Springer Proceedings in Mathematics & Statistics

Fatih Yilmaz · Araceli Queiruga-Dios ·
María Jesús Santos Sánchez ·
Deolinda Rasteiro ·
Víctor Gayoso Martínez ·
Jesús Martín Vaquero *Editors*

Mathematical Methods for Engineering Applications

ICMASE 2021, Salamanca, Spain, July 1–2

Springer

Springer Proceedings in Mathematics & Statistics

Volume 384

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

More information about this series at https://link.springer.com/bookseries/10533

Fatih Yilmaz · Araceli Queiruga-Dios ·
María Jesús Santos Sánchez · Deolinda Rasteiro ·
Víctor Gayoso Martínez · Jesús Martín Vaquero
Editors

Mathematical Methods for Engineering Applications

ICMASE 2021, Salamanca, Spain, July 1–2

Springer

Editors
Fatih Yilmaz
Faculty of Arts and Sciences
Ankara Hacı Bayram Veli University
Ankara, Turkey

María Jesús Santos Sánchez
Faculty of Sciences
University of Salamanca
Salamanca, Spain

Víctor Gayoso Martínez
Institute of Physical and Information Technologies (ITEFI)
Spanish National Research Council (CSIC)
Madrid, Spain

Araceli Queiruga-Dios
Institute of Fundamental Physics and Mathematics
University of Salamanca
Salamanca, Spain

Deolinda Rasteiro
Polytechnic Institute of Coimbra (ISEC)
Coimbra, Portugal

Jesús Martín Vaquero
Institute of Fundamental Physics and Mathematics
University of Salamanca
Salamanca, Spain

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-030-96400-9 ISBN 978-3-030-96401-6 (eBook)
https://doi.org/10.1007/978-3-030-96401-6

Mathematics Subject Classification: 97Uxx, 08Axx, 62-xx, 00B25

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Committees

Honour Committee

Prof. Dr. José Miguel Mateos Roco, Vice Chancellor for Research and Transfer, University of Salamanca, Spain
Prof. Dr. Yusuf Tekin, Rector of Ankara Hacı Bayram Veli University, Turkey

Organizing Committee

Conference Chair:
Araceli Queiruga Dios, University of Salamanca, Spain

Organizing Chairs:
Fatih Yilmaz, Ankara Hacı Bayram Veli University, Turkey
Deolinda Rasteiro, Coimbra Engineering Institute-ISEC, Portugal
Emel Karaca, Ankara Hacı Bayram Veli University, Turkey
Jesús Martín-Vaquero, Universidad de Salamanca, Spain
María Jesús Santos Sanchez, Universidad de Salamanca, Spain
Melek Sofyalioğlu, Ankara Hacı Bayram Veli University, Turkey
Mücahit AKBIYIK, Beykent University, Turkey
Mustafa Özkan, Gazi University, Turkey
Seda Yamaç Akbıyık, İstanbul Gelişim University, Turkey
Víctor Gayoso Martínez, Spanish National Research Council, Spain

Scientific Committee

Agustín Martín Muñoz, Spanish National Research Council, Spain
Abdullah Alazemi, Kuwait University, Kuwait

Alexander Kushpel, Çankaya University, Turkey
Arihant Jain, Shanxi Normal University, P. R. China
Ascensión Hernández Encinas, University of Salamanca, Spain
Ayman Badawi, American University of Sharjah, UAE
Aynur Keskin Kaymakçı, Selçuk University, Turkey
Bipan Hazarika, Gauhati University, India
Carlos Martins da Fonseca, Kuwait College of Science and Technology, Kuwait
Cristina R. M. Caridade, Instituto Superior de Engenharia de Coimbra, Portugal
Daniela Richtarikova, Slovak University of Technology in Bratislava, Slovakia
Daniela Velichova, Slovak University of Technology, Slovakia
Dursun Taçcı, Gazi University, Turkey
Emília Bigotte, Instituto Superior de Engenharia de Coimbra, Portugal
Gheorghe Morosanu, Babes-Bolyai University, Romania
Ion Mierlus-Mazilu, Technical University of Civil Engineering Bucharest, Romania
Ji-TengJia, Xidian University, China
Kadir Kanat, Ankara Hacı Bayram Veli University, Turkey
Luis Hernández Encinas, Spanish National Research Council, Spain
Luis Hernández Álvarez, Spanish National Research Council, Spain
Marie Demlova, Czech Technical University in Prague, Czech Republic
María José Cáceres García, Universidad de Salamanca, Spain
Michael Carr, Technological University Dublin, Ireland
Milica Andjelic, Kuwait University, Kuwait
Mohammad Sal Moslehian, Ferdowsi University of Mashad, Iran
Mustafa Çalışkan, Gazi University, Turkey
Nenad P. Cakic, University of Belgrade, Serbia
Nursel Çetin, Ankara Hacı Bayram Veli University, Turkey
Praveen Agarwal, Anand International College of Engineering, India
Sapna Jain, Shanxi Normal University, P. R. China
Snezhana Gocheva-Ilieva, University of Plovdiv Paisii Hilendarski, Bulgaria
Süha Yılmaz, Dokuz Eylül Üniversitesi, Turkey
Tomohiro Sogabe, Nagoya University, Japan
Vildan Öztürk, Ankara Hacı Bayram Veli University, Turkey
Zhibin Du, South China Normal University, China

Preface

This Springer Nature includes the papers that have been accepted at International Conference on Mathematics and its Applications in Science and Engineering (ICMASE 2021) which was held in Salamanca, Spain, in July 2021. This conference was organized by the University of Salamanca (Spain) and the Ankara Hacı Bayram Veli University (Turkey).

The aim of this conference is to exchange ideas, discuss new developments in mathematics, promote collaborations and interact with professionals and researchers from all over the world in the following interesting topics: Functional Analysis, Approximation Theory, Real Analysis, Complex Analysis, Harmonic and non-Harmonic Analysis, Applied Analysis, Numerical Analysis, Geometry, Topology and Algebra, Modern Methods in Summability and Approximation, Operator Theory, Fixed Point Theory and Applications, Sequence Spaces and Matrix Transformation, Spectral Theory and Differential Operators, Boundary Value Problems, Ordinary and Partial Differential Equations, Discontinuous Differential Equations, Convex Analysis and its Applications, Optimization and its Application, Mathematics Education, Application on Variable Exponent Lebesgue Spaces, Applications on Differential Equations and Partial Differential Equations, Fourier Analysis, Wavelet and Harmonic Analysis Methods in Function Spaces, Applications on Computer Engineering, and Flow Dynamics.

After the peer-review process, the scientific committee selected 27 papers, from authors from 11 different countries.

The invited speakers of the ICMASE 2021 conference were Miguel Ángel González León from the University of Salamanca (Spain), Tin-Yau Tam from the University of Nevada (USA), and Carla M. A. Pinto from the Polytechnic Institute of Porto (Portugal).

Thanks to all ICMASE 2021 committee members.

Salamanca, Spain

Araceli Queiruga-Dios
Organizing Chair

Ankara, Turkey

Fatih Yilmaz
Conference Chair

Contents

Sylvester Sums on the Frobenius Set in Arithmetic Progression

Takao Komatsu

Abstract Let $a_1, a_2, \dots, a_k$ be positive integers with $\gcd(a_1, a_2, \dots, a_k) = 1$. The concept of the weighted sum $\sum_{n \in \mathrm{NR}} \lambda^n n$ is introduced in [1, 2], where $\mathrm{NR} = \mathrm{NR}(a_1, a_2, \dots, a_k)$ denotes the set of positive integers nonrepresentable in terms of $a_1, a_2, \dots, a_k$. When $\lambda = 1$, such a sum is often called Sylvester sum. The main purpose of this paper is to give explicit expressions of the Sylvester sum ($\lambda = 1$) and the weighed sum ($\lambda \ne 1$), where $a_1, a_2, \dots, a_k$ forms arithmetic progressions. As applications, various other cases are also considered, including weighted sums, almost arithmetic sequences, arithmetic sequences with an additional term, and geometric-like sequences. Several examples illustrate and confirm our results.

Keywords Frobenius problem · Weighted sums · Sylvester sums · Arithmetic sequences

MR Subject Classifications: Primary 11D07 · Secondary 05A15 · 05A17 · 05A19 · 11B68 · 11D04 · 11P81

1 Introduction

Given positive integers $a_1, \dots, a_k$ with $\gcd(a_1, \dots, a_k) = 1$, it is well-known that all sufficiently large n can be represented as a nonnegative integer combination of $a_1, \dots, a_k$. The *Frobenius Problem* is to determine the largest positive integer that is NOT representable as a nonnegative integer combination of given positive integers that are coprime (see [3] for general references). This number is denoted by $g(a_1, \dots, a_k)$ and often called Frobenius number. Let $n(a_1, \dots, a_k)$ be the number of positive integers with no nonnegative integer representation by $a_1, \dots, a_k$. It is sometimes called Sylvester number. According to Sylvester, for positive integers a and b with $\gcd(a, b) = 1$,

T. Komatsu (✉)
Department of Mathematical Sciences, School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China
e-mail: komatsu@zstu.edu.cn

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*, Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_1

$$g(a,b)=(a-1)(b-1)-1 \quad [4]\,,$$
$$n(a,b)=\frac{1}{2}(a-1)(b-1) \quad [5]\,.$$

One of other famous characters is on the so-called *Sylvester sums*

$$s(a_1,\dots,a_k):=\sum_{n\in\mathrm{NR}(a_1,\dots,a_k)} n$$

(see, e.g., [3, §5.5], [6] and references therein), where $\mathrm{NR}(a_1,\dots,a_k)$ denotes the set of positive integers without nonnegative integer representation by $a_1,\dots,a_k$. In addition, denote the set of positive integers with nonnegative integer representation by $a_1,\dots,a_k$ by $\mathrm{R}(a_1,\dots,a_k)$. For example,

$$\mathrm{R}(4,7,11)=\{0,4,7,8,11,12,14,15,16,18,19,20,21,\dots\} \quad \text{(infinite)}\,,$$
$$\mathrm{NR}(4,7,11)=\{1,2,3,5,6,9,10,13,17\} \quad \text{(finite)}\,,$$

so $g(4,7,11)=17$. Brown and Shiue [7] found the exact value for positive integers a and b with $\gcd(a,b)=1$,

$$s(a,b)=\frac{1}{12}(a-1)(b-1)(2ab-a-b-1)\,. \tag{1}$$

Rødseth [8] generalized Brown and Shiue's result by giving a closed form for the sum of powers.

When $k=2$, there exist beautiful closed forms for Frobenius numbers, Sylvester numbers and Sylvester sums, but when $k\ge 3$, exact determination of these numbers is difficult. The Frobenius number cannot be given by closed formulas of a certain type [9], the problem to determine $F(a_1,\dots,a_k)$ is NP-hard under Turing reduction (see, e.g., [3]). One analytic approach to the Frobenius number can be seen in [10, 11]. The problems may be also solved from the aspects of the sum of integer powers of values the gaps in numerical semigroups (e.g., [7, 12]). Recently in [1, 2], we consider more general sums called *(Sylvester) weighted sums*, defined by

$$s^{(\lambda)}(a_1,a_2,\dots,a_k):=\sum_{n\in\mathrm{NR}(a_1,a_2,\dots,a_k)} \lambda^n n\,.$$

When $\lambda=1$, $s(a_1,a_2,\dots,a_k)=s^{(1)}(a_1,a_2,\dots,a_k)$ is the usual Sylvester sum, though the obtained formulas are not included in the case of $\lambda\ne 1$. When $\lambda\ne 1$, $s^{(-1)}(a_1,a_2,\dots,a_k)$ is the so-called alternate sum, which has been studied in [6, 13]. Notice that the case of two variables is given by using the Apostol-Bernoulli numbers in [1].

In fact, by introducing the other numbers, it is possible to determine the functions $g(A)$, $n(A)$ and $s(A)$ for the set of positive integers $A := \{a_1, a_2, \ldots, a_k\}$ with $\gcd(a_1, a_2, \ldots, a_k) = 1$.

For each integer i with $1 \le i \le a_1 - 1$, there exists a least positive integer $m_i \equiv i \pmod{a_1}$ with $m_i \in \mathrm{R}(a_1, a_2, \ldots, a_k)$. For convenience, we set $m_0 = 0$. With the aid of such a congruence consideration modulo a_1, very useful results are established.

Lemma 1 *We have*

$$g(a_1, a_2, \ldots, a_k) = \left(\max_{1 \le i \le a_1 - 1} m_i\right) - a_1\,, \quad [14]$$

$$n(a_1, a_2, \ldots, a_k) = \frac{1}{a_1} \sum_{i=1}^{a_1 - 1} m_i - \frac{a_1 - 1}{2}\,, \quad [15]$$

$$s(a_1, a_2, \ldots, a_k) = \frac{1}{2a_1} \sum_{i=1}^{a_1 - 1} m_i^2 - \frac{1}{2} \sum_{i=1}^{a_1 - 1} m_i + \frac{a_1^2 - 1}{12}\,. \quad [16]$$

Note that the third formula appeared with a typo in [16], and it has been corrected in [17, 18].

When the number of variables is two, similarly to the case of Frobenius number, Sylvester number and Sylvester sums, the results for weighted sums may be explicitly given [1]. However, the results become complicated when the number of variables is bigger than or equal to three. Nevertheless, if the sequence $a_1, a_2, \ldots, a_k$ has some good regularities, the results are possible to be expressed explicitly. In this paper, we consider the weighted sum and (simple) sum of nonrepresentable numbers where $a_1, a_2, \ldots, a_k$ forms arithmetic progressions. Some more varieties case are also given, including almost arithmetic sequences, arithmetic sequences with an additional term, and geometric-like sequences.

2 Simple Sum

Let us begin from the simple sums.

Let a, d and k be positive integers with $\gcd(a, d) = 1$ and $2 \le k \le a$. Selmer [15] generalized Roberts' result ([19]; $h = 1$) and Brauer's result ([20]; $h = d = 1$) by giving a formula for almost arithmetic sequences. For a positive integer h,

$$n(a, ha + d, \ldots, ha + (k-1)d) = \left(\left\lfloor \frac{a-2}{k-1} \right\rfloor + h - 1\right) a + (a-1)d\,.$$

Let $a-1=q(k-1)+r$ with $0\le r<k-1$. Selmer [15] generalized Grant's result ([21]; $h=1$) by giving a formula for almost arithmetic sequences. For a positive integer h,

$$n(a, ha+d, \dots, ha+(k-1)d)=\frac{1}{2}\bigl((a-1)(hq+d+h-1)+hr(q+1)\bigr)\,.$$

Note that $q>0$ because $k\le a$. The sum of nonrepresentable numbers in arithmetic progression are given explicitly as follows.

Theorem 1 *Let a, d and k be positive integers with $\gcd(a,d)=1$ and $2\le k\le a$. Let q and r be nonnegative integers with $a-1=q(k-1)+r$ and $0\le r<k-1$. Then,*

$$\begin{aligned}
&s(a, a+d, \dots, a+(k-1)d)\\
&=\frac{1}{12q}\Bigl(2aq^3(a+2r-1)+q^2\bigl(ad(4a+4r-5)-d(2r-1)(r+1)+6ar\bigr)\\
&\quad -q\bigl(3dr(r+1)-(a-1)((a-1)(d^2-1)+ad^2)-2ar(3d+1)\bigr)-d(a-r-1)^2\Bigr)\,.
\end{aligned}$$

Substituting (2) and (3) in Lemma 2 below into the third formula in Lemma 1, we can get Theorem 1.

Lemma 2 *When $a_1=a$, $a_2=a+d$, $\dots$, $a_k=a+(k-1)d$ with $d>0$, $\gcd(a,d)=1$ and $k\le a_1$, we have*

$$\sum_{i=1}^{a-1}m_i=\frac{a}{2}\bigl((a-1)(q+d+1)+r(q+1)\bigr)\,, \tag{2}$$

$$\begin{aligned}
\sum_{i=1}^{a-1}m_i^2&=\frac{(q+1)\bigl((2q+1)(a-r-1)+6r(q+1)\bigr)}{6}a^2+\frac{(a-1)a(2a-1)}{6}d^2\\
&\quad +2ad(q+1)\left(\frac{(a-r-1)\bigl(q(4a-4r-1)-(a-r-1)\bigr)}{12q}+\frac{r(2a-r-1)}{2}\right)\,.
\end{aligned} \tag{3}$$

Proof Since the minimal residue system $\{m_i\}$ $(1 \le i \le a_1 - 1)$ is given by

$$\begin{array}{lllll}
a_2 & a_3 & \dots\dots & a_{k-1} & a_k \\
a_2 + a_k & a_3 + a_k & \dots\dots & a_{k-1} + a_k & 2a_k \\
\dots & \dots & \dots\dots & \dots & \dots \\
a_2 + (q-1)a_k & a_3 + (q-1)a_k & \dots\dots & a_{k-1} + (q-1)a_k & qa_k \\
a_2 + qa_k & a_3 + qa_k & \dots \ a_{r+1} + qa_k & &
\end{array} \tag{4}$$

([15, (3.8)]), the summation of all the elements is

$$\begin{aligned}
\sum_{i=1}^{a-1} m_i &= \bigl(1 + 2 + \cdots + (q(k-1) + r)\bigr)d \\
&\quad + \bigl((k-1)(1 + 2 + \cdots + q) + r(q+1)\bigr)a \\
&= \frac{(q+1)\bigl(q(k-1) + 2r\bigr)}{2}a + \frac{\bigl(q(k-1) + r\bigr)\bigl(q(k-1) + r + 1\bigr)}{2}d\,.
\end{aligned} \tag{5}$$

Since $q(k-1) + r = a - 1$, we have

$$\begin{aligned}
\sum_{i=1}^{a-1} m_i &= \frac{(q+1)(a-1+r)}{2}a + \frac{(a-1)a}{2}d \\
&= \frac{a}{2}\bigl((a-1)(q+d+1) + r(q+1)\bigr)\,.
\end{aligned}$$

Similarly, in order to obtain (3), we sum up all the elements

$$\begin{array}{lllll}
a_2^2 & a_3^2 & \dots\dots & a_{k-1}^2 & a_k^2 \\
(a_2 + a_k)^2 & (a_3 + a_k)^2 & \dots\dots & (a_{k-1} + a_k)^2 & (2a_k)^2 \\
\dots & \dots & \dots\dots & \dots & \dots \\
\bigl(a_2 + (q-1)a_k\bigr)^2 & \bigl(a_3 + (q-1)a_k\bigr)^2 & \dots\dots & \bigl(a_{k-1} + (q-1)a_k\bigr)^2 & (qa_k)^2 \\
(a_2 + qa_k)^2 & (a_3 + qa_k)^2 & \dots \ (a_{r+1} + qa_k)^2 & &
\end{array}$$

Then, we have

$$
\begin{aligned}
\sum_{i=1}^{a-1} m_i^2 &= \big(1^2+2^2+\cdots+(q(k-1)+r)^2\big)d^2\\
&\quad+\big((k-1)(1^2+2^2+\cdots+q^2)+r(q+1)^2\big)a^2\\
&\quad+2ad\Bigg(1+2+\cdots+(k-1)+2\big(k+(k+1)+\cdots+(2k-2)\big)\\
&\quad+3\big((2k-1)+(2k)+\cdots+(3k-3)\big)+\cdots\\
&\quad+q\big(((q-1)k-q+2)+\cdots+(qk-q)\big)\\
&\quad+(q+1)\big((qk-q+1)+\cdots+(qk-q+r)\big)\Bigg) \qquad (6)\\
&= \frac{\big(q(k-1)+r\big)\big(q(k-1)+r+1\big)\big(2q(k-1)+2r+1\big)}{6}d^2\\
&\quad+\left(\frac{(k-1)q(q+1)(2q+1)}{6}+r(q+1)^2\right)a^2\\
&\quad+2ad\left(\frac{k(k-1)q(q+1)}{4}+\frac{(k-1)^2(q-1)q(q+1)}{3}\right.\\
&\quad\left.+(q+1)\left((qk-q)r+\frac{r(r+1)}{2}\right)\right). \qquad (7)
\end{aligned}
$$

Since $q(k-1)+r=a-1$, we have

$$
\begin{aligned}
\sum_{i=1}^{a-1} m_i^2 &= \frac{(q+1)\big((2q+1)(a-r-1)+6r(q+1)\big)}{6}a^2+\frac{(a-1)a(2a-1)}{6}d^2\\
&\quad+2ad(q+1)\left(\frac{(a-r-1)\big(q(4a-4r-1)-(a-r-1)\big)}{12q}\right.\\
&\quad\left.+\frac{r(2a-r-1)}{2}\right).
\end{aligned}
$$

2.1 Examples

When $a=7$, $r=2$ and $k=3$, we get $q=3$ and $r=0$. Hence, by Theorem 1, we have $s(7,9,11,13)=165$. In fact, the sum of nonrepresentable numbers is

$$1+2+3+4+5+6+8+10+12+13+15+17+19+24+26=165\,.$$

When $a=6$, $r=5$ and $k=4$, we get $q=1$ and $r=2$. Hence, $s(6,11,16,21)=212$. The sum of nonrepresentable numbers is

$$1+2+3+4+5+7+8+9+10+13+14+15+19+20+25+26+31=212\,.$$

2.2 Almost Arithmetic Progressions

Theorem 1 can be extended to that of almost arithmetic progressions. Nonnegative integers q and r are similarly determined by $a-1=q(k-1)+r$ with $0\le r<k-1$.

Theorem 2 *For $a, d, h>0$ with $\gcd(a,d)=1$ and $k\le a$, we have*

$$
\begin{aligned}
& s(a, ha+d, \dots, ha+(k-1)d)\\
&= \frac{1}{12q}\Big(2h^2aq^3(a+2r-1)\\
&\quad + q^2\big(3h^2a(a+3r-1)+h(ad(4a+4r-5)-d(2r-1)(r+1)-3a(a+r-1))\big)\\
&\quad + q\big(h^2a(a+5r-1)+(a-1)(d-1)(2ad-a-d-1)\\
&\qquad + 3h((a+r-1)(d(a+r)-a)-2dr^2)\big)-hd(a-r-1)^2\Big)\,.
\end{aligned}
$$

This result is similarly proved from the following sums, which are analogous ones of those in Lemma 2.

Lemma 3 *When $a_1=a$, $a_2=ha+d$, $\dots$, $a_k=ha+(k-1)d$ with $d,h>0$, $\gcd(a,d)=1$ and $k\le a_1$, we have*

$$
\begin{aligned}
\sum_{i=1}^{a-1} m_i &= \frac{a}{2}\big((a-1)(h(q+1)+d)+hr(q+1)\big)\,,\\
\sum_{i=1}^{a-1} m_i^2 &= \frac{(q+1)\big((2q+1)(a-r-1)+6r(q+1)\big)}{6}h^2a^2+\frac{(a-1)a(2a-1)}{6}d^2\\
&\quad + 2had(q+1)\left(\frac{(a-r-1)\big(q(4a-4r-1)-(a-r-1)\big)}{12q}+\frac{r(2a-r-1)}{2}\right).
\end{aligned}
$$

3 Weighted Sums

In this section, we give a formula for the weighted sums

$$
s^{(\lambda)}(a, a+d, \dots, a+(k-1)d) := \sum_{n\in \mathrm{NR}(a,a+d,\dots,a+(k-1)d)} \lambda^n n\,.
$$

As in the previous sections, let $a_1=a$, $a_2=a+d$, $\dots$, $a_k=a+(k-1)d$ with $d>0$, $\gcd(a,d)=1$ and $k\le a_1$. Let $a-1=q(k-1)+r$ with $0\le r<k-1$.

Theorem 3 *For $\lambda \neq 0$ with $\lambda^d \neq 1$, $\lambda^a \neq 1$ and $\lambda^{a_k} \neq 1$, we have*

$$
\begin{aligned}
& s^{(\lambda)}(a, a+d, \dots, a+(k-1)d) \\
& = \frac{1}{\lambda^a - 1}\Bigg(\frac{\lambda^d(\lambda^{qa_k} - \lambda^{a_k})}{\lambda^{a_k} - 1}\left(\frac{a_k\lambda^{a_k} - a\lambda^a}{\lambda^d - 1} - \frac{d(\lambda^{a_k} - \lambda^a)}{(\lambda^d - 1)^2}\right) \\
& \quad + \frac{\lambda^d(\lambda^{a_k} - \lambda^a)}{\lambda^d - 1}\left(\frac{qa_k\lambda^{qa_k}}{\lambda^{a_k} - 1} - \frac{a_k\lambda^{a_k}(\lambda^{qa_k} - 1)}{(\lambda^{a_k} - 1)^2}\right) \\
& \quad + \lambda^{qa_k+d}\left(\frac{a_{r+1}\lambda^{a_{r+1}} - a\lambda^a}{\lambda^d - 1} - \frac{d(\lambda^{a_{r+1}} - \lambda^a)}{(\lambda^d - 1)^2}\right) + qa_k\lambda^{qa_k+d}\frac{\lambda^{a_{r+1}} - \lambda^a}{\lambda^d - 1}\Bigg) \\
& \quad - \frac{a\lambda^a}{(\lambda^a - 1)^2}\left(1 + \frac{\lambda^d(\lambda^{a_k} - \lambda^a)}{\lambda^d - 1}\frac{\lambda^{qa_k} - 1}{\lambda^{a_k} - 1} + \frac{\lambda^{qa_k+d}(\lambda^{a_{r+1}} - \lambda^a)}{\lambda^d - 1}\right) \\
& \quad + \frac{\lambda}{(\lambda - 1)^2}\,.
\end{aligned}
$$

The proof is based upon the following results.

Lemma 4

$$
\begin{aligned}
\sum_{i=1}^{a_1-1} m_i\lambda^{m_i} &= \frac{\lambda^d(\lambda^{qa_k} - \lambda^{a_k})}{\lambda^{a_k} - 1}\left(\frac{a_k\lambda^{a_k} - a\lambda^a}{\lambda^d - 1} - \frac{d(\lambda^{a_k} - \lambda^a)}{(\lambda^d - 1)^2}\right) \\
&\quad + \frac{\lambda^d(\lambda^{a_k} - \lambda^a)}{\lambda^d - 1}\left(\frac{qa_k\lambda^{qa_k}}{\lambda^{a_k} - 1} - \frac{a_k\lambda^{a_k}(\lambda^{qa_k} - 1)}{(\lambda^{a_k} - 1)^2}\right) \\
&\quad + \lambda^{qa_k+d}\left(\frac{a_{r+1}\lambda^{a_{r+1}} - a\lambda^a}{\lambda^d - 1} - \frac{d(\lambda^{a_{r+1}} - \lambda^a)}{(\lambda^d - 1)^2}\right) \\
&\quad + qa_k\lambda^{qa_k+d}\frac{\lambda^{a_{r+1}} - \lambda^a}{\lambda^d - 1}\,, \qquad (8)
\end{aligned}
$$

$$
\sum_{i=1}^{a_1-1} \lambda^{m_i} = \frac{\lambda^d(\lambda^{a_k} - \lambda^a)}{\lambda^d - 1}\frac{\lambda^{qa_k} - 1}{\lambda^{a_k} - 1} + \frac{\lambda^{qa_k+d}(\lambda^{a_{r+1}} - \lambda^a)}{\lambda^d - 1}\,. \qquad (9)
$$

Remark The last two terms in (8) and the last term in (9) are equal to 0 when $r = 0$.

Proof *(Proof of Lemma 4.)* (8) is obtained by summing up all the elements

$$
\begin{array}{llll}
a_2\lambda^{a_2} & \cdots\cdots & a_{k-1}\lambda^{a_{k-1}} & a_k\lambda^{a_k} \\
(a_2+a_k)\lambda^{a_2+a_k} & \cdots\cdots & (a_{k-1}+a_k)\lambda^{a_{k-1}+a_k} & (2a_k)\lambda^{2a_k} \\
\cdots & \cdots\cdots & \cdots & \cdots \\
\big(a_2+(q-1)a_k\big)\lambda^{a_2+(q-1)a_k} & \cdots\cdots & \big(a_{k-1}+(q-1)a_k\big)\lambda^{a_{k-1}+(q-1)a_k} & (qa_k)\lambda^{qa_k} \\
(a_2+qa_k)\lambda^{a_2+qa_k} & \cdots\ (a_{r+1}+qa_k)\lambda^{a_{r+1}+qa_k} & &
\end{array}
$$

Similarly, (9) is obtained by summing up all the elements

$$
\begin{array}{llll}
\lambda^{a_2} & \cdots\cdots & \lambda^{a_{k-1}} & \lambda^{a_k} \\
\lambda^{a_2+a_k} & \cdots\cdots & \lambda^{a_{k-1}+a_k} & \lambda^{2a_k} \\
\cdots & \cdots\cdots & \cdots & \cdots \\
\lambda^{a_2+(q-1)a_k} & \cdots\cdots & \lambda^{a_{k-1}+(q-1)a_k} & \lambda^{qa_k} \\
\lambda^{a_2+qa_k} & \cdots & \lambda^{a_{r+1}+qa_k} &
\end{array}
$$

We also need the following formula in [2].

Lemma 5 *If $\lambda \neq 0$ and $\lambda^{a_1} \neq 1$, then*

$$
\begin{aligned}
& s^{(\lambda)}(a_1, a_2, \ldots, a_k) \\
& = \frac{1}{\lambda^{a_1}-1}\sum_{i=0}^{a_1-1} m_i \lambda^{m_i} - \frac{a_1\lambda^{a_1}}{(\lambda^{a_1}-1)^2}\sum_{i=0}^{a_1-1}\lambda^{m_i} + \frac{\lambda}{(\lambda-1)^2}\,.
\end{aligned}
$$

Proof *(Proof of Theorem 3.)* Substituting (8) and (9) in Lemma 4 into the formula in Lemma 5, we get the desired result. Notice that we do need an additional quantity $\lambda^{m_0} = \lambda^0 = 1$ in the second term in Lemma 5.

3.1 Examples

When $a = 7$, $r = 2$ and $k = 3$, we get $q = 3$ and $r = 0$. If $\lambda = 2$ and $d = 2$, then by Theorem 3, we have $s^{(2)}(7, 9, 11) = 2160333442$. In fact, the sum of nonrepresentable numbers is

$$
\begin{aligned}
& 2\cdot 1 + 2^2\cdot 2 + 2^3\cdot 3 + 2^4\cdot 4 + 2^5\cdot 5 + 2^6\cdot 6 + 2^8\cdot 8 + 2^{10}\cdot 10 \\
& \quad + 2^{12}\cdot 12 + 2^{13}\cdot 13 + 2^{15}\cdot 15 + 2^{17}\cdot 17 + 2^{19}\cdot 19 + 2^{24}\cdot 24 + 2^{26}\cdot 26 \\
& = 2160333442\,.
\end{aligned}
$$

When $a = 6$, $r = 5$ and $k = 4$, we get $q = 1$ and $r = 2$. If $\lambda = \sqrt{-1}$ and $d = 5$, then $s^{(\sqrt{-1})}(6, 11, 16, 21) = -20 - 22\sqrt{-1}$.

4 Arithmetic Sequences with an Additional Term

Consider the case

$$a_1 = a,\ a_2 = a + d,\ a_3 = a + 2d,\ \ldots,\ a_k = a + (k-1)d,\ a_{k+1} = a + Kd\,,$$

where $\gcd(a, d) = 1$, $K > k$ and $a \ge k$. Put

$$\begin{aligned} K - 1 &= q(k-1) + r, \quad 0 \le r < k-1\,, \\ a &= \alpha K + \beta, \quad 0 \le \beta < K\,. \end{aligned} \tag{10}$$

In [15, (3.16)], as $d = 1$, it is shown that

$$\begin{aligned} &n(a, a+1, a+2, \ldots, a+k-1, a+K) \\ &= \frac{\alpha\big(a + (q+1)(r-1) + qK + \beta +\big)}{2} + \frac{(\gamma+1)(\beta+\delta-1)}{2}\,. \end{aligned}$$

Some more special cases of the number of nonrepresentable numbers are given in [15].

The minimal residue system $1, 2, \ldots, a-1 \pmod a$, where all residues appear once and only once, can be constructed as follows. The case $d = 1$ is illustrated in [15], but we explain here again as the general d case.

The first line of this minimal residue system is the same as the whole numbers in (4). There are totally $(k-1)q + r = K - 1$ elements, which consist the residue system $\{d, 2d, \ldots, (K-1)d\} \pmod a$. The second line is of a_{k+1}, and a_{k+1} plus each number in (4), representing the residue system $\{Kd, (K+1)d, \ldots, (2K-1)d\} \pmod a$. Hence, there are totally K elements in the second line. Similarly, the jth line $(1 \le j \le \alpha - 1)$ is of $(j-1)a_{k+1}$, and $(j-1)a_{k+1}$ plus each number in (4), representing the residue system $\{(j-1)Kd, ((j-1)K+1)d, \ldots, (jK-1)d\} \pmod a$. Hence, there are totally K elements in the jth line $(2 \le j \le \alpha - 1)$. The αth line ends with the element

$$t' = \begin{cases} a_{r+1} + qa_k + (\alpha-1)a_{k+1} = (q+\alpha+1)a - (\beta+1)d & \text{if } r > 0; \\ qa_k + (\alpha-1)a_{k+1} = (q+\alpha)a - (\beta+1)d & \text{if } r = 0\,. \end{cases}$$

If $\beta = 0$, then we have already gotten the minimal residue system. Otherwise, put

$$\beta - 1 = \gamma(k-1) + \delta, \quad 0 \le \delta < k-1\,. \tag{11}$$

The final line is of β elements and ends with the element

$$t'' = \begin{cases} a_{\delta+1} + \gamma a_k + \alpha a_{k+1} = (\gamma+\alpha+2)a - d & \text{if } \delta > 0; \\ \gamma a_k + \alpha a_{k+1} = (\gamma+\alpha+1)a - d & \text{if } \delta = 0\,. \end{cases}$$

In (5), set $q(k-1) + r = K - 1$ instead of $q(k-1) + r = a - 1$. Then we have

$$
\begin{aligned}
S_1 &= S_1(a,d,q,r)\\
&:= \frac{(q+1)(K-1+r)}{2}a + \frac{(K-1)K}{2}d\,.
\end{aligned}
$$

In (6), set $q(k-1)+r = K-1$ instead of $q(k-1)+r = a-1$. Then we have

$$
\begin{aligned}
S_2 &= S_2(a,d,q,r)\\
&:= \frac{(q+1)\big((2q+1)(K-r-1)+6r(q+1)\big)}{6}a^2 + \frac{(K-1)K(2K-1)}{6}d^2\\
&\quad + 2ad(q+1)\left(\frac{(K-r-1)\big(q(4K-4r-1)-(K-r-1)\big)}{12q} + \frac{r(2K-r-1)}{2}\right).
\end{aligned}
$$

When $\beta = 0$, the whole summation of the least elements modulo $i \pmod a$ $(1 \le i \le a-1)$ is equal to

$$
\begin{aligned}
\sum_{i=1}^{a-1} m_i &= \sum_{j=1}^{\alpha}\big((j-1)a_{k+1}K + S_1\big)\\
&= \frac{\alpha(\alpha-1)K(a+Kd)}{2} + \alpha S_1\,.
\end{aligned}
$$

(When $\beta > 0$, we need more additional elements from the $(\alpha+1)$-st line, whose sum is denoted by T_1.) The whole square summation is equal to

$$
\begin{aligned}
\sum_{i=1}^{a-1} m_i^2 &= \sum_{j=1}^{\alpha}\big((j-1)^2 a_{k+1}^2 K + 2(j-1)a_{k+1}S_1 + S_2\big)\\
&= \frac{\alpha(\alpha-1)(2\alpha-1)(a+Kd)^2 K}{6} + \alpha(\alpha-1)(a+Kd)S_1 + \alpha S_2\,.
\end{aligned}
$$

(When $\beta > 0$, we need more additional elements from the $(\alpha+1)$-st line, whose sum is denoted by T_2.) Substituting them into the third formula in Lemma 1, by $\alpha = aK$, we get

$$
\begin{aligned}
&s(a, a+d, a+2d, \dots, a+(k-1)d, a+Kd)\\
&= \frac{1}{2a}\left(\frac{\alpha(\alpha-1)(2\alpha-1)(a+Kd)^2 K}{6} + \alpha(\alpha-1)(a+Kd)S_1 + \alpha S_2\right)\\
&\quad - \frac{1}{2}\left(\frac{\alpha(\alpha-1)K(a+Kd)}{2} + \alpha S_1\right) + \frac{a^2-1}{12}\\
&= \frac{\alpha(\alpha-1)(a+Kd)K\bigl(2(\alpha-2)a+(2\alpha-1)Kd\bigr)}{12a}\\
&\quad + \frac{\alpha\bigl((\alpha-2)a+(\alpha-1)Kd\bigr)}{2a}S_1 + \frac{\alpha S_2}{2a} + \frac{a^2-1}{12}\\
&= \frac{(a-K)(a+Kd)\bigl(2a^2+2aK(d-2)-K^2d\bigr)}{12\,K^2} + \frac{a^2+aK(d-2)-K^2d}{2\,K^2}S_1\\
&\quad + \frac{S_2}{2\,K} + \frac{a^2-1}{12}\,.
\end{aligned}
$$

When $\beta > 0$, from (11) the sum of additional terms

$$
\begin{array}{llll}
\alpha a_{k+1} & & & \\
\alpha a_{k+1} + a_2 & \alpha a_{k+1} + a_3 & \dots & \alpha a_{k+1} + a_k \\
\alpha a_{k+1} + a_k + a_2 & \alpha a_{k+1} + a_k + a_3 & \dots & \alpha a_{k+1} + 2a_k \\
\dots & \dots & \dots & \dots \\
\alpha a_{k+1} + (\gamma-1)a_k + a_2 & \alpha a_{k+1} + (\gamma-1)a_k + a_3 & \dots & \alpha a_{k+1} + \gamma a_k \\
\alpha a_{k+1} + \gamma a_k + a_2 & \dots \quad \alpha a_{k+1} + \gamma a_k + a_{\delta+1} & &
\end{array}
$$

is given by

$$
\begin{aligned}
&T_1 = T_1(a, d, K, \alpha, \beta, \delta, \gamma)\\
&:= \alpha\beta a_{k+1} + \gamma(a_2 + a_3 + \cdots + a_k)\\
&\quad + (k-1)a_k\bigl(1+2+\cdots+(\gamma-1)\bigr) + \gamma\delta a_k + (a_2 + \cdots + a_{\delta+1})\\
&= \alpha\beta(a+Kd) + \gamma\left((k-1)a + \frac{(k-1)k}{2}d\right) + (k-1)\bigl(a+(k-1)d\bigr)\frac{\gamma(\gamma-1)}{2}\\
&\quad + \gamma\delta\bigl(a+(k-1)d\bigr) + \delta a + \frac{\delta(\delta+1)}{2}d\\
&= \left(\alpha\beta + \frac{(\beta+\delta-1)(\gamma+1)}{2}\right)a\\
&\quad + \left(\alpha\beta K + \frac{(\beta-\delta-1)(\beta-\delta)}{2} + \frac{\delta(2\beta-\delta-1)}{2}\right)d\,.
\end{aligned}
$$

The sum of the square of additional terms is given by

$$T_2 = T_2(a, d, K, \beta, \delta, \gamma) := \beta(\alpha a_{k+1})^2 + 2\alpha a_{k+1}(T_1 - \alpha\beta a_{k+1}) + T_3\,,$$

where

$$\begin{aligned}
T_3 &= \gamma(a_2^2 + a_3^2 + \cdots + a_k^2)\\
&\quad + 2a_k\bigl(1 + 2 + \cdots + (\gamma-1)\bigr)(a_2 + a_3 + \cdots + a_k)\\
&\quad + (k-1)a_k^2\bigl(1^2 + 2^2 + \cdots + (\gamma-1)^2\bigr)\\
&\quad + \delta\gamma^2 a_k^2 + 2\gamma a_k(a_2 + a_3 + \cdots + a_{\delta+1}) + (a_2^2 + a_3^2 + \cdots + a_{\delta+1}^2)\\
&= \gamma\left((k-1)a^2 + (k-1)kad + \frac{(k-1)k(2k-1)}{6}d^2\right)\\
&\quad + \bigl(a + (k-1)d\bigr)(\gamma-1)\gamma\left((k-1)a + \frac{(k-1)k}{2}d\right)\\
&\quad + (k-1)\bigl(a + (k-1)d\bigr)^2\frac{(\gamma-1)\gamma(2\gamma-1)}{6}\\
&\quad + \delta\gamma^2\bigl(a + (k-1)d\bigr)^2 + 2\gamma\bigl(a + (k-1)d\bigr)\left(\delta a + \frac{\delta(\delta+1)}{2}d\right)\\
&\quad + \delta a^2 + \delta(\delta+1)ad + \frac{\delta(\delta+1)(2\delta+1)}{6}d^2\\
&= \left(\frac{(\beta-\delta-1)(\gamma+1)(2\gamma+1)}{6} + \delta(\gamma+1)^2\right)a^2\\
&\quad + \left(\frac{(\beta-\delta-1)\bigl(4(\beta-\delta)-k\bigr)}{6} + \delta(2\beta-\delta-1)\right)(\gamma+1)ad\\
&\quad + \frac{(\beta-1)\beta(2\beta-1)}{6}d^2\,.
\end{aligned}$$

When $\beta > 0$, the whole summation of the least elements modulo $i \pmod a$ $(1 \le i \le a-1)$ is equal to

$$\begin{aligned}
\sum_{i=1}^{a-1} m_i &= \sum_{j=1}^{\alpha}\bigl((j-1)a_{k+1}K + S_1\bigr) + T_1\\
&= \frac{\alpha(\alpha-1)K(a+Kd)}{2} + \alpha S_1 + T_1\,.
\end{aligned}$$

The whole square summation is equal to

$$\begin{aligned}
\sum_{i=1}^{a-1} m_i^2 &= \sum_{j=1}^{\alpha}\bigl((j-1)^2 a_{k+1}^2\,K + 2(j-1)a_{k+1}S_1 + S_2\bigr) + T_2\\
&= \frac{\alpha(\alpha-1)(2\alpha-1)(a+Kd)^2\,K}{6} + \alpha(\alpha-1)(a+Kd)S_1 + \alpha S_2 + T_2\,.
\end{aligned}$$

Substituting them into the third formula in Lemma 1, we get

$$
\begin{aligned}
&s(a, a+d, a+2d, \dots, a+(k-1)d, a+Kd)\\
&= \frac{1}{2a}\left(\frac{\alpha(\alpha-1)(2\alpha-1)(a+Kd)^2 K}{6} + \alpha(\alpha-1)(a+Kd)S_1 + \alpha S_2 + T_2\right)\\
&\quad - \frac{1}{2}\left(\frac{\alpha(\alpha-1)K(a+Kd)}{2} + \alpha S_1 + T_1\right) + \frac{a^2-1}{12}\\
&= \frac{\alpha(\alpha-1)(a+Kd)K\big(2(\alpha-2)a+(2\alpha-1)Kd\big)}{12a}\\
&\quad + \frac{\alpha\big((\alpha-2)a+(\alpha-1)Kd\big)}{2a}S_1 + \frac{\alpha S_2}{2a} - \frac{T_1}{2} + \frac{T_2}{2a} + \frac{a^2-1}{12}\,.
\end{aligned}
$$

Theorem 4 *Assume that* $\gcd(a,d)=1$, $K>k$ *and* $a\ge k$. *Integers* q, r, α, β, γ *and* δ *are decided as in (10) and (11). Then we have*

$$
\begin{aligned}
&s(a, a+d, a+2d, \dots, a+(k-1)d, a+Kd)\\
&= \frac{\alpha(\alpha-1)(a+Kd)K\big(2(\alpha-2)a+(2\alpha-1)Kd\big)}{12a}\\
&\quad + \frac{\alpha\big((\alpha-2)a+(\alpha-1)Kd\big)}{2a}S_1 + \frac{\alpha S_2}{2a} - \frac{T_1}{2} + \frac{T_2}{2a} + \frac{a^2-1}{12}\,,
\end{aligned}
$$

where

$$
\begin{aligned}
S_1 &= \frac{(q+1)(K-1+r)}{2}a + \frac{(K-1)K}{2}d\,,\\
S_2 &= \frac{(q+1)\big((2q+1)(K-r-1)+6r(q+1)\big)}{6}a^2 + \frac{(K-1)K(2K-1)}{6}d^2\\
&\quad + 2ad(q+1)\left(\frac{(K-r-1)\big(q(4K-4r-1)-(K-r-1)\big)}{12q}\right.\\
&\quad \left. + \frac{r(2K-r-1)}{2}\right).
\end{aligned}
$$

When $a \mid K$, $T_1 = T_2 = 0$. *When* $a \nmid K$,

$$
\begin{aligned}
T_1 &= T_1(a, d, K, \alpha, \beta, \delta, \gamma)\\
&= \left(\alpha\beta + \frac{(\beta+\delta-1)(\gamma+1)}{2}\right)a\\
&\quad + \left(\alpha\beta K + \frac{(\beta-\delta-1)(\beta-\delta)}{2} + \frac{\delta(2\beta-\delta-1)}{2}\right)d
\end{aligned}
$$

and

$$T_2 = T_2(a, d, K, \beta, \delta, \gamma) := \beta(\alpha a_{k+1})^2 + 2\alpha a_{k+1}(T_1 - \alpha\beta a_{k+1}) + T_3\,,$$

where

$$\begin{aligned}T_3 &= \left(\frac{(\beta-\delta-1)(\gamma+1)(2\gamma+1)}{6} + \delta(\gamma+1)^2\right)a^2\\ &\quad + \left(\frac{(\beta-\delta-1)\big(4(\beta-\delta)-k\big)}{6} + \delta(2\beta-\delta-1)\right)(\gamma+1)ad\\ &\quad + \frac{(\beta-1)\beta(2\beta-1)}{6}d^2\,.\end{aligned}$$

Remark When $K = a$ and $\beta = 0$, by $\sum_{i=1}^{a-1} m_i = S_1$ and $\sum_{i=1}^{a-1} m_i^2 = S_2$, Theorem 4 is reduced to Theorem 1.

