not stated in this section.
.

3. Numerical results

In this section, we discuss the optimal control problem numerically to investigate the effects of chemotherapy on the proposed model. The model parameters are given in [3]. Also, $\epsilon_w = 500$is considered. The simulation is performed with initial conditions and the back-and-forth sweep method, and finally, the optimal system is solved using an iterative process based on the fourth-order Rang-Kutta approach. The control w^* is also as follows:

$$w^* = \begin{cases} 1 & for \quad 0 \le t \le t_2 \\ 0 & for \quad t_2 \le t \le t_f \end{cases} \qquad (5)$$

Which is true at $t_2 = 22.02$.
The numerical results of the optimization system are as follows:
$J(w^*) = 11.4801 \times 10^5. T(t_f) = 0.9999 \times 10^3. I(t_f) = 1.99 \times 10^6. H(t_f) = 1.015 \times 10^3$
It is observed that by administering chemotherapy drugs, cancer cells will decrease significantly.

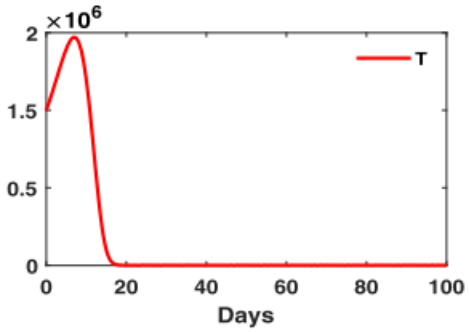


Figure 1: Tumor cell changes with chemotherapy

4. Conclusion

In this paper, we investigated the effect of chemotherapy on the model and demonstrated the dynamics of therapeutic strategies with interacting cells and their environment. For this purpose, we used the optimal therapeutic policy considering chemotherapy as a control variable and applied it to the model of De Pilis et al. The results showed that in the presented model, there was a significant reduction in cancer cells.

References

1. Sweilam N.H., Al-Mekhlafi S.M., Assiri T. and AtanganaA. (2020) Optimal control for cancer treatment mathematical model using Atangana–Baleanu–Caputo fractional derivative, Advances in Difference Equations, https://doi.org/10.1186/s13662-020-02793-9

2. Abdulrashid I., Han X., (2020) A mathematical model of chemotherapy with variable infusion, American institute of Mathematics Science, Vol. 19(4), 1875-1890.
Doi: 10.3934/cpaa.2020082

3. Das P., Mukherjee S., Das P., (2019) An investigation on Michaelis - Menten kinetics based complex dynamics of tumor - immune interaction, https://doi.org/10.1016/j.chaos.2019.08.006

Research on Controllability and Observability of Discrete-time Linear System with Iterval Coefficients

Hadi Shokohi Amiri[1,*] and Akbar Hashemi Borzabadi[2]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: first@mail.com

[2]Department of Mathematics, University of Science and Technology of Mazandaran, Behshar, Iran.
Email: borzabadi@mazust.ac.ir

Abstract. In this paper, the controllability and observability of (time-invariant) discrete-time linear system with interval coefficients is investigated. They is researched by a (time-invariant) discrete-time linear system with real coefficients and the full rank of the matrix is used in them.

Keywords: Controllability - Observability - (time-invariant) discrete-time linear system with interval coefficients

1. Introduction

Controllability and observability in the control systems are very important. In this paper, controllability and observability of (time-invariant) discrete-time linear system with interval coefficients are researched. $\mathsf{E} = \left\{x \in \mathbb{R} | e^1 \leq x \leq e^2\right\}$ for $e^1, e^2 \in \mathbb{R}$ is a closed bounded interval. Each member of E is stated as $e(\lambda) = e^1 + \lambda\left(e^2 - e^1\right), 0 \leq \lambda \leq 1$ and $e^1 = \min e(\lambda), e^2 = \max e(\lambda), 0 \leq \lambda \leq 1$ are respectively the begining and end points of E. [2]
Suppose that $\mathsf{E} = \left[e^1, e^2\right]$ is an interval number. All elements of a matrix are interval numbers then it is an interval matrix. $E = [e_{ij}]_{m\times n}$ and $\mathcal{E} = [\mathsf{E}_{ij}]_{m\times n}, \mathsf{E}_{ij} = \left[e^1_{ij}, e^2_{ij}\right]$ are respectively real and interval matrices. $E \in \mathcal{E}$ if and only if $e_{ij} \in \mathsf{E}_{ij}$ for $i = 1, \ldots, m, j = 1, \ldots, n$. The following explanations are considered:
$I(\mathbb{R})$ =The set of all interval numbers in $\mathbb{R}$.
$I(\mathbb{R})^n$ =The product space $I(\mathbb{R}) \times I(\mathbb{R}) \times \cdots \times I(\mathbb{R})$.
$I(\mathbb{R})^{m\times n}$ =The set of all interval matrices with m rows and n columns.
$J[0,1]^{m\times n}$ = The set of all real matrices with m rows and n columns such that all elements of these matrices belong to $[0,1]$.

Proposition 1.1. *An interval matrix $\mathcal{E}$ can be presented by an infinite set of real matrices, i.e.*

*Speaker.

$\mathcal{E} = \{E_\Lambda | E_\Lambda = [e_{ij}(\lambda_{ij})]_{m\times n}\ ,\ \Lambda = [\lambda_{ij}]_{m\times n} \in J[0,1]^{m\times n}\ ,\ e_{ij}(\lambda_{ij}) = e^1_{ij} + \lambda_{ij}(e^2_{ij} - e^1_{ij})\ ,\ i = 1,\dots,m\ ,\ j = 1,\cdots,n\}$.

Proof. The proof of this proposition is stated in [1] □

Definition 1.2. $\mathcal{E} \in I(\mathbb{R})^{m\times n}$ is an interval matrix. If all real matrices $E_\Lambda \in \mathcal{E}$ $(\Lambda \in J[0,1]^{m\times n})$ have full rank, then the interval matrix $\mathcal{E}$ is full rank.

The discrete-time linear system with interval coefficients is considered as follows:

$$\begin{cases} x_{k+1} = \mathcal{A}x_k + \mathcal{B}u_k \\ y_k = \mathcal{C}x_k + \mathcal{D}u_k \end{cases} \tag{1}$$

x_k, u_k and y_k are respectively state, control and output vectors and they have n, m and p dimentions. $\mathcal{A} \in (I(\mathbb{R}))^{n\times n}$, $\mathcal{B} \in (I(\mathbb{R}))^{n\times m}$, $\mathcal{C} \in (I(\mathbb{R}))^{p\times n}$, $\mathcal{D} \in (I(\mathbb{R}))^{p\times m}$ are interval matrices and they are defined in accordance with Proposition 1.1 and also $A_\Lambda \in \mathcal{A}$, $B_\Gamma \in \mathcal{B}$, $C_\Theta \in \mathcal{C}$, $D_\Omega \in \mathcal{D}$ such that A_Λ, B_Γ, C_Θ, D_Ω are real matrices that $\Lambda \in J[0,1]^{n\times n}, \Gamma \in J[0,1]^{n\times m}$, $\Theta \in J[0,1]^{p\times n}$, $\Omega \in J[0,1]^{p\times m}$.
The (time-invariant) discrete-time linear system with real coefficients that has the similar structure to the system (1) as following:

$$\begin{cases} x_{k+1} = A_\Lambda x_k + B_\Gamma u_k \\ y_k = C_\Theta x_k + D_\Omega u_k \end{cases} \tag{2}$$

Consider the first equation of (2) for $k = 0, 1, 2, \dots$ and put this equation for $k = 0$ into this equation for $k = 1$ and the resultant equation is replaced into this equation for $k = 3$ and so the same procedure continues, the equation (3) will be reached

$$x_k = A^k_\Lambda x_0 + \sum_{i=1}^{k} A^{k-i}_\Lambda B_\Gamma u_{i-1} \tag{3}$$

2. Controllability and Observability

The controllability and observability of the system (1) will be expressed via the system (2).

Definition 2.1. A system with state space equation (2) is given that $A_\Lambda : \Lambda \in J[0,1]^{n\times n}, B_\Gamma : \Gamma \in J[0,1]^{n\times m}, C_\Theta : \Theta \in J[0,1]^{p\times n}$ and $D_\Omega : \Omega \in J[0,1]^{p\times m}$ are coefficient matrices. y_1, y_2 are any position in $\mathbb{R}$. The state sequence $\{x_k\}$ can be brought from the position y_1 to y_2 by a certain control sequence $\{u_k\}$, the system (2) is called controllable. Otherwise, consider any positions $y_1, y_2 \in \mathbb{R}$, there exists a sequence control $\{u_k\}$ such that: $y_2 = A^k_\Lambda y_1 + \sum_{i=1}^{k} A^{k-i}_\Lambda B_\Gamma u_{i-1}$, then the system (2) is controllable.

Definition 2.2. The interval matrices $\mathcal{A} \in (I(\mathbb{R}))^{n\times n}, \mathcal{B} \in (I(\mathbb{R}))^{n\times m}, \mathcal{C} \in (I(\mathbb{R}))^{p\times n}$ and $\mathcal{D} \in (I(\mathbb{R}))^{p\times m}$ are the coefficient matrices of the system (1). If the system (2) is controllable for all real matrices $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ and $D_\Omega \in \mathcal{D}$ such that $\Lambda \in J[0,1]^{n\times n}, \Gamma \in J[0,1]^{n\times m}, \Theta \in J[0,1]^{p\times n}$ and $\Omega \in J[0,1]^{p\times m}$, then the system (1) is controllable.

Proposition 2.3. *The linear system (1) with interval coefficients is assumed and*

$$\mathcal{M}_{\mathcal{AB}} = \begin{bmatrix} \mathcal{B} & \mathcal{AB} & \mathcal{A}^2\mathcal{B} & \dots & \mathcal{A}^{n-1}\mathcal{B} \end{bmatrix}$$

is an interval compound matrix that $\mathcal{M}_{\mathcal{AB}} \in (I(\mathbb{R}))^{n\times mn}$. *This system is controllable if and only the interval matrix* $\mathcal{M}_{\mathcal{AB}}$ *is full rank.*

Proof. Let the system (1) is controllable, the system (2) is controllable for all real matrices $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ and $D_\Omega \in \mathcal{D}$. The real matrix

$$M_{A_\Lambda B_\Gamma} = \begin{bmatrix} B_\Gamma & A_\Lambda B_\Gamma & A_\Lambda^2 B_\Gamma & \dots & A_\Lambda^{n-1} B_\Gamma \end{bmatrix}$$

is full rank for all real matrices $\Lambda \in J[0,1]^{n\times n}, \Gamma \in J[0,1]^{n\times m}$ [3]. Then the interval compound matrix $\mathcal{M}_{\mathcal{AB}}$ is full rank.
Now suppose the interval compound matrix $\mathcal{M}_{\mathcal{AB}}$ is full rank, the real compound matrix $M_{A_\Lambda B_\Gamma}$ is full rank for all real matrices $\Lambda \in J[0,1]^{n\times n}, \Gamma \in J[0,1]^{n\times m}$. The system (2) is controllable for all real matrices $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ and $D_\Omega \in \mathcal{D}$ [3]. Then the system (1) is controllable. □

Definition 2.4. Suppose the initial time of the system (2) is l. There exists an $q > 0$, that $C_\Theta A_\Lambda^i x_l = 0$ is satisfied for $i = l, \dots, q$ it is resulted in $x_l = 0$, the system is observable at initial time l. The system is observable at every initial time l then, it is called observable.

Definition 2.5. The system (1) is assumed with interval coefficient matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and $\mathcal{D}$. The system (2) is observable for all real matrices $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ and $D_\Omega \in \mathcal{D}$ then, the system (1) is called observable.

Proposition 2.6. *The interval compound matrix*

$$\mathcal{N}_{\mathcal{CA}} = \begin{bmatrix} \mathcal{C} \\ \mathcal{CA} \\ \mathcal{CA}^2 \\ \vdots \\ \mathcal{CA}^{n-1} \end{bmatrix}$$

that $\mathcal{N}_{\mathcal{CA}} \in (I(\mathbb{R}))^{np\times n}$ *is assumed. The system (1) is observable if and only if the interval compound matrix* $\mathcal{N}_{\mathcal{CA}}$ *is full rank.*

Proof. Suppose the system (1) is observable, then the system (2) is observable for all real matrices $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ and $D_\Omega \in \mathcal{D}$. The real compound matrix

$$N_{C_\Theta A_\Lambda} = \begin{bmatrix} C_\Theta \\ C_\Theta A_\Lambda \\ C_\Theta A_\Lambda^2 \\ \vdots \\ C_\Theta A_\Lambda^{n-1} \end{bmatrix}$$

is full rank for all real matrices $\Lambda \in J[0,1]^{n\times n}, \Theta \in J[0,1]^{p\times n}$ [3]. So the interval compound matrix $\mathcal{N}_{\mathcal{CA}}$ is full rank.
If the interval compound matrix $\mathcal{N}_{\mathcal{CA}}$ is full rank, the real compound matrix $N_{C_\Theta A_\Lambda}$ is full rank for all real matrices $\Lambda \in J[0,1]^{n\times n}, \Theta \in J[0,1]^{p\times n}$. Then the system (2) is observable for all real matrices $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ and $D_\Omega \in \mathcal{D}$ [3]. So the system (1) is observable □

EXAMPLE 2.7. The interval matrices $\mathcal{A} = \begin{bmatrix} [2,5] & [-2,1] \\ 0 & [0,4] \end{bmatrix}$, $\mathcal{B} = \begin{bmatrix} [3,4] \\ [-5,7] \end{bmatrix}$, $\mathcal{C} = \begin{bmatrix} 0 & [4,7] \end{bmatrix}$, $\mathcal{D} = O$, are the coefficient matrices of the system (1) .There exist the real matrices $A_\Lambda = \begin{bmatrix} 2+3\lambda_{11} & -2+3\lambda_{12} \\ 0 & 4\lambda_{22} \end{bmatrix}$, $B_\Gamma = \begin{bmatrix} 3+\gamma_{11} \\ -5+12\gamma_{21} \end{bmatrix}$, $C_\Theta = \begin{bmatrix} 0 & 4+3\theta_{12} \end{bmatrix}$, such that $A_\Lambda \in \mathcal{A}, B_\Gamma \in \mathcal{B}, C_\Theta \in \mathcal{C}$ for all real matrices $\Lambda \in J[0,1]^{2\times 2}, \Gamma \in J[0,1]^{2\times 1}, \Theta \in J[0,1]^{1\times 2}$. The elements of the compound matrix $M_{A_\Lambda B_\Gamma} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$ are as follows:$m_{11} = 3+\gamma_{11}, m_{12} = 16+2\gamma_{11}+9\lambda_{11}+3\lambda_{11}\gamma_{11}-24\gamma_{21}-15\lambda_{12}+36\lambda_{12}\gamma_{21}, m_{21} = -5+12\gamma_{21}$ and $m_{22} = -20\lambda_{22}+48\lambda_{22}\gamma_{21}$. $det\left(M_{A_\Lambda B_\Gamma}\right) = 80-60\lambda_{22}+144\lambda_{22}\gamma_{21}-20\lambda_{22}\gamma_{11}+48\lambda_{22}\gamma_{21}\gamma_{11}+10\gamma_{11}+45\lambda_{11}+15\lambda_{11}\gamma_{11}-312\gamma_{21}-75\lambda_{12}+360\lambda_{12}\gamma_{21}-24\gamma_{11}\gamma_{21}-108\lambda_{11}\gamma_{21}-36\lambda_{11}\gamma_{11}\gamma_{21}+288\gamma_{21}^2-432\lambda_{12}\gamma_{12}^2$. Considering that the minimum and maximum amount of $M_{A_\Lambda B_\Gamma}\left(\Lambda \in J[0,1]^{2\times 2}, \Gamma \in J[01]^{2\times 1}\right)$ are respectively negative and positive, therefore there exist $\Lambda_1 \in J[0,1]^{2\times 2}, \Gamma_1 \in J[0,1]^{2\times 1} : det\left(M_{A_{\Lambda_1}} B_{\Gamma_1}\right) = 0$. So the system (1) is not controllable.

The compound matrix $N_{C_\Theta A_\Lambda} = \begin{bmatrix} 0 & 4+3\theta_{12} \\ 0 & 16\lambda_{22}+12\theta_{12}\lambda_{22} \end{bmatrix}$ is singular and the system (1) is not observable.

3. Conclusion

The controllability and observability of the (time-invariant) discrete-time linear system with interval coefficients were researched. The investigation of this system that it is time-variant will suggested.

References

1. Amiri, H.S, Borzabadi, A.H. and Heydari, A. (2020) *A different view on controllability and observability of continuous time linear systems with interval coefficients*, Iranian Journal of Numerical Analysis and Optimization, **10(1)**, 107-120.
2. Bhurjee, A.K. and Panda, G. (2012), *Efficient solution of interval optimization problem*, Math. Methods Oper. Res. **76**, 273-288.
3. Chui, C.K. and Chen, G. (1989), *Linear systems and optimal control*, Springer-Verlag Berlin Heidelberg.


12th National Mathematics Conference of Payame Noor University
Sari—Iran
6–7 November 2025

Statistics

Bayesian and E-Bayesian Estimation for Rayleigh Distribution Using Progressive Type-II Censored

Kazem Fayyaz Heidari[1], Ph. D, Department of Statistics,

Payame Noor University, P.O. Box, 19395-3697, Tehran, Iran

fayyaz@pnu.ac.ir

Abstract: In this paper, investigates the estimation of the scale parameter for Rayleigh distribution based on progressive type-II censoring samples. Bayesian and E-Bayesian estimators are produced using symmetric loss function, such as the squared error (SE) loss function. Then, these methods are compared through Monte Carlo simulation study.

Keywords: E-Bayesian, Rayleigh distribution, progressive type-II censoring.

1. Introduction

The Rayleigh distribution is due to the fact that in different fields of science and technology such as the modeling of sea wave heights in oceanography, communications engineering, distribution of industrial components life, clinical studies of cancer patients, reliability theory and analysis survival is used, its parameter estimation is recommended in various ways [6]. Also, Censoring is a common practice in longevity and reliability studies. If the experimenter determines that, after observing the first failure, R_1 units of healthy test units and at the time of the second failure, R_2 unit of healthy test units will be cached out of the test and continue until mth failure, all the remaining units of the test $R_m = n - R_1 - R_2 - \ldots - R_{m-1} - m$ outside Then The Progressive Type-II sensor will take place. In this case, the failure times of units are random variables and R_i s are predetermined constants. As a result of this censorship plan, m ordered amount is obtained which ordinal statistics are called progressive type-II censored [1].

The probability density function and cumulative distribution function of the Rayleigh distribution are respectively, as follows.

$$f(x,\lambda) = 2\lambda x e^{-\lambda x^2}, x > 0, \lambda > 0 \quad (1)$$

$$F(x,\lambda) = 1 - e^{-\lambda x^2}, x > 0, \lambda > 0 \quad (2)$$

In the second section, we will obtain the E-Bayesian estimation of the Rayleigh distribution parameter under SE and the Progressive Type-II censored. In the third section, they will be compared Bayesian and E-Bayesian estimators using Monte Carlo simulation. And the fourth part will be dedicated to the results.

2. Estimation of λ

It is assumed that the Prior distribution of λ is the gamma distribution with the Hyperparameter a and b as follows.

$$\pi(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-\lambda b}, \lambda > 0, a, b > 0 \quad (3)$$

[1] . Corresponding Author

According to Han (12), the Hyperparameter a and b are considered to be $\pi(\lambda|a,b)$ decreasing relative to λ.

Thus obtaine $\frac{d[\pi(\lambda|a,b)]}{d\lambda} = \frac{b^a\lambda^{a-2}e^{-b\lambda}}{\Gamma(a)}[(a-1)-b\lambda]$ That should be $b>0$ and $0<a\leq 1$. Berger [4] showed that increasing b would decrease the efficiency of the Bayes estimator λ.Therefore, the Hyperparameter b must be bounded above and be $0<b<c$. Showed that the most appropriate distribution b is uniform distribution.Therefore, in this paper, $\pi_1(b)$ is a continuous uniform distribution in the interval $(0,c)$ and $a=1$. In this case, relation (3) becomes $\pi(\lambda|b)=be^{-\lambda}$.

2.1 E-Bayesian estimation of λ

In this section, we will first obtain, the Bayes estimation and then the E-Bayesian estimation of λ with the loss function $L(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2$. If Rayleigh distribution with probability density function (1), the distribution of the failure time, and $Y_{1:m:n}, Y_{2:m:n}, Y_{m:m:n}$ is Progressive Type-II censored sample of it, and y_i is the finding of $Y_{1:m:n}$, then according to [2], the likelihood function is obtained as follows

$$L(\lambda\,|\,\boldsymbol{Y}) = C\left(\prod_{i=1}^{m} y_i\right) 2^m \lambda^m e^{-\lambda\sum_{i=1}^{m}(1+R_i)y_i^2} \tag{4}$$

where, $\boldsymbol{Y}=(Y_{1:m:n},Y_{2:m:n},\ldots,Y_{m:m:n})$, $C = n(n-R_1-1)\ldots(n-R_1-\ldots-R_{m-1}-m+1)$

According to (4), posterior distribution of λ is obtained as follows.

$$\pi^*(\lambda\,|\,\boldsymbol{Y}) = \frac{(b+\sum_{i=1}^{m}(1+R_i)y_i^2)^{m+1}}{m!}\lambda^m e^{-\lambda(b+\sum_{i=1}^{m}(1+R_i)y_i^2)} \tag{5}$$

Therefore, the Bayes estimation of the parameter λ under the squared error loss function is as follows

$$\hat{\lambda}_{Bay}(b) = E(\lambda|Y) = \frac{m+1}{b+\sum_{i=1}^{m}(1+R_i)y_i^2} \tag{6}$$

E-Bayesian estimation of the parameter λ, is defined $\hat{\lambda}_{EBay} = \int_{\Lambda}\hat{\lambda}_{Bay}(b)\pi_1(b)db,\ b\epsilon\Lambda$.

According to equation (6), definition 2-1 and distribution b, E-Bayesian estimation λ is obtained as follows

$$\hat{\lambda}_{EBay} = \frac{1}{c}\int_0^c \hat{\lambda}_{Bay}(b)\pi_1(b)db = \frac{m+1}{c}\left[log\frac{c+\sum_{i=1}^{m}(1+R_i)y_i^2}{\sum_{i=1}^{m}(1+R_i)y_i^2}\right] \tag{7}$$

3. Simulation study

Using [3], Progressive Type-II censored samples are obtained from the Rayleigh distribution with probability density function (1) and with parameter $\lambda=2$ as follows.

Step 1: First, we generate $m$random number independent of the standard uniform distribution and shown with $W_1, W_2, \ldots, W_m$.

Step 2: For, $i=1.2\ldots..m, V_i = W_i^{\left\{1/(i+\sum_{j=m+1-i}^{m}R_j)\right\}}$.

Step 3: For, $i=1.2\ldots..m, U_{i:m:n}=1-\Pi_{j=1}^{i}V_{m+1-j}$, in this case, $U_{1:m:n}, U_{2:m:n}, \ldots, U_{m:m:n}$ is the Progressive Type-II censored sample of standard uniform distribution.

Step 4: $Y_{i:m:n}=F^{-1}(U_{i:m:n})$ is obtained for $i=1.2\ldots..m$.

In this case, $Y_{1:m:n}, Y_{2:m:n}, \ldots, Y_{m:m:n}$ are the Progressive Type-II censored order statistics of Rayleigh distribution with probability density function (1). Now, we obtain the Bayesian and E-

Bayesian estimators by using the generated sample $Y_{1:m:n}, Y_{2:m:n}, \ldots, Y_{m:m:n}$, and relations (6), (7) and (10).The first to fourth steps are repeated 1000 times, and then the average values of estimators and the root mean square error of the estimators are calculated. The results are presented in Table indicates that the Bayesian estimation is better.

Table. Average estimates and $(\sqrt{MSE})$ under quadratic loss function for $b = 0.5$ and $c = 1.5$

$\hat{\lambda}_{EBay}$	$\hat{\lambda}_{Bay}$	$design$	M	N
3/693(1/187)	0/978(0/085)	(5,0,…..,0)	5	10
3/420(1/144)	1/886(0/325)	(2,3,0,...,0)	5	10
5/553(1/538)	0/959(0/487)	(10,0,....,0)	10	20
5/026(1/338)	2/384(0/297)	(3,4,3,0,..,0)	10	20
9/004(2/200)	0/963(0/026)	(0,20,….,0)	30	50
5/093(1/390)	4/350(0/972)	(1,4,5,0,..,0)	40	50
7/613(1/684)	6/954(1/378)	(1,……..,1)	25	50
11/97(2/681)	1/820(0/066)	(15,15,0,..,0)	50	80

4. Conclusion

In this study, Bayesian and E-Bayesian estimations of the Rayleigh distribution parameter based on the Progressive Type-II censored samples were obtained. By calculating the MSE and the average estimation, the Bayesian and E-Bayesian estimations based on the Rayleigh distribution were compared using Monte Carlo simulation. It has been shown that the Bayesian estimation has better efficiency.

References

1. Asgharzadeh, a. and Moradnejad, p. (2018). Inference for a Skew Normal Distribution Based on Progressively Type-II Censored Samples, Journal of Statistical Research of Iran, 5(1), 33-56.
2. Balakrishnan, N, and Aggarwala, R. (2020). Progressive Censoring: Theory. Methods and Applications. Birkhauser, Boston.
3. Balakrishnan. N. and Sandhu, R.A.A. (1995). Simple simulational algorithm for generating progressive Type-II censored samples, The Amer. Statist, 49, 229-230.
4. Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis, second ed., Springer -Verlag, New York.
5. Han, M. (1997). The structure of hierachical prior distribution and its applications, Chinese Operations Research and Management Science. 6(3), 31-40.
6. Rayleigh, J. W.S. (1880), On the resultant of a Large number of vibration of the somepith and of arbitrary phase. Philosophical Magazine. 5th series, 10, 73-78.

Bayesian Inference for the Power Modified Lindley Distribution based on Symmetric and Asymmetric Balanced Loss Functions

Adeleh Fallah[1,*]

[1]Department of Statistics, Payame Noor University (PNU), Tehran, Iran.
Email: adelehfallah@pnu.ac.ir

Abstract. In this paper, maximum likelihood and Bayes estimators of the parameters have been obtained for power modified Lindley distribution when sample is available from progressive Type-II censoring scheme. The Bayes estimators are obtained under symmetric and asymmetric balanced loss functions, specifically the balanced squared error loss function, the balanced linear exponential loss function. Because the integrals of the Bayes estimates do not possess closed forms, the Metropolis-Hastings algorithm is applied to approximate these integrals. One real data sets have been analyzed to demonstrate how the proposed methods can be used in practice.

Keywords: power modified Lindley distribution, Estimation, progressive Type-II censoring, Metropolis-Hastings algorithm

AMS Mathematics Subject Classification [2020]: 62N01, 62N02, 62N05

1. Introduction

Under Progressive censoring scheme, from a total of n units placed simultaneously on a life test, only $(m < n)$ are completely observed until failure. Then, given a censoring plan $R_1, \ldots, R_m$: At the time $X_{1:m:n}$ of the first failure, R_1 of the $n-1$ surviving units are randomly withdrawn (or censored) from the life-testing experiment. At the time $X_{2:m:n}$ of the next failure, R_1 of the $n - R_1 - 2$ surviving units are censored, and so on. Finally, at the time $X_{m:m:n}$ of the mth failure, all the remaining $R_m = n - m - R_1 - R_2 - \ldots - R_{m-1}$ surviving units are censored, see; [2]. The cumulative distribution function and probability density function of the power modified Lindley distribution (PML) are given by [3]

$$(1)\quad f_X(x;\theta,\alpha) = \frac{\alpha\theta}{1+\theta}\, e^{-2\theta x^{\alpha}}[(1+\theta)e^{\theta x^{\alpha}}x^{\alpha-1} + 2\theta x^{2\alpha-1} - x^{\alpha-1}],\ x>0,\ \theta>0\ \alpha>0,$$

$$(2)\qquad F_X(x;\theta,\alpha) = 1 - \left[1 + \frac{\theta x^{\alpha}}{1+\theta}e^{-\theta x^{\alpha}}\right] e^{-\theta x^{\alpha}},\ x>0,\ \theta>0,\ \alpha>0.$$

In this paper, we consider the estimation of the PML distribution based on progressively Type-II censoring, Bayes and maximum likelihood estimators are considered.

*Speaker.

2. Estimation of the model parameter

2.1. Maximum likelihood estimation. Under the progressively Type-II censored sample $\mathbf{X} = (X_1, \ldots, X_m)$, the likelihood function for $\zeta = (\theta, \alpha)$ is given by

$$L(\mathbf{x}; \zeta) = A \prod_{i=1}^{m} f(x_i; \zeta)[1 - F(x_i; \zeta)]^{R_i}, \tag{3}$$

where $A = n(n - R_1 - 1)(n - R_1 - R_2 - 2)\ldots(n - R_1 - \ldots - R_{m-1} - m + 1)$ and $\mathbf{x} = (x_1, \ldots, x_m)$ is the vector of observation. By substituting Eqs. (1) and (2) into Eq. (3), the likelihood function is

$$\begin{aligned} L(\mathbf{x}; \theta, \alpha) &= A \frac{\alpha^m \theta^m}{(1+\theta)^m} \, e^{-\theta\left[2\sum_{i=1}^{m} x_i^{\alpha} + \sum_{i=1}^{m} x_i^{\alpha} R_i\right]} \prod_{i=1}^{m} \left(1 + \frac{\theta x_i^{\alpha}}{1+\theta} e^{-\theta x_i^{\alpha}}\right)^{R_i} \\ &\times \prod_{i=1}^{m} \left[(1+\theta) e^{\theta x_i^{\alpha}} x_i^{\alpha-1} + 2\theta x_i^{2\alpha-1} - x_i^{\alpha-1}\right]. \end{aligned} \tag{4}$$

Therefore, the maximum likelihood estimators (MLE) of θ and α, can be obtained by maximizing the log-likelihood function with respect to θ and α.