If $\alpha \nmid K$, then by $a = \alpha K + \beta$, we have

$$\begin{aligned}&s(a, a+d, a+2d, \dots, a+(k-1)d, a+Kd)\\ &= \frac{(a-K)(a+Kd)\big(2a^2+2aK(d-2)-K^2d\big)}{12\,K^2} + \frac{a^2+aK(d-2)-K^2d}{2\,K^2}S_1\\ &\quad + \frac{S_2}{2\,K} + \frac{a^2-1}{12}\\ &= \frac{(a-\beta)(a-\beta-K)(a+Kd)\big(2(a-\beta-2\,K)a+(2a-2\beta-K)d\big)}{12aK^2}\\ &\quad + \frac{(a-\beta)\big((a-\beta-2\,K)a+(a-\beta-K)d\big)}{2aK^2}S_1\\ &\quad + \frac{a-\beta}{2aK}S_2 - \frac{T_1}{2} + \frac{T_2}{2a} + \frac{a^2-1}{12}\,.\end{aligned}$$

As a special case,

$$g(a, a+1, a+2, a+4) = (a+1)\left\lfloor\frac{a}{4}\right\rfloor + \left\lfloor\frac{a+1}{4}\right\rfloor + 2\left\lfloor\frac{a+2}{4}\right\rfloor - 1\,,\quad [22]$$

$$n(a, a+1, a+2, a+4) = \left\lfloor\frac{a(a+4)}{8}\right\rfloor \quad [15]$$

are found, where $\lfloor x \rfloor$ denotes the integer part of a real x. We can give the correspondence summations.

Corollary 1

$$s(a,a+1,a+2,a+4) = \begin{cases} \frac{1}{96}(a^4+8a^3+26a^2+16a) & \textit{if } a \equiv 0 \pmod 4;\\ \frac{1}{96}(a^4+8a^3+11a^2-38a+18) & \textit{if } a \equiv 1 \pmod 4;\\ \frac{1}{96}(a^4+8a^3+14a^2-32a+24) & \textit{if } a \equiv 2 \pmod 4;\\ \frac{1}{96}(a^4+8a^3+11a^2-50a+42) & \textit{if } a \equiv 3 \pmod 4. \end{cases}$$

Proof Here, $d=1, k=3, K=4, q=r=1$ and $\alpha = \lfloor a/4 \rfloor$. When $a \equiv 0 \pmod 4$, $\beta = 0$. When $a \equiv 1 \pmod 4$, $\beta = 1$, $\gamma = \delta = 0$. When $a \equiv 2 \pmod 4$, $\beta = 2$, $\gamma = 0$ and $\delta = 1$. When $a \equiv 3 \pmod 4$, $\beta = 3$, $\gamma = 1$ and $\delta = 0$. The results follow from Theorem 4.

In [22], some more special cases are found:

$$\begin{aligned} &g(a,a+1,a+2,a+5)\\ &= a\left\lfloor\frac{a+1}{5}\right\rfloor + \left\lfloor\frac{a}{5}\right\rfloor + \left\lfloor\frac{a+1}{5}\right\rfloor + \left\lfloor\frac{a+2}{5}\right\rfloor + 2\left\lfloor\frac{a+3}{5}\right\rfloor - 1,\\ &g(a,a+1,a+2,a+6)\\ &= a\left\lfloor\frac{a}{6}\right\rfloor + 2\left\lfloor\frac{a}{6}\right\rfloor + 2\left\lfloor\frac{a+1}{6}\right\rfloor + 5\left\lfloor\frac{a+2}{6}\right\rfloor + \left\lfloor\frac{a+3}{6}\right\rfloor + \left\lfloor\frac{a+4}{6}\right\rfloor + \left\lfloor\frac{a+5}{6}\right\rfloor - 1. \end{aligned}$$

We can similarly derive the correspondence result by Theorem 4.

Corollary 2

$$\begin{aligned} &s(a,a+1,a+2,a+5)\\ &= \begin{cases} \frac{1}{150}(a^4+13a^3+65a^2+35a) & \textit{if } a \equiv 0 \pmod 5;\\ \frac{1}{150}(a^4+13a^3+41a^2-85a+30) & \textit{if } a \equiv 1 \pmod 5;\\ \frac{1}{150}(a^4+13a^3+41a^2-97a+60) & \textit{if } a \equiv 2 \pmod 5;\\ \frac{1}{150}(a^4+13a^3+35a^2-139a+120) & \textit{if } a \equiv 3 \pmod 5;\\ \frac{1}{150}(a^4+13a^3+53a^2-19a+90) & \textit{if } a \equiv 4 \pmod 5. \end{cases} \end{aligned}$$

Corollary 3

$$\begin{aligned} &s(a,a+1,a+2,a+6)\\ &= \begin{cases} \frac{1}{216}(a^4+21a^3+189a^2+126a) & \textit{if } a \equiv 0 \pmod 6;\\ \frac{1}{216}(a^4+21a^3+150a^2-169a-3) & \textit{if } a \equiv 1 \pmod 6;\\ \frac{1}{216}(a^4+21a^3+141a^2-290a+48) & \textit{if } a \equiv 2 \pmod 6;\\ \frac{1}{216}(a^4+21a^3+126a^2-441a+189) & \textit{if } a \equiv 3 \pmod 6;\\ \frac{1}{216}(a^4+21a^3+141a^2-322a+240) & \textit{if } a \equiv 4 \pmod 6;\\ \frac{1}{216}(a^4+21a^3+150a^2-281a+237) & \textit{if } a \equiv 5 \pmod 6. \end{cases} \end{aligned}$$

4.1 Examples

Consider the sequence $14, 17, 20, 23, 38$. Then, $a = 14$, $d = 3$ $k = 4$, $K = 8$, $q = 2$, $r = 1$, $\alpha = 1$, $\beta = 6$, $\gamma = 1$ and $\delta = 2$. By Theorem 4, we have $s(14, 17, 20, 23, 38) = 953$. In fact, the sum of nonrepresentable numbers is

$$\begin{aligned}
&1+2+3+4+5+6+7+8+9+10+11+12+13+15+16+18\\
&+19+21+22+24+25+26+27+29+30+32+33+35+36+39\\
&+41+44+47+49+50+53+64+67\\
&=953\,.
\end{aligned}$$

By Corollary 1, when $a = 8, 9, 10, 11$, we have

$$\begin{aligned}
s(8, 9, 10, 12) &= 104\,,\\
s(9, 10, 11, 13) &= 135\,,\\
s(10, 11, 12, 14) &= 199\,,\\
s(11, 12, 13, 15) &= 272\,.
\end{aligned}$$

For example, the first sum of nonrepresentable numbers is

$$1+2+3+4+5+6+7+11+13+14+15+23 = 104\,.$$

5 Geometric-Like Sequence

Consider the case

$$a_1 = a,\ a_2 = a+1,\ a_3 = a+2,\ a_4 = a+2^2,\ \dots,\ a_{k+2} = a+2^k \quad (k \ge 2)\,.$$

Put $a = 2^k q + r$ with $0 \le r < 2^k$. The first line is the sequence $1, 2, \dots, 2^k - 1$ (mod a), that is

$$a_2 = a+1,\ a_3 = a+2,\ a_2+a_3 = 2a+3,\ a_4 = a+4,\ \dots,\\
a_2+a_3+\cdots+a_{k+1} = ka+2^k-1\,. \quad (12)$$

The second line is the sequence

$$a_{k+2},\ a_{k+2}+a_2,\ a_{k+2}+a_3,\ a_{k+2}+a_2+a_3,\ a_{k+2}+a_4,\ \dots,\ a_{k+2}+a_2+a_3+\cdots+a_{k+1}\,.$$

Similarly, the qth line is the sequence

$$(q-1)a_{k+2},\ (q-1)a_{k+2}+a_2,\ (q-1)a_{k+2}+a_3,\ (q-1)a_{k+2}+a_2+a_3,\\ (q-1)a_{k+2}+a_4,\ \dots,\ (q-1)a_{k+2}+a_2+a_3+\dots+a_{k+1}\,.$$

If $r=0$, then the line is finished. If $r>0$, then the $(q+1)$th line consists from r terms beginning from

$$qa_{k+2},\ qa_{k+2}+a_2,\ qa_{k+2}+a_3,\ \cdots\,.$$

In order to find the sum of the elements, consider the exact term which is congruent to n modulo a $(1\le n\le 2^k-1)$ in the sequence (12). If $s_1(n)$ is the exponent of 2 in the canonical representation of $n!$ and $s_2(n)$ is the number of ones in the binary representation of n, then

$$\begin{aligned} s_2(n)&=n-s_1(n)\\ &=n-\sum_{i=1}^{\infty}\left\lfloor\frac{n}{2^i}\right\rfloor=n-\sum_{i=1}^{\lfloor\log_2 n\rfloor}\left\lfloor\frac{n}{2^i}\right\rfloor \end{aligned}$$

(see, e.g., [23, Theorem 3.16]). Hence, the exact expression of the 2^k-1 terms in the first line is given by

$$a+1=s_2(1)a+1,\ a+2=s_2(2)a+2,\ 2a+3=s_2(3)a+3,\ \dots,\\ s_2(j)a+j,\ \dots,\ s_2(2^k-1)a+2^k-1\,.$$

Since

$$\left\lfloor\frac{2^i}{2^i}\right\rfloor=\dots=\left\lfloor\frac{2\cdot2^i-1}{2^i}\right\rfloor=1,\quad \left\lfloor\frac{2\cdot2^i}{2^i}\right\rfloor=\dots=\left\lfloor\frac{3\cdot2^i-1}{2^i}\right\rfloor=2\,,\\ \left\lfloor\frac{3\cdot2^i}{2^i}\right\rfloor=\dots=\left\lfloor\frac{4\cdot2^i-1}{2^i}\right\rfloor=3,\quad \dots,\left\lfloor\frac{(2^{k-i}-1)\cdot2^i}{2^i}\right\rfloor=\dots=\left\lfloor\frac{2^k-1}{2^i}\right\rfloor=2^{k-i}\,,$$

the sum of all the elements in the first line is equal to

$$\begin{aligned} &a_2+a_3+(a_2+a_3)+\dots+(a_2+a_3+\dots+a_{k+1})\\ &=\left(\sum_{j=1}^{2^k-1}s_2(j)\right)a+\sum_{j=1}^{2^k-1}j\\ &=\left(\sum_{j=1}^{2^k-1}j-\sum_{j=1}^{2^k-1}\sum_{i=1}^{\lfloor\log_2 j\rfloor}\left\lfloor\frac{j}{2^i}\right\rfloor\right)a+2^{k-1}(2^k-1)\\ &=2^{k-1}ka+2^{k-1}(2^k-1)\,. \end{aligned}$$

In general, the sum of all the elements in the jth line $(1 \le j \le q)$ is

$$\begin{aligned}&(j-1)a_{k+2}2^k + 2^{k-1}ka + 2^{k-1}(2^k-1)\\&= (j-1)(a+2^k)2^k + 2^{k-1}ka + 2^{k-1}(2^k-1)\,.\end{aligned}$$

Hence, if $r = 0$, then

$$\begin{aligned}\sum_{i=1}^{a-1} m_i &= \sum_{j=1}^{q} (j-1)(a+2^k)2^k + q2^{k-1}ka + q2^{k-1}(2^k-1)\\&= 2^{k-1}q\big((q+k-1)a + (2^kq-1)\big)\,.\end{aligned}$$

Next, by

$$\begin{aligned}\sum_{j=1}^{2^k-1} js_1(j) &= 2^{k-1}(2^k-1)(2^{k+2}-3k-5)\,,\\\sum_{j=1}^{2^k-1}\big(s_1(j)\big)^2 &= \frac{2^{k-2}}{3}\big(2^{k+1}(2^{k+1}-3k-6) + (3k^2+9k+8)\big)\,,\end{aligned}$$

the sum of all the square of elements in the first line is

$$\begin{aligned}&a_2^2 + a_3^2 + (a_2+a_3)^2 + \cdots + (a_2+a_3+\cdots+a_{k+1})^2\\&= 2^{k-1}\left(\frac{k(k+1)}{2}a^2 + (2^k-1)(k+1)a + \frac{(2^k-1)(2^{k+1}-1)}{3}\right)\,.\end{aligned}$$

The sum of all the square of elements in the jth line $(1 \le j \le q)$ is

$$\begin{aligned}&\big((j-1)a_{k+2}+a_2\big)^2 + \big((j-1)a_{k+2}+a_3\big)^2 + \cdots + \big((j-1)a_{k+2}+a_2+a_3\big)^2\\&\qquad + \cdots + \big((j-1)a_{k+2}+a_2+a_3+\cdots+a_{k+1}\big)^2\\&= (j-1)^2 2^k(a+2^k)^2 + 2(j-1)(a+2^k)\times(\text{sum of the first line})\\&\qquad + (\text{sum of square of the first line})\\&= (j-1)^2 2^k(a+2^k)^2\\&\quad + 2(j-1)(a+2^k)\big(2^{k-1}ka + 2^{k-1}(2^k-1)\big)\\&\quad + 2^{k-1}\left(\frac{k(k+1)}{2}a^2 + (2^k-1)(k+1)a + \frac{(2^k-1)(2^{k+1}-1)}{3}\right)\,.\end{aligned}$$

Hence, if $r = 0$, then

$$
\begin{aligned}
\sum_{i=1}^{a-1} m_i^2 &= \sum_{j=1}^{q}(j-1)^2 2^k (a+2^k)^2 + \sum_{j=1}^{q} 2(j-1)(a+2^k)\bigl(2^{k-1}ka + 2^{k-1}(2^k-1)\bigr)\\
&\quad + q2^{k-1}\left(\frac{k(k+1)}{2}a^2 + (2^k-1)(k+1)a + \frac{(2^k-1)(2^{k+1}-1)}{3}\right)\\
&= \frac{2^{k-2}q}{3}\biggl(\bigl(3k(2q+k-1) + 2(q-1)(2q-1)\bigr)a^2\\
&\quad + 2\bigl(2^k(4q^2+3kq-3q+2) - 3(q+k)\bigr)a + 2(2^{2k+1}q^2 - 3\cdot 2^k q + 1)\biggr).
\end{aligned}
$$

Next, assume that $r > 0$. The exact expression of the r terms in the $(q+1)$th line is given by

$$
\begin{aligned}
&qa_{k+2},\ qa_{k+2} + (a+1),\ qa_{k+2} + (a+2),\ qa_{k+2} + (2a+3),\ \dots\\
&\qquad qa_{k+2} + \bigl(s_2(j)a + j\bigr),\ \dots,\ qa_{k+2} + \bigl(s_2(r-1)a + r - 1\bigr).
\end{aligned}
$$

Thus, by $a - r = 2^k q$, the sum of additional terms from the $(q+1)$th line is

$$
\begin{aligned}
\mathfrak{T}_1 &:= rqa_{k+2} + \left(\sum_{j=1}^{r-1} s_2(j)\right)a + \sum_{j=1}^{r-1} j\\
&= \left(\frac{r(r+2q+1)}{2} - \sum_{j=1}^{r-1} s_1(j)\right)a - \frac{r(r+1)}{2}.
\end{aligned}
$$

The sum of square of additional terms from the $(q+1)$th line is

$$
\begin{aligned}
\mathfrak{T}_2 &:= rq^2a_{k+2}^2 + 2qa_{k+2}\sum_{j=1}^{r-1}\bigl(s_2(j)a + j\bigr) + \sum_{j=1}^{r-1}\left(\bigl(s_2(j)\bigr)^2 a^2 + 2js_2(j)a + j^2\right)\\
&= rq^2(a+2^k)^2 + 2q(a+2^k)\left(\sum_{j=1}^{r-1} s_2(j)\right)a + q(a+2^k)r(r-1)\\
&\quad + \sum_{j=1}^{r-1}\left(j^2 - 2js_1(j) + \bigl(s_1(j)\bigr)^2\right)a^2 + 2\left(\sum_{j=1}^{r-1} j\bigl(j - s_1(j)\bigr)\right)a + \frac{r(r-1)(2r-1)}{6}\\
&= \Biggl(\sum_{j=1}^{r-1}\bigl((s_1(j))^2 - 2js_1(j) - 2(q+1)s_1(j)\bigr)\\
&\quad + \frac{r\bigl(6q(q+r+1) + (r+1)(2r+1)\bigr)}{6}\Biggr)a^2\\
&\quad + \left(2\sum_{j=1}^{r-1}(r-j)s_1(j) - \frac{r(r+1)(3q+r+2)}{3}\right)a + \frac{r(r+1)(2r+1)}{6}.
\end{aligned}
$$

By using the third formula in Lemma 1, we get

$$
\begin{aligned}
& s(a, a+1, a+2, a+2^2, \dots, a+2^k) \\
& = \frac{1}{2a}\Bigg(\frac{2^{k-2}q}{3}\bigg(\big(3k(2q+k-1)+2(q-1)(2q-1)\big)a^2 \\
& \quad + 2\big(2^k(4q^2+3kq-3q+2)-3(q+k)\big)a \\
& \quad + 2(2^{2k+1}q^2-3\cdot 2^k q+1)+\mathfrak{T}_2\bigg) \\
& \quad - \frac{1}{2}\Big(2^{k-1}q\big((q+k-1)a-(2^k q-1)\big)+\mathfrak{T}_1\Big)+\frac{a^2-1}{12} \\
& = \frac{2^{k-3}q}{3}\big(4(q-1)(q-2)+3k(2q+k-3)\big)a \\
& \quad + \frac{1}{12}\big(2^{2k}q(4q^2+3kq-3q+2)-3\cdot 2^k q(q+k-2)-6\big) \\
& \quad + \frac{1}{a}(2^{2k+1}q^2-3\cdot 2^k q+1)-\frac{\mathfrak{T}_1}{2}+\frac{\mathfrak{T}_2}{2a}+\frac{a^2-1}{12}\,.
\end{aligned}
$$

Theorem 5 *For integers $k \ge 2$, q and r, satisfying $a = 2^k q + r$ with $0 \le r < 2^k$,*

$$
\begin{aligned}
& s(a, a+1, a+2, a+2^2, \dots, a+2^k) \\
& = \frac{2^{k-3}q}{3}\big(4(q-1)(q-2)+3k(2q+k-3)\big)a \\
& \quad + \frac{1}{12}\big(2^{2k}q(4q^2+3kq-3q+2)-3\cdot 2^k q(q+k-2)-6\big) \\
& \quad + \frac{1}{a}(2^{2k+1}q^2-3\cdot 2^k q+1)-\frac{\mathfrak{T}_1}{2}+\frac{\mathfrak{T}_2}{2a}+\frac{a^2-1}{12}\,,
\end{aligned}
$$

where

$$
\mathfrak{T}_1 = \left(\frac{r(r+2q+1)}{2} - \sum_{j=1}^{r-1} s_1(j)\right) a - \frac{r(r+1)}{2}
$$

and

$$
\begin{aligned}
\mathfrak{T}_2 = & \Bigg(\sum_{j=1}^{r-1}\big((s_1(j))^2-2js_1(j)-2(q+1)s_1(j)\big) \\
& + \frac{r\big(6q(q+r+1)+(r+1)(2r+1)\big)}{6}\Bigg)a^2 \\
& + \left(2\sum_{j=1}^{r-1}(r-j)s_1(j)-\frac{r(r+1)(3q+r+2)}{3}\right)a+\frac{r(r+1)(2r+1)}{6}\,.
\end{aligned}
$$

5.1 Examples

Consider the sequence 16, 17, 18, 20, 24. Then, $a = 16$, $k = 3$ $q = 2$ and $r = 0$. By Theorem 4, we have $s(16, 17, 18, 20, 24) = 684$. In fact, the sum of nonrepresentable numbers is

$$\begin{aligned}&1+2+3+4+5+6+7+8+9+10+11+12+13+14+15+19\\&+21+22+23+25+26+27+28+29+30+31+39+43+45\\&+46+47+63\\&=684\,.\end{aligned}$$

Consider the sequence $25, 26, 27, 29, 33$. Then, $a = 25$, $k = 3$ $q = 3$ and $r = 1$. By Theorem 4, we have $s(25, 26, 27, 29, 33) = 2557$. If the sequence is $25, 26, 27, 29, 33, 41$, then $a = 25$, $k = 4$ $q = 1$ and $r = 9$. By Theorem 4, we have $s(25, 26, 27, 29, 33, 41) = 1827$.

References

1. Komatsu, T., Zhang, Y.: Weighted Sylvester sums on the Frobenius set. Irish Math. Soc. Bull. 87, 35–43 (2021).
2. Komatsu, T., Zhang, Y.: Weighted Sylvester sums on the Frobenius set in more variables. Kyushu J. Math. **76**(1) (2022, to appear). arXiv:2101.04298 (2021)
3. Ramìrez Alfonsìn, J. L.: The Diophantine Frobenius Problem. Oxford University Press, Oxford, 2005.
4. Sylvester, J. J.: Mathematical questions with their solutions. Educational Times 41, 21 (1884).
5. Sylvester, J. J.: On subinvariants, i.e. semi-invariants to binary quantics of an unlimited order. Am. J. Math. 5, 119–136 (1882).
6. Tuenter, H. J. H.: The Frobenius problem, sums of powers of integers, and recurrences for the Bernoulli numbers. J. Number Theory 117, 376–386 (2006).
7. Brown T. C., Shiue, P. J.: A remark related to the Frobenius problem. Fibonacci Quart. 31, 32–36 (1993).
8. Rødseth, Ø. J.: A note on Brown and Shiue's paper on a remark related to the Frobenius problem. Fibonacci Quart. 32, 407–408 (1994).
9. Curtis, F.: On formulas for the Frobenius number of a numerical semigroup. Math. Scand. 67, 190–192 (1990).
10. Beck, M., Gessel, I. M., Komatsu, T.: The polynomial part of a restricted partition function related to the Frobenius problem. Electron. J. Combin. 8 (No.1), #N7 (2001).
11. Komatsu, T.: On the number of solutions of the Diophantine equation of Frobenius–General case. Math. Commun. 8, 195–206 (2003).
12. Fel, L. G., Komatsu, T., Suriajaya, A. I.: A sum of negative degrees of the gaps values in 2 and 3-generated numerical semigroup. Ann. Math. Inform. 52, 85–95 (2020).
13. Wang W., Wang, T.: Alternate Sylvester sums on the Frobenius set. Comput. Math. Appl. 56, 1328–1334 (2008).
14. Brauer, A., Shockley, B. M.: On a problem of Frobenius. J. Reine. Angew. Math. 211, 215–220 (1962).
15. Selmer, E. S.: On the linear diophantine problem of Frobenius. J. Reine Angew. Math. 293/294, 1–17 (1977).

16. Tripathi, A. : On sums of positive integers that are not of the form $ax + by$. Amer. Math. Monthly 115, 363–364 (2008).
17. Punyani, P., Tripathi, A.: On changes in the Frobenius and Sylvester numbers. Integers 18B, #A8 (2018), 12 p.
18. Tripathi, A.: On a special case of the Frobenius problem. J. Integer Seq. 20, Article 17.7.2 (2017), 12 pp.
19. Roberts, J. B.: Notes on linear forms. Proc. Amer. Math. Soc. 7, 465–469 (1956).
20. Brauer, A.: On a problem of partitions. Amer. J. Math. 64, 299–312 (1942).
21. Grant, D. D.: On linear forms whose coefficients are in arithmetic progression. Israel J. Math. 15, 204–209 (1973).
22. Dulmage, A. L., Mendelsohn, N. S.: Gaps in the exponent set of primitive matrices. Illinois J. Math. 8, 642–656 (1964).
23. Tattersall, J. J.: Elementary number theory in nine chapters. Second ed. Cambridge Univ. Press, Cambridge, 2005. xii+430 pp. ISBN: 978-0-521-61524-2; 0-521-61524-0
24. Sloane, N. J. A.: The On-Line Encyclopedia of Integer Sequences. oeis.org

Lattice Structure of Some Closed Classes for Non-binary Logic and Its Applications

Elmira Yu. Kalimulina

Abstract The paper provides a brief overview of modern applications of multi-valued logic models, where the design of heterogeneous computing systems with small computing units based on three-valued logic gives the mathematically better and more effective solution compared to binary models. It is necessary for applications to implement circuits comprised from chipsets, the operation of which is based on three-valued logic. To be able to implement such schemes, a fundamentally important theoretical problem must be solved: the problem of completeness of classes of functions of three-valued logic. From a practical point of view, the completeness of the classes of such functions ensures that circuits with the desired operations can be produced from on an arbitrary (finite) set of chipsets. In this paper, the closure operator on the set of functions of three-valued logic, that strengthens the usual substitution operator has been considered. It was shown that it is possible to recover the sublattice of closed classes in the general case of closure of functions with respect to the classical superposition operator. The problem of the lattice of closed classes for the class of functions T_2 preserving two is considered. The closure operator $\mathcal{R}_1$ for which functions that differ only by dummy variables are considered to be equivalent is considered in this paper. A lattice is constructed for closed subclasses in $T_2 = \{f \,|\, f(2, \ldots, 2) = 2\}$ – class of functions preserving two

Keywords Three-valued logic application · Three-valued logic · Closure operator · Lattice structure · Closed subclasses · Substitution operator

The publication has been prepared with the support of the Russian Foundation for Basic Research according to the research project No.20-01-00575 A.

E. Yu. Kalimulina (✉)
V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
e-mail: elmira.yu.k@gmail.com
URL: http://www.ipu.ru/staff/elmira

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_2

1 Introduction

A ternary system is the most optimal from the point of view of information density [9]. The generalization for multi-valued logic is the ternary logic [2, 3]. Further, without loss of generality instead of multivalued case a ternary logic model may be considered. In ternary logic, a statement is assigned one of three values: "true", "false", "undefined" [2, 4, 9]; in binary logic—two: either "true" or "false". Symmetric form of number representation based on three-valued logic simplifies a processing of negative numbers, since it requires an extra bit to store the sign [4].

Some features of the operation logic of a ternary computer, for example, the representation of negative numbers, give possibilities for design more reliable and high-performance modern systems, that will be useful for many modern applications. Mathematically, ternary logic is more efficient than binary logic [2, 4, 9]. Research and development of algorithms based on three-valued logic are very relevant [8], for example, in telecommunications [7, 10], in the field of artificial intelligence (AI) [6], quantum computing [7, 11–13], medicine, physics [14]. This is confirmed by a significant increase of the number of scientific publications in leading scientific journals related to various applications of three-valued logic over the past few years [17].

1.1 A Brief Overview of Modern Applications of Multivalued Logic

Here are examples of several applications where the construction of algorithms based on three-valued logic provides greater efficiency and turns out to be preferable in comparison with two-valued logic. For more detailed overview, you can read references.

Reliability analysis of structural processes and factors assessment of technical systems Multi-valued logic allows to consider qualitative variables instead of quantitative ones. Quantitative indicators (factors) are discretized by mapping into a certain m-interval scale. This approach allows you to combine quantitative and qualitative indicators within the single model. The reliability of the factors decreases minimally with such discretisation. This allows to investigate the model as fully as possible. This is especially effective in situations where there is no way to quantify the impact of a particular factor on the process. The use of qualitative variables provides additional opportunities for assessing factors.

Simulation of processes and modern design languages Simulation is the only available way to check the quality and reliability of complicated and expensive technical systems at their design stage. Automated design tools allow you to assess quality based on real-world operating conditions. Temporary simulation of circuits in an automated simulation system is often based on the principles of three-valued logic.

Design of data transmission and processing systems Ternary logic is effective in constructing computing units for equipment of data transmission networks. Potentially, the transmission of three states instead of two bits at a time can increase the data transfer rate by 1.5 times. With an increase of the number of trits (instead of bit) the speed can grow exponentially [10, 18, 19]. It is possible to implement solutions for data aggregation and transmission based on multivalued logic. These solutions will provide a single high-dimensional space for network addressing—both for standard purposes of data transmission [15] and for new tasks for controlling robotic devices for the Internet Of Things [7].

Three-valued logic is also effective both for solving problems of image processing [5] and for problems of cryptography. Quantum computing for data security is the most effective method of protecting mobile robots, the Internet of Things (IoT) and security of distributed applications. That also uses multi-valued logic models. With the rapid growth of quantum computers, ternary computing has become relevant again [5, 12, 13]. The leading IT companies have introduced their quantum computers operating on several dozen of qubits in the last decade: IBM quantum processors consist of 65 qubits, Google has 72 [20]. The developers plan to release a 1112-qubit processor called “Condor” by 2023, that should bring quantum technologies to a new commercial level [20].

Also, at present, the multi-valued logic toolkit is widely used in tasks related to data analysis and the construction of AI models, for example, in the tasks of hierarchical data clustering for arbitrary complicated data sets [6, 7]. Interpretation models via 3-valued logic allows to overcome exiting limitations on the ability to create fully automatic program-analysis algorithms [1].

At the end of this short overview of multivalued logic models, the application in economic research should be mentioned: models of collective behavior and the problem of collective choice, where “cyclical logic” arises as a special case of k-valued logic [21].

2 Theoretical Aspects of Designing of Computing Systems Based on Three-Valued Logic

All applied problems considered above are reduced to the problem of determining the factors that have an influence on the process and considering a countable set P_3 of states of these factors. Any countable number of states can be approximated by basically three states [23]: 0, 1, 2.

And for a decision making someone need to find the value of the output function Y that depends on this set. Accordingly, the output function Y can be represented as a combination of predicates on the set P_3 [22]. For this purpose complicated predicates and superpositions of these predicates on P_3 will be considered.

These predicates can be implemented (from practical point of view) as circuits of chips, the operation of which is based on three-valued logic.

2.1 Completeness of Functions Classes of Three-Valued Logic

A fundamentally important problem—**the problem of completeness of classes of functions of three-valued logic** [22]—must be solved to make this implementation possible. From the practical point of view, the completeness of the classes of functions guarantees that a circuit with the desired functional diagram can be produced based on an arbitrary finite number of chipsets. For two-valued logic, this problem was also solved by Emil Post, which led to the explosive growth of electronics [24].

Post's classical theorem describes five precomplete classes in the set of Boolean functions [24].

For the case of three-valued logic, the problem was solved by Yablonsky in 1958 [22, 23]. He proved that there are 18 precomplete classes for functions of three-valued logic. In the papers [22, 23], the closure of the set of functions with respect to the substitution operator was considered.

Unfortunately, for three-valued logic it was proved that this problem cannot be solved in a general case [23]. If the lattice of closed classes is countable in the case of two-valued logic, then it is exponential in the case of three-valued logic. However, its closure operators on the set of three-valued logic functions can be considered, which are a strength of the common substitution operator.

Solving the completeness problems for this new closure operator and finding the structure of the lattice of closed classes will help not only to restore the sublattice of closed classes in the general case of closure of functions with respect to the classical superposition operator, but also will optimize the possible production of chips for functional circuits for solving the problem described above in the Introduction.

Consider a variant of the closure operator $\mathcal{R}_\infty$, for which functions that differ only in dummy variables are considered equivalent. Let us construct a lattice for closed subclasses in $T_1 = \{f | f(1, \dots, 1) = 1\}$ — in the class of functions preserving two.

2.2 Lattice of Closed Subclasses T_2 with Respect to $\mathcal{R}_\infty$

Definition 1 Let $f(x_1, \dots, x_i, \dots, x_n) \in P_3$, $|X_f| = n$, then x_i called $\mathcal{R}_\infty$-essential for f, if there are sets $\alpha_1^n = (a_1, \dots, a_{i-1}, b^1, a_{i+1}, \dots, a_n)$, $\alpha_2^n = (a_1, \dots, a_{i-1}, b^2, a_{i+1}, \dots, a_n)$ such that $f(\alpha_1^n) \sim f(\alpha_2^n)$.

Completeness in T_2

Definition 2 Use the following notation $T^{02} \overset{\text{def}}{=} \{f | \exists i \in \{1, X_f\} : \alpha = (a_1, \dots, a_{X_f}), a_i \in \{0, 2\} \Rightarrow f(\alpha) = 2\}$

$T^{12} \overset{\text{def}}{=} \{f | \exists i \in \{1, X_f\} : \alpha = (a_1, \dots, a_{X_f}), a_i \in \{1, 2\} \Rightarrow f(\alpha) = 2\}$

$T^{02} \overset{\text{def}}{=} \{f | \alpha = (a_1, \dots, a_{X_f}); a_i \in \{0, 2\}, i \in \{1, X_f\} \Rightarrow f(\alpha) = 2\}$

$T^{12} \overset{\text{def}}{=} \{f | \alpha = (a_1, \dots, a_{X_f}); a_i \in \{1, 2\}, i \in \{1, X_f\} \Rightarrow f(\alpha) = 2\}$

Lemma 1 *The class* $T^{02}-$ *is* $\mathcal{R}_\infty$*-closed.*

Lemma 2 *The class* $T^{12}- \mathcal{R}_\infty$ *is closed.*

Proof of Lemma 1. Note that neither the permutation of variables nor identification or addition of inessential (dummy) ones affect the property functions belong to class T^{02}. This follows obviously from the class definitions.

It is also obvious that if $f \in T^{02}$, then for any function $g(f \sim g)$ it's true that $g \in T^{02}$.

Now show that the superposition of functions from the class T^{02} will also lie in class T^{02}.

Let $f \in T^{02}$, $f = f(x_1, \dots, x_n)$. Consider the function $h = f(g_1, \dots, g_n)$, where g_i- are either free variables or functions from the set T^{02}.

By contradiction, let $h \notin T^{02}$, then there is a set $\alpha = (a_1, \dots, a_{|X_h|})$, $a_i \in \{0, 2\}$, $1 \le i \le |X_h|$, such that it's true that $h(\alpha) \neq 2$.

And by the construction of the function h, and under the condition that $f \in T^{02}$ there is such i that the function $g_i(\beta) \neq$, where $\beta = (b_1, \dots, b_{|X_{g_i}|})$, $1 \le b_i \le |X_{g_i}|$—projection of vector α on the coordinate axes corresponding to free variables of the function g_i.

Thus the function $g_i \notin T^{02}$, but that contradicts the choice of function g_i. Thus $h \in T^{02}$.

The lemma 1 is proved.

The Lemma 2 can be proved by repeating the sketch of the proof of lemma 1 (by formal replacing of T^{02} by T^{12}).

Lemma 3 *The class* $T^{02}- \mathcal{R}_\infty$ *is pre-complete in the class* T_2.

Proof Note that the class $T_2 = \mathcal{R}_\infty(\{, \})$, where $f(|X_f| = 2)\&(f(\alpha) = 2$ if and only if when $\alpha = (2, 2))$, $g(|X_g| = 1)\&(g \in T_2)\&(g \notin T_\sim)$.

Let there be a function $w(w \notin T^{02})$. Then by definition there is a set $\alpha = (a_1, \dots, a_{|X_w|})$, $a_i \in \{0, 2\}$, $1 \le i \le |X_w|$ such that $w(\alpha) \neq 2$.

Let's move on from the function w to function w', derived from w by identifying variables according to the set α. Namely, variables in the set α will be identified with the same values. Thus, the whole set of variables of the function w may be split into two groups: with respect to 0 and with respect to 2. By identification, that gives the function $w'(|X_{w'}| = 2)\&(w' \notin T^{02})$.

Let without loss of generality $w'(0, 2) = 1$. If this is not true, then by rearranging the variables and moving to function $w''(w'' \sim w')$ the function with the specified property can be obtained easily.

If the vector α does not contain elements equal to 2, then the function that $\sim$ a function w' and satisfies the required properties may be considered.

Note that a function $g(g \in T^{02})\&(|X_g| = 1)\&(g \notin T_\sim)$ exists. Consider a function $w''(w'' \sim w)$ such that:

$$w''(\alpha) = \begin{cases} 1, \ \alpha = (2, 0) \\ 2, \ w'(\alpha) = 2 \\ 0, \ \text{otherwise.} \end{cases}$$

Consider a function $v_1(x, y) = g(w''(x, y))$. The property $v_1(\alpha) = 1$ for this function holds if and only if when $\alpha = (0, 2)$. Also consider a function $v_2 = v_2(x, y) = v_1(y, x)$. It is easy to see that by construction it gives $\{v_1, v_2\} \subseteq \mathcal{R}_\infty(\mathcal{T}'^{\in} \cap \sqsupseteq)$.

Consider the function d such that:

$$d(\alpha) = \begin{cases} 2, \ a_i \in \{0, 2\}, 1 \le i \le 2 \\ 1, \qquad\qquad \text{otherwise.} \end{cases}, \alpha = (a_1, a_2).$$

It's obviously that $d \in T^{02}$. Let's construct a function m:

$$m(x, y) = d(d(v_1(x, y), d(x, y)), v_2(x, y))$$

$$m(\alpha) = \begin{cases} 2, \ a_1 = 1, 1 \le i \le 2 \\ 1, \qquad\qquad \text{otherwise.} \end{cases}, \alpha = (a_1, a_2).$$

By the fact that the function $2 \in T^{02}$ a function f can be constructed such that:

$$f(x, y) = m(m(x, 2), m(y, 2))$$

$$f(\alpha) = \begin{cases} 2, \ a_i = 2, 1 \le i \le 2 \\ 1, \qquad\qquad \text{otherwise.} \end{cases}, \alpha = (a_1, a_2).$$

It was mentioned above that $\mathcal{R}_\infty(\{\{,\}\}) = \mathcal{T}_{\in}$. But by construction it can be obtained that $f \in \mathcal{R}_\infty(\Updownarrow) \subseteq \mathcal{R}_\infty(\sqsupseteq, \mathcal{T}'^{\in})$, and by definition $g \in T^{02}$, therefore $T_2 = \mathcal{R}_\infty(\sqsupseteq, \mathcal{T}'^{\in})$.

The lemma is proved.

Lemma 4 *Let $f \in T_2$ and $f \notin T_{01}$. Then $2 \in \mathcal{R}_\infty(\{)$*

Proof Consider the function $h(x) = f(x, \ldots, x)$. It is easy to show that if $h \notin T_{01}$, then $2 \in \mathcal{R}_\infty(\langle)$.

Let $h \in T_{01}$. Note that for any $g(|X_g| = 1)\&(g \in T_{01})$ it holds that $g \in \mathcal{R}_\infty(\langle)$. by condition $f \notin T_{01}$, hence there is a set $\alpha = (\alpha_1, \ldots, \alpha_n)$, $n = |X_f|$, $\alpha_i \in \{0, 1\}$, $1 \le i \le n$ such that $f(\alpha) = 2$. Construct a function $f' = f(g_1, \ldots, g_n)$, $|X_{g_i}| = 1$, $g_i \in T_{01}$, $1 \le i \le n$, at that $g_i(0) = \alpha_i$. Note that $\{g_i, h\} \subset \mathcal{R}_\infty(\{)$ therefore $f' \in \mathcal{R}_\infty(\{)$. Consider a function $h'(x) = f'(x, \ldots, x)$. By construction it can be obtained that $h'(0) = h'(2) = 2$, and therefore according to the already considered case we have $2 \in \mathcal{R}_\infty(\langle') \subset \mathcal{R}_\infty(\{)$.