2.2. Bayesian Approaches. [4] presented a generalized balanced loss function, denoted as the $L_{\rho,\omega,\delta_0}(\delta, \widehat{\delta}) = \omega\rho(\widehat{\delta}, \delta_0) + (1-\omega)\rho(\delta, \widehat{\delta})$. In this context, ω (with $0 \leqslant \omega \leqslant 1$) serves as a weight parameter, and ρ denotes a user-defined loss function. The target estimator, denoted as δ_0 , is typically derived using methods such as maximum likelihood or least squares or unbiasedness. By choosing $\rho(\delta, \widehat{\delta}) = (\widehat{\delta} - \delta)^2$, reduced to the balanced-squared error loss (BSEL) function, in the form $L_{\omega,\delta_0}(\theta, \delta) = \omega(\widehat{\delta} - \delta_0)^2 + (1-\omega)(\widehat{\delta} - \delta)^2$, where $\widehat{\delta}$ is an estimator of δ. The Bayes estimator of δ under the BSEL loss function is given by $\widehat{\delta}\omega, \delta_0(x) = \omega\delta_0 + (1-\omega)E(\delta|\mathbf{x})$. The balanced linear-exponential (BLINEX) loss function, incorporating a shape parameter c (where $c \neq 0$), is formulated by defining $\rho(\delta, \widehat{\delta}) = e^{c(\widehat{\delta}-\delta)} - c(\widehat{\delta} - \delta) - 1$, see; [5]. The Bayes estimator of δ under the BLINEX loss function is given by $\widehat{\delta}\omega, \delta_0(x) = \frac{-1}{c}\log[\omega e^{-c\delta_0} + (1-\omega)E(e^{-c\delta}|\mathbf{x})]$. It is assumed that θ and α have independent gamma priors with the pdfs

$$\pi(\theta) = \frac{b_1^{a_1}\theta^{a_1-1}\, e^{-b_1\theta}}{\Gamma(a_1)}, \pi(\alpha) = \frac{b_2^{a_2}\alpha^{a_2-1}\, e^{-b_2\alpha}}{\Gamma(a_2)}, \tag{5}$$

where a_1, b_1 and a_2, b_2 are positive hyperparameters. From Eqs. (4) and (5), the joint posterior density of θ and α becomes

$$\begin{aligned} \pi(\theta, \alpha|\mathbf{x}) &= \frac{\theta^{m+a_1-1}\alpha^{m+a_2-1}e^{-b_2\alpha}}{K\ (1+\theta)^m} e^{-\theta[\sum_{i=1}^{m} x_i^{\alpha}(2+R_i)+b_1]} \prod_{i=1}^{m} \left(1 + \frac{\theta x_i^{\alpha}}{1+\theta} e^{-\theta x_i^{\alpha}}\right)^{R_i} \\ &\times \prod_{i=1}^{m} \left[(1+\theta) e^{\theta x_i^{\alpha}} x_i^{\alpha-1} + 2\theta x_i^{2\alpha-1} - x_i^{\alpha-1}\right], \end{aligned} \tag{6}$$

where

$$\begin{aligned} K = \int_0^{\infty}\int_0^{\infty} &\frac{\theta^{m+a_1-1}\alpha^{m+a_2-1}}{e^{b_2\alpha}(1+\theta)^m} e^{-\theta[\sum_{i=1}^{m} x_i^{\alpha}(2+R_i)+b_1]} \\ &\times \prod_{i=1}^{m} \left[(1+\theta) e^{\theta x_i^{\alpha}} x_i^{\alpha-1} + 2\theta x_i^{2\alpha-1} - x_i^{\alpha-1}\right] \prod_{i=1}^{m} \left(1 + \frac{\theta x_i^{\alpha}}{1+\theta} e^{-\theta x_i^{\alpha}}\right)^{R_i} d\theta d\alpha. \end{aligned}$$

The Bayes estimator of θ under the BSE and BLINEX functions are given by

$$\widehat{\theta}_{BS} = \omega\widehat{\theta}_{ML} + (1-\omega)\int_0^\infty\int_0^\infty \frac{\theta^{m+a_1}\alpha^{m+a_2-1}e^{-b_2\alpha}}{K\ (1+\theta)^m} e^{-\theta[\sum_{i=1}^m x_i^\alpha(2+R_i)+b_1]}$$

$$\times\prod_{i=1}^m\left[(1+\theta)e^{\theta x_i^\alpha}x_i^{\alpha-1}+2\theta x_i^{2\alpha-1}-x_i^{\alpha-1}\right]\prod_{i=1}^m\left(1+\frac{\theta x_i^\alpha}{1+\theta}e^{-\theta x_i^\alpha}\right)^{R_i} d\theta d\alpha, \tag{7}$$

$$\widehat{\theta}_{BL} = \frac{-1}{c}\log\left[\omega e^{-c\widehat{\theta}_{ML}} + (1-\omega)\int_0^\infty\int_0^\infty \frac{\theta^{m+a_1-1}\alpha^{m+a_2-1}e^{-b_2\alpha}}{K\ (1+\theta)^m}\prod_{i=1}^m\left(1+\frac{\theta x_i^\alpha}{1+\theta}e^{-\theta x_i^\alpha}\right)^{R_i}\right.$$

$$\left.\times e^{-\theta[\sum_{i=1}^m x_i^\alpha(2+R_i)+b_1+c]}\prod_{i=1}^m\left[(1+\theta)e^{\theta x_i^\alpha}x_i^{\alpha-1}+2\theta x_i^{2\alpha-1}-x_i^{\alpha-1}\right]d\theta d\alpha\right], \tag{8}$$

Similay, we obtained the approximate Bayes estimates of the unknown parameter α. Since the Bayes estimate $\widehat{\theta}_{BS}$ and $\widehat{\theta}_{BL}$ cannot be obtained analytically, we adopt the Metropolis-Hastings algorithm to compute the Bayes estimate of θ.The Metropolis-Hastings algorithm steps are given below:

(1) Set the initial value $\zeta_0 = (\theta_0, \alpha_0)$ and $j = 1$.

(2) Using the Metropolis-Hastings algorithm, generate $\zeta_j = (\theta_j, \alpha_j)$ from $\pi(\theta_{(j-1)}, \alpha_{(j-1)}|\mathbf{x})$ with the bivariate normal distribution $N_2(\zeta_{(j-1)}, \Sigma)$ as proposal distribution, where Σ is the inverse of the estimated information matrix around the parameters θ and α. In other words, we put (in each repetition);

$$\Sigma = \left[\begin{array}{cc} -\frac{\partial^2 \ln L(x;\theta,\alpha)}{\partial\theta^2} & -\frac{\partial^2 \ln L(x;\theta,\alpha)}{\partial\theta\partial\alpha} \\ -\frac{\partial^2 \ln L(x;\theta,\alpha)}{\partial\alpha\partial\theta} & -\frac{\partial^2 \ln L(x;\theta,\alpha)}{\partial\alpha^2} \end{array}\right]^{-1}_{(\theta,\alpha)=(\widehat{\theta}_{MLE},\widehat{\alpha}_{MLE})}$$

(3) Set $j = j+1$ and Repeat Steps (2), N times to compute the Markov Chain Monte Carlo (MCMC) samples $\{(\theta_1, \alpha_1), \ldots, (\theta_N, \alpha_N)\}$.

The approximate Bayes estimates of θ under BSEL and BLINEX functions as

$$\widehat{\theta}_{BS} = \omega\widehat{\theta}_{ML} + \frac{(1-\omega)}{N-M}\sum_{j=M+1}^N \theta_j,\ \widehat{\theta}_{BL} = \frac{-1}{c}\log\left(\omega e^{-c\widehat{\theta}_{ML}} + \frac{(1-\omega)}{N-M}\sum_{j=M+1}^N e^{-c\theta_j}\right),$$

Similay, we obtained the approximate Bayes estimates of the unknown parameter α. M is the burn-in period.

3. Numerical results

Example 3.1. This data represents the life of fatigue fracture of Kevlar 373/epoxy subjected to constant pressure at 90% stress level until all had failed. This data set was reported by [1]

0.0251	0.0886	0.0891	0.2501	0.3113	0.3541	0.4763	0.5650	0.5761	0.6566	0.6748
0.6751	0.6753	0.7696	0.8375	0.8391	0.8425	0.8645	0.8851	0.9113	0.9120	0.9836
1.0483	1.0596	1.0773	1.1733	1.2570	1.2766	1.2985	1.3211	1.3503	1.3551	1.4595
1.4880	1.5728	1.5733	1.7083	1.7263	1.7460	1.7630	1.7746	1.8275	1.8375	1.8503
1.8808	1.8878	1.8881	1.9316	1.9558	2.0048	2.0408	2.0903	2.1093	2.1330	2.2100
2.2460	2.2878	2.3203	2.3470	2.3513	2.4951	2.5260	2.9911	3.0256	3.2678	3.4045
3.4846	3.7433	3.7455	3.9143	4.8073	5.4005	5.4435	5.5295	6.5541	9.096	

The $K-S$ statistics of the distance between the fitted and the empirical distribution functions (based on the parameters $\theta = 0.5324$ and $\alpha = 1.1182$ obtained by MLEs) is 0.0964 and the corresponding $p-value$ is 0.4516. Therefore, it is reasonable to use the PML distribution for fitting the data set. Based on the progressively Type-II censored schemes $(0, 0, 0, 0, 0, 0, 0, \ldots, 0, 20)$, we analyze the given data set and obtain the point estimates of parameters θ and α as described in Section 2. For computing the Bayes estimates, it is assumed that the priors of θ and α are improper, i.e. $a_1 = b_1 = a_2 = b_2 = 0$, since we do not have any prior information. For MCMC method using Metropolis-Hastings algorithm, we sample $N = 50000$ values and discard the initial $M = 5000$ as burn-in sample and calculate the Bayes estimates based on the remaining $N - M = 45000$ samples. The results are presented in Table 1.

TABLE 1. MLE and Bayes estimates under BSEL and BLINEX

	MLEs	ω	BSEL	BLINEX		
				$c=-3$	$c=0.0001$	$c=5$
θ	0.5106	0.0	0.5126	0.5190	0.5126	0.5025
		0.3	0.5120	0.5165	0.5120	0.5048
		0.6	0.5111	0.5136	0.5111	0.5071
		0.9	0.5108	0.5115	0.5108	0.5098
		1.0	0.5106	0.5106	0.5106	0.5106
α	1.2951	0.0	1.2915	1.3254	1.2915	1.2388
		0.3	1.2941	1.3180	1.2941	1.2563
		0.6	1.2945	1.3093	1.2951	1.2720
		0.9	1.2949	1.2984	1.2949	1.2889
		1.0	1.2951	1.2951	1.2951	1.2951

4. Conclusion

In this paper, maximum likelihood and Bayes estimators of the parameters have been obtained for PML distribution when sample is available from progressive Type-II censoring scheme. The Bayesian estimation is studied with respect to BSEL and BLINEX functions. From Table 1, we observe that with the increase ω, Bayes estimates of θ and α under the BSEL and BLINEX functions are close to the maximum likelihood estimation. For $\omega = 1$, Bayes estimates are equal to corresponding MLEs. Also observed that, the Bayes estimates under the BLENX function for $c = 0.0001$ is the same as the Bayes estimates under the BSEL function. For $\omega = 0$, the Bayes estimate are obtained under the SEL and LINEX function.

References

1. Abdul-Moniem, I. B. and Seham, M. (2015) *Transmuted Gompertz distri- bution*, Computational and Applied Mathematics, **1**, 88–96.
2. Balakrishnan, N. and Cramer, E. (2014) *The art of progressive censoring. Applications to reliability and auality*, New York: Birkhuser.
3. Chesneau, C. Tomy, L. and Jose, M. (2021) *Power modified Lindley Distribution: Theory and Applications*, Journal of Mathematical Extension, **16**, 1–32.
4. Jozani, M.J. Marchand, E. and Parsian, A. (2012) *Bayes and robust Bayesian estimation under a general class of balanced loss functions*, Statistical Papers, **53**, 51–60.
5. Zellner, A. (1986) *Bayesian estimation and prediction using asymmetric loss function*, J Am Stat Assoc, **81**, 446–451.

Confidence Intervals for the Power Modified Lindley Distribution based on Progressive Type-II Censoring Samples

Adeleh Fallah[1,*]
[1]Department of Statistics, Payame Noor University (PNU), Tehran, Iran.
Email: adelehfallah@pnu.ac.ir

Abstract. In this paper, maximum likelihood estimators of the parameters, reliability and hazard functions have been obtained for power modified Lindley distribution when sample is available from progressive Type-II censoring scheme. We compute confidence intervals based on the asymptotic method based on the MLE, credible interval, delta method and Bootstrap methods. In order to construct the asymptotic confidence intervals of the reliability and hazard functions, we need to find the variance of them, which are approximated by delta. One real data sets have been analyzed to demonstrate how the proposed methods can be used in practice.

Keywords: power modified Lindley distribution, confidence intervals, delta method, Bootstrap methods, Metropolis-Hastings algorithm

AMS Mathematics Subject Classification [2020]: 62N01, 62N02, 62N05

1. Introduction

The probability density function and cumulative distribution function of the power modified Lindley (PML) distribution are given by [3]

$$f_X(x;\theta,\alpha)=\frac{\alpha\theta}{1+\theta}\ e^{-2\theta x^{\alpha}}[(1+\theta)e^{\theta x^{\alpha}}x^{\alpha-1}+2\theta x^{2\alpha-1}-x^{\alpha-1}],\ x>0,\ \theta>0\ \alpha>0, \tag{1}$$

$$F_X(x;\theta,\alpha)=1-\left[1+\frac{\theta x^{\alpha}}{1+\theta}e^{-\theta x^{\alpha}}\right]e^{-\theta x^{\alpha}},\ x>0,\ \theta>0,\ \alpha>0. \tag{2}$$

The corresponding reliability function and hazard rate function are, respectively, given by

$$S_X(x;\theta,\alpha)=\left[1+\frac{\theta x^{\alpha}}{1+\theta}e^{-\theta x^{\alpha}}\right]e^{-\theta x^{\alpha}},\ x>0,\ \theta>0,\ \alpha>0. \tag{3}$$

$$H_X(x;\theta,\alpha)=\alpha\theta x^{\alpha-1}\left[\frac{\theta x^{\alpha}-1}{(1+\theta)e^{\theta x^{\alpha}}+\theta x^{\alpha}}+1\right],\ x>0,\ \theta>0,\ \alpha>0. \tag{4}$$

Under Progressive censoring scheme, from a total of n units placed simultaneously on a life test, only $(m<n)$ are completely observed until failure. Then, given a censoring plan

*Speaker.

$R_1, \ldots, R_m$: At the time $X_{1:m:n}$ of the first failure, R_1 of the $n-1$ surviving units are randomly withdrawn (or censored) from the life-testing experiment. At the time $X_{2:m:n}$ of the next failure, R_1 of the $n-R_1-2$ surviving units are censored, and so on. Finally, at the time $X_{m:m:n}$ of the mth failure, all the remaining $R_m = n-m-R_1-R_2-\ldots-R_{m-1}$ surviving units are censored, see; [2]. In this paper, we computed the 95% confidence intervals for θ, α, $S(t)$ and $H(t)$ based on the asymptotic distributions of the MLE, the Bayesian credible intervals, delta method and bootstrapping.

2. Interval Estimation

2.1. Asymptotic confidence interval. Under the progressively Type-II censored sample $\mathbf{X} = (X_1, \ldots, X_m)$, the likelihood function for $\zeta = (\theta, \alpha)$ is given by

$$L(\mathbf{x};\zeta) = A\prod_{i=1}^{m} f(x_i;\zeta)[1-F(x_i;\zeta)]^{R_i}, \tag{5}$$

where $A = n(n-R_1-1)(n-R_1-R_2-2)\ldots(n-R_1-\ldots-R_{m-1}-m+1)$ and $\mathbf{x} = (x_1, \ldots, x_m)$ is the vector of observation. By substituting Eqs. (1) and (2) into Eq. (5), the likelihood function is

$$\begin{aligned} L(\mathbf{x};\theta,\alpha) &= A\frac{\alpha^m\theta^m}{(1+\theta)^m}\, e^{-\theta\left[2\sum_{i=1}^m x_i^\alpha + \sum_{i=1}^m x_i^\alpha R_i\right]} \prod_{i=1}^{m}\left(1+\frac{\theta x_i^\alpha}{1+\theta}e^{-\theta x_i^\alpha}\right)^{R_i} \\ &\times \prod_{i=1}^{m}\left[(1+\theta)e^{\theta x_i^\alpha}x_i^{\alpha-1} + 2\theta x_i^{2\alpha-1} - x_i^{\alpha-1}\right]. \end{aligned} \tag{6}$$

Therefore, the maximum likelihood estimators (MLE) of θ and α, can be obtained by maximizing the log-likelihood function with respect to θ and α. Using the invariance property, the corresponding MLE of the reliability function $\hat{S}_{ML}(t)$ and hazard rate function $\hat{H}_{ML}(t)$ are obtained from Eqs. (3) and (4) after replacing θ and α by their MLEs $\hat{\theta}_{ML}$ and $\hat{\alpha}_{ML}$. Under some regularity conditions, the asymptotic joint distribution of the estimators is as follows $\begin{pmatrix} \widehat{\theta}_{MLE}-\theta \\ \widehat{\alpha}_{MLE}-\alpha \end{pmatrix} \xrightarrow{d} N_2(0, I^{-1}(\theta,\alpha))$, where matrix $I^{-1}(\theta,\alpha)$ is

$$I^{-1}(\theta,\alpha) = \begin{bmatrix} -\frac{\partial^2 \ln L(\mathbf{x};\theta,\alpha)}{\partial\theta^2} & -\frac{\partial^2 \ln L(\mathbf{x};\theta,\alpha)}{\partial\theta\partial\alpha} \\ -\frac{\partial^2 \ln L(\mathbf{x};\theta,\alpha)}{\partial\alpha\partial\theta} & -\frac{\partial^2 \ln L(\mathbf{x};\theta,\alpha)}{\partial\alpha^2} \end{bmatrix}^{-1}_{(\theta,\alpha)=(\widehat{\theta}_{MLE},\widehat{\alpha}_{MLE})} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}.$$

Thus, the $100(1-\gamma)\%$ approximate confidence intervals (CIs) for θ and α are

$$[\widehat{\theta}_l, \widehat{\theta}_u] = \hat{\theta}_{MLE} \pm z_{1-\frac{\gamma}{2}}\sqrt{I_{11}},\ [\widehat{\alpha}_l, \widehat{\alpha}_u] = \hat{\alpha}_{MLE} \pm z_{1-\frac{\gamma}{2}}\sqrt{I_{22}},$$

where, $z_{\gamma/2}$ is the upper $(\gamma/2)$ percentile of the standard normal distribution. Furthermore; to construct the asymptotic confidence intervals of the reliability and hazard functions, we need to find the variance them. In order to find the approximate estimates of the variance of $\hat{S}_t$ and $\hat{H}_t$, we use the delta method, see; [5]. Let $G_1' = \left(\frac{\partial S(t)}{\partial\theta}, \frac{\partial S(t)}{\partial\alpha}\right)$, $G_2' = \left(\frac{\partial H(t)}{\partial\theta}, \frac{\partial H(t)}{\partial\alpha}\right)$. Then the approximate estimates of $\widehat{S}$ and $\widehat{H}$ are given, respectively, by $\widehat{Var}(\widehat{S}) \simeq [G_1' I^{-1} G_1]_{(\theta,\alpha)=(\widehat{\theta}_{MLE},\widehat{\alpha}_{MLE})}$, $\widehat{Var}(\widehat{H}) \simeq [G_2' I^{-1} G_2]_{(\theta,\alpha)=(\widehat{\theta}_{MLE},\widehat{\alpha}_{MLE})}$. Thus, the $100(1-\gamma)\%$ approximate confidence intervals for $S(t)$ and $H(t)$ are

$$\widehat{S}_{(t)} \pm z_{1-\frac{\gamma}{2}}\sqrt{\widehat{Var}(\widehat{S})},\ \widehat{H}_{(t)} \pm z_{1-\frac{\gamma}{2}}\sqrt{\widehat{Var}(\widehat{H})},$$

2.2. Parametric bootstrap method. We propose two confidence intervals based on bootstrapping. The two bootstrap methods that are widely used in practice are; (i) percentile bootstrap method (Boot-p) and (ii) bootstrap-t method (Boot-t), see, for example, [4]

2.3. Bayesian Credible Intervals. It is assumed that θ and α have independent gamma priors with the pdfs

$$\pi(\theta) = \frac{b_1^{a_1}\theta^{a_1-1}\ e^{-b_1\theta}}{\Gamma(a_1)}, \pi(\alpha) = \frac{b_2^{a_2}\alpha^{a_2-1}\ e^{-b_2\alpha}}{\Gamma(a_2)}, \tag{7}$$

where a_1, b_1 and a_2, b_2 are positive hyperparameters. From Eqs. (6) and (7), the joint posterior density of θ and α becomes

$$\pi(\theta,\alpha|\mathbf{x}) = \frac{\theta^{m+a_1-1}\alpha^{m+a_2-1}e^{-b_2\alpha}}{K\ (1+\theta)^m} e^{-\theta[\sum_{i=1}^m x_i^\alpha(2+R_i)+b_1]} \prod_{i=1}^m \left(1+\frac{\theta x_i^\alpha}{1+\theta}e^{-\theta x_i^\alpha}\right)^{R_i}$$
$$\times \prod_{i=1}^m \left[(1+\theta)e^{\theta x_i^\alpha}x_i^{\alpha-1} + 2\theta x_i^{2\alpha-1} - x_i^{\alpha-1}\right], \tag{8}$$

where

$$K = \int_0^\infty \int_0^\infty \frac{\theta^{m+a_1-1}\alpha^{m+a_2-1}}{e^{b_2\alpha}(1+\theta)^m} e^{-\theta[\sum_{i=1}^m x_i^\alpha(2+R_i)+b_1]}$$
$$\times \prod_{i=1}^m \left[(1+\theta)e^{\theta x_i^\alpha}x_i^{\alpha-1} + 2\theta x_i^{2\alpha-1} - x_i^{\alpha-1}\right] \prod_{i=1}^m \left(1+\frac{\theta x_i^\alpha}{1+\theta}e^{-\theta x_i^\alpha}\right)^{R_i} d\theta d\alpha.$$

The joint posterior density of θ and α in Eq. (8) is unknown. For this, we will use the Metropolis-Hastings algorithm with a bivariate normal distribution as the proposal density distribution as proposal distribution to generate random variates from the joint posterior distribution. The Metropolis-Hastings algorithm steps are given below:

(1) Set the initial value $\zeta_0 = (\theta_0, \alpha_0)$ and $j = 1$.
(2) Using the Metropolis-Hastings algorithm, generate $\zeta_j = (\theta_j, \alpha_j)$ from $\pi(\theta_{(j-1)}, \alpha_{(j-1)}|\mathbf{x})$ with the bivariate normal distribution $N_2(\zeta_{(j-1)}, \Sigma = I^{-1}(\theta,\alpha))$ as proposal distribution.
(3) From (3) and (4), we Compute S_j and H_j.
(4) Set $j = j+1$ and Repeat (2)-(3), N times to compute the Markov Chain Monte Carlo (MCMC) samples $\{(\theta_1,\alpha_1),\ldots,(\theta_N,\alpha_N)\}$ and $\{(S_1,H_1),\ldots,(S_N,H_N)\}$.

We order the MCMC sample after burn-in in ascending order to obtain $\theta_{[M+1]} < \theta_{[M+2]} < \ldots < \theta_{[N]}$, a credible interval can be used to construct $\left(\theta_{[(\frac{\gamma}{2})(N)]}, \theta_{[(1-\frac{\gamma}{2})(N)]}\right)$, where $\theta_{[(\frac{\gamma}{2})(N)]}$ and $\theta_{[(1-\frac{\gamma}{2})(N)]}$ are the $[\frac{\gamma}{2}(N)]$-th smallest integer and the $[(1-\frac{\gamma}{2})(N)]$-th smallest integer of $\{\theta_j :\ j = M+1, M+2, \cdots, N\}$, respectively. Similay, we obtained the credible interval of the unknown parameter α, reliability and hazard functions. M is the burn-in period.

3. Numerical results

Example 3.1. This data represents the life of fatigue fracture of Kevlar 373/epoxy subjected to constant pressure at 90% stress level until all had failed. This data set was reported by [1]

0.0251	0.0886	0.0891	0.2501	0.3113	0.3541	0.4763	0.5650	0.5761	0.6566	0.6748
0.6751	0.6753	0.7696	0.8375	0.8391	0.8425	0.8645	0.8851	0.9113	0.9120	0.9836
1.0483	1.0596	1.0773	1.1733	1.2570	1.2766	1.2985	1.3211	1.3503	1.3551	1.4595
1.4880	1.5728	1.5733	1.7083	1.7263	1.7460	1.7630	1.7746	1.8275	1.8375	1.8503
1.8808	1.8878	1.8881	1.9316	1.9558	2.0048	2.0408	2.0903	2.1093	2.1330	2.2100
2.2460	2.2878	2.3203	2.3470	2.3513	2.4951	2.5260	2.9911	3.0256	3.2678	3.4045
3.4846	3.7433	3.7455	3.9143	4.8073	5.4005	5.4435	5.5295	6.5541	9.096	

The $K-S$ distance and its respective $p-value$ are computed to be $K-S=0.0964$ and $p-value=0.4516$, respectively. Therefore, it is quite reasonable to indicate that the PML distribution is fitting this data well. Based on the progressively Type-II censored schemes $(0,0,0,0,0,0,0,\ldots,0,20)$, we analyze the given data set and obtain the interval estimates of parameters θ, α, reliability and hazard functions as described in Section 2. The MLEs of parameters, reliability and hazard functions are obtained to be $(\hat{\theta}_{ML},\hat{\alpha}_{ML},\hat{S}_{ML}(t=2),\hat{H}_{ML}(t=2))=(0.5106,1.2951,0.3532,0.8429)$. For computing the credible intervals, it is assumed that the priors of θ and α are improper, i.e. $a_1=b_1=a_2=b_2=0$, since we do not have any prior information. For MCMC method using Metropolis-Hastings algorithm, we sample $N=50000$ values and discard the initial $M=5000$ as burn-in sample and calculate the credible intervals based on the remaining $N-M=45000$ samples. The results are presented in Table 1.

TABLE 1. The 95% confidence intervals for θ, α, $S(t=2)$, and $H(t=2)$

Method	θ	α	$S(t)$	$H(t)$
ML	(0.3651, 0.6561)	(0.9559, 1.6343)	(0.1783, 0.528)	(0.3203, 1.3654)
MCMC	(0.3976, 0.6453)	(1.0158, 1.5848)	(0.2633, 0.4506)	(0.5736, 1.1871)
Boot-p	(0.3771, 0.6365)	(1.0571, 1.6596)	(0.2466, 0.4487)	(0.6173, 1.2851)
Boot-t	(0.3898, 0.6530)	(1.0567, 1.6647)	(0.2430, 0.4430)	(0.6215, 1.2983)

4. Conclusion

In this paper, we have discussed an iterative procedure for obtaining the MLEs based on progressively Type-II censored samples from a two-parameter PML distribution. This article also studied the construction of CIs for the reliability and hazard functions by using some methods as parametric bootstrap, delta method. We have proposed to use the MCMC technique to compute the credible interval. The interval estimates obtained by ML, MCMC and bootstrap methods are also similar.

References

1. Abdul-Moniem, I. B. and Seham, M. (2015) *Transmuted Gompertz distri- bution*, Computational and Applied Mathematics, **1**, 88–96.
2. Balakrishnan, N. and Cramer, E. (2014) *The art of progressive censoring. Applications to reliability and auality*, New York: Birkhuser.
3. Chesneau, C. Tomy, L. and Jose, M. (2021) *Power modified Lindley Distribution: Theory and Applications*, Journal of Mathematical Extension, **16**, 1–32.
4. Efron, B. and Tibshirani, R.J. (1994) *An Introduction to the Bootstrap*, CRC Press, Boca Raton, FL.
5. Greene, W.H. (2000) *Econometric Analysis*, 4th ed., Prentice-Hall, NewYork.

Hierarchical estimation of Rayleigh distribution

Kazem Fayyaz Heidarì, Ph. D, Department of Statistics,
Payame Noor University, P.O. Box, 19395-3697, Tehran, Iran
fayyaz@pnu.ac.ir

Abstract: This paper addresses the problems of estimation when the lifetime data following Rayleigh distribution are observed under progressive type-II censoring. We obtain Bayes and Hierarchical estimates using squared error (SE) loss function. Finally, we conduct a simulation study to compare the performance of the proposed methods of estimation.

Keywords: Hierarchical Bayesian, Rayleigh distribution, Progressive type-II censoring.

1. Introduction

Rayleigh distribution was originally introduced by Rayleigh [7] in the field of acoustics; since its introduction, many researchers have used Rayleigh distribution in various fields of science and technology. Nowadays, Rayleigh distribution is widely used in statistical model, survival analysis and reliability theory. Rayleigh distribution is the foundation of much of the treatment of meteorological radar signals statistics. Moreover, it is often applied in actuarial science and in engineering work to model population lifetimes whose failure rate increases linearly. Also, Censoring is a common practice in longevity and reliability studies. If the experimenter determines that, after observing the first failure, R_1 units of healthy test units and at the time of the second failure, R_2 unit of healthy test units will be cached out of the test and continue until mth failure, all the remaining units of the test $R_m = n - R_1 - R_2 - \ldots - R_{m-1} - m$ outside Then The Progressive Type-II sensor will take place. In this case, the failure times of units are random variables and R_i s are predetermined constants. As a result of this censorship plan, m ordered amount is obtained which ordinal statistics are called progressive type-II censored [1].

Hierarchical Bayesian prior distribution was initially introduced by Lindley and Smith [6], Then examined by Han [5].

The probability density function and cumulative distribution function of the Rayleigh distribution are respectively, as follows.

$$f(x,\lambda) = 2\lambda x e^{-\lambda x^2}, x > 0, \lambda > 0 \quad (1)$$

$$F(x,\lambda) = 1 - e^{-\lambda x^2}, x > 0, \lambda > 0 \quad (2)$$

First, we will obtain the Bayesian and Hierarchical Bayesian estimation of the Rayleigh distribution parameter under SE and the Progressive Type-II censored. In continuation, these estimations will be obtained using Monte Carlo simulation. Finally, will be compared them.

2. Estimation of λ

It is assumed that the Prior distribution of λ is the gamma distribution with the Hyper-parameter

[1] . Corresponding Author

a and b as follows.

$$\pi(\lambda|a,b)=\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-\lambda b},\lambda>0,a,b>0 \quad (3)$$

According to Han [5], the Hyper-parameter a and b are considered to be $\pi(\lambda|a,b)$ decreasing relative to λ.

Thus, obtain $\frac{d[\pi(\lambda|a,b)]}{d\lambda}=\frac{b^a\lambda^{a-2}e^{-b\lambda}}{\Gamma(a)}[(a-1)-b\lambda]$ That should be $b>0$ and $0<a\leq 1$. Berger [4] showed that increasing b would decrease the efficiency of the Bayes estimator λ.Therefore, the Hyper-parameter b must be bounded above and be $0<b<c$. Showed that the most appropriate distribution b is uniform distribution.Therefore, in this paper, $\pi_1(b)$ is a continuous uniform distribution in the interval $(0,c)$ and $a=1$. In this case, relation (3) becomes $\pi(\lambda|b)=be^{-\lambda}$.

2.1 Bayesian estimation of λ

First, we will obtain the Bayes estimation and then the Hierarchical Bayesian estimation of λ with the loss function $L(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2$. If Rayleigh distribution with probability density function (1), the distribution of the failure time, and $Y_{1:m:n},Y_{2:m:n},Y_{m:m:n}$ is Progressive Type-II censored sample of it, and y_i is the finding of $Y_{1:m:n}$, then according to [2], the likelihood function is obtained as follows

$$L(\lambda\mid\boldsymbol{Y})=C\left(\prod_{i=1}^{m}y_i\right)2^m\lambda^m e^{-\lambda\sum_{i=1}^{m}(1+R_i)y_i^2} \quad (4)$$

where, $\boldsymbol{Y}=(Y_{1:m:n},Y_{2:m:n},\ldots,Y_{m:m:n})$, $C=n(n-R_1-1)\ldots(n-R_1-\ldots-R_{m-1}-m+1)$

According to (4), posterior distribution of λ is obtained as follows.

$$\pi^*(\lambda\mid\boldsymbol{Y})=\frac{(b+\sum_{i=1}^{m}(1+R_i)y_i^2)^{m+1}}{m!}\lambda^m e^{-\lambda(b+\sum_{i=1}^{m}(1+R_i)y_i^2)} \quad (5)$$

Therefore, the Bayes estimation of the parameter λ under the squared error loss function is as follows

$$\hat{\lambda}_{Bay}(b)=E(\lambda|Y)=\frac{m+1}{b+\sum_{i=1}^{m}(1+R_i)y_i^2} \quad (6)$$

2.2 Hierarchical Bayesian estimation of λ

If b is a Hyper-parameter in the parameter θ and the prior density function θ, $\pi(\theta|b)$ and the prior density function of the Hyper-parameter b, is $\pi_1(b)$, then the Hierarchical prior density function θ is defined as follows

$$\pi_2(\theta)=\int_{\Lambda}\pi(\theta|b)\pi_1(b)db,\qquad b\epsilon\Lambda$$

Therefore, the Hierarchical prior density function λ is obtained as follows

$$\pi_2(\theta)=\int_0^c\pi(\lambda|b)\pi_1(b)db=\frac{1}{c}\int_0^c be^{-b\lambda}db=\frac{1-(1+\lambda c)e^{c\lambda}}{c\lambda^2} \quad (7)$$

As a result, Bayesian posterior density function λ is obtained.

$$\pi^{**}(\lambda \mid \mathbf{Y}) = \frac{\lambda^{m-2}[1-(1+\lambda c)e^{-c\lambda}]e^{-\lambda\sum_{i=1}^{m}(1+R_i)y_i^2}}{\int_0^{\infty}\lambda^{m-2}[1-(1+\lambda c)e^{-c\lambda}]e^{-\lambda\sum_{i=1}^{m}(1+R_i)y_i^2}d\lambda} \quad (8)$$

Now, using the relation (8), the Hierarchical Bayesian estimation of the parameter λ is obtained as follows.

$$\hat{\lambda}_{HBay} = E_{\pi^{**}}(\lambda \mid Y) = \frac{m-1}{T(c+T)} \cdot \frac{(c+T)^{m+1} - T^{m+1} - (m+1)cT^{m}}{(c+T)^{m} - T^{m} - mcT^{m-1}} \quad (9)$$

where $T = \sum_{i=1}^{m}(1+R_i)y_i^2$.

3. Simulation

Using [3], Progressive Type-II censored samples are obtained from the Rayleigh distribution with probability density function (1) and with parameter $\lambda = 2$ as follows.

Step 1: First, we generate $m$random number independent of the standard uniform distribution and shown with $W_1, W_2, \ldots, W_m$.

Step 2: For, $i = 1.2.\ldots..m, V_i = W_i^{\left\{1/(i+\sum_{j=m+1-i}^{m} R_j)\right\}}$.

Step 3: For, $i = 1.2.\ldots..m, U_{i:m:n} = 1 - \Pi_{j=1}^{i} V_{m+1-j}$, in this case, $U_{1:m:n}, U_{2:m:n}, \ldots, U_{m:m:n}$ is the Progressive Type-II censored sample of standard uniform distribution.

Step 4: $Y_{i:m:n} = F^{-1}(U_{i:m:n})$ is obtained for $i = 1.2.\ldots..m$.

In this case, $Y_{1:m:n}, Y_{2:m:n}, \ldots, Y_{m:m:n}$ are the Progressive Type-II censored order statistics of Rayleigh distribution with probability density function (1). Now, we obtain the Bayesian and H-Bayesian estimators by using the generated sample $Y_{1:m:n}, Y_{2:m:n}, \ldots, Y_{m:m:n}$, and relations (6), (7) and (9).The first to fourth steps are repeated 1000 times, and then the average values of estimators and the root mean square error of the estimators are calculated. The results are presented in Table indicates that the Bayesian estimation is better.

Table. Average estimates and $(\sqrt{MSE})$ under quadratic loss function for $b = 1.5$ and $c = 2$

$\hat{\lambda}_{HBay}$	$\hat{\lambda}_{Bay}$	$design$	M	N
4/551(2/570)	0/845(0/129)	(5,0,…..,0)	5	15
4/157(2/371)	1/452(0/325)	(2,3,0,...,0)	5	15
7/119(2/952)	1/216(0/487)	(10,0,....,0)	10	25
6/153(2/374)	2/705(0/297)	(3,4,3,0,..,0)	10	25
5/727(4/634)	2/542(0/326)	(0,20,….,0)	20	60
6/275(2/474)	3/950(0/917)	(1,4,5,0,..,0)	20	60
7/037(2/374)	5/424(1/548)	(1,……..,1)	20	60
9/252(4/850)	6/815(0/376)	(15,15,0,..,0)	50	90

4. Conclusion

Based on censored samples, some researches on parameter estimation of Rayleigh distribution have already been conducted. In this study, Bayesian and H-Bayesian estimations of the Rayleigh distribution parameter based on the Progressive Type-II censored samples with squared

error loss function were obtained. By calculating the MSE and the average estimation, the Bayesian and H-Bayesian estimations based on the Rayleigh distribution were compared using Monte Carlo simulation. It has been shown that the Bayesian estimation has better efficiency.

References

1. Asgharzadeh, a. and Moradnejad, p. (2018). Inference for a Skew Normal Distribution Based on Progressively Type-II Censored Samples, Journal of Statistical Research of Iran, 5(1), 33-56.
2. Balakrishnan, N, and Aggarwala, R. (2020). Progressive Censoring: Theory. Methods and Applications.
3. Balakrishnan. N. and Sandhu, R.A.A. (1995). Simple simulational algorithm for generating progressive Type-II censored samples, The Amer. Statist, 49, 229-230.
4. Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis, second ed., Springer -Verlag, New York.
5. Han, M. (1997). The structure of hierachical prior distribution and its applications, Chinese Operations Research and Management Science. 6(3), 31-40.
6. Lindley, D.V. and Smith, A.F. (1972). Baycs estimation for the linear model. Journal of the Royal Statistical Society Series B, 34, 1-41.
7. Rayleigh, J. W.S. (1880), On the resultant of a Large number of vibration of the somepith and of arbitrary phase. Philosophical Magazine. 5th series, 10, 73-78.

The effect of the loss function on Bayes Estimator and the posterior risk for the exponential distribution

Elham Basiri*

Department of Mathematics, Kosar University of Bojnord, Bojnord, Iran.
Email: elhambasiri@kub.ac.ir

Abstract. In this paper, mathematical properties of the exponential distribution via Bayesian approach are derived under different loss functions. These properties include Bayes estimators and the corresponding posterior risks function for the parameter. Numerical computations are presented to show the usefulness of the results.

Keywords: Progressively Type II censoring, Exponential distribution, Posterior risk function

AMS Mathematics Subject Classification [2020]: 62F15, 62N01

1. Introduction

The scheme of progressive Type-II censoring is of importance in life-testing experiments. It allows the experimenter to remove units from a life test at various stages during the experiment. Suppose n units are placed on a lifetime test. At the first failure time, R_1 surviving items are randomly withdrawn from the test. At the second failure time, R_2 surviving items are selected at random and taken out of the experiment, and so on. Finally, at the time of the m-th failure, the remaining R_m objects are removed, where $\sum_{i=1}^{m} R_i = n - m$. We suppose that $X_{1:m:n}, \ldots, X_{m:m:n}$ are the progressively Type-II censored order statistics associated with a random sample of size m with censoring scheme $\tilde{R} = (R_1, \cdots, R_m)$ from the one-parameter exponential distribution with the probability density function (pdf) and the cumulative distribution function (cdf) given by

$$f_\theta(x) = \theta e^{-\theta x} \qquad F_\theta(x) = 1 - e^{-\theta x}, \qquad x > 0,\ \theta > 0. \tag{1}$$

For convenience, we will use throughout this paper X_i for $X_{i:m:n}$.

First we recall that the joint pdf of the progressively Type-II censored order statistics $X_1, \ldots, X_m$ with pdf and cdf in (1) is (see, for example, [1], [2], [3])

$$f_{X_1,\ldots,X_m;\theta}(x_1, \ldots, x_m) = C \prod_{i=1}^{m} (1 - F_\theta(x_i))^{R_i} f_\theta(x_i), \tag{2}$$

in which $C = \prod_{i=1}^{m} \left(n - i + 1 - \sum_{j=1}^{i-1} R_j\right)$, with $\sum_{j=1}^{0} R_j \equiv 0$.

*Speaker.

From (1) and (2), the likelihood function for a random sample $X_1, \cdots, X_m$ which is taken from the exponential distribution is given by

$$L(\theta) = C\theta^m e^{-\theta t}, \tag{3}$$

where t is the observed value for T.

In the following, we consider the uniform prior, which has the form

$$\pi_1(\theta) \propto 1, \quad \theta > 0.$$

The posterior distribution for θ associated with the uniform prior takes the form

$$\pi_1(\theta|x_1, \cdots, x_m) = \frac{\theta^m e^{-\theta t}}{\int_0^\infty \theta^m e^{-\theta t} d\theta} = \frac{t^{m+1}}{m!}\theta^m e^{-\theta t}, \quad \theta > 0.$$

So, for $s = \cdots, -2, -1, 1, 2, \cdots$, we have

$$E(\theta^s|x_1, \cdots, x_m) = \frac{t^{m+1}}{m!}\int_0^\infty \theta^{m+s} e^{-\theta t} d\theta = \frac{(m+s)!}{m!}\frac{1}{t^s}. \tag{4}$$

Moreover, we consider different loss functions as squared error loss function (SELF), weighted squared error loss function (WSELF), precautionary loss function (PLF), modified (quadratic) squared error loss function (M/Q SELF) and K-Loss function (KLF). In the following under the different loss functions the point estimator for as well as the their posterior risk for θ are obtained.

1. SELF: Under the SELF the loss function is $L_1 = (\hat{\theta} - \theta)^2$, where $\hat{\theta}$ is an estimator of θ. In this case, based on (4), the point estimator for θ and the associated posterior risk function are respectively given by

$$\hat{\theta}_{SELF} = E(\theta|x_1, \cdots, x_m) = \frac{m+1}{T},$$

and

$$\begin{aligned} R_1\left(\hat{\theta}_{SELF}, \theta\right) &= V(\theta|x_1, \cdots, x_m) = E(\theta^2|x_1, \cdots, x_m) - (E(\theta|x_1, \cdots, x_m))^2 \\ &= \frac{(m+2)(m+1)}{t^2} - \frac{(m+1)^2}{t^2} = \frac{m+1}{t^2}. \end{aligned}$$

2. WSELF: For this case, the loss function is $L_2 = \frac{(\hat{\theta}-\theta)^2}{\theta}$ and the point estimator for θ and the associated posterior risk function are

$$\hat{\theta}_{WSELF} = \left[E(\theta^{-1}|\mathbf{x})\right]^{-1} = \frac{m}{T},$$

and

$$R_2\left(\hat{\theta}_{WSELF}, \theta\right) = E(\theta|\mathbf{x}) - \left[E(\theta^{-1}|\mathbf{x})\right]^{-1} = \frac{m+1}{T} - \frac{m}{T} = \frac{1}{T}.$$

3. PLF: Here, the loss function is $L_3 = \frac{(\hat{\theta}-\theta)^2}{\hat{\theta}}$ and the point estimator for θ and the associated posterior risk function are

$$\hat{\theta}_{PLF} = \sqrt{E(\theta^2|\mathbf{x})} = \frac{\sqrt{(m+2)(m+1)}}{T},$$

and

$$R_3\left(\hat{\theta}_{PLF}, \theta\right) = 2\left[\sqrt{E(\theta^2|\mathbf{x})} - E(\theta|\mathbf{x})\right] = \frac{2}{T}\left[\sqrt{(m+2)(m+1)} - (m+1)\right].$$

4. M/ Q SELF: For this case, the loss function is $L_4 = \left(1 - \frac{\hat{\theta}}{\theta}\right)^2$ and the point estimator for θ and the associated posterior risk function are

$$\hat{\theta}_{MQSELF} = \frac{E(\theta^{-1}|\mathbf{x})}{E(\theta^{-2}|\mathbf{x})} = \frac{m-1}{T},$$

and

$$R_4\left(\hat{\theta}_{MQSELF}, \theta\right) = 1 - \frac{\left[E(\theta^{-1}|\mathbf{x})\right]^2}{E(\theta^{-2}|\mathbf{x})} = 1 - \frac{(m-1)(m-1)!}{m!} = \frac{1}{m}.$$

5. KLF: In this case, the loss function is $L_5 = \left[\sqrt{\frac{\hat{\theta}}{\theta}} - \sqrt{\frac{\theta}{\hat{\theta}}}\right]^2$ and the point estimator for θ and the associated posterior risk function are respectively given by

$$\hat{\theta}_{KLF} = \sqrt{\frac{E(\theta|\mathbf{x})}{E(\theta^{-1}|\mathbf{x})}} = \frac{\sqrt{m(m+1)}}{T},$$

and

$$R_5\left(\hat{\theta}_{KLF}, \theta\right) = 2\left[E(\theta|\mathbf{x})E(\theta^{-1}|\mathbf{x}) - 1\right] = \frac{2}{m}.$$

2. Numerical results

Here, we present the analysis of real data, partially considered in [4] for illustrative purposes. Records were kept for the log-times of breakdown of an insulating fluid in an accelerated test conducted in various test voltages. The values for 38 KV are
0.0899, 0.3899, 0.4699, 0.7299, 0.7399, 1.1299, 1.3998, 2.3799.
By applying the Kolmogorov-Smirnov (K-S) test for this data set for fitting exponential distribution, we have found p-value 0.8268 and K-S distance 0.2216. The maximum likelihood estimations (MLEs) is $\hat{\theta} = 1.0914$. So, the exponential distribution with parameter $\theta = 1.0914$, is an adequate model for this data set. Based on this data set, we have computed the values of the posterior risk function and the results are presented in Table 1. Different choices for m and $\tilde{R}$ ($R_1 = (n-m, 0, \cdots, 0)$ and $R_2 = (0, \cdots, 0, n-m)$) are considered. Table 1 confirms that the posterior risk for MQSELF and KLF is independent of the censoring scheme., as we expected. Comparing the two censoring schemes, the scheme $R_1 = (n-m, 0, \cdots, 0)$ yields lower posterior risk across all loss functions (excluding MQSELF and KLF, which are constant for a fixed m) than the scheme $R_2 = (0, \cdots, 0, n-m)$. Also, one can observe that the posterior risk is a decreasing function of m, when all other parameters are kept fixed. These observations support the general conclusion that the posterior risk is a decreasing function of t (and consequently n and m).

3. Conclusion

This study compared various Bayes estimators under different loss functions. Numerical results confirm that the posterior risk is a decreasing function of the total observed time (t), and consequently, a decreasing function of the number of failures (m). Furthermore, the posterior risk for the $\hat{\theta}_{MQSELF}$ and $\hat{\theta}_{KLF}$ estimators is demonstrably independent of the specific censoring scheme chosen.

TABLE 1. The values of the posterior risk function and the Bayes estimators under different loss functions

m	$\tilde{R}$	$(x_1, \cdots, x_m)$	t	Loss Function	Posterior Risk	$\hat{\theta}$
3	$(5, 0, 0)$	$(0.0899, 1.3998, 2.3799)$	4.3191	SELF	0.2144	0.9261
				WSELF	0.2315	0.6945
				PLF	0.2186	1.0354
				MQSELF	0.3333	0.4630
				KLF	0.6666	0.8020
3	$(0, 0, 5)$	$(0.0899, 0.3899, 0.4699)$	3.2992	SELF	0.3674	1.2124
				WSELF	0.3031	0.9093
				PLF	0.2862	1.3555
				MQSELF	0.3333	0.6062
				KLF	0.6666	1.0499
5	$(3, 0, 0, 0, 0)$	$(0.0899, 0.7399, 1.1299, 1.3998, 2.3799)$	6.0091	SELF	0.1661	0.9984
				WSELF	0.1664	0.8320
				PLF	0.1600	1.0784
				MQSELF	0.2000	0.6656
				KLF	0.4000	0.9114
5	$(0, 0, 0, 0, 3)$	$(0.0899, 0.3899, 0.4699, 0.7299, 0.7399)$	4.6392	SELF	0.2787	1.2933
				WSELF	0.2155	1.0777
				PLF	0.2072	1.3969
				MQSELF	0.2000	0.8622
				KLF	0.4000	1.1806

References

1. Balakrishnan, N. and Aggarwala, R. (2000) *Progressive censoring: Theory, methods, and applications*, Birkhuser, Springer, Boston.
2. Balakrishnan, N. (2007) *Progressive censoring methodology: an appraisal*, Test, **16**(2), 211–259.
3. Balakrishnan, N. and Cramer, E. (2014) *The art of progressive censoring*, Birkhuser, Springer, New York.
4. Nelson, W. (1982) *Applied life data analysis*, New York: Wiley.

Comparing three different censoring schemes from the perspective of experimental design cost

Elham Basiri* and Elham Hosseinzadeh

Department of Mathematics, Kosar University of Bojnord, Bojnord, Iran.
Email: elhambasiri@kub.ac.ir

Abstract. Type I and Type II hybrid censoring schemes are designed to overcome the drawbacks of conventional Type I and Type II censoring schemes by setting both a maximum test duration and a minimum required number of failures. The key difference lies in the termination rule, which determines the experiment's stopping time. On the other hand, cost is always an important criterion in decision-making because we often face cost constraints in experiments. This article attempts to compare three different censoring methods, as the conventional type II and type I and II hybrid censoring schemes, from a cost perspective.

Keywords: Cost function, Conventional Type II censoring, Type I hybrid censoring, Type II hybrid censoring

AMS Mathematics Subject Classification [2020]: 62N01, 62N05

1. Introduction

Hybrid censoring schemes are widely used in life-testing and reliability experiments to balance the need for a fixed experiment duration with the desire to observe a certain number of failures for better statistical inference. Type I hybrid censoring scheme prioritizes controlling the experiment duration. It is terminated at or before a pre-fixed time T, even if the target number of failures r has not been reached. Type II hybrid censoring scheme prioritizes controlling the number of observations (failures). It is terminated at or after a pre-fixed time T to ensure at least r failures are observed. Both schemes aim to balance the conflicting goals of the Type I and Type II censoring schemes to ensure the experiment is not too long and yields reasonably sufficient data for analysis. However, they achieve this balance through different priorities.

On the other hand, practical life-testing and survival analysis experiments are conducted under budgetary and resource constraints. An optimal experimental design, including the choice of a censoring scheme, must strike a balance between statistical efficiency (precision of estimation) and the total cost of the experiment.

Here, we compare three mentioned censoring schemes based on the cost of experiment.

*Speaker.

2. Main Results

In this paper, we assume that $\tilde{X} = (X_1, \ldots, X_n)$ is a sample of n units, from the one parameter exponential distribution with the probability density function (pdf) as

$$f_\theta(x) = \theta e^{-\theta x}, \qquad x > 0, \theta > 0. \tag{1}$$

Here, the corresponding order statistics are shown by $X_{1:n} \leq \cdots \leq X_{n:n}$ with the corresponding observed values as $x_{1:n} \leq \cdots \leq x_{n:n}$. We intend to compare three different censoring schemes, as the conventional Type II, Type I hybrid censoring and Type II hybrid censoring schemes. The criterion is the cost of experiment which is defined as

$$C(n, r, T) = C_0 + C_n n + C_d E(D) + C_t E(T^*), \tag{2}$$

where C_0, C_n, C_d and C_t are the set-up cost or any other related cost involved in sampling, the cost per unit, the cost per failed unit item, and the cost per unit of duration of life-testing, respectively. Also, D is the number of failures and T^* is the duration of the experiment.

Conventional Type II censoring: For conventional Type II censoring, the experiment stops at the r-th failure time, so $T^* = X_{r:n}$, and the number of failures is $D = r$, which r is a pre-fixed value. In this case, D is a fixed and T^* is a random variable. In this case, we have $E(D) = r$ and (see for example, [1])

$$E(T^*) = E(X_{r:n}) = \frac{1}{\theta} \sum_{j=1}^{r} \frac{1}{n-j+1}. \tag{3}$$

Type I hybrid censoring: In the context of Type I hybrid censoring, we have D and $T^* = min(X_{r:n}, T)$ as the number of failures and the duration of the test, respectively, in which r and T are pre-fixed values. Clearly, is this case D and T are both random variables. On the other hand, for Type I hybrid censoring we get two cases as

$$\begin{cases} \text{Case I}: & \{x_{1:n}, \ldots, x_{r:n}\}, \quad \text{if} \quad x_{r:n} < T, \\ \text{Case II}: & \{x_{1:n}, \ldots, x_{D:n}\}, \quad \text{if} \quad T < x_{r:n}, \quad 0 \leq D < r. \end{cases}$$

So, we obtain (see, for example, [3])

$$\begin{aligned} P(D = j) &= \binom{n}{j} (F_\theta(T))^j \left(\bar{F}_\theta(T)\right)^{n-j}, \quad j = 0, 1, \ldots, r-1, \\ P(D = r) &= \sum_{j=r}^{n} \binom{n}{j} (F_\theta(T))^j \left(\bar{F}_\theta(T)\right)^{n-j}. \end{aligned}$$

So, we can write (see, for example, [3])

$$E(D) = \sum_{j=0}^{r-1} j \binom{n}{j} (F_\theta(T))^j \left(\bar{F}_\theta(T)\right)^{n-j} + r \sum_{j=r}^{n} \binom{n}{j} (F_\theta(T))^j \left(\bar{F}_\theta(T)\right)^{n-j}.$$

From (1) and the binomial expansion , we find that

$$\begin{aligned} E(D) &= \sum_{j=0}^{r-1} \sum_{k=0}^{j} j \binom{n}{j} \binom{j}{k} (-1)^k \left(\bar{F}_\theta(T)\right)^{n-j+k} + r \sum_{j=r}^{n} \sum_{k=0}^{j} \binom{n}{j} \binom{j}{k} (-1)^k \left(\bar{F}_\theta(T)\right)^{n-j+k} \\ &= \sum_{j=0}^{r-1} \sum_{k=0}^{j} j \binom{n}{j} \binom{j}{k} (-1)^k e^{-\theta(n-j+k)T} + r \sum_{j=r}^{n} \sum_{k=0}^{j} \binom{n}{j} \binom{j}{k} (-1)^k e^{-\theta(n-j+k)T}. \end{aligned} \tag{4}$$

On the other hand, we have

$$\begin{aligned} E(T^*) &= E(\min(X_{r:n}, T)) \\ &= T.\bar{F}_{X_{r:n}}(T) + \int_0^T x f_{X_{r:n}}(x)dx \\ &= T.\bar{F}_{X_{r:n}}(T) - T.\bar{F}_{X_{r:n}}(T) + \int_0^T \bar{F}_{X_{r:n}}(x)dx \\ &= \int_0^T \bar{F}_{X_{r:n}}(x)dx, \end{aligned}$$

where the last equality is obtained by integrating by parts. Based on [4] and (1), we get

$$\begin{aligned} E(T^*) &= \sum_{j=0}^{r-1} \binom{n}{j} \int_0^T (F_\theta(x))^j \left(\bar{F}_\theta(x)\right)^{n-j} dx \\ &= \sum_{j=0}^{r-1}\sum_{k=0}^{j} \binom{n}{j}\binom{j}{k}(-1)^k \int_0^T e^{-\theta(n-j+k)x}dx \\ &= \sum_{j=0}^{r-1}\sum_{k=0}^{j} \binom{n}{j}\binom{j}{k}(-1)^k \frac{1-e^{-\theta(n-j+k)T}}{\theta(n-j+k)}. \end{aligned} \tag{5}$$

Type II hybrid censoring: Assuming Type II hybrid censoring, we have D and $T^* = max(X_{r:n}, T)$ as the number of failures and the duration of the test, respectively. On the other hand, for Type II hybrid censoring there are three cases as

$$\begin{cases} \text{Case I}: & \{x_{1:n}, \ldots, x_{r:n}\}, \quad \text{if} \quad T < x_{r:n}, \\ \text{Case II}: & \{x_{1:n}, \ldots, x_{D:n}\}, \quad \text{if} \quad x_{r:n} < \cdots < x_{D:n} < T < x_{D+1:n}, \quad r \le D < n, \\ \text{Case III}: & \{x_{1:n}, \ldots, x_{n:n}\}, \quad \text{if} \quad x_{n:n} < T. \end{cases}$$

So, we have

$$\begin{aligned} P(D=r) &= \sum_{j=0}^{r-1} \binom{n}{j} (F_\theta(T))^j \left(\bar{F}_\theta(T))\right)^{n-j}, \\ P(D=j) &= \binom{n}{j} (F_\theta(T))^j \left(\bar{F}_\theta(T)\right)^{n-j}, \quad j = r, \cdots, n. \end{aligned}$$

Using (1) and the binomial expansion leads to

$$E(D) = r\sum_{j=0}^{r-1}\sum_{k=0}^{j} \binom{n}{j}\binom{j}{k}(-1)^k e^{-\theta(n-j+k)T} + \sum_{j=r}^{n}\sum_{k=0}^{j} j\binom{n}{j}\binom{j}{k}(-1)^k e^{-\theta(n-j+k)T}. \tag{6}$$

On the other hand, similar to the previous process, we obtain (see, for example, [2])

$$\begin{aligned} E(T^*) &= E(\max(X_{r:n}, T)) \\ &= T + \sum_{j=0}^{r-1}\sum_{k=0}^{j} \binom{n}{j}\binom{j}{k}(-1)^k \frac{e^{-\theta(n-j+k)T}}{\theta(n-j+k)}. \end{aligned} \tag{7}$$

Based on Equations (2)-(7), we have computed the values of $C(n, r, T)$, for different choices of n, r and T, when $\theta = 1$, $C_0 = 1$, $C_n = 3$, $C_d = 1$ and $C_t = 2$. These values

are reported in Table 1. From Table 1, by an empirical evidence, we get the values of $C(n, r, T)$ increase along with the values of r, n and T increase. In fact, it was expected because increasing the values of r, n and T means that we can have more failed items. The corresponding cost to the Type I hybrid censoring for the same n and r is lower when $T = 1$ but is nearly the same to the Type II censoring when $T = 2$. So, we conclude that, when the fixed time T is short the Type I hybrid censoring terminates the test early, resulting in a significant cost reduction compared to the conventional Type II and Type II hybrid censoring schemes. When T is longer the r-th failure likely occurs before time T, making the Type I hybrid censoring behave like the standard Type II scheme. This demonstrates Type I hybrid censoring has the ability to control costs by capping the test duration. For all cases the Type II hybrid censoring cost is significantly higher than the standard Type II scheme. Type II hybrid censoring scheme is preferred when the cost of poor statistical inference (due to too few failures) is the main concern, as it guarantees a minimum amount of data.

TABLE 1. The values of $C(n, r, T)$

		Type II	Type I hybrid		Type II hybrid	
n	r		$T = 1$	$T = 2$	$T = 1$	$T = 2$
5	1	99	98.90	99	136.70	234.66
	3	130.66	125.21	130.66	142.06	234.66
	5	180.66	136.13	179.66	181.14	235.66
10	1	172	171.99	172	243.21	359.32
	5	22.91	220.80	22.91	245.32	359.32
	10	318.57	243.18	316.57	318.60	361.32

3. Conclusion

From the results in this paper we find that Type I hybrid censoring scheme is preferred when the cost of time is the primary constraint, as it caps the duration. However, Type II hybrid censoring scheme is preferred when the cost of poor statistical inference (due to too few failures) is the main concern, as it guarantees a minimum amount of data.

References

1. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (2008) *A first course in order statistics*, Philadelphia: SIAM.
2. Basiri, E. and Hosseinzadeh, E. (2025) *Optimal design of life testing plans under Type II hybrid censoring scheme with random sample size*, Statistics, Optimization and Information Computing., **14**, 584–601.
3. Bhattacharya, R., Pradhan, B. and Dewanji, A. (2014) *Optimum life testing plans in presence of hybrid censoring: a cost function approach*, Applied Stochastic Models in Business and Industry, **30**(5), 519–528.
4. David, H.A. and Nagaraja, H.N. (2003) *Order Statistics*, 3rd ed. Hoboken, New Jersey: John Wiley and Sons.

A new distribution as a combination of Rayleigh and log-series distributions

Sajjad Piradl, Faculty member, Department of Statistics, Payame Noor University, Tehran, Iran
sajjadpiradl@pnu.ac.ir

Abstract: This paper derives a new distribution as a combination of Rayleigh and log-series distributions with increasing failure rate. First, the probability density function (PDF), Cumulative distribution function (CDF), s-th moment, survival function, and hazard function of the proposed distribution are calculated. Then, the estimation of the new distribution parameters is presented using the maximum likelihood (ML) method. Finally, the new distribution is fitted on a real dataset and it is shown that this proposed distribution performs better compared to some other known distributions.
Keywords: Combined distribution, Rayleigh distribution, Log-series distribution, ML method, Hazard function.

1. Introduction

In many cases, known statistical distributions do not provide a good fit to real data. Therefore, various methods for obtaining new probability distributions have recently been studied in the statistical literature [13]. Adamidis and Loukas [1] presented two-parameter exponential-geometric distribution that has a decreasing failure rate. Kus [8] introduced the exponential-Poisson distribution and examined several of its characteristic properties. On the other hand, the Weibull-Poisson distribution, which is a generalization of the exponential-Poisson distribution was introduced by [11]. Also, the binomial-exponential distribution is presented by [2]. Increasing failure rate (IFR) distributions are of interest in many real-word data systems [2]. Cancho et al. [4] obtained the exponential-Poisson distribution with IFR. Distributions with IFR have also been studied by [9], [3], and [12]. The Rayleigh distribution has been commonly used in reliability theory and survival analysis, because its failure rate is a linear function of time. This distribution plays an important role in real-word applications because it is related to well-known distributions such as the Weibull and Chi-square distributions. In statistical research, a significant amount of work has been devoted to Rayleigh distribution. Several authors, such as [6], [14] and the references cited therein, have conducted out extensive studies on estimation, prediction, and several other inferences regarding the Rayleigh distribution [5]. The aim of this paper is to obtain a new combined distribution with IFR properly conducted as a distribution of independent Rayleigh random variables when the sample size N has a log-series distribution. Also, various characteristics of the proposed distribution and its parameter estimators are presented using the ML method.