The lemma is proved.

Lemma 5 *A class $T_\sim \cap T_2-$ is $\mathcal{R}_\infty-$precomplete in T_2.*

Proof Let $f \notin T_\sim$, $f \in T_2$, $|X_f| = n$. Let us show that $\mathcal{R}_\infty(\{\{ \cup \mathcal{T}_\sim \cap \mathcal{T}_\in\}) = \mathcal{T}_\in$. Note that, by definition, there are at least two sets $\alpha_1 = (a_1^1, \ldots, a_n^1)$ and $\alpha_2 =$

$(a_1^2, \ldots, a_n^2)$ such that $\alpha_1 \sim \alpha_2$, and $f(\alpha_1) \sim f(\alpha_2)$. Identify variables in f according to the coincidence of identical pairs in vectors α_1 and α_2. Concretely if $(a_i^1, a_i^2) = (a_j^1, a_j^2)$, then $i-$th and $j-$th variables are identified. Thus the function f' of five variables satisfying the following condition has been obtain

$$f'(0, 1, 2, 0, 1) \sim f'(0, 1, 2, 1, 0)$$

Without loss of generality, it can be assumed that after identification variables the function f' will have exactly this order variables. Otherwise, the variables will be reordered. Also note that some of the variables of the function f' can be dummy.

Note that there is $2, g \in T_\sim \cap T_2(g(0) = 1, g(1) = 0)$. Let's move on from the function f' to a function $f'' = f'(g(x_1), x_1, 2, x_2, x_3)$, $f'' \in \mathcal{R}_\infty(\mathcal{T}_\sim \cap \mathcal{T}_\in)$. A function f'' satisfies the property

$$f''(0, 0, 1) \sim f''(1, 0, 1)$$

Let without loss of generality $f''(1, 0, 1) = 2$.

There are functions $f \in T_\sim \cap T_2$, such that $f(\alpha) = 2$ if $\alpha = (2, \ldots, 2)$. Denote the set of such functions as N. Let us show by a construction that $\mathcal{R}_\infty(\{\{'',\mathrm{N}\}) = T_2$.

Let $h \in T_2$ – arbitrary function. Consider the functions $g_0, g_1, g_2 \in N(|X_{g_i}| = |X_h| = n)$.

$$g_0(\alpha) = \begin{cases} 2, & \alpha = (2, \ldots, 2) \\ 0, & \text{otherwise.} \end{cases}$$

$$g_1(\alpha) = \begin{cases} 2, & \alpha = (2, \ldots, 2) \\ 1, & \text{otherwise.} \end{cases}$$

$$g_2(\alpha) = \begin{cases} 2, & \alpha = (2, \ldots, 2) \\ 1, & h(\alpha) = 2, \\ 0, & h(\alpha) \neq 2, \end{cases}$$

Consider the function $h'(x_1, \ldots, x_n) = f''(g_2(x_1, \ldots, x_n), g_1(x_1, \ldots, x_n), g_0(x_1, \ldots, x_n))$. By construction $h' \sim h$. Thereby $\mathcal{R}_\infty(\langle') = \mathcal{R}_\infty(\langle) \subseteq \mathcal{R}_\infty(\{\{'', \mathcal{N}\}) \subseteq \mathcal{R}_\infty(\{, \mathcal{T}_\sim \cap \mathcal{T}_\in)$. Due to the arbitrariness of the function $h \in T_2$ we get $T_2 \in \mathcal{R}_\infty(\{\{, \mathcal{T}_\sim \cap \mathcal{T}_\in\})$.

The lemma is proved.

Lemma 6 *A class $T_{01} \cap T_2 - \mathcal{R}_\infty$-precomplete in T_2.*

Proof Consider the function $f \notin T_{01} \cap T_2$, $f \in T_2$. By Lemma 4 $\mathcal{R}_\infty(\in, \mathcal{T}_{/\infty} \cap \mathcal{T}_\in) \subseteq \mathcal{R}_\infty(\{, \mathcal{T}_{/\infty} \cap \mathcal{T}_\in)$. Let $h \in T_2$ — arbitrary function from T_2. Note that there is a function $g \in T_{01} \cap T_2$, satisfying the following property:

$$g(0, 2) \sim g(1, 2)$$

Without loss of generality $g(1, 2) = 2$.

Consider the function $m \in T_{01} \cap T_2(|X_m| = |X_h| = n)$ such that:

$$m(\alpha) = \begin{cases} 2, \ \alpha = (2, \dots, 2) \\ 1, \qquad h(\alpha) = 2, \\ 0, \qquad h(\alpha) \neq 2, \end{cases}$$

The function $h'(x_1, \dots, x_n) = g(m(x_1, \dots, x_n), 2)$ satisfies the property $h \sim h'$ by construction. In this way $h \in \mathcal{R}_\infty(\langle') \subseteq \mathcal{R}_\infty(\{\in, \mathcal{T}_{/\infty} \cap \mathcal{T}_\in\}) \subseteq \mathcal{R}_\infty(\{\{, \mathcal{T}_{/\infty} \cap \mathcal{T}_\in\})$. By the arbitrary function h we have $T_2 \in \mathcal{R}_\infty(\{, \mathcal{T}_{/\infty} \cap \mathcal{T}_\in)$.

The lemma is proved.

Now it is possible to formulate the main result that follows from these lemmas

Theorem 1 (Completeness) *There are five pre-complete classes in* T_2.

2.3 The Completeness Problem for the Operator $\mathcal{R}_\infty$

Let M be a given set of functions from P_3. Denote the result of the closure of the set of functions M with respect to operation of substitution and transition of the function g to the equivalent function $f \sim g$ as $\mathcal{R}_\infty(\mathcal{M})$, where

$$f \sim g \Leftrightarrow \forall \mathbf{x} \, [\ (f(\mathbf{x}) = g(\mathbf{x})) \vee (f(\mathbf{x}), g(\mathbf{x}) \in \{0, 1\})\ ].$$

Consider classes: T_{01} — class of functions preserving the set $\{0, 1\}$, T_2 — function class preserving two, and class $T_\sim$ (also $T_{\{01\},\{2\}}$ $(U(R))$ — function class, preserving the relation $\sim$.

It is easy to see that with passing from the function f to the function g property of belonging to classes T_2, T_{01}, $T_\sim$ is preserved. In this way due to the fact that classes T_2, T_{01}, $T_\sim$ are precomplete with respect to the substitution, and completion does not add new functions, then the following lemma is obtained:

Lemma 7 *Classes* T_2, T_{01}, $T_\sim$ *are* $\mathcal{R}_\infty$*-precomplete.*

Lemma 8 *Let* $f \notin T_{01}$, *Then* $2 \in \mathcal{R}_\infty(\{)$.

Proof It is easy to check that if $h(x) \notin T_{01}$ is a 1-place function, then $2 \in \mathcal{R}_\infty(\langle)$. Thus, if the function $g(x) = f(x, \dots, x)$, $g \notin T_{01}$, then the lemma is proved.

Let $g \in T_{01}$. By condition, there is a set $\boldsymbol{\alpha} = (a_1, \dots, a_n)$, $a_i \in \{0, 1\}$ such that $f(\boldsymbol{\alpha}) = 2$. Consider a function f' such that

$$g'(x) = f(g_1(x), \dots, g_n(x))$$

where $g_i(g_i(0) = a_i) \& (g_i \sim g)$. notice, that $g' \in \mathcal{R}_\infty(\{)$ and $g'(0) = 2$, and then $2 \in \mathcal{R}_\infty(\}') \subseteq \mathcal{R}_\infty(\{)$.

The lemma is proved.

Theorem 2 (Completeness) *There are three* $\mathcal{R}_\infty$*—pre-complete classes* T_2.

2.4 Conclusion

In this paper, the closure operators on the set of functions of three-valued logic, which are a strength of the usual substitution operator was considered. It was proved that the completeness problem for this operator has a solution; it is possible to recover the sublattice of closed classes in the general case of closure of functions with respect to the classical superposition operator, which will optimize possible production of chipsets for new functional circuits for transmission and data processing tasks. Also a brief overview of modern applications of three-valued logic models was given.

References

1. Reps T.W., Sagiv M., Wilhelm R. (2004) Static Program Analysis Via 3-valued Logic. In: Alur R., Peled D.A. (Eds) Computer Aided Verification. Cav 2004. Lecture Notes In Computer Science, Vol 3114. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-27813-9_2.
2. Trogemann, Georg; Nitussov, Alexander Y.; Ernst, Wolfgang (2001), Computing In Russia: The History Of Computer Devices And Information Technology Revealed, Vieweg+Teubner Verlag, Pp. 19, 55, 57, 91, 104–107, Isbn 978-3-528-05757-2.
3. Rumyantsev, Dmitri. Interviews With The Designer Of The Ternary Computer. Upgrade 33:175 (2004). An Interview With Nikolai Brusentsov, Designer Of The Setun Ternary Computer. In Russian.
4. The Ternary Calculating Machine Of Thomas Fowler: www.Mortati.Com/Glusker/Fowler/Index.Htm
5. Abiri Ebrahim, Darabi Abdolreza, Salem Sanaz. Design Of Multiple-valued Logic Gates Using Gate-diffusion Input For Image Processing Applications . Computers Electrical Engineering. 2018. Vol.69. Pages 142–157. 0.1016/J.Compeleceng.2018.05.019.
6. Aizenberg I. Complex-valued Neural Networks With Multi-valued Neurons. Studies In Computational Intelligence (Vol. 353). Springer-verlag Berlin Heidelberg. 2011. 273 C. https://doi.org/10.1007/978-3-642-20353-4.
7. Bykovsky Alexey Yu. Heterogeneous Network Architecture For Integration Of Ai And Quantum Optics By Means Of Multiple-valued Logic . Quantum Rep. 2020. -2. Pp. 126–165. https://doi.org/10.3390/Quantum2010010.
8. Cobreros Pablo, Égré Paul, Ripley David, Van Rooij Robert. Three-valued Logics And Their Applications . Journal Of Applied Non-classical Logics. 2014. Vol.24, Iss.1-2. P. 1–11. https://doi.org/10.1080/11663081.2014.909631.
9. Connelly Jeff. Ternary Computing Testbed 3-trit Computer Architecture. Phd Thesis. Computer Engineering Department. California Polytechnic State University. 2008. P.192. Url: Http:.Xyzzy.Freeshell.Org/Trinary/Cpe%20report%20-%20ternary%20computing%20testbed%20-%20rc6a.Pdf.
10. Yi Jin, Huacan He, Yangtian Lü. Ternary Optical Computer Architecture . Physica Scripta. 2005. T118. https://doi.org/10.1238/Physica.Topical.118a00098.
11. Hu Zhengbing, Deibuk Vitaly. Design Of Ternary Reversible/Quantum Sequential Elements . Journal Of Thermoelectricity. 2018. -1. C. 5–16.
12. Muthukrishnan Ashok, Stroud C. R. Jr. Multivalued Logic Gates For Quantum Computation . Phys. Rev. A. 2000. Iss. 5 (Vol. 62). https://doi.org/10.1103/Physreva.62.052309.
13. Muthukrishnan Ashok. Classical And Quantum Logic Gates: An Introduction To Quantum Computing. - Rochester Center For Quantum Information (Rcqi). Quantum Information Seminar, 1999. P. 22.

14. Warzecha M., Oszajca M., Pilarczyk K., Szaclowski K. A Three-valued Photoelectrochemical Logic Device Realising Accept Anything And Consensus Operations . Chemical Communications. 2015. Vol.51, Iss.17. P. 3559–3561. https://doi.org/10.1039/C4cc09980j.
15. Esin A., Yavorskiy R., Zemtsov N. Brief Announcement Monitoring Of Linear Distributed Computations. In: Dolev S. (Eds) Distributed Computing. Disc 2006. Lecture Notes In Computer Science, Vol 4167. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11864219-47.
16. Esin A.A. On Function Classes In P3 Precomplete With Respect To A Strengthened Closure Operator . Math Notes. 2008, 83:5. C. 594603. https://doi.org/10.1134/S0001434608050027.
17. Kak Subhash. On Ternary Coding And Three-valued Logic. 2018. Arxiv.Org/Abs/1807.06419.
18. Gaudet V. A Survey And Tutorial On Contemporary Aspects Of Multiple-valued Logic And Its Application To Microelectronic Circuits . Ieee Journal On Emerging And Selected Topics In Circuits And Systems. 2016. Vol. 6, March, No. 1. Pp. 5–12. https://doi.org/10.1109/Jetcas.2016.2528041.
19. Wu Haixia, Bai Yilong, Li Xiaoran, Wang Yiming. Design Of High-speed Quaternary D Flip-flop Based On Multiple-valued Current-mode . Journal Of Physics: Conference Series. 2020, October. Vol. 1626. https://doi.org/10.1088/1742-6596/1626/1/012067.
20. Ibm Quantum Summit 2020: Exploring The Promise Of Quantum Computing For Industry, www.Ibm.Com/Blogs/Research/2020/09/Quantum-industry/.
21. J. B. Nation. Logic On Other Planets. Preprint, 2005. www.Math.Hawaii.Edu/~Jb/Planets.Pdf.
22. S.V. Yablonskiy, G.P. Gavrilov And V.B. Kudryavtsev “Logical Algebra Functions And Post Classes”. Moscow.: “Nauka”, 1966.
23. S.V. Yablonskii, ’Functional Constructions In A K-valued Logic” , Collection Of Articles On Mathematical Logic And Its Applications To Some Questions Of Cybernetics, Trudy Mat. Inst. Steklov., 51, Acad. Sci. Ussr, Moscow, 1958, 5–142
24. Post E.L. Two-valued Iterative Systems Of Mathematical Logic . Annals Of Math. Studies. Princeton Univ. Press. 1941. V. 5.

Content Based Image Retrieval Using HDMR Constant Term Based Clustering

Ayşegül Karcılı and Burcu Tunga

Abstract The studies related with the content-based image retrieval (CBIR) has increased because of both necessity for efficient image retrieval and the limitations in large-scale systems. Efficient image retrieval refers to finding accurate image from the database with high speed. This paper presents a new efficient image retrieval method using High Dimensional Model Representation (HDMR). The method has two main steps, clustering and retrieval. In clustering part, we use k-means method on HDMR constant term while in the subsequent part, we retrieve the most similar images to a given query image from a relevant cluster. We experiment the efficiency and effectiveness of the new algorithm on Columbia Object Image Library (COIL-100) and get conspicuous results. These results are tabulated in the paper.

Keywords HDMR · Image retrieval · Image decomposition · Clustering · k-means

1 Introduction

Content Based Image Retrieval (CBIR) is a method for finding the most similar image to a given query image. It uses mathematical representations of a digital image to perform the retrieval task. Because the image itself is quite expensive to use and is quite difficult in large-scale libraries, we use HDMR components instead of image itself for image representation.

HDMR method uses divide-and-conquer philosophy. It decomposes a multivariate function with N independent variables into 2^N functions having low-variate

A. Karcılı · B. Tunga (✉)
Mathematics Engineering Department Faculty of Science and Letters, Istanbul Technical University Maslak, 34469 İstanbul, Turkey
e-mail: tungab@itu.edu.tr
URL: http://www.web.itu.edu.tr/tungab/

A. Karcılı
Huawei Technologies, Huawei Turkey R&D Center, İstanbul, Turkey

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_3

terms. That is, HDMR can be referred as an algorithm based on divide-and-conquer philosophy. These 2^N number of functions are named as constant, univariate, bivariate and higher variate ones. Although the given multivariate function can exactly be re-obtained by summation of all these functions, this procedure is very expensive. Consequently HDMR is considered as an approximation method and there are many articles in the literature that include this subject [1–3]. HDMR is also used to decompose a given multivariate dataset to the low variate datasets [4–8].

HDMR method has recently started to be used in image processing [9–11]. If a digital image in RGB format is assumed as three dimensional array, HDMR decomposes the image into constant, univariate, bivariate components. There are certain studies on different subjects in the literature using these components [12, 13].

In this study, we develop a new algorithm using these HDMR components in image retrieval issue. This algorithm has two main step: The first step is based on k-means clustering [14] and the usage of the constant term of HDMR. The second step is to realize the retrieval process to find the most similar image to a given query image and it uses different variations of univariate HDMR terms. Clustering part is used to make the retrieval process faster.

The proposed algorithm is tested on Columbia Object Image Library (COIL-100) to show its performance [15]. This library contains 7200 color images. These images belong to 100 different objects. That is, there are 72 poses per object of 5 degrees.

This paper is organized as follows. The second section contains a brief explanation of the HDMR method and its mathematical expression. This section also includes how HDMR is used for the image decomposition. The third section covers the clustering method which is used in order to speed up the system, while the new algorithm for the retrieval process is given in the fourth section. The fifth section consists of obtained results and comparisons. Certain concluding remarks are given in the last section.

2 Image Decomposition with HDMR

HDMR is a method used to decompose multivariate functions. The basic mathematical expression of the method, where N is the number of independent variables, is given as follows

$$f(x_1, \ldots, x_N) = f_0 + \sum_{i=1}^{N} f_i(x_i) + \sum_{\substack{i_1,i_2=1 \\ i_1<i_2}}^{N} f_{i_1,i_2}(x_{i_1}, x_{i_2}) + \cdots \\ + f_{12\cdots N}(x_1, x_2, \cdots, x_N) \tag{1}$$

The basic philosophy of this method is to write a multivariate function as the sum of a finite number of less variable functions. So, the right hand side of the expansion consists of a constant term, univariate terms, bivariate terms and the functions with increasing number of independent variables. When all these terms are added, the multivariate function is exactly obtained.

To find the right hand side terms uniquely, some operators and conditions are defined [1–3]. The following operator, $\mathcal{I}_0$ is utilized to determine the general structure of the constant component

$$\mathcal{I}_0 f(x_1, \ldots, x_N) \equiv \int_{a_1}^{b_1} dx_1 W_1(x_1) \cdots \int_{a_N}^{b_N} dx_N W_N(x_N) f(x_1, \ldots, x_N) \tag{2}$$

while the following $\mathcal{I}_m$ and $\mathcal{I}_{m_1 m_2}$ operators $(1 \le m \le N)$, $(1 \le m_1 < m_2 \le N)$ are used to determine the univariate and bivariate terms.

$$\begin{aligned}\mathcal{I}_m f(x_1, \ldots, x_N) \equiv& \int_{a_1}^{b_1} dx_1 W_1(x_1) \cdots \int_{a_{m-1}}^{b_{m-1}} dx_{m-1} W_{m-1}(x_{m-1}) \cdots \\ &\times \int_{a_{m+1}}^{b_{m+1}} dx_{m+1} W_{m+1}(x_{m+1}) \cdots \int_{a_N}^{b_N} dx_N W_N(x_N) f(x_1, \ldots, x_N)\end{aligned} \tag{3}$$

$$\begin{aligned}\mathcal{I}_{m_1 m_2} f(x_1, \ldots, x_N) \equiv& \int_{a_1}^{b_1} dx_1 W_1(x_1) \cdots \int_{a_{m_1-1}}^{b_{m_1-1}} dx_{m_1-1} W_{m_1-1}(x_{m_1-1}) \\ &\times \int_{a_{m_1+1}}^{b_{m_1+1}} dx_{m_1+1} W_{m_1+1}(x_{m_1+1}) \cdots \int_{a_{m_2-1}}^{b_{m_2-1}} dx_{m_2-1} W_{m_2-1}(x_{m_2-1}) \\ &\times \int_{a_{m_2+1}}^{b_{m_2+1}} dx_{m_2+1} W_{m_2+1}(x_{m_2+1}) \cdots \int_{a_N}^{b_N} dx_N W_N(x_N) f(x_1, \ldots, x_N)\end{aligned} \tag{4}$$

The following mathematical expressions called vanishing and normalization conditions respectively

$$\int_{a_1}^{b_1} dx_1 \cdots \int_{a_N}^{b_N} dx_N W(x_1, \ldots, x_N) f_i(x_i) = 0. \tag{5}$$

$$\int_{a_i}^{b_i} dx_i \, W_i(x_i) = 1, \quad 1 \le i \le N. \tag{6}$$

where $W_i(x_i)$s are named as weight factors and satisfy the following rule.

$$W(x_1, \ldots, x_N) \equiv \prod_{i=1}^{N} W_i(x_i), \quad x_i \in [\, a_i \,,\, b_i \,], \quad 1 \le i \le N \tag{7}$$

If all these operators and conditions are applied to the both sides of the HDMR expansion given in Eq. (1), then the constant, univariate and bivariate components are obtained as follows

$$\begin{aligned} f_0 &= \mathcal{I}_0 f(x_1, \ldots, x_N) \\ f_{m_1}(x_{m_1}) &= \mathcal{I}_{m_1} f(x_1, \ldots, x_N) - f_0 \\ f_{m_1 m_2}(x_{m_1}, x_{m_2}) &= \mathcal{I}_{m_1 m_2} f(x_1, \ldots, x_N) - f_{m_1}(x_{m_1}) - f_{m_2}(x_{m_2}) - f_0 \end{aligned} \tag{8}$$

where $1 \le m_1 < m_2 \le N$. The higher variate components can be determined similarly.

HDMR is also used for multivariate data problems by adapting the method to the discrete data [4–7]. For this purpose, the weight factors located in the weight function are taken as

$$W_j(x_j) \equiv \sum_{k_j=1}^{n_j} \alpha_{k_j}^{(j)} \delta\left(x_j - \xi_j^{(k_j)}\right), \quad x_j \in \left[\, a_j \,,\, b_j \,\right], \quad 1 \le j \le N \tag{9}$$

where δ is Dirac delta function and α's are the parameters to give a different importance for each node of data set [4]. They have to satisfy the normalization condition given in the Eq. (6). We chose these parameters equally in our study to give each data point same importance. When these weight factors are placed into the Eq. (8) by using the normalization and vanishing conditions all right hand side terms are obtained uniquely. The constant component of multivariate data is obtained as follows.

$$f_0 = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \cdots \sum_{k_N=1}^{n_N} \left(\prod_{i=1}^{N} \alpha_{k_i}^{(i)} \right) f(\xi_1^{(k_1)}, \ldots, \xi_N^{(k_N)}) \tag{10}$$

Similarly univariate components can be written as

$$\begin{aligned} f_m\left(\xi_m^{(k_m)}\right) &= \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \cdots \sum_{k_{m-1}=1}^{n_{m-1}} \sum_{k_{m+1}=1}^{n_{m+1}} \\ &\cdots \sum_{k_N=1}^{n_N} \left(\prod_{\substack{i=1 \\ i \ne m}}^{N} \alpha_{k_i}^{(i)} \right) f(\xi_1^{(k_1)}, \ldots, \xi_N^{(k_N)}) - f_0, \; 1 \le k_m \le n_m, \quad 1 \le m \le N \end{aligned} \tag{11}$$

Finally, bivariate components are as follows

$$f_{m_1 m_2}\left(\xi_{m_1}^{(k_{m_1})}\xi_{m_2}^{(k_{m_2})}\right) = \sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2}\cdots\sum_{k_{m_1-1}=1}^{n_{m_1-1}}\sum_{k_{m_1+1}=1}^{n_{m_1+1}}\cdots\sum_{k_{m_2-1}=1}^{n_{m_2-1}}\sum_{k_{m_2+1}=1}^{n_{m_2+1}}\cdots$$

$$\sum_{k_N=1}^{n_N}\left(\prod_{\substack{i=1\\ i\neq m_1 \wedge i\neq m_2}}^{N}\alpha_{k_i}^{(i)}\right) f(\xi_1^{(k_1)},\ldots,\xi_N^{(k_N)}) - f_{m_1}\left(\xi_{m_1}^{(k_{m_1})}\right) - f_{m_2}\left(\xi_{m_2}^{(k_{m_2})}\right) - f_0 \quad (12)$$

Therefore a given N dimensional multivariate data set is decomposed into constant univariate and bivariate data set.

In this study we deal with the image retrieval problem and an image with RGB format is considered as a 3 dimensional data set. That is, when the concerned problem is related to image processing then N is taken as 3 since RGB format. Taking this situation into consideration we can simply write the constant term of image data as

$$f_0 = \sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2}\sum_{k_3=1}^{n_3=3}\alpha_{k_1}^{(1)}\alpha_{k_2}^{(2)}\alpha_{k_3}^{(3)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) \quad (13)$$

Here $f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)})$ is the pixel value of the image at the (k_1, k_2, k_3) coordinates. The α_{k_i} coefficients are weight factor as previous mentioned and they are taken as $\frac{1}{n_i}$, $i = 1, 2, 3$. These coefficients satisfy the normalization conditions given in (6). For the univariate components, we can write the following 3 formulas.

$$f_1\left(\xi_1^{(k_1)}\right) = \sum_{k_2=1}^{n_2}\sum_{k_3=1}^{3}\alpha_{k_2}^{(2)}\alpha_{k_3}^{(3)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) - f_0$$

$$f_2\left(\xi_2^{(k_2)}\right) = \sum_{k_1=1}^{n_1}\sum_{k_3=1}^{3}\alpha_{k_1}^{(1)}\alpha_{k_3}^{(3)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) - f_0$$

$$f_3\left(\xi_3^{(k_3)}\right) = \sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2}\alpha_{k_1}^{(1)}\alpha_{k_2}^{(2)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) - f_0 \quad (14)$$

To determine the bivariate terms of the given image we derive the formulas given as

$$f_{12}\left(\xi_1^{(k_{m_1})},\xi_2^{(k_{m_2})}\right) = \sum_{k_3=1}^{3}\alpha_{k_3}^{(3)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) - f_1\left(\xi_1^{(k_1)}\right) - f_2\left(\xi_2^{(k_2)}\right) - f_0$$

$$f_{13}\left(\xi_1^{(k_{m_1})},\xi_3^{(k_{m_3})}\right) = \sum_{k_2=1}^{n_2}\alpha_{k_2}^{(2)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) - f_1\left(\xi_1^{(k_1)}\right) - f_3\left(\xi_3^{(k_3)}\right) - f_0$$

$$f_{23}\left(\xi_2^{(k_{m_2})},\xi_3^{(k_{m_3})}\right) = \sum_{k_1=1}^{n_1}\alpha_{k_1}^{(1)} f(\xi_1^{(k_1)},\xi_2^{(k_2)},\xi_3^{(k_3)}) - f_2\left(\xi_2^{(k_2)}\right) - f_3\left(\xi_3^{(k_3)}\right) - f_0$$

(15)

Now, for an RGB image as a three-way array, the HDMR decomposition gives us one constant value as constant term, three vectors as univariate terms and three matrices as bivariate terms. We use the constant term and the univariate terms in our study. The next sections will explain how we use these terms in our two-step algorithm.

3 Image Clustering Using k-Means

In this study, before starting the retrieval process we intend to apply k-means clustering to be able to search the query image in a determined small part of the large-scale database instead of the whole database. Clustering approach leads to speed up the image retrieval process. The k-means clustering method is used with constant HDMR component, f_0. We choose this term to represent an image because it is even lower dimensional than the other HDMR terms.

When applying the clustering process, constant terms for every images in the database are calculated. These terms representing the images are clustered using k-means to obtain sub-databases. K-means clustering needs starting centroids which can be calculated by the following formula

$$\begin{aligned} &centroid_j = min(x_i) + (2j-1)\frac{max(x_i) - min(x_i)}{2K}, \\ &j = 1, \cdots, K \quad i = 1, \cdots, n \end{aligned} \tag{16}$$

where K is the number of clusters n is the number of images. This formula constructs starting centroids with equal intervals. These centroids are updated through the k-means clustering and clusters are constructed around these centroids. Hence these clusters are expected to contain uncertain number of similar images. They will be used in retrieval step to speed up the process.

4 Image Retrieval Through Univariate HDMR Term

The aim of this study is to retrieve images. For this purpose we use HDMR since it is an effective method in data partitioning and dimension reduction. Each image in the database are represented by low dimensional terms through HDMR therefore an efficient algorithm is constructed.

In retrieval process a query image which is not contained in the database is investigated by firstly determining the closest cluster by calculating its f_0 component as explained in the previous step. So we have a chance to work on only a small group of images. We compute univariate HDMR terms of both query image and the images in the determined cluster. Using these values, we calculate distances between the query image and the cluster's images and the images which gives the lowest distance according to squared Euclidean distance measure are retrieved expecting them to be the most similar images. To be able to assess better, the retrieved images should be ranked according to distance values.

In the retrieval algorithm, we use three different approaches based on univariate HDMR terms, f_1, f_2, f_3. These approaches include three different vectors generated with these terms. The first vector is plain f_1 while the second one is end-to-end concatenation $f_1 - f_2 - f_3$ and the third one is end-to-end concatenation $f_1 - f_3$. The reason why these vectors are constructed is that they are significantly low in dimension and good representers of high dimensional data.

To evaluate the performance of the proposed system, we use Discounted Cumulative Gain (DCG) which is a measure of ranking quality [16]. This is a method of accumulating gains of retrieved items by discounting each gain with respect to their rank in the result list. The gain of each retrieved image is taken as 1 if the image is of the same object with the query image and as 0 if it is a different object. To report the results as success rate, we take average of DCG values obtained from all the query experiments and we divided the average value by the value obtained for hypothetical ideal case. The next section will give the experimental findings of the proposed system.

5 Findings

Columbia Object Image Library COIL-100 is used for experiments to measure the success rates and computation times. This library has 100 different objects and 72 images per object. So there are 7200 images in the library. Images of an object are taken by rotating it 5 degrees around its own axis.

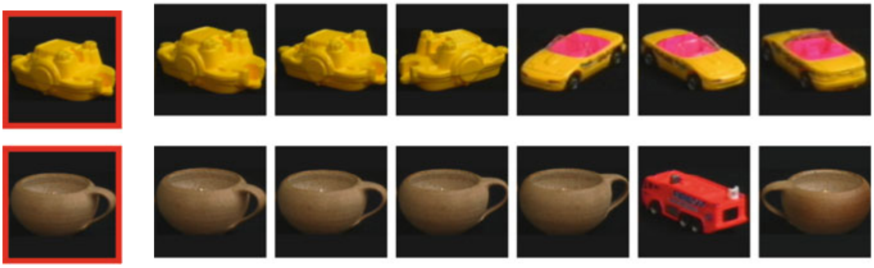
For experimental process we divide the database into training and test sets. The test set contains randomly selected 100 images where each belongs to different object. All the remaining images are reserved for training purpose. Before starting the retrieval process, the constant HDMR term is calculated for each image in the training set. Then the training set is divided into subsets by using the k-means clustering method according to the constant HDMR term. Then the constant HDMR term is calculated for each of the query images in the test set, and the image retrieval process is applied in corresponding subset for each image in the test set.

For the retrieval process the HDMR components, f_1, $f_1 - f_2 - f_3$ and $f_1 - f_3$ are applied separately. For analysis purpose, the tests are run for different number of clusters: 20, 30, 40, 50 and 60. Also for deep evaluation, different number of retrieved images are examined, which are 6, 2 and 1 images.

The results of the first retrieval experiments with f_1 term are given in Table 1. The values in this table are average of the success values for 100 test images expressed as percentage. As it is seen on the table, retrieval is considerably successful. Intra-table comparison indicates that success is increasing when the number of clusters decrease. Images of two randomly selected results from the first experiment are shown in Fig. 1.

Table 1 Success rates for the k-means clustering and f_1 retrieval

#Retrieved	20 Clusters	30 Clusters	40 Clusters	50 Clusters	60 Clusters
6	% 84.9	% 80.0	% 76.8	% 73.2	% 66.0
2	% 92.1	% 88.1	% 87.3	% 83.3	% 77.7
1	% 94.0	% 90.0	% 90.0	% 86.0	% 82.0



(a) Query (b) Retrieved images

Fig. 1 Two examples for the k-means clustering and f_1 retrieval

Table 2 Success rates for the k-means clustering and $f_1 - f_2 - f_3$ retrieval

#Retrieved	20 Clusters	30 Clusters	40 Clusters	50 Clusters	60 Clusters
6	% 82.9	% 79.0	% 74.2	% 70.8	% 64.8
2	% 93.7	% 88.5	% 86.7	% 82.4	% 76.6
1	% 96.0	% 92.0	% 91.0	% 87.0	% 82.0

(a) Query (b) Retrieved images

Fig. 2 Two examples for the k-means clustering and $f_1 - f_2 - f_3$ retrieval

Table 3 Success rates for the k-means clustering and $f_1 - f_3$ retrieval

#Retrieved	20 Clusters	30 Clusters	40 Clusters	50 Clusters	60 Clusters
6	% 85.7	% 80.7	% 77.6	% 74.1	% 66.9
2	% 94.1	% 89.5	% 87.9	% 83.9	% 78.4
1	% 96.0	% 93.0	% 91.0	% 87.0	% 83.0

(a) Query (b) Retrieved images

Fig. 3 Two examples for the k-means clustering and $f_- f_3$ retrieval

Second retrieval experiment is conducted using $f_1 - f_2 - f_3$ term. This term is tested with the expectation of better performance with respect to the previous one since it contains more features using three different aspects of image. The obtained results are given in Table 2 and two example retrieval results are shown in Fig. 2. Contrary to expectations $f_1 - f_2 - f_3$ term does not perform much better than f_1 term. New features extracted using the same data do not contribute to the performance, on the contrary they cause drawbacks with noise effect.

The last component of the proposed algorithms, which uses $f_1 - f_3$ term, also experimented and the results are tabulated on Table 3. Arbitrary two experiment results are illustrated in Fig. 3.

When we compare Tables 2 and 3, it is clearly seen that reducing the noise effect of the additional features works well in terms of performance. When Tables 1 and 3 are compared, it is observed that the success rates obtained for all clustering types in Table 3 is much higher than the rates given in Table 1. The three experiment results with different numbers of clusters are also given in terms of computation time in Table 4. In this case, $f_1 - f_3$ seems to be the best option at the image retrieval stage. Because the term $f_1 - f_3$ is a vector containing much fewer elements than the term $f_1 - f_2 - f_3$, it increases the speed efficiency.

Table 4 Computation time values in seconds

Retrieval	20 Clusters	30 Clusters	40 Clusters	50 Clusters	60 Clusters
f_1	1.51	0.77	0.58	0.47	0.38
$f_1 - f_2 - f_3$	1.24	0.92	0.69	0.57	0.46
$f_1 - f_3$	1.18	0.87	0.65	0.53	0.43

6 Conclusion

In this study, we tried to develop a new two-stage algorithm for image retrieval method. We used the HDMR method in both stages. In the first stage, we obtained the HDMR constant term then we use it in k-means method to be able to reach the image in a determined small part of the database. In the second stage, we retrieved the most similar images to a given query image from a cluster, which is determined in the first stage, utilizing the univariate HDMR terms. We Columbia Object Image Library to test the efficiency of the our algorithm.

Acknowledgements This work was supported by Research Fund of the Istanbul Technical University with project number 41411.

References

1. Demiralp, M.: High dimensional model representation and its application varieties. Mathematical Research. 9, 146–159 (2003).
2. Korkmaz E., Demiralp M.: Small scale High Dimensional Model Representation. New Aspects of Microelectronics, Nanoelectronics, Optoelectronic. 232–236 (2008).
3. Tunga, B., Demiralp, M.: Constancy maximization based weight optimization in high dimensional model representation. Numerical Algorithms. 52(3), 435–459 (2009).
4. Tunga, M.A., Demiralp, M.: A new approach for data partitioning through high dimensional model representation. International Journal of Computer Mathematics. 85(12), 1779–1792 (2008).
5. Tunga, M.A.: An approximation method to model multivariate interpolation problems: Indexing HDMR. Mathematical and Computer Modelling. 29, 743–765 (2012).
6. Tunga, B., Demiralp, M.: A novel approximation method for multivariate data partitioning fluctuation free integration based HDMR. Engineering Computations. 29(7), 743–765 (2012).
7. Tunga, M.A.: A Matrix Based Indexing HDMR Method for Multivariate Data Modelling. Journal of Mathematical Chemistry. 49(5), 1092–1114 (2011).
8. Kasap, Ö.Y., Tunga, M.A.: A polynomial modeling based algorithm in top-n recommendation. Expert Systems with Applications. 79, 313–321 (2017).
9. Altın, E.M., Tunga, B.: High dimensional model representation in image processing, Proceedings of the International Conference on Computational and Mathematical Methods in Science and Engineering, 55–64. IEEE, Cadiz, Spain, 2014.
10. Tunga, B. and Koçanaoğulları, A.: Digital image decomposition and contrast enhancement using high-dimensional model representation, Signal, Image and Video Processing, 12(2): 299–306, (2017).

11. Karcılı, A.and Tunga, B.: High dimensional model representation (HDMR) with clustering for image retrieval American Institute of Physics Conference Proceedings (AIP), 1798(020072), 1–8: (2017).
12. Karacan K. and Tunga B.: Breast Cancer Diagnosis in Digital Mammograms Using High Dimensional Model Representation, 12. İstanbul informatics congress, (IBK 2018 National), İstanbul, Turkey, (2018).
13. Karaca, E. and Tunga, M. A.: An interpolation-based texture and pattern preserving algorithm for inpainting color images, Expert Systems with Applications, 91: 223–234, (2018).
14. Hartigan J.A., Wong M.A.: Algorithm as 136: A k-means clustering algorithm, Applied Statistics, 28: 100–108, (1979).
15. Nene, S. A., Nayar, S. K., Murase, H.: Columbia Object Image Library (COIL-100), Technical Report, CUCS-006-96, (1996).
16. Discounted cumulative gain, www.en.wikipedia.org/wiki/Discounted_cumulative_gain

Neutrosophic Soft e-Open Maps, Neutrosophic Soft e-Closed Maps and Neutrosophic Soft e-Homeomorphisms in Neutrosophic Soft Topological Spaces

Palaniswamy Revathi, Kulandaivelu Chitirakala, and Appachi Vadivel

Abstract In this article, the concepts of N_sSe-open and N_sSe-closed mappings in neutrosophic soft topological spaces are introduced and their related properties are studied. Also, the work is developed to N_sS homeomorphism, N_sSe-homeomorphism, N_sSe-C homeomorphism and $N_sSeT_{\frac{1}{2}}$-space and some of their characteristics are discussed.

Keywords N_sSe-open map · N_sSe-closed map · N_sSe-homeomorphism · $N_sSeT_{\frac{1}{2}}$-space · N_sSe-C homeomorphism

1 Introduction

In Mathematics, the concept of fuzzy set was first introduced by Zadeh [1] and its topological structure was undertaken by Chang [2]. Atanassov [3–5] introduced intuitionistic fuzzy set in 1983 and its topological structure was introduced by Coker [6]. Molodstov [7] initiated the soft set theory as a new mathematical tool in 1999. Shabir and Naz [8] presented soft topological spaces in soft sets.

Smarandache [9] introduced the concepts of neutrosophy and neutrosophic set and its topological structure was given by Salama and Alblowi [10] in 2012. Maji [11] defined the Neutrosophic soft sets and the same was modified by Deli and

P. Revathi (✉)
Government Polytechnic College for Women, Madurai 625 011, India

Mathematics Section (FEAT) Annamalai University, Annamalai Nagar 608 002, India

K. Chitirakala
Department of Mathematics M.Kumarasamy College of Engineering, Karur 639 113, India

K. Chitirakala (✉) · A. Vadivel (✉)
Department of Mathematics, Annamalai University, Annamalai Nagar 608 002, India

A. Vadivel
PG and Research Department of Mathematics Government Arts College (Autonomous), Karur 639 005, India

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_4

Broumi [12]. Its topological structures were presented by Bera [13]. δ-open sets were defined by Saha [14] in fuzzy topological spaces and Vadivel et al. [15] in neutrosophic topological spaces. In 2019, Ahu Acikgoz and Ferhat Esenbel [16] defined neutrosophic soft δ-topology.

The notion of e-open sets were introduced by Ekici [17] in a general topology, Seenivasan et. al. [18] in fuzzy topological spaces, Chandrasekar et al. [19] in intuitionistic fuzzy topological spaces, Vadivel et al. [20] in neutrosophic topological spaces and recently, Revathi et al. [21] in neutrosophic soft topological spaces. In 2021, Vadivel et al. [22, 23] developed the concepts of neutrosophic e-Continuity, e-Irresolute maps, e-Open maps, e-Closed maps and e-Homeomorphisms in neutrosophic topological spaces. Recently, Revathi et al. [24] developed the concepts of neutrosophic soft e-Continuity and e-Irresolute maps.

The aim of this article is to introduce neutrosophic soft e-open and neutrosophic soft e-closed mappings in neutrosophic soft topological spaces. Moreover, neutrosophic soft e-homeomorphism, neutrosophic soft e-C homeomorphism and neutrosophic soft $eT_{\frac{1}{2}}$-space are introduced and their basic properties are obtained.

2 Preliminaries

The basic definitions and the properties of neutrosophic soft topological spaces are discussed in this section.

Definition 1 ([12]) Let Y be an initial universe, Q be a set of parameters. Let $P(Y)$ denote the set of all neutrosophic sets of Y. Then a neutrosophic soft set $(\tilde{H}, Q)$ over Y (in short, N_sSs) is defined by $(\tilde{H}, Q) = \{(q, \langle y, \mu_{\tilde{H}(q)}(y), \sigma_{\tilde{H}(q)}(y), \nu_{\tilde{H}(q)}(y)\rangle : y \in Y) : q \in Q\}$, where $\mu_{\tilde{H}(q)}(y), \sigma_{\tilde{H}(q)}(y), \nu_{\tilde{H}(q)}(y) \in [0, 1]$ are respectively called the degree of membership function, the degree of indeterminacy function and the degree of non-membership function of $\tilde{H}(q)$. Since the supremum of each μ, σ, ν is 1, the inequality $0 \leq \mu_{\tilde{H}(q)}(y) + \sigma_{\tilde{H}(q)}(y) + \nu_{\tilde{H}(q)}(y) \leq 3$ is obvious.