2. The new combined distribution

Let $X_1, \dots, X_N$ be a random sample from the Rayleigh distribution with pdf:

$$(1)\ g(x;\alpha) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}}, \begin{cases} x \geq 0 \\ \alpha > 0, \end{cases}$$

and N is a random variable with a log-series distribution with probability mass function (pmf):

$$(2)\ h(n;\beta) = -\frac{\beta^n}{nln(1-\beta)}, \begin{cases} 0<\beta<1 \\ n\epsilon\{1,2,3,\dots\} \end{cases}.$$

Also, let $Y = min(X_1, \dots, X_N)$. Then, the pdf of the new distribution as a combination of the Rayleigh and log-series distributions is as follows:

$$(3)\ p(y;\alpha,\beta) = -\frac{\beta y}{\alpha^2 ln(1-\beta)}\left(\frac{e^{-\frac{y^2}{2\alpha^2}}}{1-\beta e^{-\frac{y^2}{2\alpha^2}}}\right), y \geq 0.$$

Figure 1 shows the PDFs of the combined distribution for different parameter values.

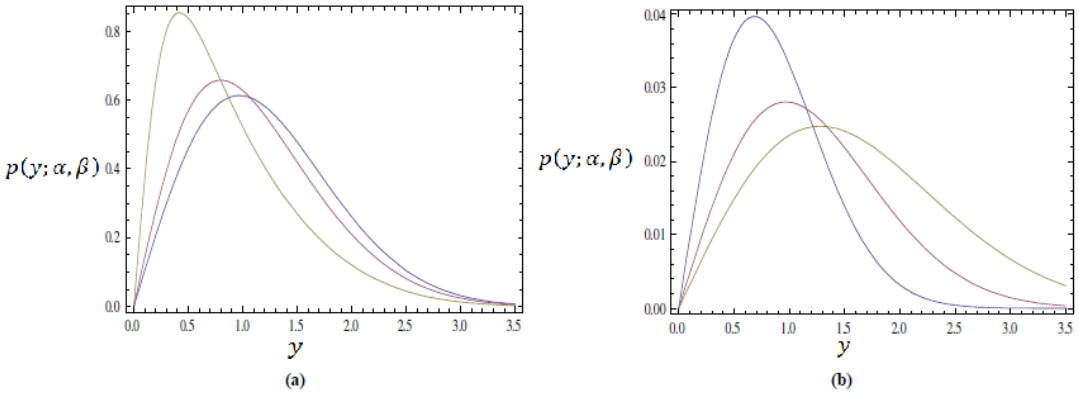


Figure 1. (a) PDFs of the combined distribution for $\alpha = 1, \beta = 0.1, 0.5, 0.9$, and (b) PDFs of the combined distribution for $\alpha = 0.5, \beta = 0.5, 1.0, 1.5$.

3. Some characteristic properties of the proposed distribution

- The CDF of Y is as follows:

$$(4)\ P(y;\alpha,\beta) = 1 - \frac{ln\left(1-ne^{-\frac{y^2}{2\alpha^2}}\right)}{ln(1-\beta)},$$

- The s-th moment of Y is given by:

$$(5) E(Y^s) = -\frac{2^{\frac{s}{2}}}{\alpha^{2s+2}\beta ln(1-\beta)}\Gamma\left(\frac{s+2}{2}\right)\sum_{j=1}^{\infty}\frac{\beta^j}{j^{\frac{s+2}{2}}},\ s = 1,2,3,\dots,$$

where the mean, and variance of Y are as follows:

$$(6)\ \mu = -\frac{\sqrt{\frac{\pi}{2}}}{\alpha^5 ln(1-\beta)}\sum_{j=1}^{\infty}\frac{\beta^j}{j^{\frac{5}{2}}},$$

$$(7)\ \sigma^2 = -\frac{2}{\alpha^6\beta ln(1-\beta)}\sum_{j=1}^{\infty}\frac{\beta^j}{j^3}, -\frac{\pi}{2\alpha^{10}[ln(1-\beta)]^2}\left(\sum_{j=1}^{\infty}\frac{\beta^j}{j^{\frac{5}{2}}}.\right)^2,$$

- The survival function of Y is as follows:

$$(8)\ S(y;\alpha,\beta) = 1 - P(y;\alpha,\beta) = \frac{ln\left(1-\beta e^{-\frac{y^2}{2\alpha^2}}\right)}{ln(1-\beta)},$$

- The hazard function of y is given by:

$$(9)\ h(y;\alpha,\beta) = -\frac{\beta y e^{-\frac{y^2}{2\alpha^2}}}{\alpha^2\left(1-\beta e^{-\frac{y^2}{2\alpha^2}}\right) ln\left(1-\beta e^{-\frac{y^2}{2\alpha^2}}\right)}.$$

Figure 2 shows the hazard functions of the combined distribution for different parameter values. The initial and long-term hazards are $h(0;\alpha,\beta)=0$ and $h(\infty;\alpha,\beta)=\infty$, respectively. Therefore, the hazard function is an increasing function.

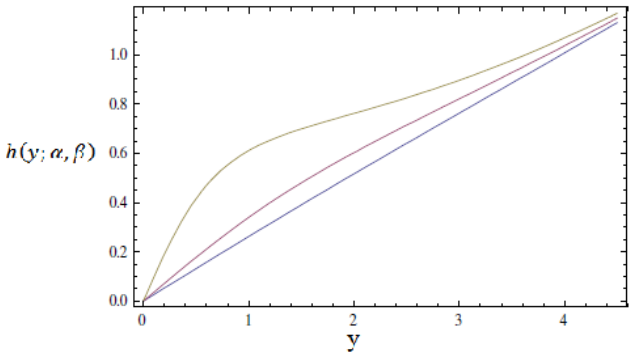


Figure 2. Hazard functions of the combined distribution for $\alpha = 2$, $\beta = 0.1, 0.5, 0.9$.

4. Estimation

ML estimators of the new distribution parameters can be calculated of the following equations:

$$(10)\ \frac{\partial l}{\partial \alpha} = \sum_{j=1}^{n}\left(\frac{{y_j}^2}{2\alpha^4}\right) - \sum_{j=1}^{n}\frac{\left(\frac{\beta {y_j}^2}{2\alpha^4}\right)e^{-\frac{{y_j}^2}{2\alpha^2}}}{\left(1-\beta e^{-\frac{{y_j}^2}{2\alpha^2}}\right)} - \left(\frac{4n}{\alpha}\right) = 0,$$

$$(11)\ \frac{\partial l}{\partial \beta} = \frac{n}{(1-\beta)ln(1-\beta)} + \sum_{j=1}^{n}\left(\frac{e^{-\frac{{y_j}^2}{2\alpha^2}}}{1-\beta e^{-\frac{{y_j}^2}{2\alpha^2}}}\right) + \left(\frac{n}{\beta}\right) = 0,$$

where l is the log-likelihood function of the new distribution based on the observed sample size n as follows:

$$(12)\ l(\alpha,\beta) = \sum_{j=1}^{n} ln\left(y_j\right) + nln(-1) - 2nln(\alpha) - nln\left(ln(1-\beta)\right) - \sum_{j=1}^{n}\left(\frac{{y_j}^2}{2\alpha^2}\right) + nln(\beta) - \sum_{j=1}^{n} ln\left(1-\beta e^{-\frac{{y_j}^2}{2\alpha^2}}\right).$$

5. Study on a real dataset

In this section, a real dataset which representing the scores of 48 students in a mathematics course on the final exams analyzed [7]. The estimated parameters using the ML method for the proposed distribution are $\hat{\alpha} = 19.28$ and $\hat{\beta} = 0.24$.

Table 1. Distributions, estimators, and values of the Kolmogorov-Smirnov (K-S) statistic for real data

Distribution	$\hat{\alpha}$	$\hat{\beta}$	K-S statistic
Proposed distribution	19.28	0.24	0.086
Binomial-exponential distribution	0.07	0.95	0.091
Weibull distribution	28.92	1.51	0.118

Exponentiated-exponential distribution	0.07	2.52	0.094
Weighted-exponential distribution	3.07	0.44	0.091
Exponential-Poisson distribution	0.07	1.00	0.093

Conclusion

In this paper, a new distribution as a combination of Rayleigh and log-series distributions is introduced. Mathematical and statistical properties of the proposed distribution are given. The estimates are examined by ML method. Also, a real dataset is analyzed. As a result, parameters of the new distribution are estimated. Obtained results are compared with other distributions. It is taken binomial-exponential, Weibull, exponentiated-exponential, weighted- exponential, and exponential-Poisson distributions for compare. It is found that the introduced distribution is found to be a good competitor according to these 5 distributions for this data set.

References

1. Lewin, L. Polylogarithms and Associated Functions, *North Holland Amsterdam*, 1981.
2. Adamidis, K., Loukas, S. A mix Distributions with Decreasing Failure Rate, *Statist. Probab. Lett.*, 39, 35-42, 1998.
3. Tugrul, O.R. Energy sector and wind energy Potential in Turkey, *Renewable Sustainable Energy Rev.*, 7, 469-484, 2003.
4. Ross, S.M., Shanthıkumar, J.G., Zhu, Z. On Increasing-Failure-Rate Random Variables, *J. Appl. Prob.*, 42, 797- 809, 2005.
5. Lariviere, M. A Note on Probability Distributions with Increasing Generalized Failure Rates, *Operations Research*, 54/3, 602-604,
2006.
6. Kus, C. A New Lifetime Distribution, *Comput. Statist. Data Anal*., 51, 4497-4509, 2007.
7. Eskin, N., Artar, H., Tolun, S. Wind Energy Potential of Gokceada Island in Turkey, *Renewable Sustainable Energy Rev*. 12,
839-851, 2008.
8. Brusset, X. Properties of Distributions with Increasing Failure Rate, *Munich Personal RePEc Archive*, 2009.
9. Gupta, R.D., Kundu, D. A New Class of Weighted Exponential Distributions, *Statistics*, 43, 621-634, 2009.
10. Cancho, V.G., Louzada-Neto, F., Barriga, G.D.C. The Poissson- Exponential Mix Distributions, *Comput. Stat. Data Anal.*, 55,
677-686, 2011.
11. Lu, W., Shi, D. A New Compounding Life Distribution: The Weibull-Poisson Distribution, *J. of Appl. Stat.*,
DOI:10.1080/02664763.2011.575126, 2011.
12. Silva, R.B., Cordeiro, G.M. The Burr XII Power Series Distributions: A New Compounding Family, *Brazilian J. of Prob. and Stat.*,
29(3), 2013.
13. Bakouch, H.S., Jazi, M.A., Nadarajah, S., Dolati, A. A Lifetime Model with Increasing Failure Rate, *Applied Mathematical*
Modelling, 38, 5392-5406, 2014.
14. Dey,S., Dey, T. Statistical Inference for the Rayleigh Distribution under Progressively Type II Censoring with Binomial Removal, *Applied Mathematical Modelling*, 38, 974-982, 2014.

A Case Study of Concomitants of Ordered Random via Generalized Exponential Distributions Variables from Bairamov Family

Reza Akbari[1], Leader Navaei[2*]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: r.akbari@pnu.ac.ir
[2]Department of Statistics, Payame Noor University (PNU), Tehran, Iran.
Email: l.navaei@pnu.ac.ir

Abstract. This article mainly introduces the Bairamov Morgenstern family under generalized exponential distribution. We adopt the concomitants of generalized order statistics for this family to construct some distributional properties.

Keywords: Bairamov family, Generalized exponential distribution.
AMS Mathematics Subject Classification [2020]: 62B10, 62G30

1. Introduction

Bairamov et al. [1] considered a generalization of the well-known bivariate FarlieGumbel-Morgenstern (FGM) distribution by inserting extra parameters. We denote this model by $BR(p_1, p_2, q_1, q_2)$. Therefore, in the current study, we transact with the distribution theory of $BR(p_1, p_2, q_1, q_2)$, which is specified by the cumulative distribution function (cdf) and probability density function (pdf), respectively, as follows

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)[1+\lambda(1-F_X^{p_1}(x))^{q_1}(1-F_Y^{p_2}(y))^{q_2}] \tag{1}$$

$$\begin{aligned} f_{X,Y}(x,y) =& f_X(x)f_Y(y)[1+\lambda(1-f_X^{p_1}(x))^{q_1-1}(1-(1+p_1q_1)F_X^{p_1}(x)) \\ & (1-F_Y^{p_2}(y))^{q_2-1}\times(1-(1+p_2q_2)F_Y^{p_2}(y))], \end{aligned} \tag{2}$$

where $p_1, p_2 \geq 1$, $q_1, q_2 \in \mathbb{N}$, $F_X(x), F_Y(y)$, $f_X(x), f_Y(y)$ are the marginal cdf's and pdf's of the random variables X and Y respectively. For $BR(p_1, p_2, q_1, q_2)$, the parameter λ has the admissible range

$$\begin{aligned} -\min\{1, \frac{1}{p_1p_2}\Big(\frac{1+p_1q_1}{p_1(q_1-1)}\Big)^{q_1-1}\Big(\frac{1+p_2q_2}{p_2(q_2-1)}\Big)^{q_2-1}\} \leq \lambda \leq \\ \min\{\frac{1}{p_1}\Big(\frac{1+p_1q_1}{p_1(q_1-1)}\Big)^{q_1-1}, \frac{1}{p_2}\Big(\frac{1+p_2q_2}{p_2(q_2-1)}\Big)^{q_2-1}\}. \end{aligned} \tag{3}$$

*Speaker.

For $BR(p_1,p_2,q_1,q_2)$, we have $\left(\frac{1+q_1}{(q_1-1)}\right)^{q_1-1}\left(\frac{1+q_2}{(q_2-1)}\right)^{q_2-1} > 1$, therefore, the admissible range of the parameter λ is:

$$-\Big(\frac{1+q_1}{(q_1-1)}\Big)^{q_1-1}\Big(\frac{1+q_2}{(q_2-1)}\Big)^{q_2-1} \le \lambda \le \min\{\Big(\frac{1+q_1}{(q_1-1)}\Big)^{q_1-1}, \Big(\frac{1+q_2}{(q_2-1)}\Big)^{q_2-1}\}. \tag{4}$$

The partnership parameter λ is known as the dependence parameter of the random variables X and Y . If $\lambda = 0$, then X and Y are independent.

The random variable X has generalized exponential GE distribution, denoted by $X \sim GE(\theta;\alpha)$, if it has the *cdf*

$$F_X(x) = (1 - exp(-\theta x))^{\alpha}, \quad x,\theta,\alpha > 0. \tag{5}$$

For the kth moment of GE distribution, $GE(\theta;\alpha)$, Kamps [3] showed that

$$\mu_k = \frac{\alpha\Gamma(k+1)}{\theta^k}\sum_{i=0}^{\chi(\alpha-1)}\frac{(-1)^i}{(i+1)^{k+1}}\binom{\alpha-1}{i}, \tag{6}$$

where $\chi(x) = x$, if x is integer and $\chi(x) = \infty$, if x is non-integer. Meanwhile, the moment generating function, mean and variance of $GE(\theta;\alpha)$ are given, respectively, by

$$M_X(t) = \alpha\beta(\alpha, 1-\frac{t}{\theta}, \quad \mu_1 = E(X) = \frac{B(\alpha)}{\theta}, \quad Var(X) = \frac{C(\alpha)}{\theta^2} \tag{7}$$

where $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $B(\alpha) = \psi(\alpha+1) - \psi(1)$, $C(\alpha) = \psi^{'}(1) - \psi^{'}(\alpha+1)$ and $\psi(.)$ is the digamma function, $\psi(1) = -\Gamma^{'}(1) = 0.57722$ is the Eulers constant, while $\psi^{'}(.)$ is its derivation. Also, they present some distributional properties of concomitants of order statistics beside record values of this *cdf*.

2. MOMENTS AND CORRELATION OF $BR - GE(\theta_1,\theta_2;\alpha_1,\alpha_2)$

In this section, we obtain the $(n,m)th$ joint moments and correlation of $BR{-}GE(\theta_1,\theta_2;\alpha_1,\alpha_2)$. From (1), let the random variables $Z \sim GE(\theta;\alpha)$ and $U \sim GE(\theta;\alpha(i+1))$, then$F_U(z) = (1-e^{-\theta z})^{\alpha(i+1)}$ and the *pdf* of U is formulated as $f_U(z) = (i+1)f_{Z(Z)}F_Z^i(Z)$. Therefore, the expectation of U^n is $E(U^n) = \int_{-\infty}^{\infty}(i+1)z^n f_{Z(Z)}F_Z^i(Z)dx$. We can also note that

$$\int_{-\infty}^{\infty} f_X(x)(1-F_X(x))^q dx = \Sigma_{i=0}^{q}\binom{q}{i}(-1)^i\int_{-\infty}^{\infty} x^n f_X(x)F_X^i(x)dx \tag{8}$$

$$\int_{-\infty}^{\infty} x^n f_X(x)F_X(x)(1-F_X(x))^q dx = \Sigma_{i=0}^{q}\binom{q}{i}(-1)^i\int_{-\infty}^{\infty} x^n f_X(x)F_X^{i+1}(x)dx \tag{9}$$

Now to obtain the moments, from the previous remark and the bivariate distribution $BR(1,1,q_1,q_2)$, where $X \sim GE(\theta_1;\alpha_1)$ and $Y \sim GE(\theta_2;\alpha_2)$. Then the $(n,m)th$ joint moments of $BR - GE(\theta_1,\theta_2;\alpha_1,\alpha_2)$ is given by

$$\begin{aligned} E(X^nY^m) =& E(x^n)E(y^m) + \lambda[\Sigma_{i_1=0}^{q_1-1}I_1E(U_1^n) - (1+q_1)\Sigma_{i_1=0}^{q_1-1}I_3E(U_2^n)] \\ &\times[\Sigma_{i_2=0}^{q_2-1}I_2E(V_1^m) - (1+q_2)\Sigma_{i_2=0}^{q_2-1}I_4E(V_2^m)] \end{aligned} \tag{10}$$

where

$$I_1 = \frac{\binom{q_1-1}{i_1}(-1)^{i_1}}{i_1+1}, \quad I_2 = \frac{\binom{q_2-1}{i_2}(-1)^{i_2}}{i_2+1}, \quad I_3 = \frac{\binom{q_1-1}{i_1}(-1)^{i_1}}{i_1+2}, \quad I_4 = \frac{\binom{q_2-1}{i_2}(-1)^{i_2}}{i_2+2} \tag{11}$$

(12) $$U_1 \sim GE(\theta_1;\alpha_1(i_1+1)), \quad U_2 \sim GE(\theta_1;\alpha_1(i_1+2)),$$
$$V_1 \sim GE(\theta_2;\alpha_2(i_1+1)), \quad V_2 \sim GE(\theta_2;\alpha_2(i_1+2)),$$

$i_1 = 0, 1, \ldots, q_1 - 1$, $i_2 = 0, 1, \ldots, q_2 - 1$. Therefore, from (10) and (7), we get

(13) $$E(XY) = \frac{B(\alpha_1)}{\theta_1}\frac{B(\alpha_2)}{\theta_2} + \lambda[\Sigma_{i_1=0}^{q_1-1} I_1 \frac{B(\alpha_1(i_1+1))}{\theta_1} - (1+q_1)\Sigma_{i_1=0}^{q_1-1} I_3 \frac{B(\alpha_1(i_1+2))}{\theta_1}]$$
$$\times [\Sigma_{i_2=0}^{q_2-1} I_2 \frac{B(\alpha_2(i_2+1))}{\theta_2} - (1+q_2)\Sigma_{i_2=0}^{q_2-1} I_4 \frac{B(\alpha_2(i_2+2))}{\theta_2}].$$

Subsequently, the correlation of X and Y is

(14)
$$\rho_{X,Y} = \frac{\lambda}{\sqrt{C(\alpha_1)C(\alpha_2)}}[\Sigma_{i_1=0}^{q_1-1} I_1 B(\alpha_1(i_1+1)) - (1+q_1)\Sigma_{i_1=0}^{q_1-1} I_3 B(\alpha_1(i_1+2))]$$
$$\times [\Sigma_{i_2=0}^{q_2-1} I_2 B(\alpha_2(i_2+1)) - (1+q_2)\Sigma_{i_2=0}^{q_2-1} I_4 B(\alpha_2(i_2+2))] = \lambda g(\alpha_1, \alpha_2, q_1, q_2),$$

where

(15)
$$g(\alpha_1, \alpha_2, q_1, q_2) = \frac{1}{\sqrt{C(\alpha_1)C(\alpha_2)}}[\Sigma_{i_1=0}^{q_1-1} I_1 B(\alpha_1(i_1+1)) - (1+q_1)\Sigma_{i_1=0}^{q_1-1} I_3 B(\alpha_1(i_1+2))]$$

(16) $$\times [\Sigma_{i_2=0}^{q_2-1} I_2 B(\alpha_2(i_2+1)) - (1+q_2)\Sigma_{i_2=0}^{q_2-1} I_4 B(\alpha_2(i_2+2))]$$
$$= \lambda g(\alpha_1, \alpha_2, q_1, q_2),$$

Clearly, for any $q_1, Q_2 \in \mathbb{N}$, the function $g(\alpha_1, \alpha_2, q_1, q_2)$ is positive and increasing function with respect to each of α_1 and α_2. Thus, $\rho_{X,Y}$ is positive and increasing function, if $\lambda > 0$, and $\rho_{X,Y}$ is negative and decreasing function, if $\lambda < 0$, with respect to each of of α_1 and α_2. Meanwhile, from Barakat et al. [2] Page (4), We have

(17) $$\lim_{\alpha\to\infty} \frac{B(\alpha(1+p)) - B(\alpha)}{\sqrt{C(\alpha)}} = \frac{\sqrt{6}}{\pi}\log(1+p)$$

thus, we can show that

(18) $$\lim_{\alpha_1,\alpha_2\to\infty} g(\alpha_1, \alpha_2, q_1, q_2) = \frac{6}{\pi_2}[\Sigma_{i_1=0}^{q_1-1}\binom{q_1-1}{i_1}(-1)^{i_1}\log(1+i_1)]$$
$$[\Sigma_{i_2=0}^{q_2-1}\binom{q_2-1}{i_2}(-1)^{i_2}\log(1+i_2)],$$

(19) $$\lim_{\alpha_1,\alpha_2\to 0^+} g(\alpha_1, \alpha_2, q_1, q_2) = 0$$

Therefore, the admissible range of $\rho_{X,Y}$ is

(20)
$$-(\frac{1+q_1}{q_1-1})^{q_1-1}(\frac{1+q_2}{q_2-1})^{q_2-1}g^*(q_1,q_2) \le \rho_{X,Y} \le \min(\frac{1+q_1}{q_1-1})^{q_1-1}g^*(q_1,q_2), (\frac{1+q_2}{q_2-1})^{q_2-1}g^*(q_1,q_2)$$

where

(21) $$g^*(q_1, q_2) = \frac{6}{\pi_2}[\Sigma_{i_1=0}^{q_1-1}\binom{q_1-1}{i_1}(-1)^{i_1}\log(1+i_1)]$$

(22) $$[\Sigma_{i_2=0}^{q_2-1}\binom{q_2-1}{i_2}(-1)^{i_2}\log(1+i_2)]$$

3. CONCLUSION

We derived the Bairamov Morgenstern family type bivariate GE distribution based on concomitant of $m-gos$. We also provided the correlation and its admissible range for such distribution.

References

[1] Bairamov, I., Kotz, S. and Bekci, M. (2001) *New generalized Farlie-GumbelMorgenstern distributions and concomitants of order statistics*, J. of App. Stat., **28(5)** , 521–536.

[2] Barakat H.M., Nigm, E.M. and Syam, A.H. (2017) *Concomitants of Ordered from HuangKotz FGM Type Bivariate Generalized Exponential Distribution*, Bull. Malays Math. Soc., **42(1)** , 337–353.

[3] Kamps, U. (1995) *A concept of generalized order statistics*, J. of Stat. Plann Inf., **48** , 1–23.

On introduction of Poisson-Pranav Distribution

Leader Navaei[1,*], Reza Akbari[2]

[1]Department of Statistics, Payame Noor University (PNU), Tehran, Iran.
Email: l.navaei@pnu.ac.ir

[2]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: r.akbari@pnu.ac.ir

Abstract. In this paper deals with formulation of Poisson-Pranav probability distribution by combining Poisson distribution and Pranav distribution for count data. Some important structural and statistical properties of this model are derived and discussed like coefficient of variation, skewness, kurtosis,reliability analysis and order statistics are beening obtained. Also for obtaining estimate of unknown parameter of this distribution maximum likelihood estimation method is used.

Keywords: Poisson distribution, Poisson-Pranav distribution, Count Data.

AMS Mathematics Subject Classification [2020]: 62B10, 62G30

1. Introduction

In our day to day life we many times deal with count data and decision making for dealing with count data becomes important. For improving decision making while dealing with count data we fit a valid probability model to count data. Mahmoudi et al [4] generalized the Poisson-Lindley distribution of Sankaran [5] and proved that his generalized distribution is more flexible for analyzing count data. Gupta and Ong [3] formulated a new generalized negative binomial distribution by considering parameter of Poisson distribution as generalized gamma variate and the resulting distribution was fitted to various data sets and proved as better alternative to negative binomial distribution. El-Monseff and Sohsah [2] obtained Poisson Weighted Lindley Distribution and studied its vital properties and applied to some real life situations. Shankar [6] introduced Aradhana distribution and its applications in real life and also studied some of its important properties. Ahmad et al. [1] obtained a new discrete compound distribution with Applications in various fields of real life and obtained its various crucial properties. The research paper which we have formulated deals with formulation of Poisson-Pranav distribution by combining Poisson distribution with Pranav distribution .

*Speaker.

2. Definition of Proposed Model (Poisson-Pranav Distribution)

If Z is a poison variate i.e., $Z|\lambda \sim P(\lambda)$, λ being itself a Pranav variate with parameter θ , then the resulting distribution determined by marginalizing over λ is Poison-Pranav distribution obtained by compounding Poisson distribution and Pranav distribution, which is denoted by $PPD(Z;\theta)$. Our proposed model i.e., Poison-Pranav model is discrete model as parent distribution (Poisson) is a discrete distribution. A random variable Z follows $PPD(\theta)$ with probability mass function given in the theorem below.

Theorem 2.1. *The probability mass function of a Poisson-Pranav Distribution i.e.,$PPD(Z;\theta)$ is given by*

$$P(Z=z)=\frac{\theta^4}{(\theta^4+6)}[\frac{\theta(1+\theta)^3+(z+3)(z+2)(z+1)}{(1+\theta)^{z+4}};\quad z=0,1,2.3\ldots.;\quad \theta>0 \tag{1}$$

The corresponding c.d.f of Poisson-Pranav distribution is obtained as:

$$\begin{aligned} F_Z(z)=\sum_{n=0}^{z}\frac{\theta^4}{(\theta^4+6)}[\frac{\theta(1+\theta)^3+(z+3)(z+2)(z+1)}{(1+\theta)^{z+4}}\ldots \\ (\theta^3z^3+9\theta^3z^2+3\theta^2z^2+26\theta^3z+21\theta^2z+\theta^7+3\theta^6+3\theta^5 \\ 1-\frac{+\theta 4+24\theta 3+36\theta^2+6z\theta+24\theta+6}{(1+\theta)^{z+4}(6+\theta)^4)} \quad z>0,\theta>0 \end{aligned} \tag{2}$$

3. STATISTICAL PROPERTIES OF POISSON-PRANAV DISTRIBUTION

In this part some vital structural properties of the Poisson-Pranav model are obtained. These include moment, generating functions (m.g.f and p.g.f).

3.1. Moments of Poisson-Pranav Distribution.

3.1.1. *Factorial Moments.* Using (1), the rth factorial moment about origin of the PPD (1) can be obtained as

$$\begin{aligned} \mu'_{(r)} =& E[E(Z^{(r)}|\lambda]; \quad Z^{(r)}=Z(Z-1)(Z-2)...(Z-r+1) \\ \mu'_{(r)} =& \int_0^\infty \Big[\sum_{z=0}^{\infty} z^r\frac{e^{-\lambda}\lambda^z}{(z!)}\Big]\frac{\theta^4}{\theta^4+6}(\theta+\lambda^3)e^{-\theta\lambda}d\lambda \\ \mu'_{(r)} =& \frac{\theta^4}{\theta^4+6}\int_0^\infty \Big[\lambda^r\Big(\sum_{z=r}^{\infty}\frac{e^{-\lambda}\lambda^{z-r}}{(z-r)!}\Big)\Big](\theta+\lambda^3)e^{-\theta\lambda}d\lambda \end{aligned}$$

Taking $u=t-r$, we get

$$\begin{aligned} \mu'_{(r)} =& \frac{\theta^4}{\theta^4+6}\int_0^\infty \Big[\lambda^r\Big(\sum_{u=0}^{\infty}\frac{e^{-\lambda}\lambda^{u}}{u!}\Big)\Big](\theta+\lambda^3)e^{-\theta\lambda}d\lambda \\ \mu'_{(r)} =& \frac{r!}{\theta^4+6}\Big[\frac{\theta^4+(r+3)(r+2)(r+1)}{\theta^r}\Big] \end{aligned} \tag{3}$$

Taking $r=1,2,3,4$ in (3), the first 4 factorial moments about origin of PoissonPranav distribution can be obtained as

$$\mu'_{(1)} =\frac{\theta^4+24}{\theta(\theta^4+6)}$$

$$\mu_{(2)}^{'} = \frac{2(\theta^4+60)}{\theta^2(\theta^4+6)}$$
$$\mu_{(3)}^{'} = \frac{6(\theta^4+120)}{\theta^3(\theta^4+6)}$$
$$\mu_{(4)}^{'} = \frac{24(\theta^4+210)}{\theta^4(\theta^4+6)}$$

3.1.2. *Moments about Origin (Raw Moments).* The first four moments about origin, using the relationship between factorial moments about origin and the moments about origin of PPD (1) are generated as

$$\mu_1^{'} = \frac{\theta^4+24}{\theta(\theta^4+6)}$$

which is the mean of Poisson-Pranav Distribution

$$\mu_2^{'} = \frac{2(\theta^4+60)+\theta(\theta^4+24)}{\theta^2(\theta^4+6)}$$
$$\mu_3^{'} = \frac{6(\theta^4+120)+6\theta(\theta^4+60)-2\theta^2(\theta^4+24)}{\theta^3(\theta^4+6)}$$
$$\mu_4^{'} = \frac{24(\theta^4+210)+18\theta(\theta^4+120)-22\theta^2(\theta^4+60)+6\theta^3(\theta^4+24)}{\theta^4(\theta^4+6)}$$

3.1.3. *Moments about the Mean (Central Moments).* Using the relationship $\mu_r = E(t-\mu_{(}^{'}1))^r = \sum_{K=0}^{r}\binom{r}{k}\mu_{(}^{'}k)(-\mu_{(}^{'}1))^{r-k}$ between moments about the mean and the moments about origin, the moments about the mean of the PPD (1) can be obtained as

$$\mu_2 = \frac{2(\theta^4+60)+\theta(\theta^4+24)(\theta^4+6)-(\theta^4+24)^2}{\theta^2(\theta^4+6)^2}$$
$$[\theta^4+6]^2 6(\theta^4+120)+6\theta(\theta^4+60)-2\theta^2(\theta^4+24)$$
$$\mu_3 = \frac{-3[\theta^4+6](2(\theta^4+60)+\theta(\theta^4+24))(\theta^4+24)+2(\theta^4+24)^3}{\theta^3(\theta^4+6)^3}$$
$$(\theta^4+6)^3 24(\theta^4+210)+18\theta(\theta^4+120)-22\theta^2(\theta^4+60)+6\theta^3(\theta^4+24)$$
$$-4(\theta^4+6)^2(6(\theta^4+120)+6\theta(\theta^4+60)-2\theta^2(\theta^4+24))(\theta^4+24)$$
$$\mu_4 = \frac{+6(\theta^4+6)(2(\theta^4+60)+\theta(\theta^4+24))(\theta^4+24)^2-3(\theta^4+24)^4}{\theta^4(\theta^4+6)^4}$$

3.2. ESTIMATION OF PARAMETERS. In this section, we estimate the unknown parameter of the Poisson-Pranav distribution by using method of maximum likelihood estimation.