Definition 2 ([11, 13]) Let Y be an initial universe & the N_sSs's $(\tilde{H}, Q)$ & $(\tilde{G}, Q)$ are in the form $(\tilde{H}, Q) = \{(q, \langle y, \mu_{\tilde{H}(q)}(y), \sigma_{\tilde{H}(q)}(y), \nu_{\tilde{H}(q)}(y)\rangle : y \in Y) : q \in Q\}$ & $(\tilde{G}, Q) = \{(q, \langle y, \mu_{\tilde{G}(q)}(y), \sigma_{\tilde{G}(q)}(y), \nu_{\tilde{G}(q)}(y)\rangle : y \in Y) : q \in Q\}$, then

(i) $0_{(Y,Q)} = \{(q, \langle y, 0, 0, 1\rangle : y \in Y) : q \in Q\}$ and $1_{(Y,Q)} = \{(q, \langle y, 1, 1, 0\rangle : y \in Y) : q \in Q\}$

(ii) $(\tilde{H}, Q) \subseteq (\tilde{G}, Q)$ iff $\mu_{\tilde{H}(q)}(y) \leq \mu_{\tilde{G}(q)}(y)$, $\sigma_{\tilde{H}(q)}(y) \leq \sigma_{\tilde{G}(q)}(y)$ & $\nu_{\tilde{H}(q)}(y) \geq \nu_{\tilde{G}(q)}(y) : y \in Y : q \in Q$.

(iii) $(\tilde{H}, Q) = (\tilde{G}, Q)$ iff $(\tilde{H}, Q) \subseteq (\tilde{G}, Q)$ and $(\tilde{G}, Q) \subseteq (\tilde{H}, Q)$.

(iv) $(\tilde{H}, Q)^c = \{(q, \langle y, \nu_{\tilde{H}(q)}(y), 1 - \sigma_{\tilde{H}(q)}(y), \mu_{\tilde{H}(q)}(y)\rangle : y \in Y) : q \in Q\}$.

(v) $(\tilde{H}, Q) \cup (\tilde{G}, Q) = \{(q, \langle y, max(\mu_{\tilde{H}(q)}(y), \mu_{\tilde{G}(q)}(y)), max(\sigma_{\tilde{H}(q)}(y), \sigma_{\tilde{G}(q)}(y)), min(\nu_{\tilde{H}(q)}(y), \nu_{\tilde{G}(q)}(y))\rangle : y \in Y) : q \in Q\}$.

(vi) $(\tilde{H}, Q) \cap (\tilde{G}, Q) = \{(q, \langle y, min(\mu_{\tilde{H}(q)}(y), \mu_{\tilde{G}(q)}(y)), min(\sigma_{\tilde{H}(q)}(y), \sigma_{\tilde{G}(q)}(y)), max(\nu_{\tilde{H}(q)}(y), \nu_{\tilde{G}(q)}(y))\rangle : y \in Y) : q \in Q\}$.

Definition 3 ([13]) A neutrosophic soft topology (in short, N_sSt) on an initial universe Y is a family τ of neutrosophic soft subsets $(\tilde{H}, Q)$ of Y where Q is a set of parameters, satisfying

(i) $0_{(Y,Q)}, 1_{(Y,Q)} \in \tau$.
(ii) $[(\tilde{H}, Q) \cap (\tilde{G}, Q)] \in \tau$ for any $(\tilde{H}, Q), (\tilde{G}, Q) \in \tau$.
(iii) $\bigcup_{\rho \in A} (\tilde{H}, Q)_\rho \in \tau, \forall \ (\tilde{H}, Q)_\rho : \rho \in A \subseteq \tau$.

Then (Y, τ, Q) is known as a neutrosophic soft topological space (in short, N_sSts) and the τ elements are known as neutrosophic soft open sets (in short, N_sSos) in Y. A N_sSs $(\tilde{H}, Q)$ is known as a neutrosophic soft closed set (in short, N_sScs) if its complement $(\tilde{H}, Q)^c$ is N_sSos.

Definition 4 ([13]) Consider a N_sSts (Y, τ, Q) and a N_sSs $(\tilde{H}, Q)$ on Y. The neutrosophic soft interior of $(\tilde{H}, Q)$ (in short, $N_sSint(\tilde{H}, Q)$) and the neutrosophic soft closure of $(\tilde{H}, Q)$ (in short, $N_sScl(\tilde{H}, Q)$) are defined as

$$N_sSint(\tilde{H}, Q) = \bigcup\{(\tilde{G}, Q) : (\tilde{G}, Q) \subseteq (\tilde{H}, Q) \text{ and } (\tilde{G}, Q) \text{ is a } N_sSos \text{ in } Y\} \tag{1}$$

$$N_sScl(\tilde{H}, Q) = \bigcap\{(\tilde{G}, Q) : (\tilde{G}, Q) \supseteq (\tilde{H}, Q) \text{ and } (\tilde{G}, Q) \text{ is a } N_sScs \text{ in } Y\}. \tag{2}$$

Definition 5 ([13, 25]) Consider a N_sSts (Y, τ, Q) and a N_sSs $(\tilde{H}, Q)$ on Y. Then $(\tilde{H}, Q)$ is known as a neutrosophic soft *regular* (resp. *pre* & *semi*) open set (in short, N_sSros (resp. $N_sS\mathcal{P}os$, & N_sSSos)) if $(\tilde{H}, Q) = N_sSint(N_sScl(\tilde{H}, Q))$ (resp. $(\tilde{H}, Q) \subseteq N_sSint(N_sScl(\tilde{H}, Q))$ & $(\tilde{H}, Q) \subseteq N_sScl(N_sSint(\tilde{H}, Q))$). The complement of the respective open sets are their respective closed sets.

Definition 6 ([16]) A set $(\tilde{H}, Q)$ is known as a neutrosophic soft δ-open set (in short, $N_sS\delta os$) if $(\tilde{H}, Q) = N_sS\delta int(\tilde{H}, Q)$.

Definition 7 ([21]) A set $(\tilde{H}, Q)$ is known as a neutrosophic soft

(i) δ-pre open set (in short, $N_sS\delta\mathcal{P}os$) if $(\tilde{H}, Q) \subseteq N_sSint(N_sS\delta cl(\tilde{H}, Q))$.
(ii) δ-semi open set (in short, $N_sS\delta Sos$) if $(\tilde{H}, Q) \subseteq N_sScl(N_sS\delta int(\tilde{H}, Q))$.
(iii) e-open set (in short, N_sSeos) if $(\tilde{H}, Q) \subseteq N_sScl(N_sS\delta int(\tilde{H}, Q)) \cup N_sSint(N_sS\delta cl(\tilde{H}, Q))$.
(iv) e^*-open set (in short, N_sSe^*os) if $(\tilde{H}, Q) \subseteq N_sScl(N_sSint(N_sS\delta cl(\tilde{H}, Q)))$.

The complement of the respective open sets are their respective closed sets.

Definition 8 ([24]) Consider any two N_sSts's (Y, τ, Q) and (Z, σ, Q). A map $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ is called neutrosophic soft

(i) continuous (in short, N_sSCts) (resp. δ-continuous, $\delta\mathcal{S}$-continuous, $\delta\mathcal{P}$-continuous, e-continuous & e^*-continuous (in short, $N_sS\delta Cts$, $N_sS\delta\mathcal{S}Cts$, $N_sS\delta\mathcal{P}Cts$, N_sSeCts & N_sSe^*Cts)) if the inverse image of every N_sSos in (Z, σ, Q) is a N_sSos (resp. $N_sS\delta os$, $N_sS\delta\mathcal{S}os$, $N_sS\delta\mathcal{P}os$, N_sSeos & N_sSe^*os) in (Y, τ, Q).

(ii) e-irresolute (in short, N_sSeIrr) if the inverse image of every N_sSeos in (Z, σ, Q) is a N_sSeos in (Y, τ, Q).

3 Neutrosophic Soft *e*-Open Mapping

Definition 9 A mapping $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$, is neutrosophic soft *e-open* (resp. *open*, δ *open*, δ-*semi open*, δ-*pre open* & e^*-*open*) (in short, N_sSeO (resp. N_sSO, $N_sS\delta O$, $N_sS\delta\mathcal{S}O$, $N_sS\delta\mathcal{P}O$ & N_sSe^*O)) if the image of every N_sS open set of (Y, τ, Q) is N_sSeo (resp. N_sSo, $N_sS\delta o$, $N_sS\delta\mathcal{S}o$, $N_sS\delta\mathcal{P}o$ & N_sSe^*o) set in (Z, σ, Q).

Theorem 1 *The statements are hold but the converse need not be true. Every*

(i) $N_sS\delta O$ *mapping is a* N_sSO *mapping.*
(ii) N_sSO *mapping is a* $N_sS\delta\mathcal{S}O$ *mapping.*
(iii) N_sSO *mapping is a* $N_sS\delta\mathcal{P}O$ *mapping.*
(iv) $N_sS\delta\mathcal{S}O$ *mapping is a* N_sSeO *mapping.*
(v) $N_sS\delta\mathcal{P}O$ *mapping is a* N_sSeO *mapping.*
(vi) N_sSeO *mapping is a* N_sSe^*O *mapping.*

Example 1 Let $Y = \{y_1, y_2, y_3\} = \{z_1, z_2, z_3\} = Z$, $Q = \{q_1, q_2\}$ and N_sSs's $(\tilde{H}_1, Q)$ in Y and $(\tilde{G}_1, Q)$ & $(\tilde{G}_2, Q)$ in Z are defined as

$$(\tilde{H}_1, q_1) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.2, 0.5, 0.8)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_1, q_2) = \{\langle y_1, (0.3, 0.4, 0.7)\rangle, \langle y_2, (0.4, 0.4, 0.6)\rangle, \langle y_3, (0.4, 0.5, 0.5)\rangle\}$$
$$(\tilde{G}_1, q_1) = \{\langle z_1, (0.2, 0.5, 0.8)\rangle, \langle z_2, (0.2, 0.5, 0.8)\rangle, \langle z_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_2) = \{\langle z_1, (0.3, 0.4, 0.7)\rangle, \langle z_2, (0.4, 0.4, 0.6)\rangle, \langle z_3, (0.4, 0.5, 0.5)\rangle\}$$
$$(\tilde{G}_2, q_1) = \{\langle z_1, (0.4, 0.5, 0.6)\rangle, \langle z_2, (0.4, 0.5, 0.6)\rangle, \langle z_3, (0.5, 0.5, 0.5)\rangle\}$$
$$(\tilde{G}_2, q_2) = \{\langle z_1, (0.4, 0.5, 0.6)\rangle, \langle z_2, (0.5, 0.5, 0.6)\rangle, \langle z_3, (0.5, 0.5, 0.5)\rangle\}$$

Then we have $\tau = \{0_{(Y,Q)}, 1_{(Y,Q)}, (\tilde{H}_1, Q)\}$ and $\sigma = \{0_{(Z,Q)}, 1_{(Z,Q)}, (\tilde{G}_1, Q), (\tilde{G}_2, Q)\}$. Let $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ be an identity mapping. Then, $(\tilde{H}_1, Q)$ is N_sSO (resp. N_sSeO) mapping in Y but not $N_sS\delta O$ (resp. $N_sS\delta\mathcal{S}O$) mapping in Z.

Example 2 Let $Y = \{y_1, y_2, y_3\} = \{z_1, z_2, z_3\} = Z$, $Q = \{q_1, q_2\}$ and N_sSs's $(\tilde{H}_1, Q)$ in Y and $(\tilde{G}_1, Q)$, $(\tilde{G}_2, Q)$ & $(\tilde{G}_3, E)$ in Z are defined as

$$(\tilde{H}_1, q_1) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.4, 0.5, 0.6)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_1, q_2) = \{\langle y_1, (0.3, 0.4, 0.7)\rangle, \langle y_2, (0.5, 0.5, 0.7)\rangle, \langle y_3, (0.5, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_1) = \{\langle z_1, (0.2, 0.5, 0.8)\rangle, \langle z_2, (0.3, 0.5, 0.7)\rangle, \langle z_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_2) = \{\langle z_1, (0.3, 0.4, 0.7)\rangle, \langle z_2, (0.4, 0.5, 0.7)\rangle, \langle z_3, (0.5, 0.4, 0.6)\rangle\}$$
$$(\tilde{G}_2, q_1) = \{\langle z_1, (0.1, 0.5, 0.9)\rangle, \langle z_2, (0.1, 0.5, 0.9)\rangle, \langle z_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_2, q_2) = \{\langle z_1, (0.2, 0.3, 0.8)\rangle, \langle z_2, (0.3, 0.5, 0.8)\rangle, \langle z_3, (0.4, 0.4, 0.7)\rangle\}$$
$$(\tilde{G}_3, q_1) = \{\langle z_1, (0.2, 0.5, 0.8)\rangle, \langle z_2, (0.4, 0.5, 0.6)\rangle, \langle z_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_3, q_2) = \{\langle z_1, (0.3, 0.4, 0.7)\rangle, \langle z_2, (0.5, 0.5, 0.7)\rangle, \langle z_3, (0.5, 0.5, 0.6)\rangle\}$$

Then we have $\tau = \{0_{(Y,Q)}, 1_{(Y,Q)}, (\tilde{H}_1, Q)\}$ and $\sigma = \{0_{(Z,Q)}, 1_{(Z,Q)}, (\tilde{G}_1, Q), (\tilde{G}_2, Q)\}$. Let $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ be an identity mapping. Then, $(\tilde{H}_1, Q)$ is $N_sS\delta\mathcal{S}O$ (resp. $N_sS\delta\mathcal{P}O$ & N_sSeO) mapping in Y but not N_sSO (resp. N_sSO & $N_sS\delta\mathcal{P}O$) mapping in Z.

Example 3 Let $Y = \{y_1, y_2\} = \{z_1, z_2\} = Z$, $Q = \{q_1, q_2\}$ and N_sSs's $(\tilde{H}_1, Q)$ in Y and $(\tilde{G}_1, Q)$ & $(\tilde{G}_2, Q)$ in Z are defined as

$$(\tilde{H}_1, q_1) = \{\langle y_1, (0.3, 0.5, 0.7)\rangle, \langle y_2, (0.5, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_1, q_2) = \{\langle y_1, (0.4, 0.5, 0.6)\rangle, \langle y_2, (0.4, 0.4, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_1) = \{\langle z_1, (0.3, 0.5, 0.5)\rangle, \langle z_2, (0.2, 0.5, 0.5)\rangle\}$$
$$(\tilde{G}_1, q_2) = \{\langle z_1, (0.4, 0.4, 0.5)\rangle, \langle z_2, (0.3, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_2, q_1) = \{\langle z_1, (0.3, 0.5, 0.7)\rangle, \langle z_2, (0.5, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_2, q_2) = \{\langle z_1, (0.4, 0.5, 0.6)\rangle, \langle z_2, (0.4, 0.4, 0.6)\rangle\}$$

Then we have $\tau = \{0_{(Y,Q)}, 1_{(Y,Q)}, (\tilde{H}_1, Q)\}$ and $\sigma = \{0_{(Z,Q)}, 1_{(Z,Q)}, (\tilde{G}_1, Q)\}$. Let $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ be an identity mapping. Then, $(\tilde{H}_1, Q)$ is N_sSe^*O mapping in Y but not N_sSeO mapping in Z.

Remark 1 The diagram shows N_sSeO mapping's in N_sSts.

$$\begin{tikzcd}
 & N_sS\delta O \text{ map} \arrow[d] & \\
 & N_sSO \text{ map} \arrow[dl] \arrow[dr] & \\
N_sS\delta SO \text{ map} \arrow[dr] & & N_sS\delta\mathcal{P}O \text{ map} \arrow[dl] \\
 & N_sSeO \text{ map} \arrow[r] & N_sSe^*O \text{ map}
\end{tikzcd}$$

Theorem 2 *A mapping* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSeO *iff for every* N_sSs $(\tilde{H}, Q)$ *of* (Y, τ, Q), $\mathfrak{f}(N_sSint(\tilde{H}, Q)) \subseteq N_sSeint(\mathfrak{f}(\tilde{H}, Q))$.

Theorem 3 *Consider a* N_sSeO *mapping* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$. *Then,* $N_sSint(\mathfrak{f}^{-1}(\tilde{H}, Q)) \subseteq \mathfrak{f}^{-1}(N_sSeint(\tilde{H}, Q))$ *for every* N_sSs $(\tilde{H}, Q)$ *of* (Z, σ, Q).

Theorem 4 *A mapping* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSeO *iff for each* N_sSs $(\tilde{G}, Q)$ *of* (Z, σ, Q) *and for each* N_sScs $(\tilde{H}, Q)$ *of* (Y, τ, Q) *containing* $\mathfrak{f}^{-1}(\tilde{G}, Q)$, *there is a* N_sSecs $(\tilde{A}, Q)$ *of* (Z, σ, Q) *such that* $(\tilde{G}, Q) \subseteq (\tilde{H}, Q)$ *and* $\mathfrak{f}^{-1}(\tilde{A}, Q) \subseteq (\tilde{H}, Q)$.

Theorem 5 *A mapping* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSeO *iff* $\mathfrak{f}^{-1}(N_sS\, ecl(\tilde{G}, Q)) \subseteq N_sScl(\mathfrak{f}^{-1}(\tilde{G}, Q))$ *for every* N_sSs $(\tilde{G}, Q)$ *of* (Z, σ, Q).

Theorem 6 *Let* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *be two neutrosophic soft mappings and* $\mathfrak{g} \circ \mathfrak{f} : (Y, \tau, Q) \to (P, \rho, Q)$ *be* N_sSeO. *If* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *is* N_sSeIrr, *then* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSeO *mapping.*

Theorem 7 *If* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSO *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *is* N_sSeO *mappings, then* $\mathfrak{g} \circ \mathfrak{f} : (Y, \tau, Q) \to (P, \rho, Q)$ *is* N_sSeO.

4 Neutrosophic Soft *e*-Closed Mapping

Definition 10 A mapping $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ is neutrosophic soft *e-closed* (resp. *closed*, *δ closed*, *δ-semi closed*, *δ-pre closed* & *e***-closed*) (in short, N_sSeC (resp.N_sSC, $N_sS\delta C$, $N_sS\delta SC$, $N_sS\delta\mathcal{P}C$ & N_sSe^*C)) if the image of every N_sS closed set of (Y, τ, Q) is N_sSec (resp. N_sSc, $N_sS\delta c$, $N_sS\delta Sc$, $N_sS\delta\mathcal{P}c$ & N_sSe^*c) set in (Z, σ, Q).

Theorem 8 *The statements are hold but the converse need not be true. Every*

(i) $N_sS\delta C$ *mapping is a* N_sSC *mapping.*
(ii) N_sSC *mapping is a* $N_sS\delta\mathcal{S}C$ *mapping.*
(iii) N_sSC *mapping is a* $N_sS\delta\mathcal{P}C$ *mapping.*
(iv) $N_sS\delta\mathcal{S}C$ *mapping is a* N_sSeC *mapping.*
(v) $N_sS\delta\mathcal{P}C$ *mapping is a* N_sSeC *mapping.*
(vi) N_sSeC *mapping is a* N_sSe^*C *mapping.*

Example 4 In Example 1, $(\tilde{H}_1, Q)^c$ is N_sSC (resp. N_sSeC) mapping in Y but not $N_sS\delta C$ (resp. $N_sS\delta\mathcal{S}C$) mapping in Z.

Example 5 In Example 2, $(\tilde{H}_1, Q)^c$ is $N_sS\delta\mathcal{S}C$ (resp. $N_sS\delta\mathcal{P}C$ & N_sSeC) mapping in Y but not N_sSC (resp. N_sSC & $N_sS\delta\mathcal{P}C$) mapping in Z.

Example 6 In Example 3, $(\tilde{H}_1, Q)^c$ is N_sSe^*C mapping in Y but not N_sSeC mapping in Z.

Remark 2 The diagram shows N_sSeC mapping's in N_sSts.

$$\begin{tikzcd}
 & N_sS\delta C \text{ map} \arrow[d] & \\
 & N_sSC \text{ map} \arrow[dl] \arrow[dr] & \\
N_sS\delta\mathcal{S}C \text{ map} \arrow[dr] & & N_sS\delta\mathcal{P}C \text{ map} \arrow[dl] \\
 & N_sSeC \text{ map} \arrow[r] & N_sSe^*C \text{ map}
\end{tikzcd}$$

Theorem 9 *A mapping* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSeC *iff for each* N_sSs $(\tilde{G}, Q)$ *of* (Z, σ, Q) *and for each* N_sSos $(\tilde{H}, Q)$ *of* (Y, τ, Q) *containing* $\mathfrak{f}^{-1}(\tilde{G}, Q)$, *there is a* N_sSeos $(\tilde{A}, Q)$ *of* (Z, σ, Q) *such that* $(\tilde{G}, Q) \subseteq (\tilde{A}, Q)$ *and* $\mathfrak{f}^{-1}(\tilde{A}, Q) \subseteq (\tilde{H}, Q)$.

Theorem 10 *If* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSC *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *is* N_sSeC. *Then* $\mathfrak{g} \circ \mathfrak{f} : (Y, \tau, Q) \to (P, \rho, Q)$ *is* N_sSeC.

Theorem 11 *If* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is* N_sSeC *map, then* $N_sSecl(\mathfrak{f}(\tilde{H}, Q)) \subseteq \mathfrak{f}(N_sScl(\tilde{H}, Q))$.

Theorem 12 *Let* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *are* N_sSeC *mappings. If every* N_sSecs *of* (Z, σ, Q) *is* N_sScs, *then* $\mathfrak{g} \circ \mathfrak{f} : (Y, \tau, Q) \to (P, \rho, Q)$ *is* N_sSeC.

Theorem 13 *Let $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ be a bijective mapping. Then the following statements are equivalent:*

(i) $\mathfrak{f}$ *is a N_sSeO mapping.*
(ii) $\mathfrak{f}$ *is a N_sSeC mapping.*
(iii) $\mathfrak{f}^{-1}$ *is N_sSeCts mapping.*

5 Neutrosophic Soft *e*-Homeomorphism

Definition 11 A bijection $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ is called a N_sS homeomorphism (in short N_sSHom) if $\mathfrak{f}$ and $\mathfrak{f}^{-1}$ are N_sSCts mappings.

Definition 12 A bijection $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ is called a N_sSe-homeomorphism (in short N_sSeHom) if $\mathfrak{f}$ and $\mathfrak{f}^{-1}$ are N_sSeCts.

Theorem 14 *Each N_sSHom is a N_sSeHom. But not conversely.*

Example 7 Let $Y = \{y_1, y_2, y_3\} = \{z_1, z_2, z_3\} = Z$, $Q = \{q_1, q_2\}$ and N_sSs's $(\tilde{H}_1, Q)$, $(\tilde{H}_2, Q)$ & $(\tilde{H}_3, Q)$ in Y and $(\tilde{G}_1, Q)$ in Z are defined as

$$(\tilde{H}_1, q_1) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.3, 0.5, 0.7)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_1, q_2) = \{\langle y_1, (0.3, 0.5, 0.7)\rangle, \langle y_2, (0.2, 0.5, 0.6)\rangle, \langle y_3, (0.4, 0.4, 0.6)\rangle\}$$
$$(\tilde{H}_2, q_1) = \{\langle y_1, (0.1, 0.5, 0.9)\rangle, \langle y_2, (0.1, 0.5, 0.9)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_2, q_2) = \{\langle y_1, (0.2, 0.4, 0.8)\rangle, \langle y_2, (0.2, 0.5, 0.7)\rangle, \langle y_3, (0.3, 0.4, 0.7)\rangle\}$$
$$(\tilde{H}_3, q_1) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.4, 0.5, 0.6)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_3, q_2) = \{\langle y_1, (0.3, 0.5, 0.6)\rangle, \langle y_2, (0.3, 0.5, 0.6)\rangle, \langle y_3, (0.5, 0.4, 0.5)\rangle\}$$
$$(\tilde{G}_1, q_1) = \{\langle z_1, (0.2, 0.5, 0.8)\rangle, \langle z_2, (0.4, 0.5, 0.6)\rangle, \langle z_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_2) = \{\langle z_1, (0.3, 0.5, 0.6)\rangle, \langle z_2, (0.3, 0.5, 0.6)\rangle, \langle z_3, (0.5, 0.4, 0.5)\rangle\}$$

Then we have $\tau = \{0_{(Y,Q)}, 1_{(Y,Q)}, (\tilde{H}_1, Q), (\tilde{H}_2, Q)\}$ and $\sigma = \{0_{(Z,Q)}, 1_{(Z,Q)}, (\tilde{G}_1, Q)\}$. Let $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ be an identity mapping. Then $\mathfrak{f}$ is N_sSeHom but not N_sSHom.

Theorem 15 *Consider a bijective mapping $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$. If $\mathfrak{f}$ is N_sSeCts, then the following statements are equivalent:*

(i) $\mathfrak{f}$ *is a N_sSeC mapping.*
(ii) $\mathfrak{f}$ *is a N_sSeO mapping.*
(iii) $\mathfrak{f}^{-1}$ *is a N_sSeHom.*

Definition 13 A N_sSts (Y, τ, Q) is known as a neutrosophic soft $eT_{\frac{1}{2}}$ (in short, $N_sSeT_{\frac{1}{2}}$)-space if every N_sSecs is N_sSc in (Y, τ, Q).

Theorem 16 *Let* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *be a* N_sSeHom. *Then* $\mathfrak{f}$ *is a* N_sSHom *if* (Y, τ, Q) *and* (Z, σ, Q) *are* $N_sSeT_{\frac{1}{2}}$*-space.*

Theorem 17 *Let* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *be a* N_sSts. *If* (Z, σ, Q) *is a* $N_sSeT_{\frac{1}{2}}$*-space, Then the following statements are equivalent:*

(i) $\mathfrak{f}$ *is* N_sSeC *mapping.*
(ii) If $(\tilde{H}, Q)$ *is a* N_sSos *in* (Y, τ, Q), *then* $\mathfrak{f}(\tilde{H}, Q)$ *is* N_sSeos *in* (Z, σ, Q).
(iii) $\mathfrak{f}(N_sSint(\tilde{H}, Q)) \subseteq N_sScl(N_sSint(\mathfrak{f}(\tilde{H}, Q)))$ *for every* N_sSs $(\tilde{H}, Q)$ *in* (Y, τ, Q).

Theorem 18 *Let* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *be* N_sSeC, *where* (Y, τ, Q) *and* (P, ρ, Q) *are two* N_sSts*'s and* (Z, σ, Q) *a* $N_sSeT_{\frac{1}{2}}$*-space, then the composition* $\mathfrak{g} \circ \mathfrak{f}$ *is* N_sSeC.

Theorem 19 *Let* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *be two* N_sSts*'s. Then the following are true:*

(i) If $\mathfrak{g} \circ \mathfrak{f}$ *is* N_sSeO *and* $\mathfrak{f}$ *is* N_sSCts, *then* $\mathfrak{g}$ *is* N_sSeO.
(ii) If $\mathfrak{g} \circ \mathfrak{f}$ *is* N_sSO *and* $\mathfrak{g}$ *is* N_sSeCts, *then* $\mathfrak{f}$ *is* N_sSeO.

6 Neutrosophic Soft *e*-C Homeomorphism

Definition 14 A bijection $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ is called a N_sSe-C homeomorphism (in short, $N_sSeCHom$) if $\mathfrak{f}$ and $\mathfrak{f}^{-1}$ are N_sSeIrr mappings.

Theorem 20 *Each* $N_sSeCHom$ *is a* N_sSeHom. *But not conversely.*

Example 8 Let $Y = \{y_1, y_2, y_3\} = \{z_1, z_2, z_3\} = Z$, $Q = \{q_1, q_2\}$ and N_sSs's $(\tilde{H}_1, Q)$, $(\tilde{H}_2, Q)$ & $(\tilde{H}_3, Q)$ in Y and $(\tilde{G}_1, Q)$ in Z are defined as

$$(\tilde{H}_1, q_1) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.3, 0.5, 0.7)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_1, q_2) = \{\langle y_1, (0.3, 0.5, 0.8)\rangle, \langle y_2, (0.2, 0.5, 0.8)\rangle, \langle y_3, (0.4, 0.5, 0.5)\rangle\}$$
$$(\tilde{H}_2, q_1) = \{\langle y_1, (0.1, 0.5, 0.9)\rangle, \langle y_2, (0.1, 0.5, 0.9)\rangle, \langle y_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_2, q_2) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.2, 0.5, 0.9)\rangle, \langle y_3, (0.3, 0.5, 0.7)\rangle\}$$
$$(\tilde{H}_3, q_1) = \{\langle y_1, (0.2, 0.5, 0.8)\rangle, \langle y_2, (0.2, 0.5, 0.6)\rangle, \langle y_3, (0.3, 0.5, 0.6)\rangle\}$$
$$(\tilde{H}_3, q_2) = \{\langle y_1, (0.1, 0.5, 0.9)\rangle, \langle y_2, (0.2, 0.5, 0.7)\rangle, \langle y_3, (0.3, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_1) = \{\langle z_1, (0.2, 0.5, 0.8)\rangle, \langle z_2, (0.2, 0.5, 0.8)\rangle, \langle z_3, (0.4, 0.5, 0.6)\rangle\}$$
$$(\tilde{G}_1, q_2) = \{\langle z_1, (0.3, 0.5, 0.8)\rangle, \langle z_2, (0.3, 0.5, 0.8)\rangle, \langle z_3, (0.4, 0.5, 0.5)\rangle\}$$

Then we have $\tau = \{0_{(Y,Q)}, 1_{(Y,Q)}, (\tilde{H}_1, Q), (\tilde{H}_2, Q)\}$ and $\sigma = \{0_{(Z,Q)}, 1_{(Z,Q)}, (\tilde{G}_1, Q)\}$. Let $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ be a mapping defined as $\mathfrak{f}(y_1) = z_1, \mathfrak{f}(y_2) = z_1$ & $\mathfrak{f}(y_3) = z_3$. Then $\mathfrak{f}$ is N_sSeHom but not $N_sSeCHom$.

Theorem 21 *If* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *is a* $N_sSeCHom$*, then* $N_sSecl(\mathfrak{f}^{-1}(\tilde{G}, Q)) \subseteq \mathfrak{f}^{-1}(N_sScl(\tilde{G}, Q))$ *(resp.* $N_sSecl(\mathfrak{f}^{-1}(\tilde{G}, Q)) = \mathfrak{f}^{-1}(N_sSecl(\tilde{G}, Q))$*) for each* N_sSs $(\tilde{G}, Q)$ *in* (Z, σ, Q)*.*

Theorem 22 *If* $\mathfrak{f} : (Y, \tau, Q) \to (Z, \sigma, Q)$ *and* $\mathfrak{g} : (Z, \sigma, Q) \to (P, \rho, Q)$ *are* $N_sSeCHom$*'s, then* $\mathfrak{g} \circ \mathfrak{f}$ *is a* $N_sSeCHom$*.*

7 Conclusion

In this paper, the concepts of N_sSeO and N_sSeC mappings in N_sSts were discussed. Furthermore, the work was extended to include N_sSHom, N_sSeHom and $N_sSeT_{\frac{1}{2}}$-space. In addition, the study demonstrated $N_sSeCHom$ and derived some of its related characteristics. This work can be used to investigate neutrosophic soft e-compactness, neutrosophic soft e-connectedness and neutrosophic soft contra e-continuous functions in future.

References

1. Zadeh, L.A.: Fuzzy sets. Information and control. 8, 338–353 (1965).
2. Chang, C.L.: Fuzzy topological space. J. Math. Anal. Appl. 24, 182–190 (1968).
3. Atanassov, K., Stoeva, S.: Intuitionistic fuzzy sets, in : polish symp. on interval and fuzzy mathematics. Poznan. 23–26 (1983).
4. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets Syst. 20, 87–96 (1986).
5. Atanassov, K.: Review and new results on intuitionistic fuzzy sets. Preprint IM-MFAIS. Sofia, 1–88 (1988).
6. Coker, D.: An introduction to intuitionistic topological spaces. Fuzzy Sets Syst. 88, 81–89 (1997).
7. Molodtsov, D.: Soft set theory-first results. Comput. Math. Appl. 37, 19–31 (1999).
8. Shabir, M., Naz, M.: On soft topological spaces. Comput. Math. Appl. 61, 1786–1799 (2011).
9. Smarandache, F.: Neutrosophic set: A generalization of the intuitionistic fuzzy sets. Inter. J. Pure Appl. Math. 24, 287–297 (2005).
10. Salama, A.A., Alblowi, S.A.: Neutrosophic Set and Neutrosophic Topological Spaces. IOSR Journal of Mathematics. 3 (4), 31–35 (2012).
11. Maji, P.K.: Neutrosophic soft set. Ann. Fuzzy Math. Inform. 5, 157–168 (2013).
12. Deli, I., Broumi, S.: Neutrosophic soft relations and some properties. Ann. Fuzzy Math. Inform. 9, 169–182 (2015).
13. Bera, T., Mahapatra, N.K.: Introduction to neutrosophic soft topological space. Opsearch. 54, 841–867 (2017).
14. Saha, S.: Fuzzy δ-continuous mappings. J. Math. Anal. Appl. 126, 130–142 (1987).
15. Vadivel, A., Seenivasan, M., John Sundar, C.: An introduction to δ-open sets in a neutrosophic topological spaces. J. Phys. Conf. Ser. 1724, 012011 (2021).
16. Acikgoz, A., Esenbel, F.: Neutrosophic soft δ-topology and neutrosophic soft compactness. AIP Conf. Proc. 2183, 030002 (2019).
17. Ekici, E.: On e-open sets, $\mathcal{DP}^{\star}$-sets and $\mathcal{DP}\bullet^{\star}$-sets and decomposition of continuity. Arab J Sci Eng. 33 (2A), 269–282 (2008).

18. Seenivasan, V., Kamala, K.: Fuzzy e-continuity and fuzzy e-open sets. Ann. Fuzzy Math. Inform. 8, 141–148 (2014).
19. Chandrasekar, V., Sobana, D., Vadivel, A.: On Fuzzy e-open Sets, fuzzy e-continuity and fuzzy e-compactness in intuitionistic fuzzy topological spaces. Sahand Communications in Mathematical Analysis (SCMA). 12 (1), 131–153 (2018).
20. Vadivel, A., John Sundar, C., Thangaraja, P.: Neutrosophic e-open sets in a neutrosophic topological spaces (Submitted).
21. Revathi, P., Chitirakala, K., Vadivel, A.: Neutrosophic Soft e-open sets in Neutrosophic Soft Topological Spaces (Submitted).
22. Vadivel, A., Thangaraja, P., John Sundar, C.: Neutrosophic e-Continuous Maps and Neutrosophic e-Irresolute Maps. TURCOMAT. 12 (1S), 369–375 (2021).
23. Vadivel, A., Thangaraja, P., John Sundar, C.: Neutrosophic e-Open Maps, Neutrosophic e-Closed Maps and Neutrosophic e-Homeomorphisms in Neutrosophic Topological Spaces (Submitted).
24. Revathi, P., Chitirakala, K., Vadivel, A.: Neutrosophic Soft e-Continuous Maps and Neutrosophic Soft e-Irresolute Maps (Submitted).
25. Ebenanjar, P.E., Immaculate, H.J., Sivaranjani, K.: Introduction to neutrosophic soft topological spatial region. Neutrosophic Sets and Systems. 31, 297–304 (2020).
26. Aras, C.G., Sadi Bayramov, S.: Neutrosophic Soft Continuity in Neutrosophic Soft Topological Spaces. Filomat. 34:10, 3495–3506 (2020).

Jointly Type-II Censored Length-Biased Exponential Distributions

Çağatay Çetinkaya

Abstract This paper deals with the jointly Type-II censored length-biased exponential populations. In this study, after introducing the jointly Type-II censoring scheme, we first obtained the maximum likelihood estimations of the unknown scale parameters with their asymptotic confidence intervals. Then, the Bayesian estimations of the parameters are obtained by using the importance sampling method. Further, the highest posterior density credible intervals of the Bayesian estimations are provided. The simulation studies are performed to evaluate the performances of the estimation methods. Finally, a numerical example is used to illustrate the theoretical outcomes.

Keywords Bayesian inference · Jointly censoring · Length-biased exponential distribution · Maximum likelihood estimation

1 Introduction

In many reliability studies, the lifetimes of the components/units may not be recorded exactly. Since the components or units are lost or removed from the experiments before they fail, the censored data sets occur. Censoring schemes (CSs) are preferable for the experimenters to provide time and cost-efficiency. The censoring schemes are constructed based on the main two classifications, Type I and Type II censoring. While the number of uncensored observations is a random variable in the Type-I censoring, the number of uncensored observations is fixed in Type-II censoring.

In the literature, most of the conventional censoring schemes deal with one sample problem. However, experimenters may need joint censoring schemes to decide for conducting comparative life tests of units. For this purpose, joint censoring schemes have been suggested in the literature. Balakrishnan and Rasouli [8] considered two exponential populations under joint Type-II censoring. Then, Rasouli and Balakrishnan [16] considered the joint progressive Type-II censored exponential populations.

Ç. Çetinkaya (✉)
Department of Accounting and Taxation, Bingöl University, Bingöl, Turkey
e-mail: ccetinkaya@bingol.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_5

The Type-II censoring scheme is the most common censoring scheme in life testing experiments. In this scheme, r units in a random sample of size n $(r < n)$ are observed and the test terminates with r-th failure. In the same manner, let's suppose two independent samples of size n and m are placed on the life testing experiment. Following recording the successive failure times and their corresponding product types, the experiment terminates with the r-th failure. This censoring plan is named a jointly Type-II censoring scheme. The jointly censored samples are handled in various studies by Ashour and Abo-Kasem [3, 6], Balakrishnan and Feng [9], Balakrishnan et al. [7], Doostparast et al. [13], Abo-Kasem et al. [1], Krishna and Goel [15]. In these studies, mostly Type-II and related CSs are handled.

On the other hand, length-biased distributions have a great importance in reliability since they provide greater flexibility in modeling data. Dara and Ahmed [12] proposed a new extension of exponential distribution denoted by moment exponential distribution. Then, it is named as the length-biased exponential (LBE) distribution by some authors such as Akhter et al. [2]. The probability density (pdf) and distribution (cdf) function of the LBE distribution are given as

$$f(t) = \frac{t}{\theta^2} \exp\left\{-\frac{t}{\theta}\right\},\ x > 0,\ \theta > 0 \tag{1}$$

$$F(t) = 1 - \left[1 + \frac{t}{\theta}\right] \exp\left\{-\frac{t}{\theta}\right\} \tag{2}$$

where θ is the scale parameter. Dara and Ahmad [12] proved that the LBE distribution is more flexible than the exponential distribution. Unlike the exponential distribution with a constant failure rate, the LBE distribution is an increasing failure rate (IFR) class of distribution. It is known that an IFR component has a better chance of surviving any shorter period and the worse chance of surviving any longer period.

In this paper, we deal with the jointly Type-II censored length-biased exponential populations. For this purpose, we first consider the maximum likelihood estimators (MLE) of the unknown scale parameters along with their asymptotic confidence intervals. Then, the Bayesian inference procedure is provided with the highest posterior density credible intervals. In Bayesian estimations, importance sampling method is used. The whole theoretical studies are illustrated with simulation and real data studies.

2 Model Description and Parameter Estimation

Let's suppose that lifetimes of n test units $X_1, X_2, \ldots, X_n$ and m test units $Y_1, Y_2, \ldots, Y_m$ be independently and identically distributed (i.i.d.) random variables (r.v.) having LBE distributions with scale parameters θ and λ, respectively with the density and distribution functions which are given in the following

$$f(x) = \frac{x}{\theta^2}\exp\left\{-\frac{x}{\theta}\right\}, \quad F(x) = 1 - \left[1 + \frac{x}{\theta}\right]\exp\left\{-\frac{x}{\theta}\right\}$$
$$g(y) = \frac{y}{\lambda^2}\exp\left\{-\frac{y}{\lambda}\right\}, \quad G(y) = 1 - \left[1 + \frac{y}{\lambda}\right]\exp\left\{-\frac{y}{\lambda}\right\} \tag{3}$$

where $x > 0, y > 0, \theta > 0, \lambda > 0$.

Then, it is assumed that $W_{(1)}, W_{(2)}, \ldots, W_{(N)}$ where $N = n + m$, denote the combined order form of test samples $(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_m)$. Then, the experiment stops with the $r-$th failure, namely at the time point $W_{(r)}$. Further, let's define an indicator variable, Z_i $(\forall \quad i = 1, 2, \ldots, r)$ that takes two values either 1 or 0 depending on whether $W_{(i)}$ is an X or Y failure, respectively.

We will prefer to use the ranking notation W_i instead of $W_{(i)}$ for simplicity in the following of the study.

2.1 Maximum Likelihood Estimation

In this subsection, we consider the maximum likelihood estimations of unknown scale parameters of the LBE distributions under jointly censoring. The likelihood function of the observed sample can be obtained by using the Eq. (3) as given in the following

$$L(\boldsymbol{w}, \boldsymbol{z}, \theta, \lambda) = C \times \prod_{i=1}^{r} \{f(w_i)\}^{z_i} \{g(w_i)\}^{(1-z_i)} \left\{\bar{F}(w_r)\right\}^{n-n_r} \left\{\bar{G}(w_r)\right\}^{m-m_r} \tag{4}$$

where $\bar{F} = 1 - F$, $\bar{G} = 1 - G$, $C = \frac{n!m!}{(n-n_r)!(m-m_r)!}$, $n_r = \sum_{i=1}^{r} z_i$ and $m_r = \sum_{i=1}^{r}(1 - z_i)$. Thus, by using Eq. (3), the likelihood function defined in Eq. (4) becomes

$$L(\boldsymbol{w}, \boldsymbol{z}, \theta, \lambda) = C \times \prod_{i=1}^{r}\left[\frac{w_i}{\theta^2}e^{-\frac{w_i}{\theta}}\right]^{z_i}\left[\frac{w_i}{\lambda^2}e^{-\frac{w_i}{\lambda}}\right]^{1-z_i}\left[1 + \frac{w_r}{\theta}\right]^{n-n_r} e^{-\frac{w_r}{\theta}(n-n_r)}$$
$$\times \left[1 + \frac{w_r}{\lambda}\right]^{m-m_r} e^{-\frac{w_r}{\lambda}(m-m_r)} \tag{5}$$

and equally

$$L(\boldsymbol{w}, \boldsymbol{z}, \theta, \lambda) \propto \frac{1}{\theta^{2n_r}}\frac{1}{\lambda^{2m_r}}\left(\prod_{i=1}^{r} w_i\right)e^{-\frac{\xi_1}{\theta}-\frac{\xi_2}{\lambda}}\left[1 + \frac{w_r}{\theta}\right]^{n-n_r}\left[1 + \frac{w_r}{\lambda}\right]^{m-m_r} \tag{6}$$

where $\xi_1 = \sum_{i=1}^{r} z_i w_i + (n - n_r)w_r$ and $\xi_2 = \sum_{i=1}^{r}(1 - z_i)w_i + (m - m_r)w_r$.

Thus, the log-likelihood function is obtained as

$$\ell(\theta,\lambda) \propto -2n_r\log(\theta) - 2m_r\log(\lambda) + \sum_{i=1}^{r}\log w_i - \frac{\xi_1}{\theta} - \frac{\xi_2}{\lambda} \\ \times +(n-n_r)\log(1+w_r/\theta) + (m-m_r)\log(1+w_r/\lambda) \tag{7}$$

The partial derivates of $\ell(\theta,\lambda)$ should be equated to zero with respect to θ and λ to obtain the MLEs of the parameters. The MLEs of the parameters are denoted by $\hat{\theta}$ and $\hat{\lambda}$ by solving the Equations

$$\frac{\xi_1}{\theta} - 2n_r - \frac{(n-n_r)w_r}{\theta+w_r} = 0 \tag{8}$$

$$\frac{\xi_2}{\lambda} - 2m_r - \frac{(m-m_r)w_r}{\lambda+w_r} = 0 \tag{9}$$

These equations can be solved using iterative methods such as Newton-Raphson.

The approximate confidence intervals for MLEs of the parameters can be obtained by using the inverse of the asymptotic Fisher information matrix. The observed inverse Fisher information matrix is given as follows

$$I\left(\hat{\theta},\hat{\lambda}\right) = -\begin{bmatrix} \frac{\partial^2\ell(\theta,\lambda)}{\partial\theta^2} & \frac{\partial^2\ell(\theta,\lambda)}{\partial\theta\partial\lambda} \\ \frac{\partial^2\ell(\theta,\lambda)}{\partial\lambda\partial\theta} & \frac{\partial^2\ell(\theta,\lambda)}{\partial\lambda^2} \end{bmatrix}_{\theta=\hat{\theta},\lambda=\hat{\lambda}} \tag{10}$$

where

$$\frac{\partial^2\ell(\theta,\lambda)}{\partial\theta^2} = \frac{2n_r}{\theta^2} - \frac{2\xi_1}{\theta^3} + \frac{(n-n_r)w_r(2\theta+w_r)}{\theta^2(\theta+w_r)^2}$$

$$\frac{\partial^2\ell(\theta,\lambda)}{\partial\lambda^2} = \frac{2m_r}{\lambda^2} - \frac{2\xi_2}{\lambda^3} + \frac{(m-m_r)w_r(2\lambda+w_r)}{\lambda^2(\lambda+w_r)^2}$$

$$\frac{\partial^2\ell(\theta,\lambda)}{\partial\theta\partial\lambda} = \frac{\partial^2\ell(\theta,\lambda)}{\partial\lambda\partial\theta} = 0$$

The estimated variances of the $\hat{\theta}$ and $\hat{\lambda}$ can be obtained as

$$\mathrm{Var}(\hat{\theta}) = -\left(\frac{\partial^2\ell(\theta,\lambda)}{\partial\theta^2}\right)^{-1}_{\theta=\hat{\theta}} \quad \text{and} \quad \mathrm{Var}(\hat{\lambda}) = -\left(\frac{\partial^2\ell(\theta,\lambda)}{\partial\lambda^2}\right)^{-1}_{\lambda=\hat{\lambda}}$$

It is known that the MLEs are consistent and asymptotically normally distributed under some regularity conditions [14]. Thus, the $100(1-\delta)\%$ asymptotic confidence intervals of λ and θ can be constructed by

$$\hat{\theta} \mp Z_{\frac{\delta}{2}}\sqrt{Var(\hat{\theta})} \text{ and } \hat{\lambda} \mp Z_{\frac{\delta}{2}}\sqrt{Var(\hat{\lambda})}$$

where Z_δ is 100 δth percentile of standard normal distribution $N(0, 1)$.

2.2 Bayesian Estimation

In this subsection, we obtained the Bayesian estimations of the unknown parameters and the corresponding credible intervals based on the jointly censoring scheme. Since θ and λ are both unknown, we may choose any specific forms of the priors. In this study, we considered inverse gamma distributions conjugate priors as $\pi(\theta) \sim \mathrm{IG}(a_1, b_1)$ and $\pi(\lambda) \sim \mathrm{IG}(a_2, b_2)$. The joint density function of the priors is given by

$$\pi(\theta, \lambda) \propto \theta^{-a_1-1} e^{-\frac{b_1}{\theta}} \lambda^{-a_2-1} e^{-\frac{b_2}{\lambda}}$$

where $\theta, \lambda > 0$ and the hyper-parameters $a_1, b_1, a_2, b_2 > 0$. Then, by using the observed censored samples and the prior distributions for the parameters, the joint posterior density function of parameters θ and λ are obtained as

$$\pi(\theta, \lambda|\mathbf{w,z}) \propto \theta^{-2n_r-a_1-1} e^{-\frac{\xi_1+b_1}{\theta}} \lambda^{-2m_r-a_2-1} e^{-\frac{\xi_2+b_2}{\lambda}} \times \left[1+\frac{w_r}{\theta}\right]^{n-n_r} \left[1+\frac{w_r}{\lambda}\right]^{m-m_r} \tag{11}$$

The joint posterior density can be written as

$$\pi(\theta, \lambda|\mathbf{w,z}) \propto \pi_1(\theta)\pi_2(\lambda)H(\theta, \lambda)$$

where

$$H(\theta, \lambda) = \left[1+\frac{w_r}{\theta}\right]^{n-n_r} \left[1+\frac{w_r}{\lambda}\right]^{m-m_r}$$

and

$$\begin{aligned} \pi_1(\theta) &\propto \mathrm{IG}(2n_r + a_1, \xi_1 + b_1) \\ \pi_2(\lambda) &\propto \mathrm{IG}(2m_r + a_2, \xi_2 + b_2) \end{aligned} \tag{12}$$

Therefore one can easily generate samples from the distribution of θ and λ given in (12). Now we would like to provide the importance sampling procedure to compute the Bayes estimates and also to construct the highest posterior density (HPD) credible intervals by using the following algorithm

Step 1: Generate θ from $\pi_1(\theta; a_1, b_1)$.
Step 2: Generate λ from $\pi_2(\lambda; a_2, b_2)$.
Step 3: Compute $H = H(\theta, \lambda)$.

Step 4: Repeat steps 1–3 M times, and generate M estimates of θ_j, λ_j and H_j, for $j = 1, 2, \ldots, M$.

Thus, an approximate Bayes estimate of θ under a squared error loss function can be obtained as

$$\hat{\theta}_B = \frac{\frac{1}{M}\sum_{i=1}^{M} \theta_i H\left(\theta_i, \lambda_i\right)}{\frac{1}{M}\sum_{i=1}^{M} H\left(\theta_i, \lambda_i\right)} \tag{13}$$

Bayesian estimate of λ can be obtained similarly. Then, the credible interval of θ and λ can be obtained by using the idea of Chen and Shao [10]. Let's assume that $U_i = H\left(\theta_{(i)}, \lambda_{(i)}\right)$ where $\theta_{(i)}$ and $\lambda_{(i)}$ for $i = 1, 2, \ldots, M$ is posterior samples generated respectively from (12) for θ and λ. Let U_i be the ordered values of U_i and define

$$\omega_i = \frac{H(\theta_i, \lambda_i)}{\sum_{i=1}^{M} H(\theta_i, \lambda_i)}$$

Then, the p-th quantile of $\hat{U} = (\hat{\theta}_B, \hat{\lambda}_B)$ can be estimated as

$$\hat{U}^{(p)} = \begin{cases} U_{(1)}, & p = 0 \\ U_{(i)}, & \sum_{j=1}^{i-1} \omega_j < p \le \sum_{j=1}^{i} \omega_j \end{cases}$$

The $100(1-\delta)\%$ HPD credible intervals are obtained as $\left\{\hat{U}^{\left(\frac{j}{M}\right)}, \hat{U}^{\left(\frac{j+[(1-\gamma)M]}{M}\right)}\right\}$ for $j = 1, 2, \ldots, M$ where [.] is the greatest integer function.

3 Simulation Studies

In this section, some simulation studies are performed for evaluating the performance of estimation methods which are developed in the previous sections. We considered different sample sizes for the two populations such as $(n, m) = (15, 15), (20, 25), (35, 30), (50, 50)$ and the corresponding different number of failure for each sample sizes as $r = (22, 26), (36, 40), (52, 60)$ and $(80, 90)$, respectively. These failure numbers are determined as providing %80 and %90 observed samples in each case. We considered three sets of parameter values such as $(\theta, \lambda) = (0.75, 0.75), (1.00, 1.50), (1.25, 0.75)$. In each cases of parameters, we determined the hyperparameter values that provide these actual values as $a_1 = a_2 = 5$, $b_1 = b_2 = 3$ for $(\theta, \lambda) = (0.75, 0.75)$, $a_1 = 4$, $a_2 = 2$, $b_1 = 3$, $b_2 = 1.5$ for $(\theta, \lambda) = (1.00, 1.50)$ and $a_1 = 2$, $a_2 = 5$, $b_1 = 1.25$, $b_2 = 3$ for $(\theta, \lambda) = (1.25, 0.75)$. The simulations were repeated 2000 times. In the MCMC, we used 3500 samples and discard the first 500 values as burn-in period and take every third variate in thinning procedure. Then, the algorithm is performed for 2000 replications. The significant level was taken as $\delta = 0.05$.

Table 1 The biases and MSEs (in parentheses) of the estimates with the corresponding average lengths and coverage probabilities (in parentheses) of their ACI and HPD intervals in the case of $\theta = 0.75$, $\lambda = 0.75$

n, m	r	$\hat{\theta}$				$\hat{\lambda}$			
		MLE	Bayes	ACI	HPD	MLE	Bayes	ACI	HPD
15, 15	22	0.00612	0.00503	0.60969	0.57467	0.01220	0.01030	0.61497	0.57818
		(0.02493)	(0.01829)	(93.15)	(95.90)	(0.02518)	(0.01845)	(93.50)	(95.95)
	26	−0.00422	−0.00376	0.56320	0.53193	−0.00177	−0.00161	0.56458	0.53313
		(0.01979)	(0.01504)	(93.25)	(96.65)	(0.02026)	(0.01539)	(93.60)	(96.40)
20, 25	36	0.00661	0.00570	0.51013	0.48912	0.00212	0.00204	0.45200	0.43930
		(0.01718)	(0.01384)	(93.95)	(95.55)	(0.01271)	(0.01068)	(94.35)	(96.25)
	40	−0.00290	−0.00287	0.48336	0.46385	−0.00089	−0.00078	0.43317	0.41957
		(0.01513)	(0.01236)	(93.95)	(95.85)	(0.01174)	(0.00999)	(94.05)	(95.65)
35, 30	52	0.00244	0.00226	0.38181	0.37572	0.00462	0.00445	0.41459	0.40476
		(0.00950)	(0.00835)	(94.35)	(94.50)	(0.01171)	(0.01007)	(93.60)	(92.95)
	60	0.00165	0.00150	0.37134	0.36462	0.00097	0.00086	0.40072	0.39143
		(0.00864)	(0.00765)	(94.10)	(95.90)	(0.01010)	(0.00876)	(94.75)	(95.95)
50, 50	80	0.00305	0.00333	0.31972	0.31507	0.00258	0.00291	0.31943	0.31392
		(0.00671)	(0.00618)	(94.85)	(95.26)	(0.00631)	(0.00580)	(95.25)	(95.80)
	90	0.00310	0.00309	0.30627	0.30244	0.00405	0.00396	0.30654	0.30300
		(0.00625)	(0.00576)	(93.95)	(95.25)	(0.00611)	(0.00563)	(94.50)	(95.20)

The performances of the estimators are evaluated with their biases and mean squared error (MSE) values in each case. We also evaluated the performances of the approximate confidence intervals according to their average lengths (AL) and coverage probabilities (CP). The whole results are presented in Tables 1, 2 and 3.

We observed that the MSEs and biases decrease with increasing sample sizes as expected. The Bayes estimates have smaller MSEs than MLEs. The difference between performances of the estimators decreases with increasing sample size. Especially in large samples, both estimation methods give almost equal estimations. In parallel to estimates, confidence intervals show similar performances. The HPD credible interval is satisfactorily better than ACI in small samples. Both methods show similar ALs and CPS with the increasing sample sizes. The confidence intervals are mostly obtained very close to their actual values, 0.95, in all cases.

4 Numerical Example

In this section, we provide a numerical example by using the generated datasets by Akhter et al. [2]. They generated some datasets from the LBE distributions and used them to evaluate the usefulness of coefficients of BLUEs of location and scale parameters. In the purpose of the jointly censoring scheme, we take the first two data

Table 2 The biases and MSEs (in parentheses) of the estimates with the corresponding average lengths and coverage probabilities (in parentheses) of their ACI and HPD intervals in the case of $\theta = 1, \lambda = 1.5$

n, m	r	$\hat{\theta}$				$\hat{\lambda}$			
		MLE	Bayes	ACI	HPD	MLE	Bayes	ACI	HPD
15, 15	22	0.00898 (0.03854)	0.00808 (0.03095)	0.77465 (93.90)	0.74331 (96.15)	0.01834 (0.11067)	0.01967 (0.10170)	1.29660 (93.10)	1.29513 (95.05)
	26	0.00120 (0.03373)	0.00085 (0.02762)	0.73361 (93.30)	0.70308 (95.85)	0.02049 (0.08692)	0.02081 (0.08108)	1.18446 (94.15)	1.18075 (95.45)
20, 25	36	0.00333 (0.02643)	0.00280 (0.02260)	0.64767 (93.70)	0.62881 (95.65)	0.00556 (0.05693)	0.00684 (0.05458)	0.94006 (94.00)	0.95478 (95.11)
	40	0.00355 (0.02620)	0.00317 (0.02261)	0.63291 (93.95)	0.61369 (95.75)	−0.00213 (0.05379)	−0.00176 (0.05171)	0.88441 (93.55)	0.88602 (94.55)
35, 30	52	−0.00056 (0.01598)	−0.00032 (0.01474)	0.48946 (94.35)	0.48312 (94.84)	0.01039 (0.04806)	0.01138 (0.04635)	0.87035 (94.75)	0.86768 (95.07)
	60	0.00308 (0.01482)	0.00294 (0.01364)	0.47523 (94.70)	0.46651 (95.50)	0.00244 (0.04046)	0.00249 (0.03908)	0.79832 (94.95)	0.80045 (95.80)
50, 50	80	0.00059 (0.01109)	0.00090 (0.01044)	0.40886 (94.35)	0.40516 (94.92)	0.00442 (0.02783)	0.00627 (0.02768)	0.66696 (94.70)	0.64280 (93.88)
	90	0.00213 (0.01042)	0.00201 (0.00983)	0.39873 (94.45)	0.39379 (95.00)	0.00087 (0.02401)	0.00137 (0.02364)	0.62454 (95.10)	0.63339 (96.05)

Table 3 The biases and MSEs (in parentheses) of the estimates with the corresponding average lengths and coverage probabilities (in parentheses) of their ACI and HPD intervals in the case of $\theta = 1.25, \lambda = 0.75$

n, m	r	$\hat{\theta}$				$\hat{\lambda}$			
		MLE	Bayes	ACI	HPD	MLE	Bayes	ACI	HPD
15, 15	22	0.01281 (0.07797)	0.01483 (0.07184)	1.09438 (93.45)	1.09150 (94.71)	0.00688 (0.02307)	0.00584 (0.01749)	0.57516 (92.40)	0.54279 (95.80)
	26	0.00125 (0.06116)	0.00171 (0.05667)	0.97951 (93.55)	0.97794 (95.00)	−0.00184 (0.01949)	−0.00170 (0.01508)	0.54606 (93.30)	0.51621 (96.05)
20, 25	36	0.00969 (0.05334)	0.01116 (0.05067)	0.90480 (93.90)	0.91106 (94.92)	0.00339 (0.01259)	0.00319 (0.01070)	0.43401 (93.85)	0.41914 (94.95)
	40	0.00387 (0.04547)	0.00452 (0.04310)	0.84179 (94.15)	0.84310 (95.95)	−0.00124 (0.01100)	−0.00121 (0.00939)	0.42143 (95.05)	0.40716 (96.35)
35, 30	52	0.00878 (0.02878)	0.00952 (0.02807)	0.67072 (94.05)	0.66917 (94.82)	0.00393 (0.00976)	0.00376 (0.00854)	0.39406 (95.45)	0.38227 (96.25)
	60	0.00194 (0.02560)	0.00230 (0.02490)	0.61458 (94.05)	0.61647 (95.00)	0.00235 (0.00980)	0.00213 (0.00861)	0.38337 (94.90)	0.37193 (95.75)
50, 50	80	0.00087 (0.01950)	0.00231 (0.01947)	0.56073 (95.50)	0.52834 (94.84)	−0.00174 (0.00612)	−0.00149 (0.00569)	0.30324 (93.40)	0.29652 (94.29)
	90	−0.00006 (0.01770)	0.00009 (0.01737)	0.52281 (93.85)	0.53156 (94.34)	0.00173 (0.00598)	0.00167 (0.00551)	0.29783 (94.50)	0.29267 (94.80)

(actual parameter values equal to 0.5) sets with sample sizes $n = 7$ and $m = 10$ as given in the following

Data Set I (X): 0.184522, 0.647899, 0.693491, 0.863154, 0.918599, 1.543884, 2.296143.

Data Set II (Y): 0.281766, 0.311408, 0.507578, 0.578070, 0.670802, 1.048554, 1.147288, 1.500197, 1.565101, 2.267506.

We considered two different values of failures as $r = 10$ and $r = 14$ and used informative hyper-parameters as $a_1 = a_2 = 3, b_1 = b_2 = 1$ to provide actual parameter values 0.5 for both parameters. In MCMC method, we used 101 000 iterations, we discard the first 1000 values in burn-in period and we take every tenth variate in thinning procedure.

In the case of $r = 10$, we obtained $\hat{\theta} = 0.45374$ and its corresponding ACI is obtained as $(0.18433, 0.72316)$ with length 0.53883. The Bayes estimate is obtained as $\hat{\theta}^B = 0.46136$ and its corresponding HPD credible interval is obtained as $(0.26759, 0.79871)$ with length 0.53113. On the other hand, we obtained $\hat{\lambda} = 0.52682$ and its corresponding ACI is obtained as $(0.22901, 0.82464)$ with length 0.59563. The Bayes estimate is obtained as $\hat{\lambda}^B = 0.52349$ and its corresponding HPD credible interval is obtained as $(0.31035, 0.88076)$ with length 0.57041.

In the case of $r = 14$, we obtained $\hat{\theta} = 0.50142$ and its corresponding ACI is obtained as $(0.22422, 0.77863)$ with length 0.55441. The Bayes estimate is obtained as $\hat{\theta}^B = 0.50197$ and its corresponding HPD credible interval is obtained as $(0.30263, 0.82543)$ with length 0.52280. On the other hand, we obtained $\hat{\lambda} = 0.52209$ and its corresponding ACI is obtained as $(0.27475, 0.76943)$ with length 0.49468. The Bayes estimate is obtained as $\hat{\lambda}^B = 52085$ and its corresponding HPD credible interval is obtained as $(0.27475, 0.76943)$ with length 0.47682.

It is observed that the estimates are getting closer to their actual values with parallel to the increasing size of the observed sample. The Bayes HPD credible intervals perform slightly better than ACI. We also assessed the convergence of the simulated Markov chains by graphical methods effectively. For this purpose, the trace plot which is a plot of the iteration number, t, against the value of the $\theta^{(t)}$ and $\lambda^{(t)}$ at each iteration and density plot of the posterior distributions of θ and λ are used. Also, the running mean (ergodic average) plot which draws the mean of sampled values up to iteration t is used. For each sample size in the simulation schemes, with their higher failure numbers, we evaluated the convergence of the Markov chains and draw all graphics and reported in Figs. 1 and 2. We observed that overall convergences of the Markov chains are satisfactory by the evaluations and observations as expectations. The trace plots fluctuate around their center with similar variation and ranges are decreasing with increasing sample sizes. The density plots seem symmetric, unimodal and the bandwidth of the densities becomes narrow with increasing sample size. Also, the running mean plot shows that the chain has achieved stationarity. All plots to check the convergence can be drawn by using **mcmcplots** package [11] in *R* software.

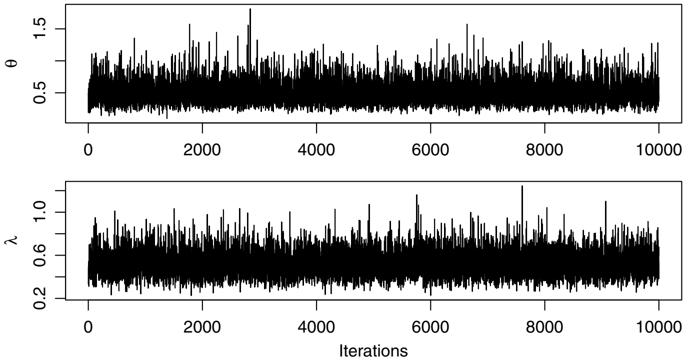


Fig. 1 The trace plots of the posterior distributions of θ and λ in the case of $r = 14$

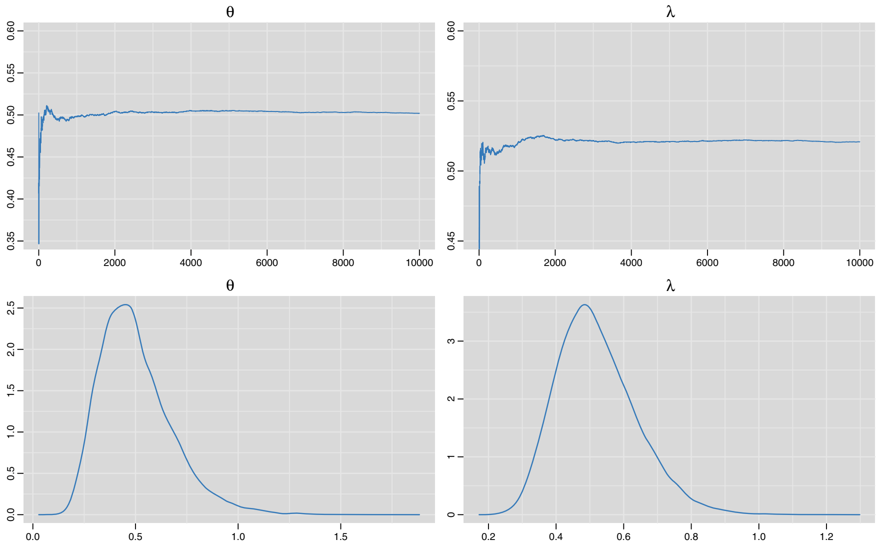


Fig. 2 The running mean and the density plots of the posterior distributions of θ and λ in the case of $r = 14$

5 Conslusions

This study deals with inference for two length-biased populations under a joint Type-II censoring scheme. We obtained the MLE and Bayes estimates with corresponding asymptotic and Bayesian HPD credible intervals. Simulations and numerical examples show that theoretical findings perform quite well. Both estimation methods provide good results with a satisfactory small MSE value. In small samples, the Bayes estimation method performs better than MLE as expected. Both estimates are getting very close in parallel to increasing sample sizes. Further, the estimations perform well even if the observed sample sizes level decrease. There is not any previous study for the LBE distribution under a jointly censoring scheme. It can be considered under more complex jointly censoring schemes for further studies.

References

1. Abo-Kasem, O. E., Nassar, M., Dey, S., Rasouli, A.: Classical and Bayesian Estimation for Two Exponential Populations based on Joint Type-I Progressive Hybrid Censoring Scheme. American Journal of Mathematical and Management Sciences, 38(4), 373–385, (2019).
2. Akhter, Z., Saran, J., Verma, K.,Pushkarna, N.: Moments of Order Statistics from Length-Biased Exponential Distribution and Associated Inference. Annals of Data Science, 1–26, (2020).
3. Ashour, S. K. Abo-Kasem, O. E.: Parameter estimation for multiple Weibull populations under joint Type-II censoring. International Journal of Advanced Statistics and Probability, 2(2), 102–107, (2014a).
4. Ashour, S. K., Abo-Kasem, O. E.: Parameter estimation for two Weibull populations under joint Type II censored scheme. International Journal of Engineering, 5(04), 8269, (2014b).
5. Ashour, S. K., Abo-Kasem, O. E.: Bayesian and non–Bayesian estimation for two generalized exponential populations under joint type II censored scheme. Pakistan Journal of Statistics and Operation Research, 57–72, (2014c).
6. Ashour, S. K., Abo-Kasem, O. E.: Statistical inference for two exponential populations under joint progressive type-I censored scheme. Communications in Statistics-Theory and Methods, 46(7), 3479–3488, (2017).
7. Balakrishnan, N., Su, F., Liu, K. Y.: Exact likelihood inference for k exponential populations under joint progressive Type-II censoring. Communications in Statistics-Simulation and Computation, 44(4), 902–923, (2015).
8. Balakrishnan, N., Rasouli, A.: Exact likelihood inference for two exponential populations under joint type-II censoring. Computational Statistics & Data Analysis, 52, 2725–2738, (2008).
9. Balakrishnan, N., Su, F.: Exact likelihood inference for k exponential populations under joint Type-II censoring. Communications in Statistics-Simulation and Computation, 44(3), 591–613, (2015).
10. Chen, M. H., Shao, Q. M.: Monte Carlo estimation of Bayesian credible and HPD intervals. Journal of Computational and Graphical Statistics, 8(1), 69–92, (1999).
11. Curtis, S. M., Goldin, I., Evangelou, E.: Package 'mcmcplots', (2018).
12. Dara, S. T., Ahmad, M.: Recent advances in moment distribution and their hazard rates. Academic Publishing GmbH KG, Lap Lambert, (2012).
13. Doostparast, M., Ahmadi, M. V., Ahmadi, J.: Bayes estimation based on joint progressive Type II censored data under LINEX loss function. Communications in Statistics-Simulation and Computation, 42(8), 1865–1886, (2013).
14. Godambe, V. P.: An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics, 31(4), 1208–1211, (1960).
15. Krishna, H., Goel, R.: Jointly Type-II censored Lindley distributions. Communications in Statistics-Theory and Methods, 1–15, (2020).
16. Rasouli, A., Balakrishnan, N.: Exact likelihood inference for two exponential populations under joint progressive type-II censoring. Communications in Statistics–Theory and Methods, 39(12), 2172–2191, (2010).

Fuzzy $\theta^*\mathcal{S}$-open and Closed Mappings in $\hat{S}$ostak's Fuzzy Topological Spaces

Odayappan Uma Maheswari, Appachi Vadivel, Arivazhagan Mughil, and Gurusamy Saravanakumar

Abstract We introduce and investigate some new classes of mappings called fuzzy $\theta^*\mathcal{S}$-open map and fuzzy $\theta^*\mathcal{S}$-closed map to the fuzzy topological spaces in $\hat{S}$ostaks sense. Also, some of their fundamental properties are studied. Moreover, we investigate the relationships between some other existing mappings.

Keywords Fuzzy open · Fuzzy θ^*-semiopen mappings · $f\theta^*\mathcal{S}$-T_1 space, $f\theta^*\mathcal{S}$-T_2 space, fuzzy $\theta^*\mathcal{S}$-connected and fuzzy $\theta^*\mathcal{S}$-compact

AMS (2000) subject classification: 03E05 · 54A05 · 54A40

1 Introduction

$\hat{S}$ostak [23] introduced the fuzzy topology as an extension of Chang's fuzzy topology [6]. It has been developed in many directions [11, 12, 22]. Weaker forms of fuzzy continuity between fuzzy topological spaces have been considered by many authors [1, 4, 10] using the concepts of fuzzy semi-open sets [1], fuzzy regular open sets [1], fuzzy preopen sets, fuzzy strongly semiopen sets [4], fuzzy γ-open sets [10], fuzzy δ-semiopen sets [3], fuzzy δ-preopen sets [3] and fuzzy semi δ-preopen sets [24]. Ganguly and Saha [9] introduced the notions of fuzzy δ-cluster points in fuzzy

O. Uma Maheswari · A. Mughil
Post Graduate and Research Department of Mathematics, J. J. College of Arts and Science (Autonomous), Bharathidasan University, Pudukkottai 622 422, India

A. Vadivel (✉)
Department of Mathematics, Annamalai University, Annamalai Nagar, Chidambaram 608 002, India

Post Graduate and Research Department of Mathematics, Government Arts college (Autonomous), Bharathidasan University, Karur 639 005, India

G. Saravanakumar
Department of Mathematics, M. Kumarasamy College of Engineering (Autonomous), Karur 639 113, India

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*, Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_6

topological spaces in the sense of Chang [6]. Kim and Park [13] introduced ι-δ-cluster points and δ-closure operators in fuzzy topological spaces in view of the definition of $\hat{S}$ostak. It is a good extension of the notions of Ganguly and Saha [9]. Park et al. [17] introduced the concept of fuzzy semi-preopen sets which is weaker than any of the concepts of fuzzy semi-open or fuzzy preopen sets. In 1968, Velicko exhibited and studied some new types of open sets called θ-open sets [25] and δ-open sets which are stronger than open sets in order to investigate the characterizations of H-closed topological spaces. Levine in 1963 initiated a new type of open set called semi-open set [14]. In 1993, Raychaudhuri and Mukherjee defined δ-preopen sets [21]. In 1997, δ-semiopen sets was obtained by Park [18] and Caldas obtained θ-semi-open sets in 2008 [5]. Shafei introduced fuzzy θ-closed [8] and fuzzy θ-open sets in 2006. Maghrabi et al. [7] introduced the notion of θ^*-semiopen sets in topological spaces in 2014. Recently, Mughil et al. [15, 16] introduced ι-fuzzy θ^*-semiopen (resp. ι-fuzzy θ^*-semiclosed) sets and maps in fuzzy topological spaces in the sense of $\hat{S}$ostak's. In this paper, we introduce the concept of fuzzy θ^*-semi open mappings in fuzzy topological spaces in the sense of $\hat{S}$ostak's. Also, some characterizations and properties of these notions are presented.

2 Preliminaries

Nonempty sets shall be indicated by $U,\ V$, $I = [0, 1]$, and $I_0 = (0, 1]$ throughout this article. For $\rho \in I$, $\overline{\rho}(u) = \rho \,\forall\, u \in U$. A fuzzy point u_t for $t \in I_0$ is an element of $I^U \ni u_t(v) = \begin{cases} t & \text{if } v = u \\ 0 & \text{if } v \neq u. \end{cases}$ $Pt(U)$ denotes the set of all fuzzy points in U. A fuzzy point $u_t \in \beta$ iff $t < \beta(u)$. A fuzzy set β is quasi-coincident with δ, denoted by $\beta q \delta$, if $u \in U \;\exists \ni \beta(u) + \delta(u) > 1$. If β is not quasi-coincident with δ, we call it $\beta \bar{q} \delta$. If $A \subset U$, we define the characteristic function χ_A on U by $\chi_A(u) = \begin{cases} 1 & \text{if } u \in A, \\ 0 & \text{if } u \notin A. \end{cases}$

Definition 1 ([23]) A function $\mathfrak{S} : I^U \to I$ is called a fuzzy topology on U if it satisfies the following conditions:

(O1) $\mathfrak{S}(\underline{0}) = \mathfrak{S}(\underline{1}) = 1$,
(O2) $\mathfrak{S}(\bigvee_{j\in\Gamma} \gamma_i) \geq \bigwedge_{j\in\Gamma} \mathfrak{S}(\gamma_j)$, for any $\{\gamma_j\}_{j\in\Gamma} \subset I^U$,
(O3) $\mathfrak{S}(\gamma_1 \wedge \gamma_2) \geq \mathfrak{S}(\gamma_1) \wedge \mathfrak{S}(\gamma_2)$, for any $\gamma_1,\ \gamma_2 \in I^U$.

The pair $(U,\ \mathfrak{S})$ is called a $\hat{S}$ostak's fuzzy topological space (for short, sfts).

Remark 1 ([19]) Let $(U,\ \mathfrak{S})$ be a sfts. Then, $\forall\, \iota \in I_0$, $\mathfrak{S}_\iota = \{\delta \in I^U : \mathfrak{S}(\delta) \geq \iota\}$ is a Change's fuzzy topology on U.

Theorem 1 ([22]) *Let $(U,\ \mathfrak{S})$ be a sfts. Then $\forall\ \beta \in I^U,\ \imath \in I_0$ we define an operator $Cl_{\mathfrak{S}}$ (resp. $Int_{\mathfrak{S}}$) $: I^U \times I_0 \to I^U$ as follows: $Cl_{\mathfrak{S}}(\beta,\ \imath) = \bigwedge\{\delta \in I^U : \beta \leq \delta,\ \mathfrak{S}(\underline{1} - \delta) \geq \imath\}$ (resp. $Int_{\mathfrak{S}}(\beta,\ \imath) = \bigvee\{\delta \in I^U : \beta \geq \delta,\ \mathfrak{S}(\delta) \geq \imath\}$).*
For $\beta,\ \delta \in I^U$ and $\imath, \jmath \in I_0$, the operator $Cl_{\mathfrak{S}}$ & $Int_{\mathfrak{S}}$ satisfies the conditions:

(1) $Cl_{\mathfrak{S}}(\underline{0}, \imath) = \underline{0}$,
(2) $\beta \leq Cl_{\mathfrak{S}}(\beta,\ \imath)$,
(3) $Cl_{\mathfrak{S}}(\beta,\ \imath) \vee Cl_{\mathfrak{S}}(\delta,\ \imath) = Cl_{\mathfrak{S}}(\beta \vee \delta,\ \imath)$,
(4) $Cl_{\mathfrak{S}}(\beta,\ \imath) \leq Cl_{\mathfrak{S}}(\beta,\ \jmath)$ *if* $\imath \leq \jmath$,
(5) $Cl_{\mathfrak{S}}(Cl_{\mathfrak{S}}(\beta,\ \imath),\ \imath) = Cl_{\mathfrak{S}}(\beta,\ \imath)$.
(6) $Int_{\mathfrak{S}}(\underline{1}, \imath) = \underline{1}$,
(7) $\beta \geq Int_{\mathfrak{S}}(\beta,\ \imath)$,
(8) $Int_{\mathfrak{S}}(\beta,\ \imath) \wedge Int_{\mathfrak{S}}(\delta,\ \imath) = Int_{\mathfrak{S}}(\beta \wedge \delta,\ \imath)$,
(9) $Int_{\mathfrak{S}}(\beta,\ \imath) \leq Int_{\mathfrak{S}}(\beta,\ \jmath)$ *if* $\jmath \leq \imath$,
(10) $Int_{\mathfrak{S}}(Int_{\mathfrak{S}}(\beta,\ \imath),\ \imath) = Int_{\mathfrak{S}}(\beta,\ \imath)$,
(11) $Int_{\mathfrak{S}}(\underline{1} - \beta, \imath) = \underline{1} - Cl_{\mathfrak{S}}(\beta,\ \imath)$ *and* $Cl_{\mathfrak{S}}(\underline{1} - \beta, \imath) = \underline{1} - Int_{\mathfrak{S}}(\beta,\ \imath)$

Definition 2 Let $(U,\ \mathfrak{S})$ be a sfts. $\beta,\ \delta \in I^U$ & $\imath \in I_0$, then

(i) $\theta Int_{\mathfrak{S}}(\beta,\ \imath) = \bigvee\{Int_{\mathfrak{S}}(\delta, \imath) : \beta \geq \delta,\ \mathfrak{S}(\overline{1} - \delta) \geq \imath\}$ is called the $\imath$-fuzzy θ-interior of β [26].
(ii) $\theta Cl_{\mathfrak{S}}(\beta,\ \imath) = \bigwedge\{Cl_{\mathfrak{S}}(\delta, \imath) : \beta \leq \delta,\ \mathfrak{S}(\delta) \geq \imath\}$ is called the $\imath$-fuzzy θ-closure of β [26].
(iii) $\imath$-fuzzy θ-open (resp. $\imath$-fuzzy θ-closed) (briefly, $\imath$-$f\theta o$ (resp. $\imath$-$f\theta c$)) [26] set if $\beta = \theta Int_{\mathfrak{S}}(\beta,\ \imath)$ (resp. $\beta = \theta Cl_{\mathfrak{S}}(\beta,\ \imath)$).
(iv) $\imath$-fuzzy θ-semiopen (resp. $\imath$-fuzzy θ-semiclosed) (briefly, $\imath$-$f\theta\mathcal{S}o$ (resp. $\imath$-$f\theta\mathcal{S}c$)) set [26] if $\beta \leq Cl_{\mathfrak{S}}(\theta Int_{\mathfrak{S}}(\beta,\ \imath),\ \imath)$ (resp. $Int_{\mathfrak{S}}(\theta Cl_{\mathfrak{S}}(\beta,\ \imath),\ \imath) \leq \beta$).
(v) $\imath$-fuzzy α-open (resp. $\imath$-fuzzy semiopen [20] & $\imath$-fuzzy γ-open [20]) (briefly, $\imath$-$f\alpha o$ (resp. $\imath$-$f\mathcal{S}o$ & $\imath$-$f\gamma o$)) set if $\beta \leq Int_{\mathfrak{S}}(Cl_{\mathfrak{S}}(Int_{\mathfrak{S}}(\beta,\ \imath),\ \imath),\ \imath)$ (resp. $\beta \leq Cl_{\mathfrak{S}}(Int_{\mathfrak{S}}(\beta,\ \imath),\ \imath))$ & $\beta \leq Cl_{\mathfrak{S}}(Int_{\mathfrak{S}}(\beta, \imath), \imath) \vee Int_{\mathfrak{S}}(Cl_{\mathfrak{S}}(\beta, \imath), \imath)$.

Definition 3 ([15]) In a fts $(U, \mathfrak{S})$, $\beta \in I^U$ is called an $\imath$-fuzzy

(i) θ^*-semiopen (resp. θ^*-semiclosed) (briefly $\imath$-$f\theta^*\mathcal{S}o$ (resp. $\imath$-$f\theta^*\mathcal{S}c$)) set if $\beta \leq Int_{\mathfrak{S}}(Cl_{\mathfrak{S}}(Int_{\mathfrak{S}}\ (\beta, \imath), \imath), \imath) \vee Cl_{\mathfrak{S}}(\theta Int_{\mathfrak{S}}(\beta, \imath), \imath)$ (resp. $\beta \geq Cl_{\mathfrak{S}}(Int_{\mathfrak{S}}\ (Cl_{\mathfrak{S}}(\beta, \imath), \imath), \imath) \wedge Int_{\mathfrak{S}}(\theta Cl_{\mathfrak{S}}(\beta, \imath), \imath)$).
(ii) Z open (briefly $\imath$-fZo) set if $\beta \leq Cl_{\mathfrak{S}}(\delta Int_{\mathfrak{S}}(\beta, \imath), \imath) \vee Int_{\mathfrak{S}}(Cl_{\mathfrak{S}}(\beta, \imath), \imath)$.
(iii) Y open (briefly $\imath$-fYo) set if $\beta \leq Cl_{\mathfrak{S}}(\theta Int_{\mathfrak{S}}(\beta, \imath), \imath) \vee Int_{\mathfrak{S}}(Cl_{\mathfrak{S}}(\beta, \imath), \imath)$.
(iv) θ^* semi closed (resp. Z closed & $\imath$-fuzzy Y closed) (briefly $\imath$-$f\theta^*\mathcal{S}c$ (resp. $\imath$-fZc & $\imath$-fYc)) set if $\underline{1} - \beta$ is an $\imath$-$f\theta^*\mathcal{S}o$ (resp. $\imath$-fZo & $\imath$-fYo) set.

Definition 4 ([15]) Let $(U, \mathfrak{S})$ be a fts, then the

(i) union of all $\imath$-$f\theta^*\mathcal{S}o$ (resp. $\imath$-fYo, $\imath$-$f\theta\mathcal{S}o$ & $\imath$-fZo) sets contained in β is called the $\imath$-$f\theta^*\mathcal{S}$ (resp. $\imath$-fY, $\imath$-$f\theta\mathcal{S}$ & $\imath$-fZ) interior of β (briefly, $\theta^*\mathcal{S}Int_{\mathfrak{S}}$ $(\beta, \imath)$ (resp. $YInt_{\mathfrak{S}}(\beta, \imath)$, $\theta\mathcal{S}Int_{\mathfrak{S}}(\beta, \imath)$ & $ZInt_{\mathfrak{S}}(\beta, \imath)$)).

(ii) intersection of all $\imath$-$f\theta^*\mathcal{S}c$ (resp. $\imath$-fYc, $\imath$-$f\theta\mathcal{S}c$ & $\imath$-fZc) sets containing β is called the $\imath$-$f\theta^*\mathcal{S}$ (resp. $\imath$-fY, $\imath$-$f\theta\mathcal{S}$ & $\imath$-fZ) closure of β (briefly, $\theta^*\mathcal{S}Cl_{\mathfrak{S}}(\beta, \imath)$ (resp. $YCl_{\mathfrak{S}}(\beta, \imath), \theta\mathcal{S}Cl_{\mathfrak{S}}(\beta, \imath)$ & $ZCl_{\mathfrak{S}}(\beta, \imath)$)).

Theorem 2 ([15]) *Let $(U, \mathfrak{S})$ be a fts. Let $\beta \in I^U$ and $\imath \in I_o$.*

(i) β is $\imath$-$f\theta^*\mathcal{S}o$ iff $\beta = \theta^*\mathcal{S}Int_{\mathfrak{S}}(\beta, \imath)$.
(ii) β is $\imath$-$f\theta^*\mathcal{S}c$ iff $\beta = \theta^*\mathcal{S}Cl_{\mathfrak{S}}(\beta, \imath)$.

Theorem 2 is satisfied by $YCl_{\mathfrak{S}}(\beta, \imath)$ & $ZCl_{\mathfrak{S}}(\beta, \imath)$.