3.2.1. *ESTIMATION OF PARAMETERS.* Method of Maximum Likelihood Estimation is simple and most efficient method of estimation. In this method unknown parameters are obtained by maximizing likelihood function. Suppose $Z_1, Z_2, \ldots, Z_n$ n is a random sample of size n taken from PoissonPranav Distribution (PPD), then the likelihood function of PPD is given as

$$L(z|\theta) = \frac{\theta^{4n}}{(\theta^4+6)^n}\prod_{i=1}^{n}\Big(\frac{(\theta(1+\theta)^3+(z_i+3)(z_i+2)(z_i+1))}{(1+\theta)^{z_i+4}}\Big)$$

The log likelihood function is

$$\log L = \sum_{i=1}^{n} \log(\theta(1+\theta)^3 + (z_i+3)(z_i+2)(z_i+1)) - (\sum_{i=1}^{n} z_i + 4n)\log(1+\theta) - n\log(\theta^4+6) + 4n\log\theta(1+\theta)^{z_i+4}\Big)$$

differentiating log likelihood function with respect to θ we get

$$\frac{\delta}{\delta\theta}\log L = \sum_{i=1}^{n} \frac{((1+\theta)^3 + 3\theta(1+\theta)^2)}{(\theta(1+\theta)^3 + (z_i+3)(z_i+2)(z_i+1))} - \frac{4n\theta^3}{(\theta^4+6)} - \frac{\sum_{i=1}^{n} z_i + 4n}{(1+\theta)} + \frac{4n}{\theta} = 0$$

The maximum likelihood of θ is obtained by solving above equation through R software.

4. CONCLUSION

We formulated a new probability model known as Poisson-Pranav distribution for count data by mechanism of compounding. Then we obtained its vital statistical properties.

References

[1] Ahmad, Z. et al. (2017) *A New Discrete Compound Distribution with Applications*, J. Stat. Appl. Pro., **6** , 233–241.

[2] El-Monsef, M.M.E. and Sohsah, N.M. (2014) *Poisson-Weighted Lindley Distribution. Jokull Journal*, Jokull Journal., **64(5)** , 192–202.

[3] Gupta, R.C. and Ong, S.H. (2004) *A New Generalization of the Negative Binomial Distribution*, Journal of Computational Statistics and Data Analysis., **45** , 287–300.

[4] Mahmoudi, E. and Zakerzadeh, H. (2010) *Generalized Poisson-Lindley Distribution*,Communications in Statistics Theory and Methods., **39(10)** , 1785–1798.

[5] Sankaran, M. (1970) *The Discrete Poisson-Lindley Distribution*, Biometrics., **26** , 145–149.

[6] Shankar, R. (2016) *Aradhna Distribution and its Applications*, International Journal of Statistics and Applications., **6(1)** , 23–34.

A new generalization of the log-Rayleigh probability distribution

Sajjad Piradl, Faculty member, Department of Statistics, Payame Noor University, Tehran, Iran
sajjadpiradl@pnu.ac.ir

Abstract: In recent years, families of probability distributions have had useful applications in different sciences. In this paper, a new generalization of the log-Rayleigh probability distribution is introduced by providing the probability density function (PDF), probability cumulative function (PCF), and some of the other main properties. Then, the maximum likelihood estimation (MLE) method is used to estimate the parameters of the proposed probability distribution. Finally, the introduced generalized log-Rayleigh probability distribution is applied to calculate the initial mass function (IMF) for stars.
Keywords: Log-Rayleigh probability distribution, Generalized log-Rayleigh probability distribution, MLE method, IMF.

1. Introduction

The log-Rayleigh probability distribution has three parameters and is a generalized version of the well-known Rayleigh probability distribution ([1], [2]). The Rayleigh family of probability distributions is used in many scientific fields, such as astrophysics ([3], [4]). In this paper, a new generalization of the log-Rayleigh probability distribution, is introduced to calculate the IMF of stars. In astrophysics, the IMF is an empirical function that describes the initial distribution of masses for a population of stars during star formation. The IMF not only in describing the formation and evolution of stars, it also serves as an important link describes the formation and evolution of galaxies ([5]). In this paper, in section 2, the main properties of the log-Rayleigh probability distribution are recalled. The generalized log-Rayleigh probability distribution is introduced in section 3. In section 4, the proposed probability distribution is applied to the analysis of the IMF of stars.

2. The log-Rayleigh probability distribution

The PDF and CDF of a continuous random variable X with a log-Rayleigh probability distribution are, respectively:

$$(1)\ f_X(x) = \left[\frac{\alpha - 2\gamma ln\left(\frac{x}{\delta}\right)}{x}\right] e^{-ln\left(\frac{x}{\delta}\right)\left[\alpha-\gamma ln\left(\frac{x}{\delta}\right)\right]}, \quad \begin{cases} \alpha \geq 0 \\ \gamma \geq 0 \\ \delta \geq 0 \\ x\epsilon[\delta,\infty) \end{cases},$$

$$(2)\ F_X(x) = 1 - e^{-\left\{\alpha ln\left(\frac{x}{\delta}\right)+b\left[ln\left(\frac{x}{\delta}\right)\right]^2\right\}}.$$

Also, the mean (μ), variance (σ^2), and r-th moment about the origin of the above random variable are, respectively:

$$(3)\ \mu = \frac{\sqrt{\pi}\alpha}{\sqrt{\gamma}}\left\{\left[\frac{1}{\sqrt{\pi}}\int_{y=\left(\frac{\alpha-1}{2\sqrt{\gamma}}\right)}^{\infty} e^{-y^2}dy\right]e^{\frac{(\alpha-1)^2}{4\gamma}} - \sqrt{\gamma}\right\},$$

$$(4)\ \sigma^2 = \frac{2\delta^2}{\sqrt[3]{\gamma^2}}\left[\gamma\left(\left\{\left[\int_{y=\left(\frac{\alpha-2}{2\sqrt{\gamma}}\right)}^{\infty} e^{-y^2}dy\right]e^{\frac{(\alpha-2)^2}{4\gamma}}\right\} - \left\{\left[\int_{y=\left(\frac{\alpha-1}{2\sqrt{\gamma}}\right)}^{\infty} e^{-y^2}dy\right]e^{\frac{(\alpha-1)^2}{4\gamma}}\right\}\right)\right.$$
$$\left. - \frac{\pi\sqrt{\gamma}}{4}\left\{\left[\int_{y=\frac{(\alpha-1)^2}{4\gamma}}^{\infty} e^{-y^2}dy\right]e^{\frac{(\alpha-1)^2}{2\gamma}}\right\}\right],$$

$$(5)\ \mu'_s = \frac{1}{\sqrt{\gamma}}\left(\delta^{\alpha s}\left\{\left[\int_{y=\left(\frac{\alpha-r}{2\sqrt{\gamma}}\right)}^{\infty} e^{-y^2}dy\right]e^{\frac{(s-\alpha)(4\gamma ln\delta-\alpha+s)}{2\gamma}}\right\} + \sqrt{\pi\gamma}\delta^s\right), \qquad s = 1,2,3,\ldots$$

3. The generalized log-Rayleigh probability distribution

The generalized log-Rayleigh probability distribution introduced in this paper is a right truncated log-Rayleigh probability distribution. The PDF and CDF of a continuous random variable X with a generalized log-Rayleigh probability distribution are, respectively:

$$(6)\ g_X(x) = \left[\frac{\alpha - 2\gamma ln\left(\frac{x}{\delta}\right)}{\left(1-\delta^{\alpha+2\gamma ln\vartheta}\vartheta^{-\alpha}e^{-\gamma ln[(\delta\vartheta)^2]}\right)x}\right]e^{-\left[ln\left(\frac{x}{\delta}\right)\right]\left[\alpha-\gamma ln\left(\frac{x}{\delta}\right)\right]}, \vartheta \geq 0,$$

$$(7)\ G_X(x) = \frac{1-e^{-\left\{\alpha\left[ln\left(\frac{x}{\delta}\right)\right]+\gamma\left[ln\left(\frac{x}{\delta}\right)\right]^2\right\}}}{1-c^{\alpha+2\gamma ln\vartheta}\vartheta^{-\alpha}e^{-\gamma ln[(\delta\vartheta)^2]}}.$$

Also, the mean (μ), and s-th moment about the origin of the above random variable are, respectively:

$$(8)\ \mu = \frac{\delta\vartheta^{\alpha}}{\sqrt{\gamma}\left(\vartheta^{\alpha}-\delta^{\alpha+2\gamma ln\vartheta}d^{a}e^{-\gamma ln[(\delta\vartheta)^2]}\right)}\left(\sqrt{\gamma}\left(1-\delta^{\alpha-1+2\gamma ln\vartheta}\vartheta^{1-\alpha}e^{-\gamma ln[(\delta\vartheta)^2]}\right)\right.$$
$$\left. - \left\{\left[\frac{2\gamma ln\left(\frac{\delta}{\vartheta}\right)-\alpha+1}{2\sqrt{\gamma}}\right]\int_{y=-\infty} e^{-y^2}dy + \int_{y=-\infty}^{\frac{(\alpha-1)}{2\gamma}} e^{-y^2}dy\right\}e^{\frac{(\alpha-1)^2}{4\gamma}}\right),$$

$$(9)\ \mu'_s = \frac{cd^a}{\sqrt{\gamma}\left(\vartheta^\alpha - \delta^{\alpha+2\gamma ln\vartheta} d^a e^{-\gamma ln[(\delta\vartheta)^2]}\right)} \left(\sqrt{\gamma}\left(\delta^{s-1} - \delta^{\alpha-1+2\gamma ln\vartheta}\vartheta^{s-\alpha} e^{-\gamma ln[(\delta\vartheta)^2]}\right) \right.$$

$$\left. - s\left\{ \int_{y=-\infty}^{\left[\frac{2\gamma ln\left(\frac{\delta}{\vartheta}\right)+\alpha-s}{2\sqrt{b}}\right]} e^{-y^2} dy + \int_{y=-\infty}^{\frac{(\alpha-s)}{2\sqrt{\gamma}}} e^{-y^2} dy \right\} e^{\frac{(-\alpha+s)[4\gamma ln\delta-\alpha+s]}{4\gamma}} \right), s = 1,2,3,\dots$$

All the four parameters of this probability distribution are found by numerically solving the following equation, which arises from the MLE method ([6], [7]):

$$(10)\ nln\delta = \sum_{i=1}^{n}\left(lnX_i + \frac{1}{\left[2\gamma ln\left(\frac{\delta}{X_i}\right) - \alpha\right]} + \frac{\delta^{\alpha+2\gamma ln\vartheta}\vartheta^{-\alpha} ln\left(\frac{\delta}{\vartheta}\right) e^{-\gamma ln[(\delta\vartheta)^2]}}{\left(\delta^{\alpha+2\gamma ln\vartheta}\vartheta^{-\alpha} ln\delta e^{-\gamma ln[(\delta\vartheta)^2]}\right) - 1} \right),$$

where $X_1, X_2, X_3, \dots, X_n$ are independent and identically distributed (IID) random variables with generalized log-Rayleigh probability distribution.

4. Using the generalized log-Rayleigh probability distribution in calculating the IMF of stars

In this section, the log-Rayleigh probability distribution and the proposed new probability distribution are applied to calculate the IMF of stars and the results are compared with those obtained from the log-Gaussian probability distribution. The log-Gaussian probability distribution is another useful probability distribution in astrophysics ([8]). The comparison between these three probability distributions is based on the star clusters Caldwell 64, Eagle Nebula, Gamma Velorum, NGC 7822, and Melotte 25 ([9]). The statistics used for the analysis and comparing are the reduced figure-of-merit function $(RFMF)$, the probability of goodness of fit (P) with an acceptable value of fit $P > 0.001$, the Akaike information criterion (AIC), the maximum distance between the theoretical and experimental degrees of freedom (MD) and the Kolmogorov-Smirnov significance level $(K - SSL)$ with an acceptable value of fit $K - SSL > 0.1$ ([10]). All results are presented in the following two tables:

Table 1. Numerical results for comparing the log-Rayleigh probability distribution with the log-Gaussian probability distribution

Star Cluster	a	b	c	RFMF	P	AIC	MD	K-SSL
Caldwell 64	6.9×10^{-3}	3.7×10^{-1}	1.2×10^{-1}	1.8×10^{1}	1.2×10^{-24}	13.4×10^{1}	2×10^{-1}	1.1×10^{-9}
Eagle Nebula	7.5×10^{-3}	1.2×10^{-1}	1.9×10^{-2}	1.2×10^{1}	6.1×10^{-16}	9.3×10^{1}	2×10^{-1}	1.2×10^{-7}
Gamma Velorum	4.9×10^{-1}	7.5×10^{-1}	1.6×10^{-1}	2.1	4×10^{-2}	2.1×10^{1}	4×10^{-2}	9×10^{-1}
NGC 7822	2×10^{-2}	9.1×10^{-1}	1.6×10^{-1}	3.1	2.5×10^{-3}	2.8×10^{1}	7×10^{-2}	4×10^{-2}
Melotte 25	9×10^{-2}	4.4×10^{-1}	1.1×10^{-1}	2.0	5.3×10^{-2}	2.0×10^{1}	4×10^{-2}	2×10^{-1}

From the values in table 1, it can be concluded that the log-Rayleigh probability distribution performances better than the L-G probability distribution in most cases.

Table 2. Numerical results for comparing the generalized log-Rayleigh probability distri-

bution with the log-Gaussian probability distribution

Star Cluster	a	b	c	RFMF	P	AIC	MD	K-SSL
Caldwell 64	6.9×10^{-3}	3.7×10^{-1}	1.2×10^{-1}	1.9×10^{1}	1.9×10^{-22}	12.3×10^{1}	2.5×10^{-2}	1×10^{-15}
Eagle Nebula	7.5×10^{-3}	1.2×10^{-1}	1.9×10^{-2}	1.3×10^{1}	3.5×10^{-14}	8.3×10^{1}	2.6×10^{-2}	6.1×10^{-13}
Gamma Velorum	4.9×10^{-1}	7.5×10^{-1}	1.6×10^{-1}	2.4	2.5×10^{-2}	2.3×10^{1}	4×10^{-2}	7.8×10^{-1}
NGC 7822	2×10^{-2}	9.1×10^{-1}	1.6×10^{-1}	3.7	1.2×10^{-3}	3.0×10^{1}	7×10^{-2}	3×10^{-2}
Melotte 25	9×10^{-2}	4.4×10^{-1}	1.1×10^{-1}	2.4	2.5×10^{-2}	2.3×10^{1}	6×10^{-2}	3.9×10^{-1}

From the values in table 2, it can be concluded that the generalized log-Rayleigh probability distribution performs better than the log-Gaussian probability distribution in most cases. Also, in general, from comparing the values in the two tables, it can be clearly seen that the generalized log-Rayleigh probability distribution is more useful than the log-Rayleigh probability distribution in accurately calculating the IMF for stars.

5. Conclusion

In this paper, a new probability distribution, called the generalized log-Rayleigh probability distribution was introduced. This probability distribution is an extension of the log-Rayleigh probability distribution and the aim was to investigate its application in astronomy. The main properties of this probability distribution were calculated and then the results of its application in the calculation of the IMF for stars were obtained. Examination of numerical results showed that this new probability distribution is more accurate in calculating the IMF for stars compared to the log-Gaussian probability distribution.

References

[1] Grashorn, J., Bittner, M., Wang, C., and Beer, M., "The log-Rayleigh distribution for local maxima of spectrally represented log-normal processes", *Proceedings of the 8th International Symposium on Reliability Engineering and Risk Management*, pp. 793-799, 2022.

[2] Rivet, B., Girin, L., and Jutten, C., "Log-Rayleigh distribution: A simple and efficient statistical representation of log-spectral coefficients", *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, Vol. 15, No. 3, pp. 796-802, 2007.

[3] Pascucci, A., "*Probability theory I: Random variables and distributions*", Springer, 2024.

[4] Anis, M. Z., and Ahsanullah, M., "Some characterizations of the Rayleigh distribution", *Afrika Statistika*, Vol. 17, No. 4, pp. 3367-3377, 2023.

[5] Guszejnov, D., and Hopkins, P. F., "Mapping the core mass function to the initial mass function", *Monthly Notice of the Royal Astronomical Society*, Vol. 450, No. 4, pp. 4137-4149, 2015.

[6] Hadi Abdul Sahib, N., "Review of maximum likelihood estimation method", *International Journal of Engineering and Information Systems (IJEAIS)*, Vol. 7, No. 8, pp. 20-23, 2023.

[7] Wu, J.-W., Hung, W.L., and Chen, C.-Y., "Approximate MLE of the scale parameter of the truncated Rayleigh distribution under the first failure-censored data", *Journal of Information and Optimization Sciences*, Vol. 25, No. 2, pp. 221-235, 2004.

[8] Adeniran, A. T., Faweya, O., Ogunlade, T. O., and Balogun, K. O., "Derivation of Gaussian probability distribution: A new Approach", *Applied Mathematics*, Vol. 11, No. 6, pp. 436-446, 2020.

[9] Krumholz, M. R., McKee, C. F., and Bland-Hawthorn, J., "Star clusters across cosmic time", *Annual Review of Astronomy and Astrophysics*, Vol. 57, pp. 227-303, 2019.

[10] Peacock, J. A., "Two-dimensional goodness-of-fit testing in astronomy", *Monthly Notice of the Royal Astronomical Society*, Vol. 202, No. 3, pp. 615-627, 1983.

A new extension of the generalized beta probability distribution

Sajjad Piradl, Faculty member, Department of Statistics, Payame Noor University, Tehran, Iran
sajjadpiradl@pnu.ac.ir

Abstract: In recent years, the beta family of probability distributions has had useful applications in astrophysics. In this paper, a new extension of the generalized beta probability distribution is introduced by presenting some of its main mathematical properties. Then, the maximum likelihood estimation (MLE) method is used to estimate the parameters of the proposed probability distribution. Finally, the introduced extended generalized beta probability distribution is applied to modeling the IMF for stars.
Keywords: Generalized beta probability distribution, Extended Generalized beta probability distribution, MLE method, IMF.

1. Introduction

The generalized beta probability distribution has two parameters and is one of the important probability distributions used in various sciences ([1], [2]). This family of probability distributions plays an important role in astrophysics ([3], [4]). In this paper, a new extension of the generalized beta probability distribution is introduced to IMF modeling for stars. In astrophysics, the IMF is an empirical function that describes the initial distribution of masses for a population of stars during star formation. The IMF not only in describing the formation and evolution of stars, it also serves as an important link describes the formation and evolution of galaxies ([5]). In this paper, in section 2, the main mathematical properties of the generalized beta probability distribution are reminded. The extended generalized beta probability distribution is introduced in section 3. In section 4, the proposed probability distribution is applied to modeling IMF for stars.

2. The generalized beta probability distribution

The PDF and CDF of a continuous random variable X with a generalized beta probability distribution are, respectively:

$$(1)\ f_X(x) = \frac{x^{r-1}(1-x)^{s-1}}{\int_{x=0}^{1} x^{r-1}(1-x)^{s-1}dx} = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}x^{r-1}(1-x)^{s-1}; \begin{cases} r>0 \\ s>0 \\ x\epsilon(0,1) \end{cases},$$

$$(2)\ F_X(x) = 2\int_{\alpha=0}^{arcsin(\sqrt{x})} (sin\alpha)^{2r-1}(cos\alpha)^{2s-1}d\alpha = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}\int_{u=\left(1-\frac{1}{x}\right)}^{\infty} \frac{u^{s-1}}{(1+u)^{r+s}}du.$$

Also, the mean (μ), variance (σ^2), and m-th moment about the origin of the above random variable are, respectively:

$$(3)\ \mu = \frac{r}{r+s},$$

$$(4)\ \sigma^2 = \frac{rs}{(r+s+1)(r+s)^2},$$

$$(5)\ \mu'_m = \frac{\Gamma(r+s)\Gamma(r+s)}{\Gamma(r+s+m)\Gamma(r)}, m = 1,2,3,\ldots,$$

where $\Gamma(.)$ is the gamma function.

3. The extended generalized beta probability distribution

The extended generalized beta probability distribution introduced in this paper is a truncated beta probability distribution of the first kind. The PDF of a continuous random variable X with an extended generalized beta probability distribution of the first kind is:

$$(6) g_X(x) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \frac{(x-r)^{\alpha-1}(s-x)^{\beta-1}}{(s-r)^{\alpha+\beta-1}}, \quad \begin{cases} \alpha > 0 \\ \beta > 0 \end{cases}.$$

Also, the mean (μ), and variance (σ^2), of the above random variable are, respectively:

$$(7)\ \mu = \frac{s\alpha + r\beta}{\alpha+\beta},$$

$$(8)\ \sigma^2 = \frac{(s-r)^2\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

All four parameters of this probability distribution are found using the MLE method, respectively ([6], [7]):

$$(9)\ \hat{r} = min(X_1, X_2, X_3, \ldots, X_n),$$

$$(10)\ \hat{v} = max(X_1, X_2, X_3, \ldots, X_n),$$

$$(11)\ \hat{\alpha} = \frac{(\bar{X}-\hat{r})[\hat{s}(\hat{r}-\bar{X}) + S^2 + \bar{X}^2 - \hat{r}\bar{X}]}{S^2(\hat{r}-\hat{s})}; \begin{cases} S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \end{cases},$$

$$(12)\ \hat{\beta} = \frac{(\hat{s}-\bar{X})[\hat{s}(\hat{r}-\bar{X}) + S^2 + \bar{X}^2 - \hat{r}\bar{X}]}{S^2(\hat{r}-\hat{s})}; \begin{cases} S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{Y})^2 \\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \end{cases},$$

where $X_1, X_2, X_3, \ldots, X_n$ are independent and identically distributed (IID) random variables with extended generalized beta probability distribution.

4. Using the extended generalized beta probability distribution to model IMF for stars

In this section, the proposed new probability distribution is applied to modeling the IMF of stars and the results are compared with those obtained from the log-Gaussian probability distribution. The log-Gaussian probability distribution is another useful probability distribution in astrophysics ([8]). The comparison between these two probability distributions is based on the star clusters Caldwell 64 and NGC 7822 ([9]). The statistics used for the analysis and compar-

ing are the reduced figure-of-merit function $(RFMF)$, the probability of goodness of fit (P) with an acceptable value of fit $P > 0.001$, the Akaike information criterion (AIC), the maximum distance between the theoretical and experimental degrees of freedom (MD) and the Kolmogorov-Smirnov significance level $(K-SSL)$ with an acceptable value of fit $K-SSL > 0.1$ ([10]). All results are presented in the following two tables:

Table 1. Numerical results for comparing the extended generalized beta probability distribution with the log-Gaussian probability distribution, based on the Caldwell 64 star cluster

Type of probability distribution	r	s	α	β	RFMF	P	AIC	MD	K-SSL
Log-Gaussian probability distribution	-0.55	0.50	-	-	1.86	0.01	37.64	0.07	0.10
Extended generalized beta probability distribution	0.12	1.47	1.67	2.77	1.31	0.20	29.10	0.06	0.29

From the values in table 1, it can be concluded that in the case of the Caldwell 64 star cluster, the extended generalized beta probability distribution performs better than the log-Gaussian probability distribution.

Table 2. Numerical results for comparing the extended generalized beta probability distribution with the log-Gaussian probability distribution, based on the NGC 7822 star cluster

Type of probability distribution	r	s	α	β	RFMF	P	AIC	MD	K-SSL
Log-Gaussian probability distribution	-1.26	1.03	-	-	3.73	1.3×10^{-7}	71.24	0.10	0.05
Extended generalized beta probability distribution	0.02	1.46	0.56	1.55	1.96	0.01	39.30	0.11	0.28

From the values in table 2, also it can be concluded that in the case of the NGC 7822 star cluster, the extended generalized beta probability distribution performs better than the log-Gaussian probability distribution. Therefore, in general, from comparing the values in the two tables, it can be clearly seen that the extended generalized beta probability distribution is more useful than the log-Gaussian probability distribution in accurately modeling the IMF for stars.

5. Conclusion

In this paper, a new probability distribution, called the extended generalized beta probability distribution was introduced. This probability distribution is a extension of the generalized beta probability distribution, and the goal was to investigate its application in astrophysics. The main mathematical properties of this probability distribution were calculated and then the results of its application in IMF modeling for stars were obtained. Examination of numerical results showed that this new probability distribution is more accurate in IMF modeling for stars compared to the log-Gaussian probability distribution.

References

[1] Mir, K. A., Ahmed, A., and Reshi, J. A., "A new class of weighted generalized beta distribution of

first kind and its structural properties", *International Journal of Research in Management*, Vol. 3, No. 6, pp. 49-58, 2013.
[2] Ahmed, A., Mir, K. A., and Reshi, J. A., "On size biased generalized beta distribution of first kind", *IOSR Journal of Mathematics (IOSR-JM)*, Vol. 5, No. 2, pp. 41-48, 2013.
[3] Adnan, M. R. S., and Kiser, H., "A class of beta second kind mixture distributions", *Journal of Statistical Theory and Applications*, Vol. 19, No. 4, pp. 518-525, 2020.
[4] Cordeiro, G. M., and Brito, R. d. S., "The beta power distribution", *Brazilian Journal of Probability and Statistics*, Vol. 26, No. 1, pp. 88-112, 2012.
[5] Lazar, A., and Bromm, V., "Probing the initial mass function of the first stars with transients", *Monthly Notice of the Royal Astronomical Society*, Vol. 511, No. 2, pp. 2505-2514, 2022.
[6] Beckman, R. J., and Tiet jen, G. L., "Maximum likelihood estimation for the beta distribution", *Journal of Statistical Computation and Simulation*, Vol. 7, No. 3-4 pp. 253-258, 1978.
[7] Feron, B., "Curvature and inference for maximum likelihood estimates", *The Annals of Statistics*, Vol. 46, No. 4, pp. 1664-1692, 2018.
[8] Andai, A., "On the geometry of generalized Gaussian distributions", *Journal of Multivariate Analysis*, Vol. 100, No. 4, pp. 777-793, 2009.
[9] Freundlich, J., "Star formation across cosmic time", *Fundamental Plasma Physics*, Vol. 11, pp. 1-12, 2024.
[10] Jiang, J., "Goodness-of-fit tests for mixed model diagnostics", *The Annals of Statistics*, Vol. 29, No. 4, pp. 1137-1164, 2001.

A 15-year Research Profile of *the Bulletin of the Iranian Mathematical Society* (2008-2022): a Scientometric Approach

Heidar Mokhtari* (Corresponding Author)
Associate professor, Department of Knowledge and Information Science, Payame Noor University, Tehran, Iran. E-mail: h.mokhtari@pnu.ac.ir, ORCID: 0000-0002-2471-0408
Mohammad Rahimi-Alangi
Assistant Professor, Department of Mathematics, Payame Noor University, Tehran, Iran
E-mail: mrahimi.al@pnu.ac.ir, ORCID: 0000-0002-4137-1997

Abstract: Scientometrics is a quantitative approach to evaluate the research performance of scientific entities such as journals. This study aimed to conduct a scientometric analysis and visualization of the Bulletin of the Iranian Mathematical Society (the Bulletin) from its being indexed in Scopus in 2008 to 2022 (a 15-year time span). Required data of 1,553 published papers was extracted from SCOPUS and underwent a scientometric analysis with applying EXCEL Microsoft Office for summarizing data and VOSviewer bibliometric package for visualizing co-authored countries and clustering keywords. The publication trend was increasing steadily in the studied years (R2=.984). The majority of papers (N=1,524, about 98.1%) were original research. Among the collaborating authors (N= 159), the first rank commonly belonged to Mahdavi-Amiri, N. from Iran, Shy, W. and Wang, J.R., both from China (each with publishing 8 papers). Out of 159 affiliations, the first to third ranks belonged to Iranian affiliations: Institute for Studies in Theoretical Physics and Mathematics (with 50 papers), Iranian Research Institute for Fundamental Sciences (with 47 papers) and University of Tabriz (with 42 papers), respectively. Among contributing countries, Iran ranked first with 684 papers (about 44.04% of all papers), followed by China (with 391 papers) and India (with 86 papers). The citation trend was increasing these years (R2=.996). Out of 802 unique keywords with 5,022 occurrences, the main clustering keywords were “fixed points”, “analytic functions”, “frames”, “derivation”, “Banach algebra”, “variational methods” and “prime graph”. This study is the first scientometric study of the Bulletin as a leading Iranian journal and its results are beneficial to the editorial team for better decision making and helpful for the audience and authors.

Keywords: *Bulletin of the Iranian Mathematical Society;* Bibliometrics; Scientometrics; Research profile

1. Introduction

As one of the main research fields with an interdisciplinary nature, bibliometrics or scientometrics has been conducted in a wide range of studies in different disciplines. Pritchard in1969 formally defined this research field as the application of mathematics and statistical methods to books, articles and all document types and other media of communication for the quantification of their research performance [2].