Definition 5 ([16]) Let $(U, \mathfrak{S}_1)$ and $(U, \mathfrak{S}_2)$ be fts's and $k : (U, \mathfrak{S}_1) \to (V, \mathfrak{S}_2)$ a mapping. k is called fuzzy θ^*-semicontinuous (resp. Y-continuous & Z-continuous) (briefly, $f\theta^*\mathcal{S}Cts$ (resp. $fYCts$ & $fZCts$)) if $k^{-1}(\delta)$ is $\imath$-$f\theta^*\mathcal{S}o$ (resp. $\imath$-fYo & $\imath$-fZo) $\forall\ \delta \in I^U,\ \imath \in I_0$ with $\mathfrak{S}_2(\delta) \geq \imath$.

3 Fuzzy $\theta^*\mathcal{S}$-open and Closed Mappings

Definition 6 Let $(U, \mathfrak{S}_1)$, $(V, \mathfrak{S}_2)$ be sfts's and $k : (U, \mathfrak{S}_1) \to (V, \mathfrak{S}_2)$ be a mapping. Then k is called fuzzy

(i) $\theta^*\mathcal{S}$-open (briefly, $f\theta^*\mathcal{S}O$) mapping if $k(\beta)$ is $\imath$-$f\theta^*\mathcal{S}o$ set of V for each $\beta \in I^U$, $\imath \in I_0$ with $\mathfrak{S}_1(\beta) \geq \imath$.
(ii) $\theta^*\mathcal{S}$-closed (briefly, $f\theta^*\mathcal{S}C$) mapping if $k(\beta)$ is $\imath$-$f\theta^*\mathcal{S}c$ set of V for each $\beta \in I^U$, $\imath \in I_0$ with $\mathfrak{S}_1(\underline{1} - \beta) \geq \imath$.

Theorem 3 *Let $(U, \mathfrak{S})$ be a fts and $\imath \in I_0$, then*

1. *Every $f\theta O$ (resp. $f\theta C$) map is fO (resp. fC).*
2. *Every $f\theta O$ (resp. $f\theta C$) map is $f\theta\mathcal{S}O$ (resp. $f\theta\mathcal{S}C$).*
3. *Every fO (resp. fC) map is $f\alpha O$ (resp. $f\alpha C$).*
4. *Every $f\alpha O$ (resp. $f\alpha C$) map is $f\theta^*\mathcal{S}O$ (resp. $f\theta^*\mathcal{S}C$).*
5. *Every $f\theta\mathcal{S}O$ (resp. $f\theta\mathcal{S}C$) map is $f\theta^*\mathcal{S}O$ (resp. $f\theta^*\mathcal{S}C$).*
6. *Every $f\theta^*\mathcal{S}O$ (resp. $f\theta^*\mathcal{S}C$) map is fYO (resp. fYC).*
7. *Every $f\theta^*\mathcal{S}O$ (resp. $f\theta^*\mathcal{S}C$) map is $f\mathcal{S}O$ (resp. $f\mathcal{S}C$).*
8. *Every $f\mathcal{S}O$ (resp. $f\mathcal{S}C$) map is $f\gamma O$ (resp. $f\gamma C$).*
9. *Every fYO (resp. fYC) map is fZO (resp. fZC).*
10. *Every fZO (resp. fZC) map is $f\gamma O$ (resp. $f\gamma C$).*

Remark 2 It is clear from the preceding definitions that the following consequences are true for $\imath \in I_0$. But not converse are shown in the following examples.

$$\begin{tikzcd}
\boxed{fO} \arrow[r] & \boxed{f\alpha O} \arrow[dr] & & \boxed{f\mathcal{S}O} \arrow[r] & \boxed{f\gamma O} \\
& & \boxed{f\theta^\star\mathcal{S}O} \arrow[ur] \arrow[dr] & & \\
\boxed{f\theta O} \arrow[uu] \arrow[r] & \boxed{f\theta\mathcal{S}O} \arrow[ur] & & \boxed{fYO} \arrow[r] & \boxed{fZO} \arrow[uu]
\end{tikzcd}$$

Example 1 Let $U = V = \{k,\ l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5$, $\rho_1(l) = 0.6$, $\rho_1(m) = 0.4$; $\rho_2(k) = 0.5$, $\rho_2(l) = 0.4$, $\rho_2(m) = 0.4$. Then $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \\ 0 & \text{ow.} \end{cases} \quad \text{and } \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then $i_U : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ is fO (resp. $f\theta^\star\mathcal{S}O$)-map but not $f\theta O$ (resp. $f\theta\mathcal{S}O$)-map.

Example 2 Let $U = V = \{k,\ l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5$, $\rho_1(l) = 0.6$, $\rho_1(m) = 0.4$; $\rho_2(k) = 0.5$, $\rho_2(l) = 0.4$, $\rho_2(m) = 0.4$; $\rho_3(k) = 0.5$, $\rho_3(l) = 0.4$, $\rho_3(m) = 0.6$ Let $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_3, \\ 0 & \text{ow.} \end{cases} \quad \text{and } \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then $i_U : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ is $f\theta\mathcal{S}O$ (resp. $f\mathcal{S}O$ & fYO)-map but not $f\theta O$ (resp. $f\theta^\star\mathcal{S}O$)-map.

Example 3 Let $U = V = \{k,\ l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5$, $\rho_1(l) = 0.6$, $\rho_1(m) = 0.4$; $\rho_2(k) = 0.5$, $\rho_2(l) = 0.4$, $\rho_2(m) = 0.4$; $\rho_4(k) = 0.5$, $\rho_4(l) = 0.4$, $\rho_4(m) = 0.5$ Let $\mathfrak{S}_i : I^U \to I$,defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_4, \\ 0 & \text{ow.} \end{cases} \quad \text{and } \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then $i_U : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ is fZO-map but not fYO-map.

Example 4 Let $U = V = \{k,\ l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5$, $\rho_1(l) = 0.6$, $\rho_1(m) = 0.4$; $\rho_2(k) = 0.5$, $\rho_2(l) = 0.4$, $\rho_2(m) = 0.4$; $\rho_5(k) = 0.5$, $\rho_5(l) = 0.5$, $\rho_5(m) = 0.5$ Let $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_5, \\ 0 & \text{ow.} \end{cases} \quad \text{and } \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then $i_U : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ is $f\theta^\star\mathcal{S}O$-map but not $f\alpha O$-map.

Example 5 Let $U = V = \{k,\ l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_6(k) = 0.5$, $\rho_6(l) = 0.5$, $\rho_6(m) = 0.4$; $\rho_7(k) = 0.5, \rho_7(l) = 0.5$, $\rho_7(m) = 0.9$; $\rho_8(k) = 0.5$, $\rho_8(l) = 0.5$, $\rho_8(m) = 0$. Let $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_8, \\ 0 & \text{ow.} \end{cases} \quad \text{and } \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_6, \rho_7, \\ 0 & \text{ow.} \end{cases}$$

Then $i_U : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ is $f\gamma O$-map but not fZO (resp. $f\mathcal{S}O$)-map.

Example 6 Let $U = V = \{k,\ l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_9(k) = 0.5$, $\rho_9(l) = 0.5$, $\rho_9(m) = 0.4$; $\rho_{10}(k) = 0.5$, $\rho_{10}(l) = 0.5$, $\rho_{10}(m) = 0.1$; $\rho_{11}(k) = 0.5$, $\rho_{11}(l) = 0.5$, $\rho_{11}(m) = 0.2$ Let $\mathfrak{S}_i : I^U \to I$,

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_{11}, \\ 0 & \text{ow.} \end{cases} \quad \text{and } \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_9, \rho_{10}, \\ 0 & \text{ow.} \end{cases}$$

Then $i_U : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ is $f\alpha O$-map but not $f0$-map.

Theorem 4 *Let* $(U,\ \mathfrak{S}_1)$, $(V,\ \mathfrak{S}_2)$ *be sfts's and* $k : U \to V$ *be a mapping. The statements are identical in this case:*

(i) k *is a* $f\theta^*\mathcal{S}O$ *mapping.*
(ii) $k(I_{\mathfrak{S}_1}(\beta,\ \imath)) \le \theta^*\mathcal{S}I_{\mathfrak{S}_2}(k(\beta),\ \imath)$ *for each* $\beta \in I^U$ *and* $\imath \in I_0$.
(iii) $I_{\mathfrak{S}_1}(k^{-1}(\mu),\ \imath) \le k^{-1}(\theta^*\mathcal{S}I_{\mathfrak{S}_2}(\mu,\ \imath))$ *for each* $\mu \in I^V$ *and* $\imath \in I_0$.

Proof (i)⇒(ii) For all $\beta \in I^U$, $\imath \in I_0$ since $\mathfrak{S}_1(I_{\mathfrak{S}_1}(\beta,\ \imath)) \ge \imath$, we see that $k(I_{\mathfrak{S}_1}(\beta,\ \imath))$ is $\imath$-$f\theta^*\mathcal{S}o$ in V. From Theorem 2, $k(I_{\mathfrak{S}_1}(\beta,\ \imath)) = \theta^*\mathcal{S}I\mathfrak{S}_2(k(I_{\mathfrak{S}_1}(\beta,\ \imath)),\ \imath) \le \theta^*\mathcal{S}I\mathfrak{S}_2(k(\beta),\ \imath)$.

(ii)⇒(iii) Let $\mu \in I^V$, $\imath \in I_0$. By (ii) we have $k(I_{\mathfrak{S}_1}(k^{-1}(\mu),\ \imath)) \le \theta^*\mathcal{S}I\mathfrak{S}_2(kk^{-1}(\mu),\ \imath) \le \theta^*\mathcal{S}I\mathfrak{S}_2(\mu,\ \imath) \Rightarrow I_{\mathfrak{S}_1}(k^{-1}(\mu),\ \imath)) \le k^{-1}(\theta^*\mathcal{S}I\mathfrak{S}_2(\mu,\ \imath))$.

(iii)⇒(i) For each $\beta \in I^U$, $\imath \in I_0$ with $\mathfrak{S}_1(\beta) \ge \imath$, since $I_{\mathfrak{S}_1}(\beta,\ \imath) = \beta$, $k(\beta) \le \theta^*\mathcal{S}I\mathfrak{S}_2(k(\beta),\ \imath) \le k(\beta)$. Thus $k(\beta) = \theta^*\mathcal{S}I\mathfrak{S}_2(k(\beta),\ \imath)$. By Theorem 2, $k(\beta)$ is $\imath$-$f\theta^*\mathcal{S}o$ in V.

Theorem 5 *Let* $(U,\ \mathfrak{S}_1)$ *and* $(V,\ \mathfrak{S}_2)$ *be sfts's and* $k : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ *be a* $f\theta^*\mathcal{S}O$ *mapping. If* $\mu \in I^V$ *and* $\beta \in I^U$, $\mathfrak{S}_1(\underline{1} - \beta) \ge r$, $\imath \in I_0 \ni k^{-1}(\mu) \le \beta$, *then there exists an* $\imath$-$f\theta^*\mathcal{S}c$ *set* ν *of* $V \ni \mu \le \nu, k^{-1}(\nu) \le \beta$.

Proof Let $\nu = \underline{1} - k(\underline{1} - \beta)$. Since $k^{-1}(\mu) \le \beta$, we have $k(\underline{1} - \beta) \le \underline{1} - \mu$. since k is $f\theta^*\mathcal{S}O$ map, then ν is ι-$f\theta^*\mathcal{S}c$ in V and $k^{-1}(\nu) = \underline{1} - k^{-1}(k(\underline{1} - \beta)) \le \underline{1} - (\underline{1} - \beta) = \beta$.

Theorem 6 *If $k : (U, \mathfrak{S}_1) \to (V, \mathfrak{S}_2)$ be a $f\theta^*\mathcal{S}O$ mapping. Then for each $\mu \in I^V$, $\iota \in I_0$,*

$$k^{-1}(C_{\mathfrak{S}_2}(\theta I_{\mathfrak{S}_2}(\mu, \iota), \iota) \wedge k^{-1}(I_{\mathfrak{S}_2}(C_{\mathfrak{S}_2}(I_{\mathfrak{S}_2}(\mu, \iota), \iota), \iota)) \le C_{\mathfrak{S}_1}(k^{-1}(\mu), \iota).$$

Theorem 7 *If $k : (U, \mathfrak{S}_1) \to (V, \mathfrak{S}_2)$ be a bijective mapping such that*

$$k^{-1}(C_{\mathfrak{S}_2}(I_{\mathfrak{S}_2}(C_{\mathfrak{S}_2}(\mu, \iota), \iota), \iota)) \wedge k^{-1}(I_{\mathfrak{S}_2}(\theta C_{\mathfrak{S}_2}(\mu, \iota), \iota)) \le C_{\mathfrak{S}_1}(k^{-1}(\mu), \iota).$$

for each $\mu \in I^V$, $\iota \in I_0$, then k is $f\theta^*\mathcal{S}O$ map.

Theorem 8 *Let $(U, \mathfrak{S})$, $(V, \mathfrak{P})$ be sfts's and let $k : U \to V$ be a $f\theta^*\mathcal{S}C$ mapping. The statements are identical in this case:*

(i) If k is a surjective map and $k^{-1}(\alpha)\ \overline{q}\ k^{-1}(\rho)$ in U, then there exists $\alpha, \rho \in I^V$ $\ni \alpha\ \overline{q}\ \rho$.
(ii) $\theta^*\mathcal{S}I_\eta(\theta^*\mathcal{S}C_\eta(k(\beta), \iota), \iota) \le k(C_{\mathfrak{S}}(\beta, \iota))$, for each $\beta \in I^U$ and $\iota \in I_0$.

Theorem 9 *Let $(U, \mathfrak{S})$, $(V, \mathfrak{P})$ be sfts's and let $k : U \to V$ be a mapping. The statements are identical in this case:*

(i) k is called $f\theta^*\mathcal{S}C$ map.
(ii) $\theta^*\mathcal{S}C_\eta(k(\beta), \iota) \le k(C_{\mathfrak{S}}(\beta, \iota))$, for each $\beta \in I^U$ and $\iota \in I_0$.
(iii) If k is surjective, then for each subset μ of V and each ι-fo set α in U containing $k^{-1}(\mu)$, $\exists$ ι-$f\theta^*\mathcal{S}o$ set ρ of V containing $\mu \ni k^{-1}(\rho) \le \alpha$.

Theorem 4 to 9 are satisfied by YCO & ZCO maps.

Definition 7 A sfts $(U, \mathfrak{S})$ is called fuzzy

(i) $\theta^*\mathcal{S}$-T_1 (resp. T_1 [2], YT_1 & ZT_1) (briefly, $f\theta^*\mathcal{S}$-T_1 (resp. fT_1, fYT_1 & fZT_1)) space if for every two distinct fuzzy points u, v of U, there exists two ι-$f\theta^*\mathcal{S}o$ (resp. ι-fo, ι-fYo & ι-fZo) sets β, $\mu \ni u \in \beta$, $v \notin \beta$ and $v \in \mu$, $u \notin \mu$.
(ii) $\theta^*\mathcal{S}$-T_2 (resp. T_2 [2] YT_2 & ZT_2) (briefly, $f\theta^*\mathcal{S}$-T_2 (resp. fT_2, fYT_2 & fZT_2)) space if for every two distinct fuzzy points u, v of U, there exists two disjoint ι-$f\theta^*\mathcal{S}o$ (resp. ι-fo, ι-fYo & ι-fZo) sets β, $\mu \ni u \in \beta$, $v \in \mu$.
(iii) $\theta^*\mathcal{S}$-connected (resp. connected [2], Y-connected & Z-connected) (briefly, *$f\theta^*\mathcal{S}$-con* (resp. *$fcon$*, *$fYcon$* & *$fZcon$*)) if it cannot be expressed as the union of two disjoint non-empty ι-$f\theta^*\mathcal{S}o$ (resp. ι-fo, ι-fYo & ι-fZo) sets of X. If U is not *$f\theta^*\mathcal{S}$-con* (resp. not *$fcon$*, *$fYcon$* & *$fZcon$*), then it is fuzzy $\theta^*\mathcal{S}$-disconnected (resp. fuzzy disconnected, fuzzy Y-disconnected & fuzzy Z-disconnected).
(iv) $\theta^*\mathcal{S}$-Lindeloff (Lindeloff [2], Y-Lindeloff & Z-Lindeloff) if every ι-$f\theta^*\mathcal{S}o$ (resp. ι-fo, ι-fYo & ι-fZo) cover of U has a countable subcover.

(v) $\theta^*\mathcal{S}$-compact (resp. compact [2], Y-compact & Z-compact) (briefly, $f\theta^*\mathcal{S}$-com (resp. $fcom$, $fYcom$ & $fZcom$)) if for every ι-$f\theta^*\mathcal{S}o$ (resp. ι-fo, ι-fYo & ι-fZo) cover of U has a finite subcover.

Definition 8 Let $(U, \mathfrak{S})$ be a sfts and $\iota \in I_0$. A fuzzy set $\mu \in I^U$ is called fuzzy compact [2] (briefly $fcom$) in $(U, \mathfrak{S})$ iff for each family $\{\beta_i \in I^U | \mathfrak{S}(\beta_i) \geq r, i \in \Gamma\} \ni \mu \leq \bigvee_{i \in \Gamma} \beta_i$ there exists a finite index set $\Gamma_0 \subset \Gamma \ni \mu \leq \bigvee_{i \in \Gamma_0} \beta_i$.$(U, \mathfrak{S})$ is called $fcom$ iff $\underline{1}$ is $fcom$ in $(U, \mathfrak{S})$.

Definition 9 Let $(U, \mathfrak{S})$ be a sfts and $\iota \in I_0$. A fuzzy set $\mu \in I^U$ is called fuzzy $\theta^*\mathcal{S}$ (resp. Y & Z)-compact (briefly, $f\theta^*\mathcal{S}$-com, fY-com & fZ-com) in $(U, \mathfrak{S})$ if for each family $\{\beta_i \in I^U | \beta_i$ is ι-$f\theta^*\mathcal{S}o$ (resp. ι-fYo & ι-fZo), $i \in \Gamma\} \ni \mu \leq \bigvee_{i \in \Gamma} \beta_i$ there exists a finite index set $\Gamma_0 \subset \Gamma \ni \mu \leq \bigvee_{i \in \Gamma_0} \beta_i$. $(U, \mathfrak{S})$ is called $f\theta^*\mathcal{S}$ (resp. Y & Z)-com if $\underline{1}$ is $f\theta^*\mathcal{S}$-com (resp. fY-com & fZ-com) in $(U, \mathfrak{S})$.

Theorem 10 *Let $(U, \mathfrak{S})$, $(V, \mathfrak{P})$ be sfts's and let $k : U \to V$ be a bijective $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) mapping. The statements are identical in this case:*

(i) If U is a fT_i-space, then V is $f\theta^*\mathcal{S}$-T_i (resp. fY-T_i & fZ-T_i) where $i = 1, 2$.
(ii) If V is an $f\theta^*\mathcal{S}$-com (resp. fY-com & fZ-com) (resp. fuzzy $\theta^*\mathcal{S}$ (resp. Y & Z)-Lindeloff) space, then U is $fcom$ (resp. fuzzy Lindeloff).

Theorem 11 *Let $(U, \mathfrak{S}_1)$ and $(V, \mathfrak{S}_2)$ be sfts's. If $k : U \to V$ is a surjective $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) mapping and V is $f\theta^*\mathcal{S}$-con (resp. fY-con & fZ-con) space, then U is $fcon$.*

Remark 3 Let $(U, \mathfrak{S}_1)$ and $(V, \mathfrak{S}_2)$ be sfts's and $k : U \to V$ be a mapping. The composition of two $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) mappings need not be $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map as shown by the following example.

Example 7 Let $U = V = W = \{k, l, m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5$, $\rho_1(l) = 0.6$, $\rho_1(m) = 0.4$; $\rho_2(k) = 0.5$, $\rho_2(l) = 0.4$, $\rho_2(m) = 0.4$; $\rho_3(k) = 0.5$, $\rho_3(l) = 0.4$, $\rho_3(m) = 0.6$ Let $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0}, \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_3, \\ 0 & \text{ow.} \end{cases} \qquad \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0}, \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \\ 0 & \text{ow.} \end{cases}$$

$$\text{and } \mathfrak{S}_3(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0}, \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then the identity mapping $k : (U, \mathfrak{S}_1) \to (V, \mathfrak{S}_2)$&$g : (V, \mathfrak{S}_2) \to (W, \mathfrak{S}_3)$ are $f\theta^\star\mathcal{S}O$-map but $g \circ k$ is not $f\theta\mathcal{S}O$-map, since the image under $g \circ k$ of the set ρ_3 is not $f\theta^\star\mathcal{S}o$ in $(W, \mathfrak{S}_3)$.

Example 8 Let $U = V = W = \{k, l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5,\ \rho_1(l) = 0.6,\ \rho_1(m) = 0.4;\ \rho_2(k) = 0.5,\ \rho_2(l) = 0.4,\ \rho_2(m) = 0.4;\ \rho_4(k) = 0.5,\ \rho_4(l) = 0.4,\ \rho_4(m) = 0.5$ Let $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_4, \\ 0 & \text{ow.} \end{cases} \quad \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \\ 0 & \text{ow.} \end{cases}$$

$$\text{and } \mathfrak{S}_3(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then the identity mapping $k : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2) \& g : (V,\ \mathfrak{S}_2) \to (W,\ \mathfrak{S}_3)$ are fYO-map but $g \circ k$ is not fYO-map, since the image under $g \circ k$ of the set ρ_4 is not fYo in $(Z,\ \mathfrak{S}_3)$.

Example 9 Let $U = V = W = \{k, l,\ m\}$ and $\rho_1, \rho_2 \in I^U$ defined as $\rho_1(k) = 0.5,\ \rho_1(l) = 0.5,\ \rho_1(m) = 0.4;\ \rho_2(k) = 0.5,\ \rho_2(l) = 0.5,\ \rho_2(m) = 0.9;\ \rho_5(k) = 0.5,\ \rho_5(l) = 0.5,\ \rho_5(m) = 0$. Let $\mathfrak{S}_i : I^U \to I$, defined as

$$\mathfrak{S}_1(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_5, \\ 0 & \text{ow.} \end{cases} \quad \mathfrak{S}_2(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \\ 0 & \text{ow.} \end{cases}$$

$$\text{and } \mathfrak{S}_3(\beta) = \begin{cases} 1 & \text{if } \beta \in \{\underline{0},\ \underline{1}\}, \\ \frac{1}{2} & \text{if } \beta = \rho_1, \rho_2, \\ 0 & \text{ow.} \end{cases}$$

Then the identity mapping $k : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2) \& g : (V,\ \mathfrak{S}_2) \to (Z,\ \mathfrak{S}_3)$ are fZO-map but $g \circ k$ is not fZO-map, since the image under $g \circ k$ of the set ρ_5 is not fYo in $(W,\ \mathfrak{S}_3)$.

Theorem 12 *Let $(U,\ \mathfrak{S}_1)$, $(V,\ \mathfrak{S}_2)$ and $(W,\ \mathfrak{S}_3)$ be sfts's. If $k : (U,\ \mathfrak{S}_1) \to (V,\ \mathfrak{S}_2)$ and $g : (V,\ \mathfrak{S}_2) \to (W,\ \mathfrak{S}_3)$ are mappings, then*

(i) If k is fO map and g is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map, then $g \circ k$ is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) mapping.
(ii) If $g \circ k$ is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) mapping and k is a surjective continuous map, then g is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map.
(ii) If $g \circ k$ is fO mapping and g is an injective $f\theta^*\mathcal{S}Cts$ (resp. $fYCts$ & $fZCts$) map, then k is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map.

Proof (i) Let $\mu \in \mathfrak{S}_1$. Since k is fO map, then $k(\mu)$ is an $\imath$-fo set in $(V, \mathfrak{S}_2)$. Since g is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map, then $g(k(\mu)) = (g \circ k)(\mu)$ is $\imath$-$f\theta^*\mathcal{S}o$ (resp. $\imath$-fYo & $\imath$-fZo) set in $(W, \mathfrak{S}_3)$. Hence $g \circ k$ is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map.

(ii) Let $\mu \in \mathfrak{S}_2$. Since k is $fCts$, then $k^{-1}(\mu)$ is an $\imath$-fo set in $(U, \mathfrak{S}_1)$. But $g \circ k$ is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map, then $(g \circ k)(k^{-1}(\mu))$ is $\imath$-$f\theta^*\mathcal{S}o$ (resp. $\imath$-fYo & $\imath$-fZo) set in $(W, \mathfrak{S}_3)$. Hence by surjective of k, we have $g(\mu)$ is $\imath$-$f\theta^*\mathcal{S}o$ set of $(W, \mathfrak{S}_3)$. Hence, g is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map.

(iii) Let $\mu \in \mathfrak{S}_1$ and $g \circ k$ be an fO map. Then $(g \circ k)(\mu) = g(k(\mu)) \in \mathfrak{S}_3$. Since g is an injective $f\theta^*\mathcal{S}Cts$ (resp. $fYCts$ & $fZCts$) map, hence $k(\mu)$ is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map in $(V, \mathfrak{S}_2)$. Therefore k is $f\theta^*\mathcal{S}O$ (resp. fYO & fZO).

4 Conclusion

In this paper, we introduce and investigate some new classes of mappings called $f\theta^*\mathcal{S}O$ (resp. fYO & fZO) map and $f\theta^*\mathcal{S}C$ (resp. fYC & fZC) map to the fuzzy topological spaces in $\hat{S}$ostak's sense. These open and closed mappings are extended to their contra open and contra closed mappings of $f\theta^*\mathcal{S}o$ (resp. fYo & fZo) set in future. Also, these mappings can be extended to some of their functions such as homeomorphism, continuous function, irresolute functions in $f\theta^*\mathcal{S}o$ (resp. fYo & fZo) set.

References

1. Azad, K.K.: On fuzzy semi continuity, fuzzy almost continuity and fuzzy weakly continuity. J. Math. Anal. Appl. 82, 14–32 (1981).
2. Bamini, S., Saraswathi, M., Vijayalakshmi, B., Vadivel, A.: Fuzzy M-open and fuzzy M-closed mappings in $\hat{S}$ostak's fuzzy topological spaces. International Journal of Innovative Technology and Exploring Engineering (IJITEE). 8 (7), 1661-1668 (2019).
3. Bhattacharyya, A., Mukherjee, M.N.: On fuzzy δ-almost continuous and δ^*-almost continuous functions. J. Tripura Math. Soc. 2, 45–57 (2000).
4. Bin Shahna, A.S.: On fuzzy strong semi-continuity and fuzzy precontinuity. Fuzzy Sets and Systems. 44, 303-308 (1991).
5. Caldas, M., Ganster, M., Georgiou, D.N., Jafari, S., Noiri, T.: On θ-semi-open sets and separation axioms in topological spaces. Carpathian. J. Math. 24 (1), 13-22 (2008).
6. Chang, C.L.: Fuzzy topological spaces. J. Math. Anal. Appl. 24, 182–189 (1968).
7. El-Maghrabi, A.I., Embaby, O.A., Mubarki, A.M.: Some properties of θ^*-open sets in topological spaces. Jokull Journal. 64 (4), 59-73 (2014).
8. El-Shafei, M.E., Zakari, A.: θ-generalized closed sets in fuzzy topological spaces. Arab. J. Sci. Eng. Sect. A Sci. 31 (2), 197–206 (2006).
9. Ganguly, S., Saha, S.: A note on δ-continuity and δ-connected sets in fuzzy set theory. Simon Stein. 62, 127-141 (1988).
10. Hanafy, I.M.: Fuzzy γ-open sets and fuzzy γ-continuity. J. Fuzzy Math. 7 (2), 419-430 (1999).

11. Hazra, R.N., Samanta, S.K., Chattopadhyay, K.C.: Fuzzy topology redefined. Fuzzy Sets and Systems. 4, 79-82 (1992).
12. Hazra, R.N., Samanta, S.K., Chattopadhyay, K.C.: Gradation of openness: fuzzy topology. Fuzzy Sets and Systems. 49 (2), 237-242 (1992).
13. Kim, Y.C., Park, J.W.: Some properties of r-generalized fuzzy closed sets. Far East J. of Math. Science. 7 (3), 253-268 (2002).
14. Levine, N.: Semi-open sets and Semi-continuity in topological spaces. American Mathematical Monthly. 70, 36-41 (1963).
15. Mughil, A., Vadivel, A., Uma Maheswari, O., Saravanakumar, G.: Fuzzy θ^*-semiopen sets in $\hat{S}$ostak's fuzzy topological spaces. (Submitted).
16. Mughil, A., Vadivel, A., Uma Maheswari, O., Saravanakumar, G.: Continuous Maps Via ι-fuzzy θ^*-semiopen Sets in $\hat{S}$ostak's Fuzzy Topological Spaces. (Submitted).
17. Park, J.H., Lee, B.Y.: Fuzzy semi-preopen sets and fuzzy semi-precontinuous mappings. Fuzzy Sets and Systems. 67, -395 (1994).
18. Park, J.H., Lee, B.Y., Son, M.J.: On δ-semi-open sets in topological spaces. Journal of the Indian Academy of Mathematics. 19 (1), 59-67 (1997).
19. Ramadan, A.A.: Smooth topological spaces. Fuzzy Sets and Systems. 48, 371-375 (1992).
20. Ramadan, A.A., Abbas, S.E., Coker, D.: Fuzzy γ-continuity in Sostak's fuzzy topology. The Journal of Fuzzy Sets and Systems. 48, 371-375 (1992).
21. Raychaudhhuri, S., Mukherjee, N.: On δ-almost continuity and δ-pre-open sets. Bull. Inst. Math. Acad. Sinica. 21, 357-366 (1993).
22. Samanta, S.K., Chattopadhyay, K.C.: Fuzzy topology, Fuzzy closure operator, Fuzzy compactness and fuzzy connectedness. Fuzzy Sets and Systems. 54, 207-212 (1993).
23. $\check{S}$ostak, A.S.: On a fuzzy topological structure. Rend. Circ. Matem. Palermo Ser. II. 11, 89-103 (1985).
24. Thakur, S.S., Khare, R.K.: Fuzzy semi δ-preopen sets and fuzzy semi δ-precontinuous mappings. Universitatea din Bacau studii si cerceturi Strintitice Seria Matematica. 14, 201–211 (2004).
25. Velicko, N.V.: H-closed topological spaces. Amer. Math. Soc. Transl. 78, 103-118 (1968).
26. Vijayalakshmi, B., Bamini, S., Saraswathi, M., Vadivel, A.: Fuzzy M-open sets in Šostak's Fuzzy Topological Spaces. Malaya Journal of Matematik. S (1), 234-242 (2019).

Nano Z Seperation Axioms

Xavier Arul Selvaraj and Ushakoyala Balakrishna

Abstract Based on nano Z open set, we discuss and study about separation axioms, connectedness and compactness in a nano topological spaces. In this paper, the types of separation axioms such as $\mathfrak{N}ZT_0$, $\mathfrak{N}ZT_1$ and $\mathfrak{N}ZT_2$ spaces are introduced and discuss in nano topological spaces. Also, nano Z regular space and nano Z normal space of nano Z open sets are established in nano topological spaces. Finally, nano Z compactness and nano Z connectedness are study and dealt with some properties in nano topological spaces.

Keywords $\mathfrak{N}ZT_0$ · $\mathfrak{N}ZT_1$ · $\mathfrak{N}ZT_2$ · $\mathfrak{N}Zreg$ · $\mathfrak{N}Znor$ spaces · nano Z compactness · nano Z connectedness

AMS (2000) subject classification: 54A05 · 54A40 · 54B05

1 Introduction

Lellis Thivagar [6] introduced the concept of Nano topology, which was defined in terms of approximations and boundary region of a subset of a universe using an equivalence relation on it, as well as Nano closed sets, Nano-interior, and Nano-closure. Magharabi and Mubarki [3] in 2011 developed the concept of Z-open sets in topological spaces and examined some of their features. Because of their applicability in numerous domains of mathematics and other real fields, the class of sets known as Z-open sets is becoming more relevant in topological spaces. In $\mathfrak{N}ts$, Khalaf and Ahmed Elmoasry [4] explored nano separation axioms. We offer the notion of nano

X. A. Selvaraj
Department of Mathematics, Government Arts and Science College for Women, Bargur, Krishnagiri 635 104, India

Department of Mathematics, DDE, Annamalai University, Annamalai Nagar 608 002, India

U. Balakrishna (✉)
H & S Department, Keshav Memorial Institute of Technology, Narayanaguda, Hyderabad 500 029, India

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_7

Z-open sets [1] and investigate their features in nano topological space based on these motives.

The purpose of this paper is to discuss the types of separation axioms such as $\mathfrak{N}ZT_0$, $\mathfrak{N}ZT_1$ and $\mathfrak{N}ZT_2$ spaces are introduced in nano topological spaces. Also, nano Z regular and nano Z normal spaces are established in nano topological spaces. Also, the concept of nano Z compactness and nano Z connectedness by using $\mathfrak{N}Zo$ sets in $\mathfrak{N}ts$'s are discussed.

2 Preliminaries

Definition 1 ([1]) Let $(U, \mathscr{T}_R(A))$ be a $\mathfrak{N}ts$. Let K be an $\mathfrak{N}s$ in $(U, \mathscr{T}_R(A))$. Then K is said to be a nano

(i) Z-open (briefly, $\mathfrak{N}Zo$) set if $K \subseteq \mathfrak{N}cl(\mathfrak{N}int_\delta(K)) \cup \mathfrak{N}int(\mathfrak{N}cl(K))$,
(ii) Z-closed (briefly, $\mathfrak{N}Zc$) set if $\mathfrak{N}int(\mathfrak{N}cl_\delta(K)) \cap \mathfrak{N}cl(\mathfrak{N}int(K)) \subseteq K$.

The family of all $\mathfrak{N}Zo$ (resp. $\mathfrak{N}Zc$) sets of a space $(U, \mathscr{T}_R(A))$ will be as always denoted by $\mathfrak{N}ZO(U, A)$ (resp. $\mathfrak{N}ZC(U, A)$).

Definition 2 ([1]) Let $(U, \mathscr{T}_R(A))$ be a $\mathfrak{N}ts$ and let $K \subseteq U$ then the

(i) nano Z-interior of K is the union of all $\mathfrak{N}Zo$ sets contained in K and denoted by $\mathfrak{N}Zint(K)$.
(ii) nano Z-closure of K is the intersection of all $\mathfrak{N}Zc$ sets containing K and denoted by $\mathfrak{N}Zcl(K)$.

Definition 3 ([2]) A function $h : (P_1, \mathscr{T}_R(A)) \to (P_2, \mathscr{S}_{R'}(B))$ is said to be Nano Z continuous (briefly, $\mathfrak{N}ZCts$), if for each $\mathfrak{N}\mathfrak{l}\mathfrak{v}c$ set M of P_2, the set $h^{-1}(M)$ is $\mathfrak{N}Zc$ set of P_1.

Definition 4 [2] A function $h : (P_1, \mathscr{T}_R(A)) \to (P_2, \mathscr{S}_{R'}(B))$ is called Nano Z irresolute (briefly, $\mathfrak{N}ZIrr$) function, if for each $\mathfrak{N}Zc$ subset M of Q, the set $h^{-1}(M)$ is $\mathfrak{N}Zc$ subset of P.

All other undefined definitions and properties in this paper are from [1, 2, 5–9].

Throughout this paper, $(P, \mathscr{T}_R(A))$ is a $\mathfrak{N}ts$ with respect to A where $A \subseteq P$, R is an equivalence relation on P. Then P/R denotes the family of equivalence classes of P by R.

3 Nano Z Separation Axioms

Definition 5 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z-T_0 (briefly, $\mathfrak{N}ZT_0$) space, if for any two points $l_1 \neq l_2$, $\exists$ $\mathfrak{N}Zo$ set containing one of them but not the other.

Clearly, every $\mathfrak{N}T_0$ space is $\mathfrak{N}ZT_0$ space.

Example 1 Let $U = \{Z_a, Z_b, Z_c, Z_d, Z_e\}$ with $U/R = \{\{Z_c\}, \{Z_a, Z_b\}, \{Z_d, Z_e\}\}$ and $P = \{Z_a, Z_c\}$. The $\mathfrak{N}t$ $\mathscr{T}_R(P) = \{U, \phi, \{Z_c\}, \{Z_a, Z_b\}, \{Z_a, Z_b, Z_c\}\}$. If $Z_a \neq Z_c$, then $\{Z_a\}$ and $\{Z_c\}$ are $\mathfrak{N}Zo$ sets but $\{Z_a\}$ is not $\mathfrak{N}o$ set. So, it is $\mathfrak{N}ZT_0$ space is $\mathfrak{N}T_0$ space.

Theorem 1 *Let $\mathfrak{N}ZC(P, A)$ is closed under arbitrary intersection and a $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is a $\mathfrak{N}ZT_0$ space iff $\mathfrak{N}Z$ closures of distinct points are distinct.*

Proof Let l_1, l_2 be distinct points of P. Since P is a $\mathfrak{N}ZT_0$ space, $\exists \mathfrak{N}Zo$ $\mu \ni l_1 \in \mu \& l_2 \notin \mu$. Consequently, $P - \mu$ is a $\mathfrak{N}Zc$ set containing l_2 but not l_1. But $\mathfrak{N}Zcl(\{l_2\})$ is the intersection of all $\mathfrak{N}Zc$ sets containing l_2. Hence $l_2 \in \mathfrak{N}Zcl(\{l_2\})$ but $l_1 \notin \mathfrak{N}Zcl(\{l_2\})$ as $l_1 \notin P - \mu$. Therefore $\mathfrak{N}Zcl(\{l_1\}) \neq \mathfrak{N}Zcl(\{l_2\})$.

Conversely, let $\mathfrak{N}Zcl(\{l_1\}) \neq \mathfrak{N}Zcl(\{l_2\})$ for $l_1 \neq l_2$. Then there exist at least one point $l_3 \in P \ni l_3 \in \mathfrak{N}Zcl(\{l_1\})$ but $l_3 \notin \mathfrak{N}Zcl(\{l_2\})$. Suppose $l_1 \notin \mathfrak{N}Zcl(\{l_2\})$ because if $l_1 \in \mathfrak{N}Zcl(\{l_2\})$ then $\{l_1\} \subset \mathfrak{N}Zcl(\{l_2\})$ implies $\mathfrak{N}Zcl$ $(\{l_1\}) \subset \mathfrak{N}Zcl(\{l_2\})$. So $l_3 \in \mathfrak{N}Zcl(\{l_2\})$ which is a contradiction. Hence $l_1 \notin \mathfrak{N}Zcl(\{l_2\})$ which implies $l_1 \in P - \mathfrak{N}Zcl(\{l_2\})$ which is a $\mathfrak{N}Zo$ set containing l_1 but not l_2. Hence P is a $\mathfrak{N}ZT_0$ space.

Theorem 2 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is a injective, $\mathfrak{N}Zirr$ function and Q is a $\mathfrak{N}ZT_0$ then P is a $\mathfrak{N}ZT_0$.*

Proof Suppose Q is a $\mathfrak{N}ZT_0$ space. Let l_1 & l_2 be any two distinct points in P. Since h is injective $h(l_1)$ & $h(l_2)$ are distinct points in Q. Since Q is a $\mathfrak{N}ZT_0$ space, $\exists$ $\mathfrak{N}Zo$ μ in $Q \in h(l_1)$ but not $h(l_2)$. Again since h is $\mathfrak{N}Zirr$, $h^{-1}(\mu)$ is a $\mathfrak{N}Zo$ set in $P \in l_1$ but not l_2. Therefore P is a $\mathfrak{N}ZT_0$ space.

Theorem 3 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is a injective, $\mathfrak{N}ZCts$ mapping and Q is a $\mathfrak{N}T_0$ space then P is a $\mathfrak{N}ZT_0$ space.*

Proof Let l_1 & l_2 be any two distinct points in P. Since h is injective $h(l_1)$ & $h(l_2)$ are distinct points in Q. Since Q is a $\mathfrak{N}T_0$ space, $\exists$ $\mathfrak{N}o$ set μ in $Q \in h(l_1)$ but not $h(l_2)$. Again since h is $\mathfrak{N}ZCts$, $h^{-1}(\mu)$ is a $\mathfrak{N}Zo$ set in $P \in l_1$ but not l_2. Hence P is a $\mathfrak{N}ZT_0$ space.

Definition 6 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z-T_1 (briefly, $\mathfrak{N}ZT_1$) space, if for any two points $l_1 \neq l_2$, $\exists$ $\mathfrak{N}Zo$ sets μ & H with $l_1 \in \mu, l_2 \notin \mu$ & $l_1 \notin H, l_2 \in H$.

Clearly, every $\mathfrak{N}T_1$ space is $\mathfrak{N}ZT_1$ space.

Example 2 Let $U = \{Z_a, Z_b, Z_c, Z_d, Z_e\}$ with $U/R = \{\{Z_c\}, \{Z_a, Z_b\}, \{Z_d, Z_e\}\}$ and $P = \{Z_a, Z_c\}$. The $\mathfrak{N}t$ $\mathscr{T}_R(P) = \{U, \phi, \{Z_c\}, \{Z_a, Z_b\}, \{Z_a, Z_b, Z_c\}\}$. If $Z_a \neq Z_c$, then $\{Z_a, Z_b\}$ and $\{Z_b, Z_c\}$ are $\mathfrak{N}Zo$ sets and $\{Z_a\} \in \{Z_a, Z_b\}$, $\{Z_c\} \notin \{Z_a, Z_b\}$ and $\{Z_a\} \notin \{Z_b, Z_c\}$, $\{Z_c\} \notin \{Z_b, Z_c\}$. But $\{Z_a\}$ is not $\mathfrak{N}o$ set. So, it is $\mathfrak{N}ZT_1$ space is $\mathfrak{N}T_1$ space.

Theorem 4 *Let $\mathfrak{N}ZO(P, A)$ is closed under a. u. then a $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is a $\mathfrak{N}ZT_1$ space iff every singleton subset $\{l_1\}$ of P is a $\mathfrak{N}Zc$ set.*

Proof Let P be a $\mathfrak{N}ZT_1$ space & $l_1 \in P$. Let $l_2 \in P - \{l_1\}$. Then for $l_1 \neq l_2$ there exist a $\mathfrak{N}Zo$ set $\mu_{l_2} \ni l_2 \in \mu_{l_2}$ & $l_1 \notin \mu_{l_2}$. Consequently, $l_2 \in \mu_{l_2} \subset P - \{l_1\}$. That is $P - \{l_1\} = \bigcup\{\mu_{l_2} : l_2 \in P - \{l_1\}\}$, which is the union of $\mathfrak{N}Zo$ sets, so $\mathfrak{N}Zo$. Hence $\{l_1\}$ is a $\mathfrak{N}Zc$ set.