Scientific journals are important channels of scientific communication and conceived as gateways to new information. They need to be evaluated from research perspective in order to identify their role in scientific development and research influence [3]. This can be done by applying scientometric methods that quantitatively study the publications of a journal. The scientometric analysis of specific journals has been done in past decades. In the past, the scientometric data on a specific journal including among others, frequencies of published papers, received citations and highly-productive authors, institutions and countries as well as those of highly-influential ones were studied. However, emerging new technologies and inventing scientometric software packages resulted in visualization of the research performance of scientific journals, including among others, keyword co-occurrence and subject clustering. Using scientometric methods in analyzing the knowledge structure and scientific features of the papers of a journal provides a good guide for its potential authors and some guiding references for its future development. It also can reveal a specific journal's current status and development trend, as a basis for further improving its quality [8].

Few scientometric studies have been conducted in mathematics. For example, a study reviewed the degree of author collaboration in China's mathematical science from 1999-2014 [3]. Researchers' contribution to the mathematical research during 2015-2019 was analyzed in Dimension database [6]. *Indian Journal of Pure and Applied Mathematics* underwent a bibliometric survey during 1998-2017 for identifying key patterns of citations of its papers [5].

Based on the information provided in its new website (https://www.springer.com/journal/41980), *the Bulletin of the Iranian Mathematical Society* (here abbreviated as *the Bulletin*) is a publication of the Iranian Mathematical Society in English that has been published since 1974. As a pioneering journal, it publishes original research papers with significant contributions of broad interest, and invited survey articles on hot topics, from distinguished mathematicians worldwide. With six issues per year, *the Bulletin* provides a platform for presenting high-level mathematical research in most areas of mathematics. From January 2018, it is published by Springer. Based on the information collected by Scopus in 2023 [4], its SJR (SCImago Journal Rank) = .383 and SNIP (Source Normalized Impact per Paper) = 888. However, this main journal has not been evaluated from the scientometric perspective yet. The bibliometric overview of the journal and visualizing its scientific patterns and trends can be beneficial to mathematics community as well as the audience, authors and editorial team of *the Bulletin*. This study aimed to conduct a scientometric analysis and visualization of this journal from its being indexed in Scopus in 2008 to 2022 (a 15-year time span).

2. Methodology

This study was a bibliometric / scientometric study, focusing on a specific scientific journal. The approach has been widely used by several scientific journals in different fields worldwide. Data was collected by using Scopus database. As one of the comprehensive abstract and indexing

databases in the world, Scopus (https://www.scopus.com) is maintained by a Netherlands' institute, named Elsevier. Including peer-reviewed literature, such as scientific journals, books and conference proceedings, it indexes scientific items of main subject fields such as science, technology, medicine, social sciences, and arts and humanities.

The time span of this study was a 15-years period, 2008--2022, as the published papers of *the Bulletin* started to be indexed in Scopus from 2008. The following formula was used for data extraction in Scopus in December 2023:

SRCTITLE (Bulletin of the Iranian Mathematical Society) AND PUBYEAR < 2023

1,553 papers were identified and their bibliographic information and scientometric data were analyzed. We used some scientometric indicators for determining the trend of annual publication and year-wise received citation and top highly-cited papers as well as identifying highly-productive authors, institutions and countries contributing to *the Bulletin*. VOSviewer, Version 1.6.19 software package, was applied for networking co-authoring countries/territories contributed to *the Bulletin* and clustering highly-frequent keywords used in its papers. The software visualizes the intended results through a wide range of selected scientometric indicators [7]. For summarization of scientometric data in tables and figures, Excell 2010 was used, too.

3. Findings

3.1. Annual frequencies of published papers and publication trend

Out of 1,553 published papers of *the Bulletin* indexed in Scopus during the studied period (2008-2022), year 2022 with 224 published papers (14.4% of all papers) and year 2008 with only 16 published papers (only .01% of all papers) had the first and last ranks, respectively. Figure 1 depicts the publication trend by the publication years from 2008 to 2022. As can be seen, the publication trend was increasing steadily in these years (R^2=.984). The number of published papers has been considerably increased in 2016 (with 138 papers) and especially in 2022 (with 224 papers).

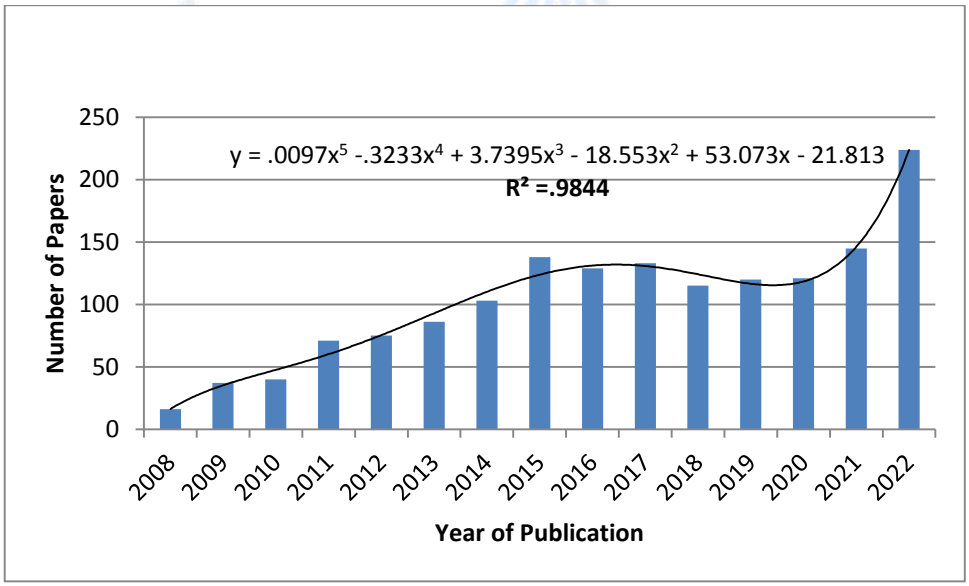


Figure 1. Year-wise frequency distribution of papers published in *the Bulletin* (2008-2022)

3.2.Types of published papers

Table 1 shows the frequency distribution of published papers by document types. As can be seen, the majority of papers (N=1,524, about 98.1%) were original research articles. Conference papers ranked second in this regard (N= 14, about .9%).

Table 1. Frequency distribution and percentage of papers published in *the Bulletin* by their document type (2008-2022)

Rank	Document type	Frequency (%)	%
1	Original article	1524	98.1
2	Conference paper	14	.9
3	Erratum	9	.6
4	Editorial	3	.2
5	Letter	1 (.1)	.1
5	Note	1(.1)	.1
5	Retracted	1 (.1)	.1
Total	-	1,553	100

3.3.Most-productive authors

Among the publishing authors that amounted to 159 individual authors, the first rank commonly belonged to Mahdavi-Amiri, N. from Iran, Shy, W., and Wang, J.R., both from China (each with publishing 8 papers). Table 2 shows some information on the top 10 highly-productive authors publishing at least six papers. These ten authors contributed to publishing 66 papers and six of them were from Iran's universities. Other contributing authors published no less than three papers.

Table 2. Most productive authors publishing in *the Bulletin* (2008-2022)

Rank	Author name	Number of papers	Affiliation	Country of origin
1	Mahdavi-Amiri, N.	8	Sharif University of Technology,	Iran
1	Shi, W.	8	Suzhou Univer-sity, Suzhou	China
1	Wang, J.R.	8	Guizhou University	China
2	Abdollahi, A.	6	University of Isfahan	Iran
2	Ebadian, A.	6	University of Isfahan	Iran
2	Jabbarzadeh, M.R.	6	University of Tabriz	Iran
2	Moori, J.	6	North-West University	South Africa
2	O'Regan, D.	6	University of Galway	Ireland
2	Shahmorad, S.	6	University of Tabriz	Iran
2	Tehranian, A.	6	Islamic Azad Univer-sity	Iran

3.4.Top contributing affiliations

Out of 159 unique affiliations (universities and research institutes) contributed to *the Bulletin*, top ten affiliations were shown in Table 3. The first to third ranks belonged to the Institute for Studies in Theoretical Physics and Mathematics (with 50 papers), Iranian Research Institute for Fundamental Sciences (with 47 papers) and University of Tabriz (with 42 papers), respectively.

These ten affiliations were all from Iran's universities and research institutes. By publishing 365 papers, the authors in these ten affiliations contributed to *the Bulletin* by publishing about 23.5% of all of its papers. The least contributions of the individual affiliations were 3 papers.

Table 3. Most active affiliations publishing in *the Bulletin* (2008-2022)

Rank	Affiliation	Number of papers	% of total papers
1	Institute for Studies in Theoretical Physics and Mathematics	50	3.22
2	Iranian Research Institute for Fundamental Sciences	47	3.03
3	University of Tabriz	42	2.70
4	Amirkabir University of Technology	41	2.64
5	Isfahan University of Technology	39	2.51
6	Ferdowsi University of Mashhad	34	2.19
7	University of Isfahan	31	1.99
8	Kharazmi University	29	1.87
9	Islamic Azad University	28	1.80
10	Shahid Bahonar University of Kerman	24	1.55

3.5.Top contributing countries /territories

Authors from 80 countries/territories contributed to *the Bulletin*. Table 4 depicts ten highly-productive countries publishing in *the Bulletin*. Iran ranked first with publishing 684 papers (about 44.04% of all papers), followed by China (with 391 papers) and India (with 86 papers). These top ten countries published the majority of papers (N=1,415, about 91.11% of total papers). Twenty contributing countries published only one paper.

Table 4. Most active countries publishing in the Bulletin (2006-2015)

Rank	Country / Territories	Number of papers	% of total papers
1	Iran	684	44.04
2	China	391	25.17
3	India	86	5.54
4	Turkey	82	5.28
5	United States of America	41	2.64
6	Vietnam	30	1.93
7	South Korea	29	1.87
8	South Africa	26	1.67
9	Saudi Arabia	25	1.61
10	Egypt	21	1.35

Figure 2 better depicts the co-authorship network of contributing countries with publishing at least three papers in *the Bulletin*. As can be seen, the most active countries in the co-authorship network were Iran, China, India and Tuekey.

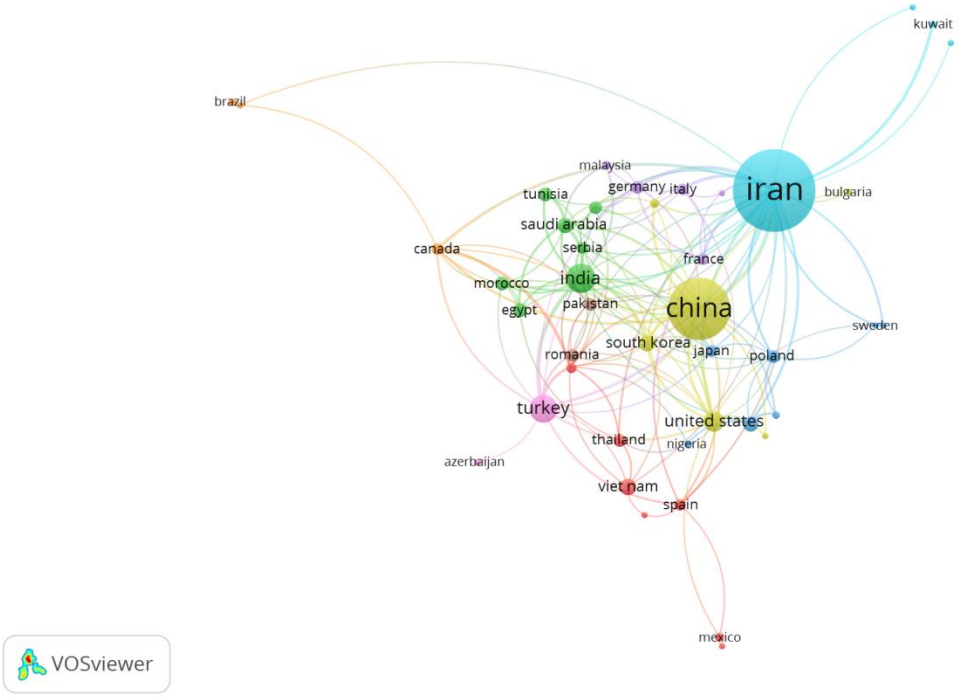


Figure 2. Co-authorship network of most productive countries (published at least 3 papers) in *the Bulletin* (2008-2022)

3.6.Highly-frequent keywords and keyword clustering

817 unique author-assigned keywords that occurred 5,065 times were used in the studied papers. These keywords had some inconsistencies as to their spelling, duplication, and repetition and so on. After removing some irrelevant keywords (such as existence) or revising similar ones (such as frame and frames), 802 unique keywords with 5,022 occurrences were identified. Table 5 shows the top highly-frequent keywords (ones with >10 occurrences in the papers). These keywords appeared 213 times in the published papers. "Fixed point", "Derivation", “Finite groups” were in top by occurring 47, 31 and 19 times, respectively.

Table 5. Top ten highly-frequent keywords used in the papers published in *the Bulletin* (2008-2022)

Rank	Keyword	Frequency	Rank	Keyword	Frequency
1	Fixed point	47	6	Frames	13
2	Derivation	31	6	Hadamard product	13
3	Finite groups	19	7	Subordination	12
4	Analytic function	17	7	Nonexpansive mapping	12
5	Banach algebra	14	8	Prime graph	11
6	Eigenvalues	13	8	Prime ring	11

107 keywords were occurred 4 times or more in the papers. These keywords were selected for depicting density visualization in order to clustering the co-occurred keywords. As only 79 keywords with obvious links and connections were proposed by VOSViewr, these keywords were visualized as a keyword clustering network (Figure 4). In the formed subject clusters, reflecting different dispersed nodes or keywords, the main nodes were fixed points, analytic

functions, frames, derivation, Banach algebra, variational methods and prime graph.

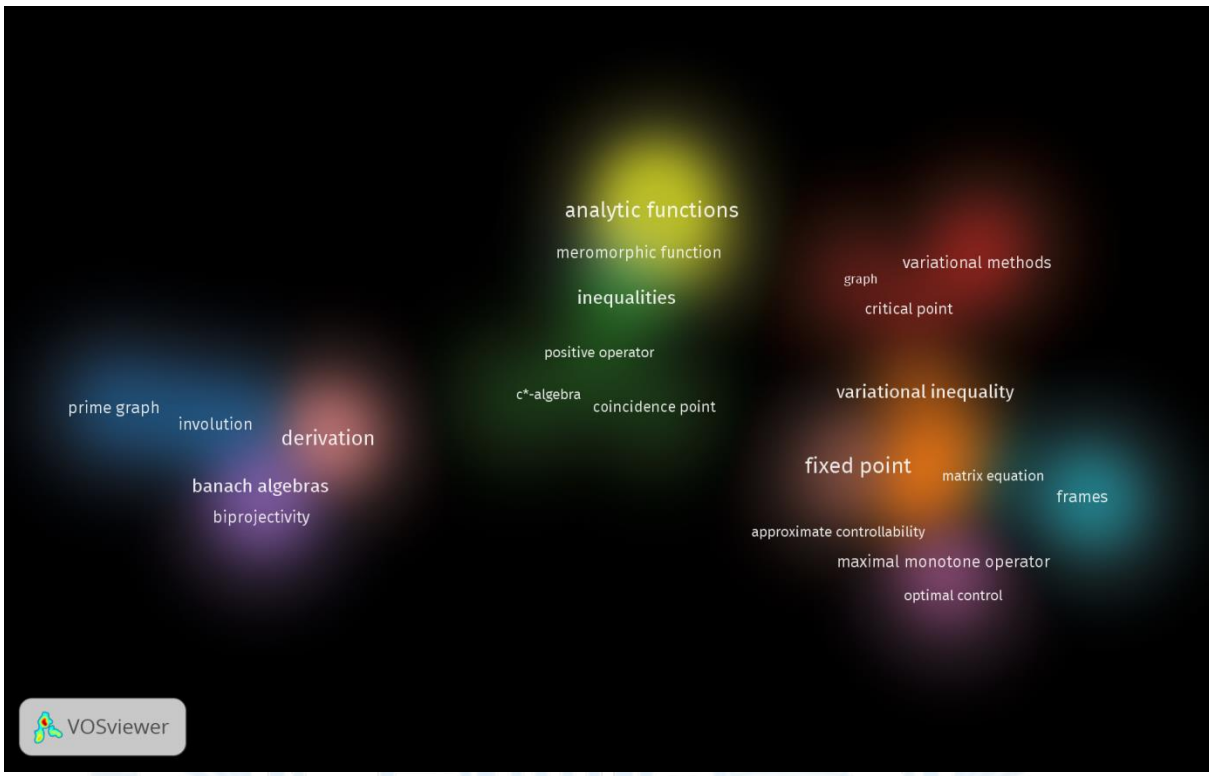


Figure 3. Keyword co-occurrence density visualization of most highly-frequent keywords (occurred four times and more) used in the papers of *the Bulletin* (2008-2022)

3.7.Annual frequencies of received citations and citation trend

The total citations amounted to 4,267 in the studied time span. Figure 4 shows the citation counts by the publication year of cited papers. The citation trend was increasing these years (R2=.996). The citation trend was ascending in the studied period by no citations in 2008 up to 773 citations in 2021 and 914 citations in 2022. About 39.5% of all received citation belonged to these two consecutive years.

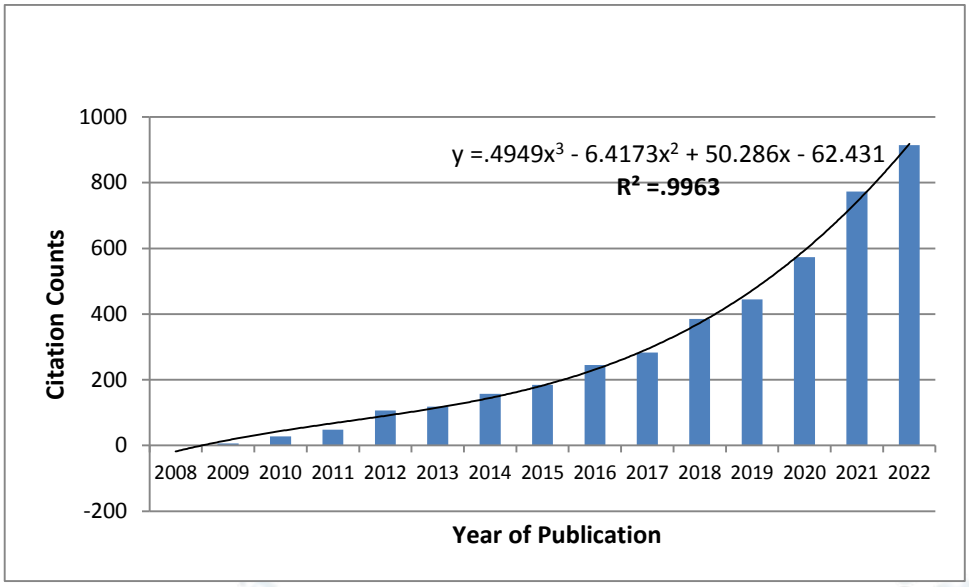


Figure 4. Year-wise frequency distribution of cited papers of *the Bulletin* (2008-2022)

3.8.Most-cited papers

Out of all 1,553 published papers, 949 papers (61.11%) were cited at least on time. The total citations amounted to 4,267. The mean rate of received citations was 2.75 per paper. Of cited papers, 108 papers (11.38%) received 10 or more citations. Table 6 depicts the bibliographic information of the top 5 highly-cited papers with citation counts>50. The first-ranked paper entitled as "Radius Problems for Starlike Functions Associated with the Sine Function" was authored by Cho, N.E., Kumar, V., Kumar, S.S., and Ravichandran, V. in 2019 and received 104 citations.

Table 6. Top five highly-cited papers published in *the Bulletin* (2008-2022)

Rank	Author name	Title	Citation counts	Publication year
1	Cho, N.E., Kumar, V., Kumar, S.S., Ravichandran, V.	Radius Problems for Starlike Functions Associated with the Sine Function	104	2019
2	Radenović, S., Kadelburg, Z., Jandrlić, D., Jandrlić, A.	Some Results on Weakly Contractive Maps	98	2012
3	Ali, R.M., Lee, S.K., Ravichandran, V., Supramaniam, S.	The Fekete-Szegö Coefficient Functional for Transforms of Analytic Functions	67	2009
4	Acar, Ö., Durmaz, G., Minak, G.	Generalized Multivalued F-contractions on Complete Metric Spaces	65	2014
5	Srivastava, H.M., Eker, S.S., Ha-	Faber Polynomial Coefficient Estimates for Bi-univalent Functions Defined by the	56	2018

	midi, S.G., Jahangiri, J.M.	Tremblay Fractional Derivative Operator		

4. Discussion and Conclusion

Aiming at analyzing the research performance of *the Bulletin* during its being indexed in Scopus in 2008 up to the end of 2022, we found that it has find its way in research on mathematics as its ever-increasing rate of annual growth of publication showed. Iranian authors were most active in authoring the papers from different universities and research institutes nationwide. This is the case when regarding the most productive affiliations (universities and research institutes). In addition, when considering the highly-contributing countries, it is obvious that the contribution of European countries is weak as a symbol of the international collaboration. For internalization of *the Bulletin* more than ever, it is needed to design for publishing more papers from authors with other countries of origin worldwide.

These papers are mostly original research articles and review articles as one of main documents in summarization and systematization of a discipline were absents and the share of other communicative media such as letters to editors was low and these types of documents need to be emphasized.

We encounter some problems in key-word clustering. The author-assigned keywords were not accurate in some cases and some keywords were vague. We recommend that *the Bulletin* request its author to be selective and more accurate in assigning keywords to their submissions or design a thesaurus-based controlled vocabulary for keywords to be assigned. As keywords are main items in searching and retrieving information from information database, being accurate in selecting them would be beneficial to more reading and citedness of the papers of *the Bulletin*.

Regarding the citation rate of *the Bulletin*, the ever-increasing rate of its received citation and about three citations per paper showed that *the Bulletin* has been emphasized in its domain and more authors use its items for more documenting their papers.

The mean growth rate of received citations of *the Bulletin* was satisfying since about 39% of papers have not received any citations. The annual growth of citation counts of *the Bulletin* is the sign of its international reach. An accurate citation analysis is needed for depicting the mere influence of the Bulletin based on its citation.

In conclusion, gradual increase in the number of published papers and their received citations shows that *the Bulletin* achieved the target of attracting the attention of researchers worldwide. Year-by-year increase in received citations of the journal indicates its promise and deep influence on research development. However, contributing authors, institutions and countries are not geographically and internationally distributed worldwide. In addition, assigned keyword should be more accurate and consistent for better visibility of papers in the scientific community. The scientometric indicators of *the Bulletin* are signs of its worldwide development, scientific quality and academic prestige. After about 50 years, the Bulletin has found its way to develop and influence the field.

This study is a relatively comprehensive and the first scientometric analysis and visualization of *the Bulletin* as a leading Iranian journal in its field. The results of the study are beneficial to its editorial team for better decision making on its further development as well as helpful for its audience and authors interesting topics in mathematics to have a better contact with and effective contributions to *the Bulletin*.

References

1. H. Mokhtari, S. Barkhan, D. Haseli, M.K. Saberi, *A bibliometric analysis and visualization of the Journal of Documentation: 1945–2018.* Journal of Documentation, 77(1), (2021), 69-92.
2. A. Pritchard. *Statistical bibliography or bibliometrics*. Journal of Documentation, 25 (4), (1969), 348-349.
3. S. Rajani, B. Ravi. *Collaboration and productivity of mathematical science research in China: a scientometric study*. International Journal of Library and Information Studies, 7(4), (2017), 412-420.
4. Scopus: *Source title search*. https://www.scopus.com/sources.uri (2023). Accessed 13 Decenber 2023.
5. S. Singh, K. Chand. *Indian Journal of Pure and Applied Mathematics: A Bibliometric Survey, 1998-2017.* Library Progress (International), 41(2), (2021), 275-286. https://doi.org/10.5958/2320-317X.2021.00031.3
6. P. Suharso, L. Setyowati, M.N. Arifah. *Bibliometric analysis related to mathematical research through Database Dimensions*. Journal of Physics: Conference Series, 1776. (2021). https://doi.org/10.1088/1742-6596/1776/1/012055
7. N.J. Van Eck, L. Waltman. *Software survey: VOSviewer, a computer program for bibliometric mapping*. Scientometrics, 84, (2010), 523–538.
8. Z. Xu, D. Yu, X. Wang. *A bibliometric overview of International Journal of Machine Learning and Cybernetics between 2010 and 2017*. International Journal of Machine Learning and Cybernetics, 10, (2019), 2375-2387. https://doi.org/10.1007/s13042-018-0875-9.


12th National Mathematics Conference of Payame Noor University
Sari—Iran
6–7 November 2025

Topology

Existence of Solutions for Sequential Liouville-Caputo Fractional Differential Equations

Rahmat Darzi[1,*], Bahram Agheli[2]

[1]Department of Mathematics, NeK.C., Islamic Azad University, Neka, Iran.
Email: Rahmat.Darzi@iau.ac.ir
[2]Department of Mathematics, QaS.C., Islamic Azad University, Qaemshahr, Iran.
Email: bahram.agheli@iau.ac.ir

Abstract. This work investigates a class of fractional differential equations of a distinct nature: sequential Liouville-Caputo FDEs, accompanied by antiperiodic and boundary conditions defined via a Riemann–Liouville integral, under appropriate assumptions.

Keywords: Sequential Liouville-Caputo derivative, Antiperiodic, Existence

AMS Mathematics Subject Classification [2020]: 34A55, 34B99.

1. Introduction

Previous investigations on Fractional Differential Equations (FDE) and Partial Differential Equations (PDE) with integral boundary conditions can be located in references [1–6].

We consider the following problem:

$$\begin{cases} \mathfrak{D}^{\alpha}\upsilon(\theta)+k\mathfrak{D}^{\alpha-1}\upsilon(\theta)=\lambda\left(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta)\right), \\ \beta_1\upsilon(0)+\psi_1\upsilon(1)+\gamma_1\mathfrak{I}^{r}\upsilon(\varsigma)=\epsilon_1, \\ \beta_2\upsilon'(0)+\psi_2\upsilon'(1)+\gamma_2\mathfrak{I}^{r}\upsilon'(\varsigma)=\epsilon_2, \\ \beta_3\upsilon''(0)+\psi_3\upsilon''(1)+\gamma_3\mathfrak{I}^{r}\upsilon''(\varsigma)=\epsilon_3, \end{cases} \tag{1}$$

the parameter $\alpha\in(2,3]$ represents a real number, while $\beta_i,\psi_i,\gamma_i,\epsilon_i\in\mathbb{R}$ for $i=1,2,3$ and $\theta\in[0,1]$, $k,r,\varsigma>0$. The operator $\mathfrak{D}^{\alpha}$ denotes the Liouville-Caputo derivative, and the boundary conditions involve antiperiodic cases.

2. Main results

Lemma 2.1. *Authorizing $\nu\in C[0,1]$ and $\upsilon\in C^2[0,1]$. Therefore, the following sequential FDE*

$$\mathfrak{D}^{\alpha}\upsilon(\theta)+k\mathfrak{D}^{\alpha-1}\upsilon(\theta)=\nu(\theta), \tag{2}$$

*Speaker.

for $\theta \in [0,1]$ *and* $k>0$ *with the boundary conditions*

$$(3)\qquad \begin{cases} \beta_1 \upsilon(0)+\psi_1\upsilon(1)+\gamma_1 \mathfrak{I}^r \upsilon(\varsigma)=\epsilon_1 \\ \beta_2 \upsilon'(0)+\psi_2\upsilon'(1)+\gamma_2 \mathfrak{I}^r \upsilon'(\varsigma)=\epsilon_2 \\ \beta_3 \upsilon''(0)+\psi_3\upsilon''(1)+\gamma_3 \mathfrak{I}^r \upsilon''(\varsigma)=\epsilon_3 \end{cases}$$

has a unique solution

$$(4)\qquad \upsilon(\theta)=f_1(\theta)\phi_1\left(\nu(1),\nu(\varsigma)\right)+f_2(\theta)\phi_2\left(\nu(1),\nu(\varsigma)\right)+f_3(\theta)\phi_3\left(\nu(1),\nu(\varsigma)\right)+\int_0^\theta e^{-k(\theta-s)}\left(\mathfrak{I}^{\alpha-1}\nu(s)\right)ds.$$

Lemma 2.2. *Authorizing that* $\nu \in C\left([0,1],\mathbb{R}\right)$. *So, we can gain*

$$i.\ \ |\phi_1\left(\nu(1),\nu(\varsigma)\right)| \le \underbrace{\left(|\psi_1|\frac{1-e^{-k}}{k\Gamma(\alpha)}+|\gamma_1|\frac{\varsigma^{\alpha+r}(1-e^{-k\varsigma})}{k\Gamma(\alpha)\Gamma(r+1)}\right)}_{L_1}\|\nu\|+|\epsilon_1| = L_1\|\nu\|+|\epsilon_1|$$

$$ii.\ \ |\phi_2\left(\nu(1),\nu(\varsigma)\right)| \le \underbrace{\left(|\psi_2|\frac{k+1-e^{-k}}{k\Gamma(\alpha)}+|\gamma_2|\frac{\varsigma^{\alpha+r}(k+1-e^{-k\varsigma})}{k\Gamma(\alpha)\Gamma(r+1)}\right)}_{L_2}\|\nu\|+|\epsilon_2|$$

$$= L_2\|\nu\|+|\epsilon_2|$$

$$iii.\ \ |\phi_3\left(\nu(1),\nu(\varsigma)\right)| \le \underbrace{\left(|\psi_3|\frac{\alpha-1+k(2-e^{-k})}{k\Gamma(\alpha)}+|\gamma_3|\frac{(\alpha-1)\varsigma^{\alpha+r-1}+\varsigma^{\alpha+r}(2-e^{-k\varsigma})}{\Gamma(\alpha)\Gamma(r+1)}\right)}_{L_3}\times$$

$$\|\nu\|+|\epsilon_3| = L_3\|\nu\|+|\epsilon_3|$$

Our hypothesis regarding λ will be elucidated prior to commencing and presenting the main results

(a): $\lambda:[0,1]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is continuous.