Conversely, suppose $\{p_1\}$ is a $\mathfrak{N}Zc$ set $\forall$ $p_1 \in P$. Let p_1 & $p_2 \in P$ with $p_1 \neq p_2$. Now $p_1 \neq p_2$ implies $p_2 \in P - \{p_1\}$. Hence $P - \{p_1\}$ is a $\mathfrak{N}Zo$ set $\in p_2$ but not p_1. Similarly $P - \{p_2\}$ is a $\mathfrak{N}Zo$ set $\in p_1$ but not p_2. Therefore, P is a $\mathfrak{N}ZT_1$ space.

Theorem 5 *If $k : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is a injective, $\mathfrak{N}Zirr$ function and Q is a $\mathfrak{N}ZT_1$ space then P is also a $\mathfrak{N}ZT_1$ space.*

Proof Let p_1 & p_2 be pair of distinct points in P. Since k is injective, there exist two distinct points m_1 & m_2 of $Q \ni k(p_1) = m_1$ & $k(p_2) = m_2$. Since Q is a $\mathfrak{N}ZT_1$ space, $\exists$ $\mathfrak{N}Zo$ sets ρ & η in $Q \ni m_1 \in \rho,\ m_2 \notin \rho$ & $m_1 \notin \eta,\ m_2 \in \eta$. That is $p_1 \in k^{-1}(\rho),\ p_1 \notin k^{-1}(\eta)$ & $p_2 \in k^{-1}(\eta),\ p_2 \notin k^{-1}(\rho)$. Since k is $\mathfrak{N}Zirr$ function, $k^{-1}(\rho)$& $k^{-1}(\eta)$ are $\mathfrak{N}Zo$ sets in P. Thus for two distinct points p_1 & p_2 of P there exists $\mathfrak{N}Zo$ sets $k^{-1}(\rho)$ & $k^{-1}(\eta) \ni p_1 \in k^{-1}(\rho),\ p_1 \notin k^{-1}(\eta)$ & $p_2 \in k^{-1}(\eta),\ p_2 \notin k^{-1}(\rho)$. Therefore P is $\mathfrak{N}ZT_1$ space.

Theorem 6 *If $k : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is a $\mathfrak{N}ZCts$ injection and Q is a $\mathfrak{N}T_1$ space then P is a $\mathfrak{N}ZT_1$ space.*

Proof For any two distinct points p_1 & p_2 of P, there exist two distinct points m_1 & m_2 of $Q \ni k(p_1) = m_1$ & $k(p_2) = m_2$. Since Q is a $\mathfrak{N}T_1$ space, there exist a $\mathfrak{N}o$ sets ρ & ν in $Q \ni m_1 \in \rho,\ m_2 \notin \rho$ & $m_1 \notin \nu,\ m_2 \in \nu$. That is $p_1 \in k^{-1}(\rho),\ p_1 \notin k^{-1}(\nu)$ & $p_2 \in k^{-1}(\nu),\ p_2 \notin k^{-1}(\rho)$. Since k is a $\mathfrak{N}ZCts$ function, $k^{-1}(\rho)$& $k^{-1}(\nu)$ are $\mathfrak{N}Zo$ sets in P. Thus for two distinct points p_1 & p_2 of P there exists $\mathfrak{N}Zo$ sets $k^{-1}(\rho)$ & $k^{-1}(\nu) \ni p_1 \in k^{-1}(\rho),\ p_1 \notin k^{-1}(\nu)$ & $p_2 \in k^{-1}(\nu),\ p_2 \notin k^{-1}(\rho)$. Therefore P is $\mathfrak{N}ZT_1$ space.

Definition 7 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano ZT_2 (briefly, $\mathfrak{N}ZT_2$) space, if for any two points $l \neq l_2$, $\exists$ disjoint $\mathfrak{N}Zo$ sets μ & ν with $l \in \mu$ & $m \in \nu$.

Clearly, every $\mathfrak{N}T_2$ space is $\mathfrak{N}ZT_2$.

Example 3 Let $U = \{Z_a, Z_b, Z_c, Z_d, Z_e\}$ with $U/R = \{\{Z_c\}, \{Z_a, Z_b\}, \{Z_d, Z_e\}\}$ and $P = \{Z_a, Z_c\}$. The $\mathfrak{N}t$ $\mathscr{T}_R(P) = \{U, \phi, \{Z_c\}, \{Z_a, Z_b\}, \{Z_a, Z_b, Z_c\}\}$. If $Z_a \neq Z_c$, then $\{Z_a\}$ and $\{Z_c\}$ are $\mathfrak{N}Zo$ sets and $\{Z_a\} \cap \{Z_c\} \neq \emptyset$. But $\{Z_a\}$ is not $\mathfrak{N}o$ set. So, it is $\mathfrak{N}ZT_2$ space is $\mathfrak{N}T_2$ space.

Theorem 7 *If $k : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is a $\mathfrak{N}ZCts$ injection and Q is a $\mathfrak{N}T_2$ space then P is a $\mathfrak{N}ZT_2$ space.*

Proof For any two distinct points l_1 & l_2 of P, there exist two distinct points m_1 & m_2 of $Q \ni k(l_1) = m_1$ & $k(l_2) = m_2$. Since Q is a $\mathfrak{N}T_2$ space, $\exists$ disjoint $\mathfrak{N}o$ sets ρ & η in $Q \ni m_1 \in \rho$ & $m_2 \in \eta$. That is $l_1 \in k^{-1}(\rho)$ & $l_2 \in k^{-1}(\eta)$. Since k is a $\mathfrak{N}ZCts$ function, $k^{-1}(\rho)$& $k^{-1}(\eta)$ are $\mathfrak{N}Zo$ sets in P. Further as k is injective $k^{-1}(\rho) \cap k^{-1}(\eta) = k^{-1}(\rho \cap \eta) = k^{-1}(\emptyset) = \emptyset$. Thus for two distinct points l_1 & l_2 of P there exists $\mathfrak{N}Zo$ sets $k^{-1}(\rho)$ & $k^{-1}(\eta) \ni l_1 \in k^{-1}(\rho)$ & $l_2 \in k^{-1}(\eta)$. Therefore P is $\mathfrak{N}ZT_2$ space.

Theorem 8 *If $k : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is a injective, $\mathfrak{N}Zirr$ function and Q is a $\mathfrak{N}ZT_2$ space then P is also a $\mathfrak{N}ZT_2$ space.*

Proof Let l_1 & l_2 be a pair of distinct points in P. Since k is injective, $\exists$ two distinct points m_1 & m_2 of $Q \ni k(l_1) = m_1$ & $k(l_2) = m_2$. Since Q is a $\mathfrak{N}ZT_2$ space, there exist a disjoint $\mathfrak{N}Zo$ sets ρ & η in $Q \ni m_1 \in \rho$ & $m_2 \in \eta$. That is $l_1 \in k^{-1}(\rho)$ & $l_2 \in k^{-1}(\eta)$. Since k is a $\mathfrak{N}Zirr$ function, $k^{-1}(\rho)$& $k^{-1}(\eta)$ are disjoint $\mathfrak{N}Zo$ sets in P. Thus for two distinct points l_1 & l_2 of P there exists disjoint $\mathfrak{N}Zo$ sets $k^{-1}(\rho)$ & $k^{-1}(\eta) \ni l_1 \in k^{-1}(\rho)$ & $l_2 \in k^{-1}(\eta)$. Therefore P is $\mathfrak{N}ZT_2$ space.

Theorem 9 *A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is a $\mathfrak{N}ZT_2$ space iff for each $l_1 \neq l_2$,$\exists$ $\mathfrak{N}Zo$ set $\rho \ni l_1 \in \rho$ & $l_2 \notin \mathfrak{N}Zcl(\rho)$.*

Proof Assume that P is a $\mathfrak{N}ZT_2$ space. Let $l_1,\ l_2 \in P$ & $l_1 \neq l_2$, then there exists disjoint $\mathfrak{N}Zo$ sets ρ & $\eta \ni l_1 \in \rho$ & $l_2 \in \eta$. Clearly, $P - \eta$ is $\mathfrak{N}Zc$ set. Since $\rho \cap \eta = \emptyset,\ \rho \subset P - \eta$. Therefore $\mathfrak{N}Zcl(\rho) \subset \mathfrak{N}Zcl(P - \eta) = P - \eta$. Now $l_2 \notin P - \eta \implies l_2 \notin \mathfrak{N}Zcl(\rho)$.

Conversely, let $l_1,\ l_2 \in P$ & $l_1 \neq l_2$. By hypothesis $\exists$ $\mathfrak{N}Zo$ set $\rho \ni l_1 \in \rho$ & $l_2 \notin \mathfrak{N}Zcl(\rho)$. This implies $\exists$ $\mathfrak{N}Zc$ set $\eta \ni l_2 \notin \eta$. Therefore $l_2 \in P - \eta$ is a $\mathfrak{N}Zo$ set. Thus $\exists$ two disjoint $\mathfrak{N}Zo$ sets ρ and $P - \eta \ni l_1 \in \rho$ & $l_2 \in P - \eta$. Therefore P is a $\mathfrak{N}ZT_2$ space.

4 Nano *Z* Regular Spaces

Definition 8 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z regular (briefly, $\mathfrak{N}Zreg$) space, if for any point η & $\mathfrak{N}c$ set $\kappa \notin \eta$, $\exists$ disjoint $\mathfrak{N}Zo$ sets μ & ν with $\eta \in \mu$ & $\kappa \subset \nu$.

Clearly, every $\mathfrak{N}reg$ space is $\mathfrak{N}Zreg$ space.

Theorem 10 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}Cts$ bijective, $\mathfrak{N}Zo$ function and P is a $\mathfrak{N}reg$ space then Q is $\mathfrak{N}Zreg$.*

Proof Let κ be a $\mathfrak{N}c$ set in Q & $l_2 \notin \kappa$. Take $l_2 = h(l_1)$ for some $l_1 \in P$. Since h is $\mathfrak{N}Cts$, $h^{-1}(\kappa)$ is $\mathfrak{N}c$ in $P \ni l_1 \notin h^{-1}(\kappa)$. Now P is a $\mathfrak{N}reg$ space, $\exists$ disjoint $\mathfrak{N}o$ sets μ & $\nu \ni l_1 \in \mu$ & $h^{-1}(\kappa) \subset \nu$. That is $l_2 = h(l_1) \in h(\mu)$ & $\kappa \subset h(\nu)$. Since h is $\mathfrak{N}Zof$, $h(\mu)$& $h(\nu)$ are $\mathfrak{N}Zo$ sets in Q & h is bijective, $h(\mu) \cap h(\nu) = h(\mu \cap \nu) = h(\emptyset) = \emptyset$. Therefore Q is a $\mathfrak{N}Zreg$ space.

Theorem 11 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}ZCts$, $\mathfrak{N}c$ injection and Q is a $\mathfrak{N}reg$ space then P is $\mathfrak{N}Zreg$.*

Proof Let κ be a $\mathfrak{N}c$ set in P & $\eta \notin \kappa$. Since h is $\mathfrak{N}c$ injection, $h(\kappa)$ is $\mathfrak{N}c$ in $Q \ni h(\eta) \notin h(\kappa)$. Now Q is a $\mathfrak{N}reg$ space, $\exists$ disjoint $\mathfrak{N}o$ sets ρ & $\eta \ni h(\eta) \in \rho$ & $h(\kappa) \subset \eta$. This implies $\eta \in h^{-1}(\rho)$ & $\kappa \subset h^{-1}(\eta)$. Since h is $\mathfrak{N}ZCts$ function, $h^{-1}(\rho)$& $h^{-1}(\eta)$ are $\mathfrak{N}Zo$ sets in P. Further, $h^{-1}(\rho) \cap h^{-1}(\eta) = \emptyset$. Hence P is a $\mathfrak{N}Zreg$ space.

Theorem 12 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}Zirr$, $\mathfrak{N}c$ injection and Q is a $\mathfrak{N}Zreg$ space then P is also a $\mathfrak{N}Zreg$ space.*

Proof Let κ be a $\mathfrak{N}c$ set in P & $\eta \notin \kappa$. Since h is $\mathfrak{N}c$ injection, $h(\kappa)$ is $\mathfrak{N}c$ in $Q \ni h(\eta) \notin h(\kappa)$. Now Q is a $\mathfrak{N}Zreg$ space, $\exists$ disjoint $\mathfrak{N}Zo$ sets ρ & $\eta \ni h(\eta) \in \rho$ & $h(\kappa) \subset \eta$. This implies $\eta \in h^{-1}(\rho)$ & $\kappa \subset h^{-1}(\eta)$. Since h is $\mathfrak{N}Zirr$, $h^{-1}(\rho)$& $h^{-1}(\eta)$ are $\mathfrak{N}Zo$ sets in P. Further, $h^{-1}(\rho) \cap h^{-1}(\eta) = \emptyset$. Hence P is a $\mathfrak{N}Zreg$ space.

Theorem 13 *Let $\mathfrak{N}ZC(P, A)$ is closed under arbitrary intersection and in any $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$, then the statement*

(i) *P is $\mathfrak{N}Zreg$.*
(ii) *For every point $\eta \in P$ & $\mathfrak{N}o$ set $\mu \subseteq \eta$ $\exists$ $\mathfrak{N}Zo$ set $\nu \ni \eta \in \nu \subset \mathfrak{N}Zcl(\nu) \subset \mu$.*
(iii) *For every $\mathfrak{N}c$ set κ, $\kappa = \cap\{\mathfrak{N}Zcl(\mu) : \kappa \subset \mu$ & $\mu \in \mathfrak{N}ZO(P, A)\}$.*
(iv) *For every set Ψ & $\mathfrak{N}o$ set $\Phi \ni \Psi \cap \Phi \neq \emptyset$, $\exists$ $\mathfrak{N}Zo$ set $O \ni \Psi \cap O \neq \emptyset$ & $\mathfrak{N}Zcl(O) \subset \Phi$.*
(v) *For every non empty set Ψ & $\mathfrak{N}c$ set $\Phi \ni \Psi \cap \Phi \neq \emptyset$, $\exists$ disjoint $\mathfrak{N}Zo$ sets L_1 & $L_2 \ni \Psi \cap L_1 \neq \emptyset$ & $\Phi \subset L_2$*

are equivalent.

Proof (i) $\Rightarrow$ (ii): Let μ be a $\mathfrak{N}o$ set $\subseteq \eta$. Then $P - \mu$ is $\mathfrak{N}c$ set $\nsubseteq \eta$. Since P is $\mathfrak{N}Zreg$, $\exists$ $\mathfrak{N}Zo$ sets S & $\nu \ni \eta \in \nu$, $P - \mu \subset S$ & $\nu \cap S = \emptyset$. This implies $\nu \subset P - S$. Therefore $\mathfrak{N}Zcl(\nu) \subset \mathfrak{N}Zcl(P - S) = P - S$, because $P - S$ is $\mathfrak{N}Zc$. Hence $\eta \in \nu \subset \mathfrak{N}Zcl(\nu) \subset P - S \subset \mu$. That is $\eta \in \mathfrak{N}Zcl(\nu) \subset \mu$.

(ii) $\Rightarrow$ (iii): Let κ be a $\mathfrak{N}c$ set & $\eta \notin \kappa$. Then $P - \kappa$ is a $\mathfrak{N}o$ set $\subseteq \eta$. By (ii) there is a $\mathfrak{N}Zo$ set $\nu \ni \eta \in \nu \subset \mathfrak{N}Zcl(\nu) \subset P - \kappa$. And so, $\kappa \subset P - \mathfrak{N}Zcl(\nu) \subset P - \nu$. Consequently, $P - \nu$ is $\mathfrak{N}Zc$ set not containing η. Put $\mu = P - \mathfrak{N}Zcl(\nu)$. This implies $\kappa \subset \mu$ & μ is $\mathfrak{N}Zo$ set of P & $\eta \notin \mathfrak{N}Zcl(\mu)$, implies $\cap\{\mathfrak{N}Zcl(\mu) : \kappa \subset \mu$ & $\mu \in \mathfrak{N}ZO(P, A)\} \subset \kappa$. But κ is $\mathfrak{N}c$ & every $\mathfrak{N}c$ set is $\mathfrak{N}Zc$. Therefore $\kappa \subset \cap\{\mathfrak{N}Zcl(\mu) : \kappa \subset \mu$ & $\mu \in \mathfrak{N}ZO(P, A)\}$ is always true. Thus $\kappa = \cap\{\mathfrak{N}Zcl(\mu) : \kappa \subset \mu \& \mu \in \mathfrak{N}ZO(P, A)\}$.

(iii) $\Rightarrow$ (iv): Let $\Psi \cap \Phi \neq \emptyset$ & Φ is $\mathfrak{N}o$. Let $\eta \in \Psi \cap \Phi$. Then $P - \Phi$ is a $\mathfrak{N}c$ set $\nsubseteq \eta$. By (iii) $\exists$ $\mathfrak{N}Zo$ set μ of $P \ni P - \Phi \subset \mu$ & $\eta \notin \mathfrak{N}Zcl(\mu)$. Put $O = P - \mathfrak{N}Zcl(\mu)$, then O is $\mathfrak{N}Zo$ set of P, $\eta \in \Psi \cap O$ & $\mathfrak{N}Zcl(O) \subset \mathfrak{N}Zcl(P - \mu) = P - \mu \subset \Phi$. Hence $\mathfrak{N}Zcl(O) \subset \Phi$.

(iv) $\Rightarrow$ (v): If $\Psi \cap \Phi \neq \emptyset$, where Ψ is non empty and Φ is $\mathfrak{N}c$, then $\Psi \cap (P - \Phi) \neq \emptyset$ & $P - \Phi$ is $\mathfrak{N}o$. By (iv), $\exists$ $\mathfrak{N}Zo$ set $L_1 \ni \Psi \cap L_1 \neq \emptyset$ & $L_1 \subset \mathfrak{N}Zcl(L_1) \subset$

$P - \Phi$. Put $P - \mathfrak{N}Zcl(L_1) = L_2$, then L_2 is $\mathfrak{N}Zo$ set of P, $P - \Phi \subset L_1$ & $\mathfrak{N}Zcl(P - L_1) \subset L_2$, $P - L_1 \subset L_2$. Hence $\Phi \subset L_2$.

(v) $\Rightarrow$ (i): Let κ be a $\mathfrak{N}c$ set $\ni \eta \notin \kappa$, then $\{\eta\} \cap \kappa = \emptyset$. By (v), $\exists$ disjoint $\mathfrak{N}o$ sets L_1 & $L_2 \ni \{\eta\} \cap L_1 = \emptyset$ & $\kappa \subset L_2$, hence P is $\mathfrak{N}Zreg$.

5 Nano Z Normal Spaces

Definition 9 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z normal (briefly, $\mathfrak{N}Znor$) space, if for any two disjoint $\mathfrak{N}c$ sets ι & κ, there exists disjoint $\mathfrak{N}Zo$ sets μ & ν with $\iota \subset \mu$ & $\kappa \subset \nu$.

Clearly, every $\mathfrak{N}nor$ space is $\mathfrak{N}Znor$ space.

Theorem 14 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}Cts$ bijective, $\mathfrak{N}Zo$ function from a $\mathfrak{N}nor$ space P to Q then Q is $\mathfrak{N}Znor$.*

Proof Let ι & κ be disjoint $\mathfrak{N}c$ sets in Q. Since h is $\mathfrak{N}Cts$ bijective, $h^{-1}(\iota)$ & $h^{-1}(\kappa)$ are disjoint $\mathfrak{N}c$ sets in P. Now P is a $\mathfrak{N}nor$ space, $\exists$ disjoint $\mathfrak{N}o$ sets μ & $\nu \ni h^{-1}(\iota) \subset \mu$ & $h^{-1}(\kappa) \subset \nu$. That is $\iota \subset h(\mu)$ & $\kappa \subset h(\nu)$. Since h is $\mathfrak{N}Zof$, $h(\mu)$& $h(\nu)$ are $\mathfrak{N}Zo$ sets in Q & h is injective, $h(\mu) \cap h(\nu) = h(\mu \cap \nu) = h(\emptyset) = \emptyset$. Therefore Q is a $\mathfrak{N}Znor$ space.

Theorem 15 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}ZCts$, $\mathfrak{N}c$ injection & Q is a $\mathfrak{N}nor$ space then P is $\mathfrak{N}Znor$.*

Proof Let ι & κ be disjoint $\mathfrak{N}c$ sets in Q. Since h is $\mathfrak{N}c$ injection, $h(\iota)$ & $h(\kappa)$ are disjoint $\mathfrak{N}c$ sets in Q. Now Q is a $\mathfrak{N}nor$ space, $\exists$ disjoint $\mathfrak{N}o$ sets ρ & $\eta \ni h(\iota) \subset \rho$ & $h(\kappa) \subset \eta$. That is $\iota \subset h^{-1}(\rho)$ & $\kappa \subset h^{-1}(\eta)$. Since h is $\mathfrak{N}ZCts$ function, $h^{-1}(\rho)$& $h^{-1}(\eta)$ are $\mathfrak{N}Zo$ sets in P. Further $h^{-1}(\rho) \cap h^{-1}(\eta) = \emptyset$. Therefore P is a $\mathfrak{N}Znor$ space.

Theorem 16 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}Zirr$, $\mathfrak{N}c$ injection and Q is a $\mathfrak{N}Znor$ then P is $\mathfrak{N}Znor$.*

Proof Let ι & κ be disjoint $\mathfrak{N}c$ sets in Q. Since h is $\mathfrak{N}c$ injection, $h(\iota)$ & $h(\kappa)$ are disjoint $\mathfrak{N}c$ sets in Q. Now Q is a $\mathfrak{N}nor$ space, $\exists$ disjoint $\mathfrak{N}Zo$ sets ρ & $\eta \ni h(\iota) \subset \rho$ & $h(\kappa) \subset \eta$. That is $\iota \subset h^{-1}(\rho)$ & $\kappa \subset h^{-1}(\eta)$. Since h is $\mathfrak{N}Zirr$, $h^{-1}(\rho)$& $h^{-1}(\eta)$ are $\mathfrak{N}Zo$ sets in P. Further $h^{-1}(\rho) \cap h^{-1}(\eta) = \emptyset$. Therefore P is a $\mathfrak{N}Znor$ space.

Theorem 17 *The statements are identical for a $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$.*

(i) *P is $\mathfrak{N}Znor$.*
(ii) *For each $\mathfrak{N}c$ set Φ and for each $\mathfrak{N}o$ set $\mu \subseteq \Phi$, $\exists$ a $\mathfrak{N}Zo$ set $S \subseteq \Phi \ni \mathfrak{N}Zcl(S) \subset \mu$*
(iii) *For each pair of disjoint $\mathfrak{N}c$ sets Φ & Ψ $\exists$ $\mathfrak{N}Zo$ set $\mu \subseteq \Phi \ni \mathfrak{N}Zcl(\mu) \cap \Psi = \emptyset$.*

Proof (i) $\Rightarrow$ (ii): Let Φ be a $\mathfrak{N}c$ set and μ be a $\mathfrak{N}o$ set $\subseteq \Phi$. Then $\Phi \cap (P - \mu) = \emptyset$. They are disjoint $\mathfrak{N}c$ sets in P. Since P is $\mathfrak{N}Znor$, $\exists$ disjoint $\mathfrak{N}Zo$ sets S & $\nu \ni \Phi \subset S$, $P - \mu \subset \nu$ that is $P - \nu \subset \mu$. Now $S \cap \nu = \emptyset$, implies $S \subset P - \nu$. Therefore $\mathfrak{N}Zcl(S) \subset \mathfrak{N}Zcl(P - \nu) = P - \nu$, because $P - \nu$ is $\mathfrak{N}Zc$ set. Thus, $\Phi \subset S \subset \mathfrak{N}Zcl(S) = P - \nu \subset \mu$, That is $\mathfrak{N}Zcl(S) \subset \mu$.

(ii) $\Rightarrow$ (iii): Let Φ & Ψ be disjoint $\mathfrak{N}c$ sets in P, then $\Phi \subset P - \Psi$ & $P - \Psi$ is $\mathfrak{N}o$ set $\subseteq \Phi$. By (ii) there exists $\mathfrak{N}Zo$ set $\mu \ni \Phi \subset \mu$ & $\mathfrak{N}Zcl(\mu) \subset P - \Psi$, which implies $\mathfrak{N}Zcl(\mu) \cap \Psi = \emptyset$.

(iii) $\Rightarrow$ (i): Let Φ & Ψ be disjoint $\mathfrak{N}c$ sets in P. By (iii), $\exists$ $\mathfrak{N}Zo$ set $\mu \ni \Phi \subset \mu$ & $\mathfrak{N}Zcl(\mu) \cap \Psi = \emptyset$ or $\Psi \subset P - \mathfrak{N}Zcl(\mu)$. Now μ & $P - \mathfrak{N}Zcl(\mu)$ are disjoint $\mathfrak{N}Zo$ sets of $P \ni \Phi \subset \mu$ & $\Psi \subset P - \mathfrak{N}Zcl(\mu)$. Hence P is $\mathfrak{N}Znor$.

6 Nano Z Compactness

Definition 10 A collection $\{\mu_i : i \in \nu\}$ of $\mathfrak{N}Zo$ sets in $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z open cover (briefly, $\mathfrak{N}Zocov$) of a subset Ω of $(P, \mathscr{T}_R(A))$, if $\Omega \subset \bigcup\{\mu_i : i \in \nu\}$.

Definition 11 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z compact (briefly, $\mathfrak{N}Zcomp$) if every $\mathfrak{N}Zocov$ of P has a finite subcover.

Definition 12 A subset Ω of $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called $\mathfrak{N}Zcomp$ relative to P if for every collection $\{\mu_i : i \in \nu\}$ of $\mathfrak{N}Zo$ subsets of $P \ni \Omega \subset \bigcup\{\mu_i : i \in \nu\}$, $\exists$ finite subset ν_0 of $\nu \ni \Omega \subset \bigcup\{\mu_i : i \in \nu_0\}$.

Definition 13 A subset Ω of $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called $\mathfrak{N}Zcomp$ if Ω is $\mathfrak{N}Zcomp$ as a subspace of P.

Theorem 18 *Every $\mathfrak{N}Zc$ subset of $\mathfrak{N}Zcomp$ space is $\mathfrak{N}Zc$ relative to P, if $\mathfrak{N}ZO(P, A)$ is closed under arbitrary union.*

Proof Let P be $\mathfrak{N}Zcomp$ space and μ is $\mathfrak{N}Zc$ subset of P. Then $P - \mu$ is $\mathfrak{N}Zo$ in P. Let $Q = \{\mu_i : i \in \nu\}$ be $\mathfrak{N}Zocov$ of μ by $\mathfrak{N}Zo$ subsets in P. Then $Q^* = Q \cup (P - \mu)$ is a $\mathfrak{N}Zocov$ of P. Qince P is $\mathfrak{N}Zcomp$, Q^* is reducible to a finite subcover of P say $P = \mu_{i_1} \cup \mu_{i_2} \cup \cdots \cup \mu_{i_n}, \mu_{i_k} \in Q$. But μ & $P - \mu$ are disjoint. Hence $\mu \subset \mu_{i_1} \cup \mu_{i_2} \cup \cdots \cup \mu_{i_n} \in Q$. $\Longrightarrow$ any $\mathfrak{N}Zocov$ Q of μ contains a finite subcover. Hence, μ is $\mathfrak{N}Zcomp$ relative to P.

Theorem 19 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is surjective, $\mathfrak{N}ZCts$ function and P is $\mathfrak{N}Zcomp$ then Q is $\mathfrak{N}comp$.*

Proof Let $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is surjective, $\mathfrak{N}ZCts$ from $\mathfrak{N}Zcomp$ space $(P, \mathscr{T}_R(A))$ to $(Q, \mathscr{S}_{R'}(B))$. Let $\{\mu_i : i \in \nu\}$ be $\mathfrak{N}ocov$ of Q. Since h is $\mathfrak{N}ZCts$, $\{h^{-1}(\mu_i) : i \in \nu\}$ is $\mathfrak{N}Zocov$ of P. Since P is $\mathfrak{N}Zcomp$ implies $\mathfrak{N}Zocov$ $\{h^{-1}(\mu_i) : i \in \nu\}$ has a finite subcover say $\{h^{-1}(\mu_1), h^{-1}(\mu_2), \cdots, h^{-1}(\mu_n)\}$. Since h is surjective $\{\mu_1, \mu_2, \cdots, \mu_n\}$ is a $\mathfrak{N}ocov$ of Q, which is finite. Hence Q is $\mathfrak{N}comp$.

Theorem 20 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}Zirr$ and a subset Ω of P is $\mathfrak{N}Zcomp$ subset relative to P, then the image $h(\Omega)$ is $\mathfrak{N}Zcomp$ relative to Q.*

Proof Let $\{\mu_i : i \in \nu\}$ be any collection of $\mathfrak{N}Zo$ subsets of $Q \ni h(\Omega) \subset \bigcup\{\mu_i : i \in \nu\}$. Then $\Omega \subset \bigcup\{h^{-1}(\mu_i) : i \in \nu\}$ where $\{h^{-1}(\mu_i) : i \in \nu\}$ is family of $\mathfrak{N}Zo$ sets in P. Since Ω is $\mathfrak{N}Zcomp$ relative to P, the $\mathfrak{N}Zocov$ $\{h^{-1}(\mu_i) : i \in \nu\}$ of P has a finite subcover say $\{h^{-1}(\mu_i) : i \in \nu_0\}$, where ν_0 is a finite subset of $\nu \ni \Omega \subset \bigcup\{h^{-1}(\mu_i) : i \in \nu_0\}$. Therefore $h(\Omega) \subset \bigcup\{\mu_i : i \in \nu_0\}$. Hence $h(\Omega)$ is $\mathfrak{N}Zcomp$ relative to Q.

7 Nano *Z* Connectedness

Definition 14 A $\mathfrak{N}ts$ $(P, \mathscr{T}_R(A))$ is called Nano Z connected (briefly, $\mathfrak{N}Zcon$) if P is not the disjoint union of two nonempty $\mathfrak{N}Zo$ subsets.

Theorem 21 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}ZCts$, surjection and P is $\mathfrak{N}Zcon$ then Q is $\mathfrak{N}con$.*

Proof Suppose Q is not $\mathfrak{N}con$. Then $Q = \mu \cup \nu$ that is, μ & ν are disjoint nonempty $\mathfrak{N}o$ sets in Q. Since h is $\mathfrak{N}ZCts$ surjection, $P = h^{-1}(\mu) \cup h^{-1}(\nu)$ where $h^{-1}(\mu)$ & $h^{-1}(\nu)$ are disjoint nonempty $\mathfrak{N}Zo$ subsets of $P \implies P$ is not $\mathfrak{N}Zcon$ space. This is a contradiction. Hence Q is $\mathfrak{N}con$.

Theorem 22 *If $k : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is $\mathfrak{N}Zirr$, surjection and P is $\mathfrak{N}Zcon$ then Q is also $\mathfrak{N}Zcon$.*

Proof Suppose Q is not $\mathfrak{N}Zcon$. Then $Q = \rho \cup \eta$ where ρ & η are disjoint nonempty $\mathfrak{N}Zo$ sets in Q. Since k is $\mathfrak{N}Zirr$ surjection, $P = k^{-1}(\rho) \cup k^{-1}(\eta)$ where $k^{-1}(\rho)$ & $k^{-1}(\eta)$ are disjoint nonempty $\mathfrak{N}Zo$ subsets of $P \implies P$ is not $\mathfrak{N}Zcon$ space. This is a contradiction. Hence Q is $\mathfrak{N}Zcon$.

Corollary 1 *If $h : (P, \mathscr{T}_R(A)) \to (Q, \mathscr{S}_{R'}(B))$ is bijective $\mathfrak{N}Zof$ and Q is $\mathfrak{N}Zcon$ then Q is $\mathfrak{N}con$.*

Theorem 23 *For a $\mathfrak{N}ts$ $(U, \mathscr{T}_R(P)$, the statements are equivalent.*

(i) *P is $\mathfrak{N}Zcon$.*
(ii) *The only subsets of P which are both $\mathfrak{N}Zo$ & $\mathfrak{N}Zc$ are the empty set $\emptyset$ & P.*
(iii) *Each $\mathfrak{N}ZCts$ function of P into a discrete space Q with at least two points is a constant function.*

Proof (i) $\Rightarrow$ (ii): Suppose (i) holds and κ is a proper subset of P which is both $\mathfrak{N}Zo$ & $\mathfrak{N}Zc$. Then $P - \kappa$ is also both $\mathfrak{N}Zo$ & $\mathfrak{N}Zc$. Therefore $P = \kappa \cup (P - \kappa)$ is disjoint union of two nonempty $\mathfrak{N}Zo$ sets. This contradicts to P is $\mathfrak{N}Zcon$. Therefore $\kappa = \emptyset$ or P.

(ii) $\Rightarrow$ (i): Suppose (ii) holds. If possible P is not $\mathfrak{N}Zcon$, then $P = \mu \cup \nu$ where μ & ν are disjoint non empty $\mathfrak{N}Zo$ sets in P. Since $\mu = P - \nu$, μ is $\mathfrak{N}Zc$ set. But by assumption, $\mu = \emptyset$ or P which is contradiction. Hence (i) hold.

(ii) $\Rightarrow$ (iii): Let h be $\mathfrak{N}ZCts$ function, where Q is discrete space with at least two points. Then $h^{-1}(\{\eta\})$ is both $\mathfrak{N}Zo$ & $\mathfrak{N}Zc$ for each $\eta \in Q$ & $P = \bigcup\{h^{-1}(\{\eta\}) : \eta \in Q\}$. By assumption, $h^{-1}(\{\eta\}) = P$ or $\emptyset$. If $h^{-1}(\{\eta\}) = \emptyset$ for all $\eta \in Q$, then h will not be function. Also there cannot exist more than one point $\eta \in Q \ni h^{-1}(\{\eta\}) = P$ & $h^{-1}(\eta_1) = \emptyset$ where $\eta \neq \eta_1 \in Q$. That is, h is constant function.

(iii) $\Rightarrow$ (ii): Let κ be both $\mathfrak{N}Zo$ & $\mathfrak{N}Zc$ in P. Suppose $\kappa \neq \emptyset$ & h be a $\mathfrak{N}ZCts$ function defined by $h(\kappa) = \{\eta\}$ & $h(P - \kappa) = \{\eta_1\}$ for some distinct points η & η_1 in Q. We know, h is constant function. Hence $\kappa = P$.

8 Conclusion

In this paper, many interesting notions on nano separation axioms via $\mathfrak{N}Zo$ sets and their respective $\mathfrak{N}Z$ compactness and $\mathfrak{N}Z$ connectedness are studied. This can be extended to locally nano Z compactness and locally nano Z connectedness in a nano topological spaces in future. Also, the notions may be extended to mappings such as open, closed, contra open, contra closed, continuous, strongly continuous, perfectly continuous and irresoluteness of nano Z open sets in a nano topological spaces.

References

1. Arul Selvaraj, X., Balakrishna, U.: Z-open Sets in Nano Topological Spaces. (Submitted).
2. Arul Selvaraj, X., Balakrishna, U.: Nano Z-Continuous Mappings and Homeomorphisms in Nano Topological Spaces. (Submitted).
3. EL-Magharabi, A.I., Mubarki, A.M.: Z-open sets and Z-continuity in topological spaces. International Journal of Mathematical Archive. 2 (10), 1819–1827 (2011).
4. Khalaf, M.M., Elmoasry, A.: Nano generalized closed sets and slightly Nano Separation Axioms. 11 (1), 91–98 (2015).
5. Lashin, E.F., Medhat, T.: Topological reduction of information systems. Chaos, Solitons and Fractals. 25, 277–286 (2015).
6. Lellis Thivagar, M., Richard, C.: On nano forms of weakly open sets. International Journal of Mathematics and Statistics Invention. 1, 31–37 (2013).
7. Pankajam, V., Kavitha, K.: δ open sets and δ nano continuity in δ nano topological space. International Journal of Innovative Science and Research Technology. 2 (12), 110–118 (2017)
8. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences. 11, 341–356 (1982).
9. Shalini, S.B.: A study on nano generalized β closed sets and its maps in nano topological spaces. Ph.D Thesis. Nirmala College for Women (A). Coimbatore, (2018).

On Wovenness of K-Fusion Frames

Animesh Bhandari and Saikat Mukherjee

Abstract In frame theory literature, there are several generalizations of frame, K-fusion frame presents a flavour of one such generalization, basically it is an intertwined replica of K-frame and fusion frame. K-fusion frames come naturally, having significant applications, when one needs to reconstruct functions (signals) from a large data in the range of a bounded linear operator. Motivated by the concept of weaving frames, in this paper we study wovenness of K-fusion frames. This article presents characterizations of weaving K-fusion frames. Paley-Wiener type perturbations and conditions on erasure of frame components are discussed to examine wovenness.

Keywords Frame · K-fusion frame · Weaving

1 Introduction

The notion of Hilbert frames was first introduced by Duffin and Schaeffer [8] in 1952. After several decades, in 1986, frame theory was popularized by the groundbreaking work by Daubechies, Grossman and Meyer [6] by showing its practical significance in distributed signal processing. Since then frame theory has been widely applicable by mathematicians and engineers in various fields.

Frame theory literature became familiarized through several generalizations, one such generalization is K-fusion frame, K-fusion frame was first studied by Liu et al. [13]. After that Neyshaburi et al. [14] and Bhandari et al. [2] produced several results on K-fusion frame.

In a sensor networking system, a frame provides a stable reconstruction of signals. If there are two frames, having similar characteristics, then absence of one or more

A. Bhandari
VIT Bhopal University, Kotrikalan 466114, Madhya Pradesh, India

S. Mukherjee (✉)
National Institute of Technology Meghalaya, Shillong 793003, India
e-mail: saikat.mukherjee@nitm.ac.in

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_8

frame elements from one of the frames still may lead to a perfect reconstruction on account of substitution by the corresponding frame elements by the other frame. In this context, basically one can think of the intertwinedness between two sets of sensors, or in general between two frames, which leads to the idea of weaving frames. Weaving frames or woven frames were first introduced by Bemrose et al. [1]. Later the concept of wovenness was studied by Deepshikha et al. [7], Bhandari et al. [3]. Furthermore, Garg et al. [10] studied weaving K-fusion frames.

This article focuses on study and characterization of wovenness of K-fusion frames. The outline of this article is organized as follows. Section 2 is devoted to the definitions and results related to K-fusion frames and woven frames. Main results of this article on weaving K-fusion frames are discussed in Sect. 3.

Throughout this paper $\mathcal{H}$ is a separable Hilbert space, $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is the space of all bounded linear operators from $\mathcal{H}_1$ into $\mathcal{H}_2$, $\mathcal{L}(\mathcal{H})$ represents $\mathcal{L}(\mathcal{H}, \mathcal{H})$, $P_{\mathcal{A}}$ is the orthogonal projection on $\mathcal{A}$, $\mathcal{I}$ is countable index set, $R(T)$ is denoted as range of a bounded linear operator T and $T^{\dagger}$ is the Moore-Penrose pseudo inverse of T.

2 Preliminaries

In this section we discuss some important results that aid us to develop the rest of this article.

Definition 1 Let $K \in \mathcal{L}(\mathcal{H})$ for which a weighted collection $\mathcal{W}_w = \{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ of closed subspaces in $\mathcal{H}$ is said to be a *K-fusion frame* for $\mathcal{H}$ if there exist constants $0 < A, B < \infty$ so that for every $f \in \mathcal{H}$ we have,

$$A\|K^* f\|^2 \leq \sum_{i\in\mathcal{I}} w_i^2 \|P_{\mathcal{W}_i} f\|^2 \leq B\|f\|^2. \tag{1}$$

A and B are called lower and upper K-fusion frame bounds, respectively. If only the right inequality is satisfied, $\mathcal{W}_w$ is said to be a K-fusion Bessel sequence or simply a fusion Bessel sequence with Bessel bound B.

Definition 2 ([10]) Let $K \in \mathcal{L}(\mathcal{H})$ and consider two K-fusion frames $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$, $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$. Then they are said to be woven if there are universal constants A, B so that for every $\sigma \subset \mathcal{I}$ and for every $f \in \mathcal{H}$ we have,

$$A\|K^* f\|^2 \leq \sum_{i\in\sigma} w_i^2 \|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2 \|P_{\mathcal{V}_i} f\|^2 \leq B\|f\|^2. \tag{2}$$

Proposition 1 *([10]) Let $K \in \mathcal{L}(\mathcal{H})$ for which $\mathcal{W}_w = \{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\mathcal{V}_v = \{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ be two K-fusion Bessel sequences in $\mathcal{H}$ with Bessel bounds B_1, B_2 respectively. Then for every $\sigma \subset \mathcal{I}$, the associated weaving between them also forms a K-fusion Bessel sequence in $\mathcal{H}$ with the universal Bessel bound $B_1 + B_2$.*

The following Lemma provides a discussion regarding Moore-Penrose pseudo-inverse. For detailed discussion regarding the same we refer [5, 11].

Lemma 1 *Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces and $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ be a closed range operator, then the followings hold:*

1. $TT^\dagger = P_{R(T)}$, $T^\dagger T = P_{R(T^*)}$
2. $\frac{\|f\|}{\|T^\dagger\|} \leq \|T^* f\|$ *for all* $f \in R(T)$.
3. $TT^\dagger T = T$, $T^\dagger T T^\dagger = T^\dagger$, $(TT^\dagger)^* = TT^\dagger$, $(T^\dagger T)^* = T^\dagger T$.

Lemma 2 *([9, 12]) Suppose $\mathcal{H}$ and $\mathcal{K}$ are two Hilbert spaces and $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. Consider $\mathcal{W}$ be a closed subspace of $\mathcal{H}$ and $\mathcal{V}$ be a closed subspace of $\mathcal{K}$. Then the following results are satisfied:*

1. $P_{\mathcal{W}} T^* P_{\overline{T\mathcal{W}}} = P_{\mathcal{W}} T^*$.
2. $P_{\mathcal{W}} T^* P_{\mathcal{V}} = P_{\mathcal{W}} T^*$ *if and only if* $T\mathcal{W} \subset \mathcal{V}$.

Applying the foregoing Lemma we fabricate an analogous result.