(b): there exist constants $a_{11},a_{12},a_{13}\in\mathbb{R}^+$ such that for all $\theta\in[0,1]$ and $\upsilon,\ \upsilon^*\in\mathbb{R}$: $|\lambda(\theta,\upsilon,\upsilon^*)|\le a_{11}|\upsilon|^{\sigma_1}+a_{12}|\upsilon^*|^{\sigma_2}+a_{13},\quad 0<\sigma_1,\sigma_2<1.$

(c): there exist constants $a_{21},a_{22}\in\mathbb{R}^+$ such that for all $\theta\in[0,1]$ and $\upsilon_1,\upsilon_2,\upsilon_1^*,\upsilon_2^*\in\mathbb{R}$: $|\lambda(\theta,\upsilon_2,\upsilon_2^*)-\lambda(\theta,\upsilon_1,\upsilon_1^*)|\le a_{21}|\upsilon_2-\upsilon_1|+a_{22}|\upsilon_2^*-\upsilon_1^*|.$

Theorem 2.3. *Suppose that conditions* (**a**) *and* (**b**) *are satisfied. Subsequently, it follows that the problem referenced as 1 possesses at least one solution.*

Theorem 2.4. *If assumptions* (**a**) *and* (**c**) *are satisfied, and the condition* $L_4\left(a_{21}+a_{22}\right)<1$ *holds, then the problem stated in reference 1 possesses a unique solution.*

Proof. Taking $sup_{\theta\in[0,1]}|\lambda(\theta,0,0)|=N<\infty$ way that $r\ge\frac{L_4N+L_5}{1-L_4(a_{21}+a_{22})}$. Firstly, it is revealed that $\mathfrak{P}(\psi_r)\subseteq\psi_r$, where $\psi_r=\{\upsilon|\ \upsilon\in\mathfrak{E};\|\upsilon\|_{\alpha-1}\le r\}$. For each υ belonging to the set ψ_r, through direct computation, it can be shown that

$$\begin{aligned}|\mathfrak{P}\upsilon(\theta)| &= |f_1(\theta)|\left|\phi_1\left(\lambda(1,\upsilon(1),\mathfrak{D}^{\alpha-1}\upsilon(1)),\lambda(\varsigma,\upsilon(\varsigma),\mathfrak{D}^{\alpha-1}\upsilon(\varsigma))\right)\right| \\ &+ |f_2(\theta)|\left|\phi_2\left(\lambda(1,\upsilon(1),\mathfrak{D}^{\alpha-1}\upsilon(1)),\lambda(\varsigma,\upsilon(\varsigma),\mathfrak{D}^{\alpha-1}\upsilon(\varsigma))\right)\right| \\ &+ |f_3(\theta)|\left|\phi_3\left(\lambda(1,\upsilon(1),\mathfrak{D}^{\alpha-1}\upsilon(1)),\lambda(\varsigma,\upsilon(\varsigma),\mathfrak{D}^{\alpha-1}\upsilon(\varsigma))\right)\right| \\ &+ \int_0^\theta e^{-k(\theta-s)}\left|\mathfrak{I}^{\alpha-1}\left(\lambda(s,\upsilon(s),\mathfrak{D}^{\alpha-1}\upsilon(s))\right)\right|ds\end{aligned}$$

$$
\begin{aligned}
&\leq M_1L_1\left(\max_{0\leq t\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|\right)\\
&+M_1|\epsilon_1|+M_2L_2\left(\max_{0\leq t\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|\right)\\
&+M_2|\epsilon_2|+M_3L_3\left(\max_{0\leq \theta\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|\right)\\
&+M_3|\epsilon_3|+\left(\max_{0\leq \theta\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|\right)\times\\
&\frac{(1-e^{-k\theta})\theta^{\alpha}}{k\Gamma(\alpha)}\leq M_1L_1\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)+M_1|\epsilon_1|\\
&+M_2L_2\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)+M_2|\epsilon_2|\\
&+M_3L_3\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)+M_3|\epsilon_3|\\
&+\frac{1}{k\Gamma(\alpha)}\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)\\
&\leq\left(M_1L_1+M_2L_2+M_3L_3+\frac{1}{k\Gamma(\alpha)}\right)\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)\\
&+\left(M_1|\epsilon_1|+M_2|\epsilon_2|+M_3|\epsilon_3|\right).
\end{aligned}
$$

Also, for any $\upsilon\in\psi_r$, we have

$$
\begin{aligned}
\left|\mathfrak{D}^{\alpha-1}\mathfrak{P}\upsilon(\theta)\right|&\leq\frac{k}{|\delta_6|}\left|\phi_3\left(\lambda(1,\upsilon(1),\mathfrak{D}^{\alpha-1}\upsilon(1)),\lambda(\varsigma,\upsilon(\varsigma),\mathfrak{D}^{\alpha-1}\upsilon(\varsigma))\right)\right|\times\\
&\int_0^{\theta}\frac{(\theta-s)^{2-\alpha}}{\Gamma(3-\alpha)}e^{-ks}ds\\
&+\int_0^{\theta}\frac{(\theta-s)^{2-\alpha}}{\Gamma(3-\alpha)}\left|\mathfrak{I}^{\alpha-1}\left(\lambda(s,\upsilon(s),\mathfrak{D}^{\alpha-1}\upsilon(s))\right)\right|ds\\
&+k\int_0^{\theta}\frac{(\theta-s)^{2-\alpha}}{\Gamma(3-\alpha)}\left(\int_0^{s}e^{-k(s-m}\left|\mathfrak{I}^{\alpha-1}\left(\lambda(m,\upsilon(m),\mathfrak{D}^{\alpha-1}\upsilon(m))\right)\right|dm\right)ds\\
&\leq\frac{k}{|\delta_6|}\left(L_3\max_{0\leq\theta\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|+|\epsilon_3|\right)\\
&+\frac{1}{\Gamma(\alpha)}\left(\max_{0\leq\theta\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|\right)\\
&+\frac{s^{\alpha}(1-e^{-ks})}{\Gamma(\alpha)}\left(\max_{0\leq\theta\leq 1}\left|\lambda(\theta,\upsilon(\theta),\mathfrak{D}^{\alpha-1}\upsilon(\theta))-\lambda(\theta,0,0)+\lambda(\theta,0,0)\right|\right)\\
&\frac{k}{|\delta_6|\Gamma(4-\alpha)}\Big(L_3\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)+|\epsilon_3|\Big)\\
&\quad+\frac{1}{\Gamma(\alpha)\Gamma(4-\alpha)}\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)\\
&+\frac{1}{\Gamma(\alpha)\Gamma(4-\alpha)}\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)\\
&=\left(\frac{kL_3}{|\delta_6|\Gamma(4-\alpha)}+\frac{2}{\Gamma(\alpha)\Gamma(4-\alpha)}\right)\left(a_{21}\|\upsilon\|+a_{22}\|\mathfrak{D}^{\alpha-1}\upsilon\|+N\right)
\end{aligned}
$$

$$= \left(M_4 L_3 + \frac{2}{\Gamma(\alpha)\Gamma(4-\alpha)} \right) \left(a_{21}\|v\| + a_{22}\|\mathfrak{D}^{\alpha-1}v\| + N \right) + M_4|\epsilon_3|.$$

Now, with the help of above discussion, we acquire

$$\begin{aligned}
\|\mathfrak{P}v\|_{\alpha-1} &= \|\mathfrak{P}v\| + \|\mathfrak{D}^{\alpha-1}\mathfrak{P}v\| \\
&\leq \left(\sum_{i=1}^{3} M_i L_i + \frac{1}{k\Gamma(\alpha)} \right) \left(a_{21}\|v\| + a_{22}\|\mathfrak{D}^{\alpha-1}v\| + N \right) + \sum_{i=1}^{3} M_i|\epsilon_i| \\
&\quad + \left(M_4 L_3 + \frac{2}{\Gamma(\alpha)\Gamma(4-\alpha)} \right) \left(a_{21}\|v\| + a_{22}\|\mathfrak{D}^{\alpha-1}v\| + N \right) + M_4|\epsilon_3| \\
&= \left(\sum_{i=1}^{3} M_i L_i + M_4 L_3 + \frac{\Gamma(4-\alpha)+2k}{k\Gamma(\alpha)\Gamma(4-\alpha)} \right) \left(a_{21}\|v\| + a_{22}\|\mathfrak{D}^{\alpha-1}v\| + N \right) \\
&\quad + \sum_{i=1}^{2} M_i|\epsilon_i| + (M_3 + M_4)\,|\epsilon_3| \\
&= L_4\,(a_{21} + a_{22})\,\|v\|_{\alpha-1} + L_4 N + L_5 \leq r.
\end{aligned}$$

where $L_4 = \sum_{i=1}^{3} M_i L_i + M_4 L_3 + \frac{\Gamma(4-\alpha)+2k}{k\Gamma(\alpha)\Gamma(4-\alpha)}$ and $L_5 = \sum_{i=1}^{2} M_i|\epsilon_i| + (M_3 + M_4)\,|\epsilon_3|$. Therefore

$$\begin{aligned}
\|\mathfrak{P}v_2 - \mathfrak{P}v_1\|_{\alpha-1} &= \|\mathfrak{P}v_2 - \mathfrak{P}v_1\| + \|\mathfrak{D}^{\alpha-1}\mathfrak{P}v_2 - \mathfrak{D}^{\alpha-1}\mathbb{F}\mathfrak{P}v_1\| \\
&\leq \left(\sum_{i=1}^{3} M_i L_i + M_4 L_3 + \frac{\Gamma(4-\alpha)+2k}{k\Gamma(\alpha)\Gamma(4-\alpha)} \right) (a_{21} + a_{22})\,\|v_2 - v_1\|_{\alpha-1} \\
&= L_4\,(a_{21} + a_{22})\,\|v_2 - v_1\|_{\alpha-1}.
\end{aligned}$$

Given that $L_4\,(a_{21} + a_{22}) < 1$, it can be deduced that the operator $\mathfrak{P}$ is constrained. Consequently, $\mathfrak{P}$ possesses a distinct FP, implying that the system 1 boasts a unique solution. This observation effectively shows that the result has been achieved. □

References

[1] Y. Li, Y. Sang, H. Zhang, Solvability of a coupled system of nonlinear fractional differential equations with fractional integral conditions, J. Appl. Math. Comput, 50(1-2) (2016) 73-91.

[2] T. Qi, Y. Liu, Y. Zou, Existence result for a class of coupled fractional differential systems with integral boundary value conditions, J. Nonli. Sci. Appl, 10(7) (2017) 4034-4045.

[3] Y. Qiao and Z. Zhou, Existence of solutions for a class of fractional differential equations with integral and anti-periodic boundary conditions, Bound. Val. Prob, 2017 (2017).

[4] H. Zhang, Y. Li, W. Lu, Existence and uniqueness of solutions for a coupled system of nonlinear fractional differential equations with fractional integral boundary conditions, J. Nonli. Sci. Appl, 9(5) (2016) 2434-2447.

[5] H. Zhang, Y. Li, J. Yang, New sequential fractional differential equations with mixed-type boundary conditions, J. Func. Spac, 2020, Article ID 6821637, 9 pages 2020.

[6] Y. Zou, G. He, The existence of solutions to integral boundary value problems of fractional differential equations at resonance, J. Func. Spac, 2017, Article ID 2785937, 7 pages, 2017.

Generalization of the Topological Interior Operator via Stacks

Amin Talabeigi[1], Ghasem Mirhosseinkhan[2] and Ali Shahrezaei[3,*]

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran.
Email: talabeigi@pnu.ac.ir, amin.talabeigi@gmail.com

[2]Department of Mathematics and Computer Sciences, Sirjan University of Technology, Sirjan, Iran.
Email: gh.mirhosseini@sirjantech.ac.ir, gh.mirhosseini@yahoo.com

[3] Ph.D. Student. Department of Mathematics, Payame Noor University, Tehran, Iran.
Email: alishahrezaee58@student.pnu.ac.ir

Abstract. We present a new operator, derived from stacks, which generalizes the conventional topological interior operator by relaxing its underlying conditions. The study focuses on the operators key properties and culminates in an examination and characterization of the generalized topological structure it defines.

Keywords: stack, stack space, generalized interior operator.

AMS Mathematics Subject Classification [2020]: 54A05, 54A99.

1. Introduction

Within the framework of general topology, any given topology τ on a set X uniquely defines a topological interior operator, commonly denoted as int_τ. This operator, when applied to an arbitrary subset $A \subseteq X$, returns the largest open set that is fully contained within A. A fundamental characterization of an open set A in the topological space (X, τ) is that it remains invariant under the interior operation, meaning $int_\tau(A) = A$.

The interior operator int_τ acts as a function $int_\tau : P(X) \to P(X)$, where $P(X)$ denotes the power set of X. This function satisfies the following four fundamental axioms for all arbitrary subsets $A, B \subseteq X$: (1): $int_\tau(X) = X$, (2): $int_\tau(A) \subseteq A$, (3): $int_\tau(int_\tau(A)) = int_\tau(A)$, (4): $int_\tau(A \cap B) = int_\tau(A) \cap int_\tau(B)$.

Conversely, any function $int : P(X) \to P(X)$ satisfying these four axioms uniquely defines a topological interior operator, and its set of fixed points, $\{A : int(A) = A\}$, constitutes a topology on X.

More recently, in [2], the authors introduced a generalized notion of topological interior operator under weaker conditions. Here we are faced with the fundamental question: is there a concrete and useful example of such a generalized interior operator, as proposed

*Speaker.

in [2], and if so, what are the salient features of this operator as well as the form and properties of the associated generalized induction structure?

Our paper aims to address this question by presenting such a generalized operator and characterizing the resulting topological structure it induces. In summary, we demonstrate a concrete instance of the generalized operator proposed in [2], implementing it with the aid of the stack $\mathcal{S}$.

In fact, considering a stack $\mathcal{S}$ on an arbitrary topological space (X, τ), we first define and analyze an operator called the $\mathcal{S}$-preinterior. Subsequently, we use this operator to construct the desired generalized interior operator, namely the $\mathcal{S}$-interior operator.

2. Preliminaries

Definition 2.1. (Generalized Interior Operator [2]): A mapping $Int : P(X) \to P(X)$ is called the generalized interior operator if for any set $A, B \subseteq X$, Int satisfies the following three axioms:

(I_1): $Int(A) \subseteq A$, (I_2): $A \subseteq B \Rightarrow Int(A) \subseteq Int(B)$, (I_3): $Int(Int(A)) = Int(A)$.

Definition 2.2. [3] A stack on a set X (or, topological space (X, τ)) is defined as a non-null collection $\mathcal{S}$ of nonempty subsets of X such that: for any $A \subseteq B \subseteq X$; if $A \in \mathcal{S}$ then $B \in \mathcal{S}$.

It is also worth noting that if the condition " if $A \cup B \in \mathcal{S}$, then $A \in \mathcal{S}$ or $B \in \mathcal{S}$ " is considered in conjunction with the aforementioned condition, we refer to $\mathcal{S}$ as a grill on the set X (or, the space (X, τ)), see [1].

Example 2.3.
(1) Let $X = \{1, 2, 3, 4\}$ and $\mathcal{S} = \{\{2\}, \{1, 2\}, \{2, 3\}, \{2, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{2, 3, 4\}, X\}$. Clearly, $\mathcal{S}$ is a stack on X.
(2) The collection $\{A \subseteq X : int_\tau(A) \neq \emptyset\}$ is a stack on any topological space (X, τ).
(3) ([5] Proposition 2.8) For any stack $\mathcal{S}$ on X, $dual(\mathcal{S}) := \{A \subseteq X \mid A^c \notin \mathcal{S}\}$ is again a stack on X.

3. Main Results

By a stack space $(X, \tau, \mathcal{S})$ we mean a topological space (X, τ) with a stack $\mathcal{S}$ on it.

3.1. Operator $(\mathcal{S}, \tau)$-preint. Let $(X, \tau, \mathcal{S})$ be a stack space. Then we define

$$(\mathcal{S}, \tau)\text{-}preint(A) := \{x \in X \mid U - A \notin \mathcal{S} \text{ for some } U \in \tau(x)\}$$

as an operator on $P(X)$, where $\tau(x) := \{U \in \tau \mid x \in U\}$. For ease of writing, instead of $(\mathcal{S}, \tau)\text{-}preint(A)$, we use one of $A^\diamond(\mathcal{S}, \tau)$, $A^\diamond(\mathcal{S})$, or $A^\diamond$.

Theorem 3.1. *Let $(X, \tau, \mathcal{S})$ be a stack space and $A, B \subseteq X$. Then*

(1) *$int(A^\diamond) = A^\diamond$, i.e., $A^\diamond$ is open in (X, τ).*
(2) *$int(A) \subseteq A^\diamond \subseteq (A^\diamond)^\diamond$.*
(3) *$(-)^\diamond$ is monotone, i.e., $A^\diamond \subseteq B^\diamond$ whenever $A \subseteq B$.*
(4) *If $A \notin \mathcal{S}$, then $(X - A)^\diamond = X$, and especially $X^\diamond = X$.*
(5) *for all $A, B \subseteq X$, $(A \cap B)^\diamond = A^\diamond \cap B^\diamond$.*
(6) *$A^\diamond(\mathcal{S}_1 \cup \mathcal{S}_2) = A^\diamond(\mathcal{S}_1) \cap A^\diamond(\mathcal{S}_2)$ for any stacks $\mathcal{S}_1$ and $\mathcal{S}_2$. So, if $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $A^\diamond(\mathcal{S}_2) \subseteq A^\diamond(\mathcal{S}_1)$, that is, $(-)^\diamond$ is decreasing with respect to stacks.*

Remark 3.2. From statements (1) and (2) in Theorem 3.1, it is clear that the set A cannot contain the set $A^\diamond$. Furthermore, the following example shows that $A^\diamond$ cannot contain A, thus proving that there is no general inclusion relation between A and $A^\diamond$. This

example also shows that the inverse inclusion stated in part (2) of Theorem 3.1 does not hold in general.

EXAMPLE 3.3. In the stack space $(X, \tau, \mathcal{S})$ with $X = \{1, 2, 3, 4\}$, $\tau = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, X\}$ and $\mathcal{S} = \{\{2\}, \{1, 2\}, \{2, 3\}, \{2, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{2, 3, 4\}, X\}$ we have:

(1) If $A = \{3, 4\}$ then $A^\diamond = \{3, 4\}^\diamond = \{1\}$, so $A \not\subseteq A^\diamond$ and $A^\diamond \not\subseteq A$.
(2) Also for $A = \{3\}$, $\{3\}^\diamond = \{1\}$ and $(\{3\}^\diamond)^\diamond = \{1\}$, so, $(A^\diamond)^\diamond \neq A$
(3) In addition, we have $\emptyset^\diamond = \{1\} \neq \emptyset$, that is, equality $\emptyset^\diamond = \emptyset$ is not generally valid.

In part 3 of Example 3.3, we saw that the equality $\emptyset^\diamond = \emptyset$ does not hold in general in every stack space. Considering the condition $\tau \setminus \{\emptyset\} \subseteq \mathcal{S}$ in the stack space in question, this equality holds.

The following theorem applies to this type of stack space.

THEOREM 3.4. *Let $(X, \tau, \mathcal{S})$ be a stack space. Then the following are equivalent:*

(1) $\tau \setminus \{\emptyset\} \subseteq \mathcal{S}$,
(2) $\emptyset^\diamond = \emptyset$,
(3) $X - H \notin \mathcal{S}$ *implies* $cl_\tau(H) = X$,
(4) $H \notin \mathcal{S}$ *implies* $int_\tau(H) = \emptyset$
(5) *For every closed set* F, $F^\diamond \subseteq F$,
(6) *For every closed set* F, $F^\diamond = int_\tau(F)$, *that is, in any stack space* $(X, \tau, \mathcal{S})$ *with* $\tau \setminus \{\emptyset\} \subseteq \mathcal{S}$, *the operators* $(-)^\diamond$ *and* int_τ *are equal on closed subsets..*

REMARK 3.5. Considering the assumptions of Example 3.3, we have:

(1) : $\emptyset^\diamond = \{1\}^\diamond = \{3\}^\diamond = \{4\}^\diamond = \{1, 3\}^\diamond = \{1, 4\}^\diamond = \{3, 4\}^\diamond = \{1, 3, 4\}^\diamond = \{1\}$,
(2) : $\{2\}^\diamond = \{1, 2\}^\diamond = \{2, 3\}^\diamond = \{2, 4\}^\diamond = \{1, 2, 3\}^\diamond = \{1, 2, 4\}^\diamond = \{2, 3, 4\}^\diamond = X^\diamond = X$.

3.2. Operator $(\mathcal{S}, \tau)$-int. It is not difficult to check that the operator $Int_\mathcal{S} A := A \cap A^\diamond(\mathcal{S})$ has the following;

(1) $(I_1) : Int_\mathcal{S}(A) \subseteq A$;
(2) $(I_2) : A \subseteq B \Rightarrow Int_\mathcal{S}(A) \subseteq Int_\mathcal{S}(B)$;
(3) $(I_3) : Int_\mathcal{S}(Int_\mathcal{S}(A)) = Int_\mathcal{S}(A)$.

So, considering any stack $\mathcal{S}$ on a topological space (X, τ) leads to the induction of a new type of generalized interior operator $Int_\mathcal{S}$ and therefore, inducing a new generalized topology in the form $\tau^*(\mathcal{S}) = \{A : Int_\mathcal{S}(A) = A\}$ (For short, we use τ^* instead of $\tau^*(\mathcal{S})$). Since, for any $A \subseteq (X, \tau)$, we have; $int_\tau A \subseteq (\mathcal{S}, \tau)\text{-}preint(A) \subseteq X$ (part (2) of Theorem 3.1), so $int_\tau A \cap A \subseteq (\mathcal{S}, \tau)\text{-}preint(A) \cap A \subseteq X \cap A$ and thus $int_\tau A \subseteq Int_\mathcal{S}(A) \subseteq A$. That means, for any stack $\mathcal{S}$ on a topological space (X, τ); $(X, \tau) \subseteq (X, \tau^*(\mathcal{S}))$.

THEOREM 3.6. *Let $(X, \tau, \mathcal{S})$ be a stack space and $A \subseteq X$. Then*

(1) $int_\tau(A) \subseteq Int_\mathcal{S}(A) \subseteq A^\diamond$, *and* $int_\tau(A) \subseteq Int_\mathcal{S}(A) \subseteq A$
(2) $A^\diamond$ *is both* τ*-open and* $\tau^*(\mathcal{S})$*-open.*
(3) *If* A *is* τ*-open, then* $Int_\mathcal{S}(A) = int_\tau(A) = A$.
(4) $Int_\mathcal{S}(A) = \emptyset \iff A^c$ *is* τ^**-dense in* $(X, \tau^*(\mathcal{S}))$.
(5) *The collection* $\beta := \{U - H \mid U \in \tau, H \notin \mathcal{S}\}$ *is a base for* τ^*.
(6) *For every* $H \notin \mathcal{S}$, $X - H$ *is* $\tau^*(\mathcal{S})$*-open.*

REMARK 3.7. Using Remark 3.5 and also considering the assumptions of Example 3.3, we have:

(1) $Int_{\mathcal{S}}(A) = \begin{cases} \emptyset & A \in \{\emptyset, \{3\}, \{4\}, \{3,4\}\} \\ \{1\} & A \in \{\{1\}, \{1,3\}, \{1,4\}, \{1,3,4\}\} \\ A & A \in \{\{2\}, \{1,2\}, \{2,3\}, \{2,4\}, \{1,2,3\}, \{1,2,4\}, \{2,3,4\}, X\} \end{cases}$
(2) $\tau^*(\mathcal{S}) = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{2,3\}, \{2,4\}, \{1,2,3\}, \{1,2,4\}, \{2,3,4\}, X\}$,
(3) $\beta = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,2,3\}, \{1,2,4\}, X\}$.

In the end, choosing $\mathcal{S} = \{A \subseteq X : int(A) \neq \emptyset\}$, we have the following two theorems;

THEOREM 3.8. *Let (X, τ) be a topological space. Considering $\{A \subseteq X : int(A) \neq \emptyset\}$ as a stack on this space, we have; $A^{\diamond}(\mathcal{S}, \tau) = intcl(A)$ and ahso, $Int_{\mathcal{S}}(A) = A \cap intcl(A)$.*

According to, Theorem 3.6, we have the following;

THEOREM 3.9. *Let (X, τ) be a topological space, then considering $\{A \subseteq X : int(A) \neq \emptyset\}$ as a stack on this space, we have;*

(1) *$\tau^*_{\mathcal{G}} = PO(X, \tau)$, where $PO(X, \tau)$ is the collection of all pre-open subsets of (X, τ).*
(2) *The collection $\beta = \{U - N \mid U \in \tau,\ int(N) = \emptyset\}$ is a base for $PO(X, \tau)$.*

PROOF. (1) $\tau^*(\mathcal{S}) = \{A \subseteq X : Int_{\mathcal{S}}(A) = A\}$ (by Theorem 3.6) $= \{A \subseteq X : A \cap intcl(A) = A\} = \{A \subseteq X : A \subseteq intcl(A)\} = PO(X, \tau)$.
(2) According to part (6) of Theorem 3.6, the proof is straightforward.

□

EXAMPLE 3.10. Let's assume that X and τ are the same as in Example 3.3. Choosing $\mathcal{S} = \{A \subseteq X : int(A) \neq \emptyset\}$, then $\mathcal{S} = \{\{1\}, \{2\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, X\}$. So, $PO(X) = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,2,3\}, \{1,2,4\}, X\}$ and $\beta_{PO(X)} = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,2,4\}, \{1,3,4\}, X\}$.

Finally, we should note that using grills instead of stacks in the above material will yield different results; see [4].

4. Conclusion

We have provided a practical example of the generalized operator introduced in [2]. Furthermore, by introducing a practical example of the generalization of the topological interior operator, we have introduced a method for generalizing the topological structure of a topological space to a type of generalized structure. We have also demonstrated that the collection of pre-open sets, a well-known and important set in topology, can be effectively calculated using our proposed method.

References

1. Choquet, G. (1947) *Sur les notions de filtre et grille*, Comptes Rendus Acad Sci Paris 224:171–173
2. Lei, Y. and Zhang, J. (2021) *Closure system and its semantics*, Axioms, **198**(10), 1–30.
3. Talabeigi, A. (2020) *Embedding topological space in a type of generalized topological spaces*, Khayyam J Math., **6(2)**, 250–256.
4. Talabeigi, A., Mirhosseinkhan, Gh., and Shahrezaei, A. (2025) *Extension of the Topological Interior Operator via Grills*, 56TH ANNUAL IRANIAN MATHEMATICS CONFERENCE.
5. Talabeigi, A. (2022) *Extracting some supra topologies from the topology of a topological space using stacks*, AUT Journal of Mathematics and Computing, **3 (1)**, 45–52.

The equivalent conditions for QHC-subspace and QHC-subset

Fatemeh Heidari[1,*],Mohammad Nikpour[2]

[1]Department of Mathematics, Kashan University, Kashan, Iran.
Email: zahraheidari@gmail.com
[2]Department of Psychology, Islamic Azad University, Mobarakeh, Iran.
Email: md.nikpou.mobin@gmail.com

Abstract. In this article, we first introduce the concepts of QHC-subspaces and QHC-subsets. Although these two concepts are independent of each other,we examine the conditions under which they become equivalent.

Keywords: QHC-subspace, QHC-subset, finite interior intersection

AMS Mathematics Subject Classification [2020]: 00A09, 54C40.

1. Introduction

A topological space on a set X with topology τ shall be denoted (X,τ). For a topological space (X,τ) the subspace topology on a subset A of X is τ_A. We shall write A^c, $cl(A)$ and $int(A)$ for the complement, closure, interior and boundary respectively of a set A with respect to a topology τ. We shall subscript these sets if several different topologies are being discussed; for example, if τ and τ^* are topologies on X and $A \subseteq X$, then $cl_\tau(A)$ and $cl_{\tau^*}(A)$ are the closures of A with respect to τ and τ^* respectively. The topological space X is called a quasi H-closed space (briefly: QHC) if for each τ-open cover $\{U_i\}_{i\in I}$ of X there exists a finite subset I_0 of I such that $X = \bigcup_{i\in I_0} cl(U_i)$. A subset A of the topological space X is said to be a QHC-subset if $\{U_i\}_{i\in I}$ is a family of τ-open sets of X that covers A, then there exists a finite subset I_0 of I such that $A \subseteq \bigcup_{i\in I_0} cl(U_i)$. A subset Y of the topological space X is a QHC-subspace if it is a QHC-space with respect to the subspace topology induced from X. It is easy to see that if A is a QHC-subspace of the topological space X, then it is a QHC-subset of X. It is not hard to find examples that show the converse of this fact is not true in general ([5] p. 161 and [3] Example 2.1).

Theorem 1.1. *Let (X,τ) be a topological space. Then the following hold:*
(1) QHC is preserved by finite unions.
(2) The closure of a QHC-subspace is a QHC-subspace.
(3) If A is a closed subset of a QHC-space X and if the boundary of A is a QHC-subspace,

*Speaker.

then so is A.
(4) A QHC-space is compact if and only if each closed subset is QHC.

Proof. It is obvious. □

Definition 1.2. A subset A of (X,τ) is said to be pre-open if $A(cl(A))$. The family of all pre-open subset of (X,τ) is denoted by $PO(X,\tau)$.

Proposition 1.3. *[4] A set $A \in PO(X,\tau)$ is a QHC-subspace of (X,τ) if and only if it is a QHC-subset.*

Definition 1.4. A family $\mathcal{A}$ of subsets of the topological space X has the finite interior intersection property, if the intersection of interiors of sets in each finite subfamily of $\mathcal{A}$ is non-empty. In other words if $A_1, \cdots, A_n \in \mathcal{A}$, then $\bigcap_{i=1}^{n} int(A_i) \neq \emptyset$.

Proposition 1.5. *The topological space X is QHC if and only if for every family of closed subsets $\{F_i\}_{i\in I}$ in X satisfying the finite interior intersection property, $\bigcap_{i\in I} F_i \neq \emptyset$.*

Proof. Necessity. Let $\{F_i\}_{i\in I}$ be a family of closed subsets of X such that $\bigcap_{i\in I} F_i = \emptyset$. Since $X = \bigcup_{i\in I} F_i^c$, $\{F_i^c\}_{i\in I}$ is an open cover of X. Thus there exists a finite subset I_0 of I such that $X = \bigcup_{i\in I_0} cl(F_i^c) = \bigcup_{i\in I_0} (int(F_i))^c$. Therefore $\bigcap_{i\in I_0} int(F_i) = \emptyset$.
Sufficiently. If $\{F_i\}_{i\in I}$ is the family of closed subset in X satisfying the interior finite intersection property, then for each finite subsets I_0 of I, $\bigcap_{i\in I_0} int(F_i) = \emptyset$. Let $\mathcal{C}$ be an open cover of X. Then $X = \bigcup_{C\in\mathcal{C}} C$ implies $\emptyset = \bigcap_{C\in\mathcal{C}} C^c$. Since $\mathcal{F} = \{C^c | C \in \mathcal{C}\}$ is a family of closed subset in X, so there is a finite subcollection $\{C_1^c, \cdots, C_n^c\}$ of $\mathcal{F}$, such that $\emptyset = \bigcap_{i=1}^{n} int_\tau(C_i^c) = \bigcap_{i=1}^{n} (cl_\tau(C_i))^c$. It follows that $\bigcup_{i=1}^{n} cl(C_i) = X$ and (X,τ) is a QHC-space. □

Using the same basic method of proof of proposition 1.5 we can also prove the following:

Proposition 1.6. *Let X be a the topological space and $Y \subseteq X$. Then Y is a QHC-subset if and only if for every family of closed subsets $\{F_i\}_{i\in I}$ in X included in Y satisfying the finite interior intersection property, $\bigcap_{i\in I} F_i \neq \emptyset$.*

2. Main results

Proposition 2.1. *Let (X,τ) be a topological space. Then the following hold:*
(1) Every dense QHC-subset is a QHC-space.
(2) Supposed that Y is a closed QHC-subset in X and $A \subseteq Y$ is a closed subset in X. If $int(A) \neq \emptyset$ in Y implies that $int_\tau(A) \neq \emptyset$, then Y is a QHC-space.

Proof. (1) Let A be a dense and QHC-subset in X and $A = \bigcup_{U\in\mathcal{U}} U \cap A$, where $\mathcal{U}$ is a τ-open in X. Then $A \subseteq \bigcup_{U\in\mathcal{U}} U$ and there are $U_1, \cdots, U_n \in \mathcal{U}$ such that $A \subseteq \bigcup_{i=1}^{n} cl(U_i) \subseteq \bigcup_{i=1}^{n} cl(U_i \cap A)$. Hence

$$A = \bigcup_{i=1}^{n} cl(U_i \cap A) \cap A = \bigcup_{i=1}^{n} cl_{\tau_A}(U_i \cap A).$$

(2) Follows from Proposition 1.6. □

It is clear that in the proposition 2.1 (2), if X is Hausdorff, the assumption that Y is closed is redundant. It can also be said that the assumption that Y is closed is not a necessary condition for the proposition to be true. For example, if X has a finite complement topology, then proper infinite subsets of X are not closed and the condition given in (2) holds. Moreover, every subset of X is compact and therefore a QHC-space.

Proposition 2.1 raised the question of whether the converse is true if X is Hausdorff. It can be easily seen that the answer to this question is negative. For example, let X be a Hausdorff space with no isolated points. Let $Y \subseteq X$ be finite. Hence Y is a QHC-space and for every $\emptyset \neq A \subseteq Y$, $int(A) = A \neq \emptyset$ in Y, while $int_\tau(A) = \emptyset$.

THEOREM 2.2. *[2] Let (X, τ) be a space and $A, B \subseteq X$. If A is a QHC-subset and B is a closed and open subset in (X, τ), then $A \cap B$ is a QHC-subset in (X, τ).*

COROLLARY 2.3. *For every QHC-space (X, τ), every closed and open subset of X is a QHC-subset in (X, τ).*

In the following, we will expand Theorem 2.2 and Corollary 2.3 using the next propositions and lemmas.

PROPOSITION 2.4. *Let (X, τ) be a topological space and $A \subseteq X$. Then the following statements hold:*
(1) If A and A^c are QHC-subsets, then X is a QHC-space.
(2) Let $A \subseteq Y \subseteq X$. If Y is a QHC-subset in X and A is a closed and open subset in X, then A is a QHC-subspace in X.
(3) If A is a closed and open subset in QHC-space X, then A is a QHC-space.
(4) Every open QHC-subset A in (X, τ) is a QHC-space.
(5) If A is closed and open subset in X, then the following hold:
(a) X is a QHC-space.
(b) Both A and A^c are QHC-spaces.
(c) Both A and A^c are QHC-subsets.

PROOF. We only prove (2), the rest of the proofs are straightforward. Let $\{A \cap U_i\}_{i \in I}$ be an arbitrary τ_A-open cover of A in Y, where U_i is an τ-open subset in X. Since $Y \subseteq (\bigcup_{i \in I} U_i) \cup A^c$ and A^c is open in X, then there are $i_1, \cdots, i_n \in I$ such that $Y \subseteq (\bigcup_{j=1}^{n} cl_\tau(U_{i_j})) \cup cl_\tau(A^c)$. Therefore

$$A = (\bigcup_{j=1}^{n} A \cap cl_\tau(U_{i_j})) \cup (A \cap cl_\tau(A^c)) = \bigcup_{j=1}^{n} A \cap cl_\tau(U_{i_j}) = \bigcup_{j=1}^{n} cl_\tau(A \cap U_{i_j}).$$

□

LEMMA 2.5. *If A, B are distinct α-open sets in (X, τ), then A, B are open subsets in $A \cup B$.*

PROOF. It is enough to show that A is an open subset in $A \cup B$. Let $a \in A$, then there is an open subset U such that $a \in U \subseteq cl(int(A))$. Hence

$$int(A) \cap int(B) = \emptyset \Rightarrow cl(int(A)) \cap int(B) = \emptyset \Rightarrow U \cap int(B) = \emptyset.$$

Therefore

$$U \cap cl(int(B)) = \emptyset \Rightarrow U \cap B = \emptyset \Rightarrow U \cap (A \cup B) = U \cap A \subseteq A.$$

As a result $a \in int(A)$ in $A \cup B$. □

DEFINITION 2.6. A subset A of (X, τ) is said to be α-open if $A \subseteq int\big(cl(int(A))\big)$. The family of all α-open subset of (X, τ) is denoted by τ^α. For every space the family τ^α forms a topology on X.

LEMMA 2.7. *If U is an open subset in (X, τ) and $A \in \tau^\alpha$, then $A \cap cl_\tau(U) = cl_{\tau_A}(A \cap U)$.*

PROOF. It is enough to prove that $A \cap cl_\tau(U) \subseteq cl_{\tau_A}(A \cap U)$. Consider a point $a \in A \setminus cl_{\tau_A}(A \cap U)$. We must show that $a \notin cl_\tau(U)$. Since $a \in A \subseteq int(cl_\tau(int(A)))$, then there is an open neighborhood W such that $(W \cap A) \cap (A \cap U) = \emptyset$ and $W \subseteq cl_\tau(int(A))$. Hence $W \cap (A \cap U) = \emptyset$ and $W \subseteq cl_\tau(int(A))$. Therefore $W \cap U \cap cl_\tau(int(A)) = \emptyset$ and $W \subseteq cl_\tau(int(A))$. This show that $W \cap U = \emptyset$ and $a \notin cl_\tau(U)$. □

We are now ready to improve 2.2 as follows:

PROPOSITION 2.8. *Let A, B be α-open sets in (X, τ). Then the following statements are equivalent:*
(1) $A \cup B$ is a QHC-space.
(2) A, B are QHC-spaces.
(3) A, B are QHC-subsets.
(4) $A \cup B$ is a QHC-subset.

PROOF. These follow from Proposition 2.4 (5), Lemma 2.5 and 2.7. □

3. Conclusion

In this article, we only addressed the limited conditions for the equivalence of QHC-subsets and QHC-subspaces. In topology, open sets and generalized open sets play an important role, such that by using other types of generalized open sets like semi-open and semi pre-open sets, it can be shown that QHC-subsets and QHC-subspaces are equivalent. For this purpose, one can refer to [relevant sources.

References

1. Arthur Steen, L., Arthur Seebach, J. (1995), *Counterexamples in Topology*, Dover Publications.
2. Duszynski, Z. (2010), *On quasi H-closed subspaces*, Institute of Mathematics, **59**, pp. 187–197.
3. Noiri, T. (1978), *On S-closed subspaces*, Atti della Accademia Nazionale dei Lincei. Rendiconti. Classe di Scienze Fisiche, Matematiche e Naturali, **64**, pp. 157–162.
4. Noiri, T., Mashhour. A.S., Hasanein. I.A , El-Deeb. S.N. (1982), *A note on S-closed subspaces*, Mathematics Seminar Notes, Kobe University, **10**, pp. 431– 435.
5. Porter, J. and Thomas, J. (1969), *On H-closed and minimal Hausdorff spaces*, Transactions of the American Mathematical Society, **138**, pp. 159–170.

When $C_{c\infty}(X)$ is an ideal of $C_c(X)$

Somayeh Soltanpour[1,*]

[1]Department of Basic Science, Petroleum University of Technology, Ahvaz, Iran.
Email: s.soltanpour@put.ac.ir

Abstract. $C_{c\infty}(X)$ denotes the set of all functions $f \in C_c(X)$ that vanish at infinity, that is, for each $n \in \mathbb{N}$, the set $\{x \in X : |f(x)| \geq 1/n\}$ is compact. It has been shown that $C_{c\infty}(X)$ coincides with the intersection of all free ideals in $C_c^*(X)$. We characterize the spaces X for which $C_{c\infty}(X)$ forms an ideal in $C_c(X)$. In particular, if X is a locally compact, zero-dimensional Hausdorff space, then $C_{c\infty}(X)$ is an ideal in $C_c(X)$ if and only if X is c-pseudocompact. Moreover, if there exists a function $f \in C_{c\infty}(X) \setminus C_{cK}(X)$ with $Z(f)$ open, then $C_{c\infty}(X)$ does not form an ideal in $C_c(X)$.

Keywords: $C_c(X)$, c-pseudocompact spaces, Compact support, $C_{cK}(X)$, $C_{c\infty(X)}$.

AMS Mathematics Subject Classification [2020]: Primary: 54C40, 54C30; Secondary: 13C11.

1. Introduction

The study of rings of continuous functions has long provided a deep interplay between algebraic and topological structures. For a completely regular Hausdorff space X, the ring $C(X)$ of real-valued continuous functions encapsulates a wide range of topological information about X. In this paper, we focus on the subring $C_{c\infty}(X)$ consisting of continuous functions whose supports are countably compact at infinity, see [1]. Our goal is to establish how various algebraic properties of $C_{c\infty}(X)$ correspond to specific topological features of X.

The support of $f \in C(X)$ is defined as the closure of $X \setminus Z(f)$, and let $C_K(X) = \{f \in C(X) : \text{support } f \text{ is compact}\}$. The equivalence between $C_K(X)$ and the intersection of free maximal ideals (where an ideal I in $C(X)$ is termed free if $\bigcap Z[I] = \emptyset$, and otherwise fixed) has been investigated, as seen in [5], [6]. Let $C_\infty(X)$ denote the set of all functions $f \in C(X)$ that vanish at infinity, meaning $\{x \in X : |f(x)| \geqslant 1/n\}$ is compact for each $n \in \mathbb{N}$. It is clear that $C_K(X) \subseteq C_\infty(X)$, and $C_K(X)$ is the intersection of the free ideals in $C(X)$, as well as the intersection of the free ideals in $C^*(X)$. As outlined in [3, 7F], $C_\infty(X)$ is the intersection of the free maximal ideals in $C^*(X)$. Therefore, both $C_K(X)$ and $C_\infty(X)$ are intersections of certain essential ideals. It has been emphasized that $C_\infty(X)$ may not be an ideal in $C(X)$. Building upon this observation, spaces X for which $C_\infty(X)$ qualifies as an ideal in $C(X)$ have been characterized in [2]. Specifically, $C_\infty(X)$

*Speaker.

is an ideal in $C(X)$ if and only if every open locally compact subset of X is bounded. In particular, for a locally compact Hausdorff space X, $C_\infty(X)$ serves as an ideal in $C(X)$ if and only if X is a pseudocompact space, see [1], [2].

When restricting to continuous functions with countable image, we denote by $C_c(X)$ and $C_c^*(X)$ the sets of all (bounded) continuous real-valued functions on X with countable image, see [4]. The subring $C_{c\infty}(X)$, consisting of functions in $C_c(X)$ vanishing at infinity, is introduced as an analogue of $C_\infty(X)$, see [5], [6]. It is shown that $C_{c\infty}(X)$ equals the intersection of all free ideals in $C_c^*(X)$. This paper investigates the ideal-theoretic properties of $C_{c\infty}(X)$ in $C_c(X)$. In particular, we characterize the topological spaces X for which $C_{c\infty}(X)$ forms an ideal in $C_c(X)$. We prove that this occurs if and only if X is c-pseudocompact.

2. Main results

The ring $C_{c\infty}(X)$ has been introduced as the set of all continuous real-valued functions on X with countable image that vanish at infinity. This ring generalizes $C_\infty(X)$ in the context of functions with countable range, and its algebraic and topological properties form the focus of the present study. For any topological space X, the set of all continuous real valued functions with countable image which vanish at infinity is a ring, which is denoted by $C_{c\infty}(X)$. In fact for every $f, g \in C_{c\infty}(X)$, we have $\{x \in X : |f(x) + g(x)| \geq \frac{1}{n}\} \subseteq \{x \in X : |f(x)| \geq \frac{1}{2n}\} \cup \{x \in X : |g(x)| \geq \frac{1}{2n}\}$ and $\{x \in X : |f(x)g(x)| \geq \frac{1}{n}\} \subseteq \{x \in X : |f(x)| \geq \frac{1}{\sqrt{n}}\} \cup \{x \in X : |g(x)| \geq \frac{1}{\sqrt{n}}\}$.

THEOREM 2.1. *$C_{c\infty}(X)$ is the intersection of all free maximal ideals in $C_c^*(X)$, i.e.,*

$$C_{c\infty}(X) = \bigcap_{p \in \beta_0 X \setminus X} M_c^{P^*} = \bigcap_{p \in \beta_0 X \setminus X} \{f \in C_c^*(X) \ : \ f^{\beta_0}(\beta_0 X \setminus X) = \{0\}\}.$$

THEOREM 2.2. *Let X be a zero dimension and Hausdorff space. The intersection of all free maximal ideals of $C_c(X)$ is contained the intersection of all free maximal ideals of $C_c^*(X)$.*

LEMMA 2.3. *Let A be an open subset of X, then $A = X \setminus Z(f)$ for some $f \in C_{c\infty}(X)$ if and only if A is σ-compact locally subset of X.*

COROLLARY 2.4. *$C_{c\infty}(X)$ contains a unit of $C_c(X)$ if and only if X is a locally compact σ-compact space.*

THEOREM 2.5. *Let X be a countably completely regular Hausdorff and zero-dimensional space. The following conditions are equivalent:*

(1) *$C_{c\infty}(X)$ is an ideal in $C_c(X)$.*
(2) *Every open locally compact subset of X is bounded.*
(3) *Every open locally compact σ-compact subset of X is bounded.*

COROLLARY 2.6. *Let X be a locally compact Hausdorff zero-dimensional space. Then $C_{c\infty}(X)$ is an ideal in $C_c(X)$ if and only if X is a c-pseudocompact space.*

COROLLARY 2.7. *Suppose that there exists $g \in C_{c\infty}(X)$ with $Z(g)$ Lindelof and bounded. If $C_{c\infty}(X)$ is an ideal in $C_c(X)$, then X is a compact space.*

LEMMA 2.8. *Let $X = Y \oplus Z$, i.e., Y and Z are disjoint open subsets of X such that $X = Y \cup Z$. $C_{c\infty}(X)$ is an ideal of $C_c(X)$ if only if $C_{c\infty}(Y)$ is an ideal of $C_c(Y)$ and $C_{c\infty}(Z)$ is an ideal of $C_c(Z)$.*

Let us recall that $C_{cK}(X)$ denotes the set of all functions in $C_c(X)$ with compact support i.e., $cl_X(X \setminus Z(f))$ is compact.

PROPOSITION 2.9. *Let $f \in C_{c\infty}(X) \setminus C_{cK}(X)$ such that $Z(f)$ be an open set. Then $C_{c\infty}(X)$ is not ideal of $C_c(X)$.*

Acknowledgement

The authors wish to thank the anonymous reviewers for their valuable suggestions.

References

1. Aliabad, A. R., Azarpanah, F., and Namdari, M. (2004). *Rings of continuous functions vanishing at infinity*, Comment. Math. Univ. Carol. **45**(3), 519–533.
2. Azarpanah, F., and Soundararajan, T. (2001). *When the family of functions vanishing at infinity is an ideal of $C(X)$*, Rocky Mountain J. Math. **31**(4), 1133–1140.
3. Gillman, L., and Jerison, M. (1960). *Rings of Continuous Functions*, Van Nostrand, New York.
4. Ghadermazi, M., Karamzadeh, O. A. S., and Namdari, M. (2019). *$C(X)$ versus its functionally countable subalgebra*, Bull. Iranian Math. Soc. **45**(1), 173–187.
5. Karamzadeh, O.A.S., Rostami, M. (1985). *O*n the intrinsic topology and some related ideals of $C(X)$, Proceedings of the American Mathematical Society **93** , no. 1, 179–184.
6. Taherifar, A. (2017). *O*n a question of Kaplansky, *Topology and its Applications* **232**, 98–101.

$C_{c\infty}(X)$ and Related Ideals in $C_c(X)$

Somayeh Soltanpour[1,*]

[1]Department of Science, Petroleum University of Technology, Ahvaz, Iran.
Email: S.soltanpour@put.ac.ir

Abstract. In this paper, we introduce and investigate the subring $C_{c\infty}(X)$, consisting of all continuous real-valued functions on X with countable image that vanish at infinity. This ring generalizes the classical $C_\infty(X)$ to the context of countable-valued functions. We prove that $C_{c\infty}(X)$ coincides with the intersection of all free ideals in $C_c^*(X)$, the ring of bounded continuous countable-valued functions. Moreover, we characterize when $C_{c\infty}(X)$ forms an ideal in $C_c(X)$, showing that this occurs if and only if X is c-pseudocompact. Relationships between $C_{c\infty}(X)$ and other subrings, including $C_{cK}(X)$ (functions with compact support) and $C_{c\psi}(X)$ (functions with c-pseudocompact support), are studied, and we establish that $C_{c\infty}(X) = C_{c\psi}(X)$ if and only if X is compact. Finally, we examine the ideal $I(X)$, defined as the intersection of all free maximal ideals of $C_c(X)$, and describe various compactness properties—μ_0-, η_0-, and $c\psi$-compactness—via inclusions among these ideals.

Keywords: $C_c(X)$, c-pseudocompact spaces, Compact support, $C_{cK}(X)$, $C_{c\infty(X)}$.

AMS Mathematics Subject Classification [2020]: Primary: 54C40, 54C30; Secondary: 13C11.

1. Introduction

Let X be a topological space and denote by $C(X)$ the ring of all continuous real-valued functions on X. Understanding how the algebraic properties of $C(X)$ reflect the topology of X is a central topic in the theory of function rings. Notable subrings include $C_K(X)$ and $C_\infty(X)$, defined respectively by

$$C_K(X) = \{f \in C(X) : \operatorname{supp}(f) \text{ is compact}\},$$

$$C_\infty(X) = \{f \in C(X) : \{x \in X : |f(x)| \geq 1/n\} \text{ is compact for all } n \in \mathbb{N}\}.$$

It is well-known that $C_K(X) \subseteq C_\infty(X)$, and that $C_K(X)$ coincides with the intersection of all free ideals in both $C(X)$ and $C^*(X)$, the ring of bounded continuous functions. Furthermore, by [3, 7F], $C_\infty(X)$ is the intersection of all free maximal ideals in $C^*(X)$. Thus, both $C_K(X)$ and $C_\infty(X)$ can be regarded as intersections of essential ideals in $C(X)$.

The structure of $C_\infty(X)$ has been widely studied (see [1, 2]). For instance, $C_\infty(X)$ has finite Goldie dimension if and only if X is finite, and for any Hausdorff space X, there

*Speaker.

exists a locally compact Hausdorff space Y with $C_\infty(X) \cong C_\infty(Y)$. Moreover, for locally compact Hausdorff spaces X and Y, $C_\infty(X) \cong C_\infty(Y)$ if and only if $X \cong Y$. However, $C_\infty(X)$ is not always an ideal in $C(X)$. A characterization of spaces for which $C_\infty(X)$ is an ideal is given in [2]: $C_\infty(X)$ is an ideal in $C(X)$ if and only if every open locally compact subset of X is bounded. In particular, if X is locally compact and Hausdorff, this holds precisely when X is pseudocompact.

To generalize these ideas, we focus on continuous functions with countable images. Denote by $C_c(X)$ and $C_c^*(X)$ the rings of all (bounded) continuous real-valued functions on X with countable range (see [4, 5]). We introduce

$$C_{c\infty}(X) = \{f \in C_c(X) : \{x \in X : |f(x)| \geq 1/n\} \text{ is compact for all } n \in \mathbb{N}\},$$

which consists of countable-valued functions vanishing at infinity. This provides a natural analogue of $C_\infty(X)$ in the countable-valued setting.

We show that $C_{c\infty}(X)$ coincides with the intersection of all free ideals in $C_c^*(X)$. We also study its ideal-theoretic properties within $C_c(X)$ and characterize the spaces X for which $C_{c\infty}(X)$ forms an ideal, proving that this occurs if and only if X is c-pseudocompact.

Furthermore, we examine the relationships between $C_{c\infty}(X)$ and other subrings:

$$C_{cK}(X) = \{f \in C_c(X) : \operatorname{supp}(f) \text{ is compact}\},$$

$$C_{c\psi}(X) = \{f \in C_c(X) : \operatorname{supp}(f) \text{ is } c\text{-pseudocompact}\}.$$

We establish that $C_{c\infty}(X) = C_{c\psi}(X)$ if and only if X is compact. We also study the ideal

$$I(X) = \bigcap\{\text{free maximal ideals of } C_c(X)\}$$

and characterize various compactness properties such as μ_0-, η_0-, and $c\psi$-compactness through inclusions among these ideals.

2. Main Results

We denote by $C_{c\psi}(X)$ the set of all functions with c-pseudocompact support. We aim to establish that $C_{c\psi}(X)$ constitutes an ideal in $C_c(X)$, see [1]. It is evident that $C_{cK}(X) \subseteq C_{c\psi}(X)$. When $C_{cK}(X) = C_{c\psi}(X)$, the space X is termed $c\psi$-compact. We assert that $C_{c\infty}(X) \subseteq C_{c\psi}(X)$ if and only if $C_{c\infty}(X)$ is an ideal of $C_c(X)$. Moreover, for a locally compact, zero-dimensional, and Hausdorff space X, $C_{c\infty}(X) = C_{c\psi}(X)$ if and only if X is compact. Another ideal associated with $C_{c\infty}(X)$ is denoted by $I(X)$ and defined as the intersection of all free maximal ideals of $C_c(X)$. For any space X, the inclusion relationships $C_{cK}(X) \subseteq I(X) \subseteq C_{c\psi}(X)$ hold. When $C_{cK}(X) = I(X)$ or $I(X) = C_{c\psi}(X)$, it is said that X is μ_0-compact or η_0-compact, respectively. In this section, we will establish that for a completely regular, zero-dimensional, and Hausdorff space X, the equality $C_{c\infty}(X) = C_{cK}(X)$ holds if and only if every open, locally compact, and σ-compact subset of X is contained within a compact subset of X.

Proposition 2.1. *Suppose X is a completely regular Hausdorff space. Then $C_{c\infty}(X) = C_{cK}(X)$ if and only if every open, locally compact, σ-compact subset of X is contained in a compact set.*

Proposition 2.2. *$C_{c\infty}(X) \subseteq C_{c\psi}(X)$ if and only if every open, locally compact subset of X is bounded.*

Corollary 2.3. *If X is locally compact and Hausdorff, then $C_{c\infty}(X) = C_{c\psi}(X)$ if and only if X is compact.*

PROPOSITION 2.4. *Let*

$$C_{cl\sigma}(X) = \{f \in C_c(X) : X \setminus Z(f) \text{ is locally compact and } \sigma\text{-compact}\}.$$

Then $C_{cl\sigma}(X)$ is either the smallest z_c-ideal containing $C_{c\infty}(X)$ or $C_{cl\sigma}(X) = C_c(X)$.

LEMMA 2.5. *For a space X, define:*

(1) $C_{cl}(X) = \{f \in C_c(X) : X \setminus Z(f) \text{ is locally compact}\}$.
(2) $C_{c\bar{l}}(X) = \{f \in C_c(X) : \overline{X \setminus Z(f)} \text{ is locally compact}\}$.
(3) $C_{c\sigma}(X) = \{f \in C_c(X) : X \setminus Z(f) \text{ is } \sigma\text{-compact}\}$.
(4) $C_{c\bar{\sigma}}(X) = \{f \in C_c(X) : \overline{X \setminus Z(f)} \text{ is } \sigma\text{-compact}\}$.
(5) $I_{c\overline{l\sigma}}(X) = \{f \in C_c(X) : \overline{X \setminus Z(f)} \text{ is contained in an open locally } \sigma\text{-compact set}\}$.
(6) $C_{c\overline{l\sigma}}(X) = \{f \in C_c(X) : \overline{X \setminus Z(f)} \text{ is locally compact and } \sigma\text{-compact}\}$.
(7) $C^*_{cl\sigma}(X) = \{f \in C^*_c(X) : X \setminus Z(f) \text{ is locally compact and } \sigma\text{-compact}\}$.

Then $C^*_{cl\sigma}(X)$ is an ideal of $C^*_c(X)$, and the others are z_c-ideals in $C_c(X)$.

LEMMA 2.6. *The following statements hold:*

(1) $I_{c\overline{l\sigma}}(X) \subseteq C_{c\overline{l\sigma}}(X) \subseteq C_{cl\sigma}(X) \subseteq C_{cl}(X)$.
(2) $I_{c\overline{l\sigma}}(X) \subseteq C_{c\infty}(X)C_c(X) \subseteq C_{cl\sigma}(X)$.
(3) $C_{cK}(X) = C_{c\bar{\sigma}}(X) \cap C_{c\psi}(X)$.
(4) $C_{cl\sigma}(X) = C_{cl}(X) \cap C_{c\sigma}(X) \subseteq C_{cl}(X) \cap C_{cR}(X)$.
(5) $C_{cK}(X) \subseteq C_{c\bar{l}}(X) \subseteq C_{cl}(X)$.

PROPOSITION 2.7. *The following statements hold:*

(1) $I(X) = C_{c\bar{\sigma}}(X)$ *if and only if X is μ_0-compact.*
(2) $C_{c\psi}(X) \subseteq C_{c\infty}(X)$ *if and only if X is η_0-compact. Hence $C_{c\psi}(X) = C_{c\infty}(X)$ if and only if X is η_0-compact and every open locally compact set is relatively c-pseudocompact.*
(3) $C_{c\psi}(X) \subseteq C_{c\bar{\sigma}}(X)$ *if and only if X is ψ_0-compact.*

In the following theorem, we characterize the space X for which the smallest z_c-ideal containing $C_{c\infty}(X)$ is a prime ideal. We designate a point $x \in X$ as an l-point if it has a compact neighborhood. It is evident that the set of l-points in X is open.

THEOREM 2.8. *$C_{cl\sigma}(X)$ is a prime ideal is and only if X has at most one non-l-point $x^* \in X$ and for any two disjoint cozeroset, one which does not contain the non-l-point, is locally compact σ-compact.*

PROPOSITION 2.9. *$C^*_{cl\sigma}(X) = C_{c\infty}(X)$ if and only if every zero-set contained in an open locally compact σ-compact subset of X is compact.*

THEOREM 2.10. *$I(X) = C_{cl\sigma}(X)$ if and only if for every open locally compact σ-compact subset A of X, $cl_X A$ is c-pseudocompact and every zero-set in A is compact.*

COROLLARY 2.11. *Let X be a c-realcompact space. Then every open locally compact σ-compact subset of X has compact closure if and only if $I(X) = C_{cl\sigma}(X)$.*

PROPOSITION 2.12. *A locally compact σ-compact open set G in X has a c-pseudocompact closure if and only if $\beta_0 X \setminus X \subseteq cl_{\beta_0 X}(X \setminus G)$. In particular, $\beta_0 X \setminus X \subseteq cl_{\beta_0 X} Z(f)$ if and only if $X \setminus Z(f)$ is locally compact σ-compact and $cl_{\beta_0 X}(X \setminus Z(f))$ is c-pseudocompact.*

PROPOSITION 2.13. $C_{cl}(X) = \bigcap_{x\in\mathbb{N}} M_{cx} = \{f \in C_c(X) : f(x) = 0,\ \forall\, x \in \mathbb{N}\}$.

PROPOSITION 2.14. *If $cl_X L = X \setminus int_X N$ is locally compact (σ-compact), then $C_{c\bar{l}}(X) = C_{cl}(X)$ ($C_{c\sigma}(X) = C_{c\bar{\sigma}}(X)$).*

PROPOSITION 2.15. *The following statements hold:*

(1) *If L is σ-compact, then $C_{cl\sigma}(X) = C_{cl}(X)$.*

(2) *If X is second countable and $C_{cl\sigma}(X) = C_{cl}(X)$, then L is σ-compact.*

PROPOSITION 2.16. *The following statements hold:*

(1) *X is locally compact if and only if $C_{c\bar{l}}(X) = C_{cl}(X) = C_c(X)$, if and only if $C_{cl\sigma}(X)$ is a free ideal, if and only if $C_{cl\sigma}(X) = C_{c\sigma}(X)$.*
(2) *X is σ-compact if and only if $C_{c\bar{\sigma}}(X) = C_{c\sigma}(X) = C_c(X)$.*
(3) *X is locally compact σ-compact if and only if*

$$C_{c\overline{l\sigma}}(X) = C_{c\infty}(X)C_c(X) = C_{cl\sigma}(X) = C_c(X).$$

PROPOSITION 2.17. *Let X be a locally compact σ-compact space. Then X is perfectly normal if and only if every open subset of X is σ-compact.*

PROPOSITION 2.18. *Let X be a normal space. If $C_{c\bar{l}}(X) = C_{cK}(X)$, then every closed subset of X contained in L is compact. Whenever L is closed the converse is also true, in fact if L is compact, then $C_{c\bar{l}}(X) = C_{cK}(X)$.*

LEMMA 2.19. *No point of $A \subseteq X$ has a compact neighborhood in X if and only if $f(A) = 0$ for all $f \in C_{c\infty}(X)$.*

PROPOSITION 2.20. *Let $\mathcal{A}$ be a commutative algebra over the rationales with unity. Let I be an ideal of $\mathcal{A}$. Then an ideal D of I is a maximal ideal of I if and only if $D = M \cap I$ for some maximal ideal M in $\mathcal{A}$.*

Acknowledgement

The authors wish to thank the anonymous reviewers for their valuable suggestions.

References

[1] Aliabad, A. R., Azarpanah, F., and Namdari, M. (2004). *Rings of continuous functions vanishing at infinity*, Comment. Math. Univ. Carol. **45**(3), 519–533.
[2] Azarpanah, F., and Soundararajan, T. (2001). *When the family of functions vanishing at infinity is an ideal of $C(X)$*, Rocky Mountain J. Math. **31**(4), 1133–1140.
[3] Gillman, L., and Jerison, M. (1960). *Rings of Continuous Functions*, Van Nostrand, New York.
[4] Ghadermazi, M., Karamzadeh, O. A. S., and Namdari, M. (2019). *$C(X)$ versus its functionally countable subalgebra*, Bull. Iranian Math. Soc. **45**(1), 173–187.
[5] Ghadermazi, M., Karamzadeh, O. A. S., and Namdari, M. (2013). *On the functionally countable subalgebra of $C(X)$*, Rend. Sem. Mat. Univ. Padova **129**, 47–69.