Lemma 3 *Let $\mathcal{H}_1, \mathcal{H}_2$ be two Hilbert spaces and $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ be one-one, closed range operator. Suppose $\mathcal{W}$ is a closed subspace of $\mathcal{H}_1$ and $T(\mathcal{W})$ is a closed subspace of $\mathcal{H}_2$. Then the following holds:*

$$P_{T(\mathcal{W})} T^{\dagger *} P_{T^\dagger T(\mathcal{W})} = P_{T(\mathcal{W})} T^{\dagger *} P_{\mathcal{W}} = P_{T(\mathcal{W})} T^{\dagger *}.$$

3 Main Results

We begin this section by providing two intertwining results on K-fusion frames between two separable Hilbert spaces.

Lemma 4 *Let $K \in \mathcal{L}(\mathcal{H}_1)$ for which $\mathcal{W}_w = \{(\mathcal{W}_i, w_i)\}_{i \in \mathcal{I}}$ be a K-fusion frame for $\mathcal{H}_1$. Suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is a closed range operator with $T^\dagger \overline{T(\mathcal{W}_i)} \subset \mathcal{W}_i$, for all $i \in \mathcal{I}$ and $\sum_{i \in \mathcal{I}} w_i^2 < \infty$. Then $\{(T\mathcal{W}_i, w_i)\}_{i \in \mathcal{I}}$ forms a TKT^*-fusion frame for $\mathcal{H}_2$.*

Proof First we prove for all $i \in \mathcal{I}$, $T(\mathcal{W}_i)$ is a closed subspace in $\mathcal{H}_2$. Since $T^\dagger \overline{T(\mathcal{W}_i)} \subset \mathcal{W}_i$, then $TT^\dagger \overline{T(\mathcal{W}_i)} \subset T(\mathcal{W}_i)$. But applying Lemma 2.5.2 of [5] we have $T^\dagger\big|_{R(T)} = T^*(TT^*)^{-1}$ and hence $\overline{T(\mathcal{W}_i)} \subset T(\mathcal{W}_i)$. Therefore, for every $i \in \mathcal{I}$, $T(\mathcal{W}_i)$ is a closed subspace in $\mathcal{H}_2$. Since $\{(\mathcal{W}_i, w_i)\}_{i \in \mathcal{I}}$ is a K-fusion frame for $\mathcal{H}_1$, there exist $A, B > 0$ so that for every $f \in \mathcal{H}_1$ we have,

$$A\|K^* f\|^2 \leq \sum_{i \in \mathcal{I}} w_i^2 \|P_{\mathcal{W}_i} f\|^2 \leq B\|f\|^2. \tag{3}$$

Again applying Lemma 2 and using Eq. (3), for every $f \in \mathcal{H}_2$ we obtain,

$$\begin{aligned}\frac{A}{\|T\|^2}\|(TKT^*)^*f\|^2 \le A\|K^*(T^*f)\|^2 &\le \sum_{i\in\mathcal{I}} w_i^2\|P_{\mathcal{W}_i}T^*f\|^2 \\ &= \sum_{i\in\mathcal{I}} w_i^2\|P_{\mathcal{W}_i}T^*P_{\overline{T\mathcal{W}_i}}f\|^2 \\ &= \sum_{i\in\mathcal{I}} w_i^2\|P_{\mathcal{W}_i}T^*P_{T\mathcal{W}_i}f\|^2 \\ &\le \|T\|^2\sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}f\|^2\end{aligned}$$

and hence $\sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}f\|^2 \ge \frac{A}{\|T\|^4}\|(TKT^*)^*f\|^2$. Furthermore, since $\sum_{i\in\mathcal{I}} w_i^2 < \infty$, for every $f \in \mathcal{H}_2$ we get, $\sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}f\|^2 \le \left(\sum_{i\in\mathcal{I}} w_i^2\right)\|f\|^2$.

Lemma 5 *Let $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ be a weighted collection of closed subspaces in $\mathcal{H}_1$ and $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ be one-one, closed range operator so that for some $K \in \mathcal{L}(\mathcal{H}_2)$, $\{(T\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ be a K-fusion frame for $R(T)$. Then $\left\{\left(\mathcal{W}_i, \frac{w_i}{\|T\|}\right)\right\}_{i\in\mathcal{I}}$ forms a $T^\dagger KT$-fusion frame for $\mathcal{H}_1$.*

Proof Since $\{(T\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ is a K-fusion frame for $R(T)$, there exist $A, B > 0$ so that for every $h_2^{(1)} \in R(T)$ we have,

$$A\|K^*h_2^{(1)}\|^2 \le \sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}h_2^{(1)}\|^2 \le B\|h_2^{(1)}\|^2. \tag{4}$$

Now since T is one-one and $R(T)$ is closed, for every $h_1 \in \mathcal{H}_1$ there exists $h_2 \in \mathcal{H}_2$ so that $h_1 = T^*h_2$ and for every $h_2 \in \mathcal{H}_2$ we have $h_2 = h_2^{(1)} + h_2^{(2)}$, where $h_2^{(1)} \in R(T)$ and $h_2^{(2)} \in R(T)^\perp$.

Therefore, $h_2^{(1)} = T^{*\dagger}(h_1 - T^*h_2^{(2)}) = T^{*\dagger}h_1$. Hence applying Lemma 3 we get,

$$\begin{aligned}\sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}h_2^{(1)}\|^2 = \sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}T^{\dagger*}h_1\|^2 &= \sum_{i\in\mathcal{I}} w_i^2\|P_{T\mathcal{W}_i}T^{\dagger*}P_{\mathcal{W}_i}h_1\|^2 \\ &\le \|T^\dagger\|^2\sum_{i\in\mathcal{I}} w_i^2\|P_{\mathcal{W}_i}h_1\|^2.\end{aligned}$$

Consequently, using Eq. (4) for every $h_1 \in \mathcal{H}_1$ we obtain,

$$\begin{aligned}\sum_{i\in\mathcal{I}}\left(\frac{w_i}{\|T\|}\right)^2\|P_{\mathcal{W}_i}h_1\|^2 &\ge \frac{A}{\|T\|^2\|T^\dagger\|^2}\|(T^\dagger K)^*h_1\|^2 \\ &\ge \frac{A}{\|T\|^4\|T^\dagger\|^2}\|(T^\dagger KT)^*h_1\|^2.\end{aligned}$$

Furthermore, applying Lemma 2 and using Eq. (4) for every $h_1 \in \mathcal{H}_1$ we get,

$$\begin{aligned}\sum_{i\in\mathcal{I}}\left(\frac{w_i}{\|T\|}\right)^2\|P_{\mathcal{W}_i}h_1\|^2 &= \sum_{i\in\mathcal{I}}\left(\frac{w_i}{\|T\|}\right)^2\|P_{\mathcal{W}_i}T^*h_2\|^2\\ &= \sum_{i\in\mathcal{I}}\left(\frac{w_i}{\|T\|}\right)^2\|P_{\mathcal{W}_i}T^*P_{T\mathcal{W}_i}h_2\|^2\\ &\le \sum_{i\in\mathcal{I}}w_i^2\|P_{T\mathcal{W}_i}h_2\|^2\\ &= \sum_{i\in\mathcal{I}}w_i^2\|P_{T\mathcal{W}_i}(h_2^{(1)}+h_2^{(2)})\|^2\\ &= \sum_{i\in\mathcal{I}}w_i^2\|P_{T\mathcal{W}_i}h_2^{(1)}\|^2\\ &\le B\|h_2^{(1)}\|^2\\ &\le B\|T^\dagger\|^2\|h_1\|^2.\end{aligned}$$

Hence our assertion is tenable.

As a consequence of Lemmas 4 and 5, the following two propositions show that K-wovenness is preserved under bounded linear operators.

Proposition 2 *Let $K\in\mathcal{L}(\mathcal{H}_1)$ for which $\mathcal{W}_w=\{(\mathcal{W}_i,w_i)\}_{i\in\mathcal{I}}$ and $\mathcal{V}_v=\{(\mathcal{V}_i,v_i)\}_{i\in\mathcal{I}}$ be K-fusion frames for $\mathcal{H}_1$. Further let us consider a closed range operator $T\in\mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)$ with $T^\dagger\overline{T(\mathcal{W}_i)}\subset\mathcal{W}_i$ and $T^\dagger\overline{T(\mathcal{V}_i)}\subset\mathcal{V}_i$, for all $i\in\mathcal{I}$ for all $i\in\mathcal{I}$. Suppose $\mathcal{W}_w$ and $\mathcal{V}_v$ are weaving K-fusion frames for $\mathcal{H}_1$, then $\{(T\mathcal{W}_i,w_i)\}_{i\in\mathcal{I}}$ and $\{(T\mathcal{V}_i,v_i)\}_{i\in\mathcal{I}}$ are weaving TKT^*-fusion frames for $\mathcal{H}_2$.*

Proof Applying Lemma 4, our assertion is tenable.

Proposition 3 *Let $\{(\mathcal{W}_i,w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i,v_i)\}_{i\in\mathcal{I}}$ be two weighted collections of closed subspaces in $\mathcal{H}_1$. Suppose $T\in\mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)$ to be one-one, closed range operator so that for some $K\in\mathcal{L}(\mathcal{H}_2)$, $\{(T\mathcal{W}_i,w_i)\}_{i\in\mathcal{I}}$ and $\{(T\mathcal{V}_i,v_i)\}_{i\in\mathcal{I}}$ are weaving K-fusion frames for $R(T)$ with the universal bounds A,B. Then $\left\{\left(\mathcal{W}_i,\frac{w_i}{\|T\|}\right)\right\}_{i\in\mathcal{I}}$ and $\left\{\left(\mathcal{V}_i,\frac{v_i}{\|T\|}\right)\right\}_{i\in\mathcal{I}}$ are weaving $T^\dagger KT$-fusion frames for $\mathcal{H}_1$ with the universal bounds $\frac{A}{\|T\|^4\|T\|^2}$, $B\|T^\dagger\|^2$.*

Proof The proof will be followed from Lemmas 1 and 5.

In the following result we discuss images of weaving fusion frames under bounded, linear operator preserve their wovenness with respect to the said operator.

Proposition 4 *Let $\{(\mathcal{W}_i,w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i,v_i)\}_{i\in\mathcal{I}}$ be weaving fusion frames for $\mathcal{H}$. Then for every $K\in\mathcal{L}(\mathcal{H})$, $\{(\overline{K\mathcal{W}_i},w_i)\}_{i\in\mathcal{I}}$ and $\{(\overline{K\mathcal{V}_i},v_i)\}_{i\in\mathcal{I}}$ are weaving K-fusion frames for $\mathcal{H}$.*

Proof Let $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ be weaving fusion frames for $\mathcal{H}$ with the universal bounds A, B. Then for every $\sigma \subset \mathcal{I}$ and $f \in \mathcal{H}$ we have,

$$A\|f\|^2 \leq \sum_{i\in\sigma} w_i^2\|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\mathcal{V}_i} f\|^2 \leq B\|f\|^2. \tag{5}$$

Therefore, using Eq. (5) and applying Lemma 2, for every $K \in \mathcal{L}(\mathcal{H})$, $\sigma \subset \mathcal{I}$ and $f \in \mathcal{H}$ we obtain,

$$\sum_{i\in\sigma} w_i^2\|P_{\overline{K\mathcal{W}_i}} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\overline{K\mathcal{V}_i}} f\|^2 \geq \frac{A}{\|K\|^2}\|K^* f\|^2.$$

The universal upper bound of the respective weaving will achieved by Proposition 1.

Next result provides a characterization of weaving fusion frames by means of weaving K-fusion frames and conversely.

Proposition 5 *Let $K \in \mathcal{L}(\mathcal{H})$ and consider two weighted collections $\mathcal{W}_w$, $\mathcal{V}_v$ of closed subspaces of $\mathcal{H}$. Then*

1. *$\mathcal{W}_w$ and $\mathcal{V}_v$ are weaving K-fusion frames for $\mathcal{H}$ whenever they are weaving fusion frames for $\mathcal{H}$.*
2. *If $R(K)$ is closed, then $\mathcal{W}_w$ and $\mathcal{V}_v$ form weaving fusion frames for $R(K)$ whenever they are weaving K-fusion frames for $R(K)$.*

Proof 1. Let $\mathcal{W}_w$ and $\mathcal{V}_v$ be weaving fusion frames for $\mathcal{H}$ with the universal bounds A, B. Then for every $\sigma \subset \mathcal{I}$ and $f \in \mathcal{H}$ we get,

$$\frac{A}{\|K\|^2}\|K^* f\|^2 \leq A\|f\|^2 \leq \sum_{i\in\sigma} w_i^2\|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\mathcal{V}_i} f\|^2 \leq B\|f\|^2.$$

2. Suppose $\mathcal{W}_w$ and $\mathcal{V}_v$ are weaving K-fusion frames for $R(K)$ with the universal bounds C, D. Then for every $\sigma \subset \mathcal{I}$ and $f \in \mathcal{H}$ we have,

$$C\|K^* f\|^2 \leq \sum_{i\in\sigma} w_i^2\|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\mathcal{V}_i} f\|^2 \leq D\|f\|^2. \tag{6}$$

Again using closed range property for every $f \in R(K)$ we have, $\|K^* f\|^2 \geq \frac{1}{\|K^\dagger\|^2}\|f\|^2$. Therefore, using Eq. (6) we obtain,

$$\frac{C}{\|K^\dagger\|^2}\|f\|^2 \leq \sum_{i\in\sigma} w_i^2\|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\mathcal{V}_i} f\|^2 \leq D\|f\|^2.$$

In the following results we discuss stability of wovenness of K-fusion frames under perturbation and erasures. Analogous erasure result for frame can be observed in [4].

Theorem 1 *Let $T, K \in \mathcal{L}(\mathcal{H})$ with K has closed range and suppose for every $f \in \mathcal{H}$ we have, $\|(T^* - K^*)f\| \leq \alpha_1\|T^* f\| + \alpha_2\|K^* f\| + \alpha_3\|f\|$, for some $\alpha_1, \alpha_2, \alpha_3 \in (0, 1)$. Then $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ are weaving T-fusion frames if they are weaving K-fusion frames for $R(K)$.*

Proof Let $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ be weaving K-fusion frames with the universal bounds A, B. Then for every $\sigma \subset \mathcal{I}$ and every $f \in R(K)$ we have,

$$A\|K^* f\|^2 \leq \sum_{i\in\sigma} w_i^2\|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\mathcal{V}_i} f\|^2 \leq B\|f\|^2. \tag{7}$$

Again for every $f \in \mathcal{H}$ we have, $\|K^* f\| \geq \|T^* f\| - \|(T^* - K^*)f\|$ and hence applying closed range property of K (see Lemma 1) and employing given perturbation condition for every $f \in R(K)$ we obtain,

$$(1 - \alpha_1)\|T^* f\| \leq (1 + \alpha_2 + \alpha_3\|K^\dagger\|)\|K^* f\|.$$

Therefore, using Eq. (7), for every $\sigma \subset \mathcal{I}$ we obtain,

$$A\left(\frac{1-\alpha_1}{1+\alpha_2+\alpha_3\|K^\dagger\|}\right)^2\|T^* f\|^2 \leq \sum_{i\in\sigma} w_i^2\|P_{\mathcal{W}_i} f\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{\mathcal{V}_i} f\|^2 \leq B\|f\|^2,$$

for every $f \in R(K)$.

Corollary 1 *Let $T, K \in \mathcal{L}(\mathcal{H})$ and suppose $\alpha_1, \alpha_2 \in (0, 1)$ so that for every $f \in \mathcal{H}$ we have, $\|T^* f - K^* f\| \leq \alpha_1\|T^* f\| + \alpha_2\|K^* f\|$. Then $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ are T-woven if and only if they are K-woven.*

Theorem 2 *Let $K \in \mathcal{L}(\mathcal{H}_1)$ for which $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ be weaving K-fusion frames for $\mathcal{H}_1$ with universal lower bound A and suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ with $T^\dagger\overline{T(\mathcal{W}_i)} \subset \mathcal{W}_i$ and $T^\dagger\overline{T(\mathcal{V}_i)} \subset \mathcal{V}_i$ for all $i \in \mathcal{I}$. Let us assume $\mathcal{J} \subset \mathcal{I}$ and $0 < C < \frac{A}{\|T\|^2}$ so that for every $f \in \mathcal{H}_2$*

$$\sum_{i\in\mathcal{J}} w_i^2\|P_{T\mathcal{W}_i}\|^2 \leq C\|TK^*T^* f\|^2. \tag{8}$$

Then $\{(T\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}\setminus\mathcal{J}}$ and $\{(T\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}\setminus\mathcal{J}}$ form weaving TKT^*-fusion frames for $\mathcal{H}_2$.

Proof Since $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ are weaving K-fusion frames for $\mathcal{H}_1$, then by Lemma 4 and Proposition 2, $\{(T\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(T\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ form weaving TKT^*-fusion frames for $\mathcal{H}_2$ with universal lower bound $\frac{A}{\|T\|^2}$ in $\mathcal{H}_2$. Therefore, applying Eq. (8), for every $\sigma \subset \mathcal{I} \setminus \mathcal{J}$ and for every $f \in \mathcal{H}_2$ we obtain,

$$\begin{aligned}
&\sum_{i\in\sigma} w_i^2\|P_{T\mathcal{W}_i}\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{T\mathcal{V}_i}\|^2 \\
&= \sum_{i\in\sigma\cup\mathcal{J}} w_i^2\|P_{T\mathcal{W}_i}\|^2 + \sum_{i\in\sigma^c} v_i^2\|P_{T\mathcal{V}_i}\|^2 - \sum_{i\in\mathcal{J}} w_i^2\|P_{T\mathcal{W}_i}\|^2 \\
&\geq \frac{A}{\|T\|^2}\|(TKT^*)^*f\|^2 - C\|(TKT^*)^*f\|^2 \\
&= \left(\frac{A}{\|T\|^2} - C\right)\|(TKT^*)^*f\|^2,
\end{aligned}$$

where σ^c is the complement of σ in $\mathcal{I}\setminus\mathcal{J}$.

The universal upper bound will be followed by Proposition 1.

By choosing $\mathcal{H}_1 = \mathcal{H}_2$ and $T = I$, we obtain the following result.

Corollary 2 *Let $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ be weaving K-fusion frames for $\mathcal{H}$ with the universal bounds A, B. Let us consider $\mathcal{J}\subset\mathcal{I}$ and $0 < C < A$ so that for every $f \in \mathcal{H}$,*

$$\sum_{i\in\mathcal{J}} w_i^2\|P_{\mathcal{W}_i}\|^2 \leq C\|K^*f\|^2,$$

then $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}\setminus\mathcal{J}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}\setminus\mathcal{J}}$ are weaving K-fusion frames $\mathcal{H}$ with the universal bounds $(A - C)$, B.

Using Proposition 3, we get the following result analogous to Theorem 2.

Theorem 3 *Let $\{(\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ be two weighted collections of closed subspaces in $\mathcal{H}_1$ and $K \in \mathcal{L}(\mathcal{H}_2)$. Suppose $T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is one-one, closed range operator so that $\{(T\mathcal{W}_i, w_i)\}_{i\in\mathcal{I}}$ and $\{(T\mathcal{V}_i, v_i)\}_{i\in\mathcal{I}}$ are weaving K-fusion frames for $R(T)$ with the universal lower bound A. Further suppose $\mathcal{J}\subset\mathcal{I}$ and $0 < C < \frac{A}{\|T\|^4\|T^\dagger\|^2}$ so that for every $f \in \mathcal{H}_1$*

$$\sum_{i\in\mathcal{J}} \left(\frac{w_i}{\|T\|}\right)^2 \|P_{\mathcal{W}_i}f\|^2 \leq C\|(T^\dagger KT)^*f\|^2. \tag{9}$$

Then $\left\{\left(\mathcal{W}_i, \frac{w_i}{\|T\|}\right)\right\}_{i\in\mathcal{I}\setminus\mathcal{J}}$ and $\left\{\left(\mathcal{V}_i, \frac{v_i}{\|T\|}\right)\right\}_{i\in\mathcal{I}\setminus\mathcal{J}}$ are weaving $T^\dagger KT$-fusion frames for $\mathcal{H}_1$.

4 Conclusion

Frames, especially the theory of K-fusion frames, and their wovenness became an important tool for big data analysis. In this paper, we presented theoretical aspects of K-fusion frames. We have discussed sufficient conditions under which wovenness of K-fusion frames is preserved between two Hilbert spaces. Stability of weaving K-fusion frames under Paley-Wiener type perturbations and erasure of frame components are also examined.

Acknowledgements The first author acknowledges the fiscal support of MHRD and the academic support of VIT Bhopal University, Government of India. The second author is supported by DST-SERB project MTR/2017/000797.

References

1. Bemrose, T., Casazza, P. G., Gröchenig, K., Lammers, M. C. and Lynch, R. G., Weaving frames, *Oper. Matrices.* 10 (4)(2016), 1093–1116.
2. Bhandari, A. and Mukherjee, S., Atomic subspaces for operators, *Indian J. Pure Appl. Math.* 51 (3)(2020), 1039–1152.
3. Bhandari, A., Mukherjee, S.: Characterizations of woven frames, *Int. J. Wavelets Multiresolut. Inf. Process* 18 (5)(2020), 2050033.
4. Casazza, P. G. and Lynch, R. G., Weaving properties of Hilbert space frames, *International Conference on Sampling Theory and Applications (SampTA)*, (2015).
5. Christensen, O., *Frames and Bases-An Introductory Course*, Birkhäuser, Boston, 2008.
6. Daubechies, I., Grossmann, A. and Meyer, Y., Painless nonorthogonal expansions, *J. Math. Phys.* 27(5)(1986), 1271–1283.
7. Deepshikha and Vashisht, L. K., Weaving K- frames in Hilbert spaces, *Results Math.* 73:81(2018).
8. Duffin, R. and Schaeffer, A. C., A class of nonharmonic Fourier series, *Trans. Amer. Math. Soc.* 72(2)(1952), 341–366.
9. Găvruţa, P., On the duality of fusion frame, *J. Math. Anal. Appl.* 333(2)(2007), 871–879.
10. Garg, S. and Vashisht, L. K., Weaving K- fusion frames in Hilbert spaces, *Ganita* 67(1)(2017), 41–52.
11. Kato, T., *Perturbation Theory for Linear Operators*, Springer, New York, 1980.
12. Li, X. B., Yang, S. Z. and Zhu, Y. C., Some results about operator perturbation of fusion frames in Hilbert spaces, *J. Math. Anal. Appl.* 421(2015), 1417–1427.
13. Liu, A. F. and Li, P. T., K-fusion frames and the corresponding generators for unitary systems, *Acta Math. Sin. (Engl. Ser.)* 34 (5)(2018), 843–854.
14. Neyshaburi, F. A. and Arefijamaal, A. A., Characterization and construction of K-fusion frames and their duals in Hilbert spaces, *Results Math.* 73:47 (2018).

Different Time Schemes with Differential Quadrature Method in Convection-Diffusion-Reaction Equations

Bengisen Pekmen Geridönmez

Abstract In this study, two-dimensional, unsteady convection-diffusion and convection-diffusion-reaction equations are numerically investigated with different time schemes. In the governing equations, the space derivatives are approximated by differential quadrature method (DQM) which gives highly accurate results using small number of grid points and the time derivatives are handled by different time schemes. The distribution of nodes is achieved by non-uniform Gauss-Chebyshev-Lobatto (GCL) grid points. The problems having the exact solutions are chosen to check the best error behavior. Computational cost in view of central processing unit time and the efficiency of different time schemes in terms of errors are examined. As expected, explicit time schemes need smaller time increments while implicit time schemes enable one to use larger time increments. In each chosen problems, Adams-Bashforth-Moulton and Runge-Kutta of order four exhibit the best error behavior.

Keywords Convection-diffusion-reaction · Differential quadrature method · Time schemes

1 Introduction and Problem Definition

Convection-diffusion equations, also adding reaction term, describe many physical problems such as heat transfer, chemical reaction processes, fluid dynamics etc. In order to be able to interpret the physical reality as much as possible, the numerical simulations should be performed not only as accurate as possible but also computationally efficient in view of less memory usage and short period of time of central processing unit (CPU) in computers.

There are many numerical studies in literature on two-dimensional, unsteady, convection-diffusion equations. Some of them may be mentioned as follows. Tian

B. P. Geridönmez (✉)
TED University, Department of Mathematics, 06420 Kolej-Cankaya Ankara, TR, Turkey

Uppsala University, Department of Information Technology, Division of Scientific Computing, Uppsala, Sweden

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*, Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_9

and Ge [1] presented compact alternating direction implicit method of exponential high-order. Dehghan and Mohebbi [2] proposed a fourth order finite difference method for space derivatives and boundary value method for time integration. Mittal and Tripathi [3] used collocation of modified bi-cubic B-spline functions for space derivatives and strong stability preserving Runge Kutta method (ssprk(4,3)) for time derivatives. Mittal an Jiwari [4] employed the cubic B-spline quasi-interpolation for space derivatives and ssprk(5,4) for time derivatives. Rashidinia et al. [5] derived a stable Gaussian radial basis function method. Hoz et al. [6] used Chebyshev differentiation matrices for space derivatives. Li et al. [7] analyzed convection-diffusion-reaction equation utilizing local mesh-independent Petrov-Galerkin method based on radial basis function collocation method. Wang et al. [8] presented a local knot method to solve two and three dimensional convection-diffusion-reaction equations in random domains. In both theoretical and analytical aspect, a novel study is carried out in [9] generalizing the scheme for discontinuous Petrov Galerkin to more general Banach spaces using a very general residual minimization approach. In papers [10–12], fractional differential equations are concerned in theoretical view based on Lie algebra. Convection-diffusion-reaction equation with fractional time derivative in the sense of Caputo derivative is also solved by Li et al. [13] using finite difference schemes for time derivative and local discontinuous Galerkin method for space derivatives.

In this study, differential quadrature method is applied to space derivatives while the different time schemes are implemented. DQM gives very accurate and fast results using small number of nodes in a small computational domain. The sensitivity of DQM depends the number of nodes. In the current study, the impact of various forms of time schemes is analyzed.

The two-dimensional unsteady convection-diffusion-reaction equation in a bounded rectangular domain $\Omega\ :\ [a, b] \times [c, d]$ is defined by [3, 5]

$$\frac{\partial u}{\partial t} = p_1 \frac{\partial^2 u}{\partial x^2} + p_2 \frac{\partial^2 u}{\partial y^2} - q_1 \frac{\partial u}{\partial x} - q_2 \frac{\partial u}{\partial y} + f(x, y, t, u), \tag{1}$$

where $x \in (a, b)$, $y \in (c, d)$, $t > 0$, q_1 and q_2 are constant speeds of convection, $p_1, p_2 > 0$ are constant diffusivity in x and y directions, respectively, and $f(x, y, t, u)$ is the reaction term. Initial and boundary conditions are

$$\begin{aligned}
&u(x, y, 0) = g(x, y)\\
&u(a, y, t) = \chi_1(y, t),\ u(b, y, t) = \chi_2(y, t)\\
&u(x, c, t) = \chi_3(x, t),\ u(x, d, t) = \chi_4(x, t),\ 0 \le t \le t_{\max},
\end{aligned}$$

where $t_{\max}$ is the peak time level. Equation (1) is reduced to convection-diffusion equation when $f(x, y, t, u) = 0$.

2 DQM Application and Time Schemes

Approximation of space derivatives at a grid point is done by using all values of grid points in the entire region of the problem in differential quadrature method.

In Eq. (1), Lagrange polynomials based DQM admits the space derivatives as [14]

$$\frac{\partial u}{\partial x} = \sum_{k=1}^{N} a_{ik} u_{kj}, \quad \frac{\partial u}{\partial y} = \sum_{k=1}^{M} \bar{a}_{jk} u_{ik}, \quad \frac{\partial^2 u}{\partial x^2} = \sum_{k=1}^{N} b_{ik} u_{kj}, \quad \frac{\partial^2 u}{\partial y^2} = \sum_{k=1}^{M} \bar{b}_{jk} u_{ik}, \tag{2}$$

by the weighting coefficients for the first and second order derivatives given explicitly

$$a_{ik} = \frac{Q^{(1)}(x_i)}{(x_i - x_k) Q^{(1)}(x_k)}, \; i \neq k, \qquad a_{ii} = - \sum_{k=1, k\neq i}^{N} a_{ik}, \tag{3}$$

$$b_{ik} = 2 a_{ik} \left(a_{ii} - \frac{1}{x_i - x_k} \right), \; i \neq k, \qquad b_{ii} = - \sum_{k=1, k\neq i}^{N} b_{ik}, \tag{4}$$

where $Q^{(1)}(x_i) = \prod_{k=1, k\neq i}^{N} (x_i - x_k)$, N, M are the number of GCL points on x- and y-axis and $i = 1, 2, \ldots, N$.

Let $D_x, \; D_y, \; D_{xx}, \; D_{yy}, \; D_2 = D_{xx} + D_{yy}$ be differentiation matrices formed by Equation (2). Additionally, let A be the matrix $pD_2 - q(D_x + D_y)$ considering the convection-diffusion part of Eq. (1), and thus,

$$u_t = Au + f(t, x, y, u) = H(u) \tag{5}$$

where subindex t refer to the time derivative of u.

Some efficient time schemes are chosen for approximation of time derivatives. An iterative system depending on time is to be developed with various time schemes as follows :

(a) The first order Backward Differentiation Formula (BDF1) (Backward-Euler):

$$\frac{u^{n+1} - u^n}{\Delta t} = Au^{n+1} + f(t_{n+1}, x, y, u^n) \tag{6}$$

(b) The second order Backward Differentiation Formula (BDF2) :

$$\frac{3u^{n+1} - 4u^n + u^{n-1}}{2\Delta t} = Au^{n+1} + f(t_{n+1}, x, y, u^n) \tag{7}$$

(c) The third order Backward Differentiation Formula (BDF3) (Houbolt method):

$$\frac{11u^{n+1} - 18u^n + 9u^{n-1} - 2u^{n-2}}{6\Delta t} = Au^{n+1} + f(t_{n+1}, x, y, u^n) \tag{8}$$

(d) Trapezoidal rule (Trap) ≡ Crank-Nicolson

$$
\begin{aligned}
\frac{u^{n+1}-u^{n}}{\Delta t} &= \frac{1}{2}\big(H(u^{n+1})+H(u^{n})\big)\\
\frac{u^{n+1}-u^{n}}{\Delta t} &= \frac{1}{2}(Au^{n+1}+f(t_{n+1},x,y,u^{n}))+\frac{1}{2}(Au^{n}+f(t_{n},x,y,u^{n}))
\end{aligned}
\tag{9}
$$

(e) Adams-Moulton formula of order 3 (AM3) :

$$
\begin{aligned}
\frac{u^{n+1}-u^{n}}{\Delta t} &= \frac{5}{12}H(u^{n+1})+\frac{2}{3}H(u^{n})-\frac{1}{12}H(u^{n-1})\\
\left(\frac{\mathcal{I}}{\Delta t}-\frac{5A}{12}\right)u^{n+1} &= \left(\frac{\mathcal{I}}{\Delta t}+\frac{2A}{3}\right)u^{n}-\frac{A}{12}u^{n-1}\\
&\quad+\frac{5}{12}f(t_{n+1},x,y,u^{n})+\frac{2}{3}f(t_{n},x,y,u^{n})\\
&\quad-\frac{1}{12}f(t_{n-1},x,y,u^{n-1})
\end{aligned}
\tag{10}
$$

(f) Heun's Method (2nd order, predictor (P) - corrector (C))

$$
\begin{aligned}
(P):\ u_1^{n+1}-u^{n} &= \Delta t H(u^{n})\\
(C):\ u_2^{n+1}-u^{n} &= 0.5\Delta t H(u^{n})+0.5\Delta t H(u_1^{n+1})\\
(C):\ u^{n+1}-u^{n} &= 0.5\Delta t H(u^{n})+0.5\Delta t H(u_2^{n+1})
\end{aligned}
\tag{11}
$$

(g) Adams-Bashforth-Moulton (4th order, PC)

$$
\begin{aligned}
(P):\ \bar{u}^{n+1} &= u^{n}+\Delta t\left(\frac{55}{24}H(u^{n})-\frac{59}{24}H(u^{n-1})+\frac{37}{24}H(u^{n-2})-\frac{9}{24}H(u^{n-3})\right)\\
(C):\ u^{n+1} &= u^{n}+\Delta t\left(\frac{9}{24}H(\bar{u}^{n+1})+\frac{19}{24}H(u^{n})-\frac{5}{24}H(u^{n-1})+\frac{1}{24}H(u^{n-2})\right)
\end{aligned}
\tag{12}
$$

(h) Runge Kutta of order 4 (RK4)

$$k_1 = Au^{n}+f(t_n,x,y,u^{n}) \tag{13a}$$

$$k_2 = A(u^{n}+0.5\Delta t k_1)+f(t_n+0.5\Delta t,x,y,u^{n}+0.5\Delta t k_1) \tag{13b}$$

$$k_3 = A(u^{n}+0.5\Delta t k_2)+f(t_n+0.5\Delta t,x,y,u^{n}+0.5\Delta t k_2) \tag{13c}$$

$$k_4 = A(u^{n}+\Delta t k_3)+f(t_{n+1},x,y,u^{n}+\Delta t k_3) \tag{13d}$$

$$u^{n+1} = u^{n}+\frac{\Delta t}{6}\big(k_1+2k_2+2k_3+k_4\big) \tag{13e}$$

(i) 3-stage Strong Stability Preserving Runge Kutta of order 3 (ssprk3) [15] :

$$
\begin{aligned}
u_1 &= u^n + \Delta t\big(Au^n + f(t_n, x, y, u^n)\big)\\
u_2 &= \frac{3}{4}u^n + \frac{1}{4}u_1 + \frac{1}{4}\Delta t\big(Au_1 + f(t_{n+1}, x, y, u_1)\big)\\
u^{n+1} &= \frac{1}{3}u^n + \frac{2}{3}u_2 + \frac{2}{3}\Delta t\big(Au_2 + f(t_n + 0.5\Delta t, x, y, u_2)\big) \qquad (14)
\end{aligned}
$$

(j) 5-stage Strong Stability Preserving Runge Kutta of order 4 (ssprk54) [16] :

$$
\begin{aligned}
u_1 &= u^n + 0.39175\Delta t\big(Au^n + f(t_n, x, y, u^n)\big)\\
u_2 &= 0.44437\, u^n + 0.55563\, u_1\\
&\quad + 0.36841\Delta t(Au_1 + f(t_n + 0.39175\Delta t, x, y, u_1))\\
u_3 &= 0.6201\, u^n + 0.3799\, u_2\\
&\quad + 0.25189\Delta t(Au_2 + f(t_n + 0.58608\Delta t, x, y, u_2))\\
u_4 &= 0.17808\, u^n + 0.82192\, u_3\\
&\quad + 0.54497\Delta t(Au_3 + f(t_n + 0.47454\Delta t, x, y, u_3))\\
u &= 0.00683\, u^n + 0.51723\, u_2 + 0.1276\, u_3\\
&\quad + 0.0846\Delta t(Au_3 + f(t_n + 0.47454\Delta t, x, y, u_3))\\
&\quad + 0.34834\, u_4 + 0.22601\Delta t(Au_4 + f(t_n + 0.93501\Delta t, x, y, u_4)) \qquad (15)
\end{aligned}
$$

In all time schemes, Δt is the time increment. In items (a)-(e), $\mathcal{I}$ is the identity matrix. Iteration is terminated at an assigned maximum time level.

3 Numerical Computations

Convection-diffusion, diffusion-reaction and convection-diffusion-reaction equations are solved in the following problems. Grid distribution based on GCL nodes is established.

Error between approximated results and the exact solution are computed by the following formulas :

$$
\begin{aligned}
\text{maximum absolute error } err_{inf} &= max\{|u_{ex} - u|\} = ||u - u_{ex}||_\infty\\
\text{relative error } err_{rel} &= \frac{||u - u_{ex}||_\infty}{||u_{ex}||_\infty}\\
\text{average absolute error } err_{avg} &= mean(|u_{ex} - u|).
\end{aligned}
$$

Table 1 Errors with 21×21 GCL grids at $t_{\max} = 1$ when $q = 1$, $p = 0.1$

Time scheme	Δt	$err_{\inf}$	err_{rel}	err_{avg}	CPU (s)
BDF1	0.01	4.819e-05	2.18e-05	1.191e-05	2.12
	$1e-4$	4.827e-07	2.184e-07	1.193e-07	108
BDF2	0.01	7.085e-08	3.205e-08	1.257e-08	2.45
	$1e-4$	9.618e-11	4.351e-11	4.248e-12	182
BDF3	0.01	3.277e-08	1.482e-08	3.826e-09	2.15
	$1e-4$	9.276e-11	4.197e-11	3.206e-12	182
Trap	0.01	8.046e-09	3.64e-09	1.988e-09	2.86
	$1e-4$	8.355e-11	3.78e-11	2.652e-12	187
AM3	$1e-4$	5.374e-11	2.431e-11	3.763e-12	135.91
Heun	$1e-4$	7.572e-13	3.426e-13	1.882e-13	257.13
ABM4	$1e-4$	1.421e-14	6.429e-15	3.761e-15	56.91
RK4	$1e-4$	1.332e-14	6.027e-15	3.492e-15	99.5
ssprk3	$1e-4$	6.004e-13	2.716e-13	1.479e-13	80.45
ssprk54	$1e-4$	1.002e-10	4.534e-11	2.476e-11	122

Problem 1. Convection-Diffusion Equation. Equation (1) is concerned in case of $f(x, y, t, u) = 0$ as

$$u_t = p\nabla^2 u - q(u_x + u_y), \quad \Omega \ : \ [0, 1] \times [0, 1] \tag{16}$$

with the analytical solution given as [3]

$$u(x, y, t) = ae^{bt}\left(e^{-xc_x} + e^{-yc_y}\right), \tag{17}$$

where $c_x = c_y = \dfrac{-q \pm \sqrt{q^2 + 4bp}}{2p} > 0$, and $a = 1,\ b = 0.1$ are fixed in computations. Initial and Dirichlet boundary conditions are taken from the analytical solution. For $q = 1$ and $p = 0.1$ at $t_{\max} = 1$, Table 1 illustrates the errors and CPU times in different time schemes. Keeping up $\Delta t = 10^{-4}$, the smallest error is obtained with RK4 and the fast results come with ABM4 having almost the same accuracy with RK4. Implicit methods BDF2, BDF3 and Trap provide good accuracy enabling one to use larger time increments such as $\Delta t = 0.01$.

Problem 2. Diffusion-Reaction Equation. Equation (1) is considered without convection terms but with a reaction term as [5]

$$u_t = u_{xx} + u_{yy} + f(t, x, y, u), \quad \Omega : [0, 0.5] \times [0, 0.5], \tag{18}$$

where

Table 2 Error analysis at $t_{\max} = 3$ using 11×11 GCL grids

Time scheme	Δt	$err_{\inf}$	err_{rel}	err_{avg}	CPU
BDF1	0.001	6.803e-07	1.366e-05	1.385e-07	4.88
BDF2	0.001	4.529e-07	9.096e-06	9.222e-08	5.41
	1e-4	4.528e-08	9.095e-07	9.221e-09	39.4
BDF3	0.001	4.53e-07	9.099e-06	9.225e-08	6.7
	1e-4	4.528e-08	9.095e-07	9.221e-09	36.7
Trap	0.001	2.264e-07	4.547e-06	4.61e-08	10.6
	1e-4	2.264e-08	4.547e-07	4.61e-09	61.8
AM3	1e-4	1.887e-08	3.789e-07	3.842e-09	96.4
Heun	1e-4	1.08e-12	2.169e-11	2.32e-13	80.3
ABM4	1e-4	1.089e-13	2.187e-12	2.36e-14	32.6
RK4	1e-4	1.66e-13	3.333e-12	3.464e-14	22
ssprk3	1e-4	1.631e-13	3.275e-12	3.418e-14	42.7
ssprk54	1e-4	1.652e-13	3.317e-12	3.49e-14	30.6

$$
\begin{aligned}
f(t, x, y, u) = {} & \frac{u}{u^2+1} \\
& + \left(2\pi^2 - 1 - \frac{1}{1 - e^{-2t}\sin^2(\pi x)\sin^2(\pi y)}\right) e^{-t}\sin(\pi x)\sin(\pi y).
\end{aligned} \tag{19}
$$

In the presence of this reaction term, the governing equation becomes nonlinear. Exact solution is $u(x, y, t) = e^{-t}\sin(\pi x)\sin(\pi y)$. Initial and boundary conditions are obtained from this analytical solution. In Table 2, 11×11 number of GCL grid points is fixed. As expected, DQM provides very good accuracy using very small number of grid points. Further, at $\Delta t = 10^{-4}$, almost the same accuracy is noted with ABM4, ssprk3 and RK4, respectively, and the fastest one is RK4.

In Ref. [5], 10×10 Chebyshev points give around 10^{-11} accuracy with ABM4 in 75.58 seconds. As can be noted in Table 2, DQM gives a little bit better accuracy and faster results with ABM4 in usage of 11×11 GCL points.

Problem 3. Convection-Diffusion-Reaction Equation. This problem searches what happens if the Neumann boundary conditions exist. Consider the problem [5, 6]

$$
u_t = \frac{1}{2}\nabla^2 u - \sin(x)u_x - \sin(y)u_y + (\cos(x) + \cos(y))u - (\cos(x) + \cos(y)), \tag{20}
$$

where $(x, y) \in [0, \pi/2]^2$.

The exact solution is given by $u(x, y, t) = 1 + e^{-t}\sin(x)\sin(y)$. The mixed boundary conditions (Dirichlet and Neumann) are concerned as $u(x, 0, t) = u(0, y, t) = 1$ and $u_y(x, \pi/2, t) = 0 = u_x(\pi/2, y, t) = 0$. Table 3 presents

Table 3 Error analysis at $t_{\max} = 1$ using 21×21 GCL grids

Time scheme	Δt	$err_{\inf}$	err_{rel}	err_{avg}	CPU
BDF1	0.001	0.0001839	0.0001344	6.56e-05	19.26
BDF2	0.001	1.534e-07	1.121e-07	5.472e-08	21.41
	1e-4	1.532e-09	1.12e-09	5.468e-10	237.94
BDF3	0.001	2.348e-07	1.717e-07	8.379e-08	24.04
	1e-4	2.352e-09	1.719e-09	8.399e-10	233.5
Trap	0.001	3.065e-08	2.241e-08	1.094e-08	28.73
	1e-4	2.951e-10	2.157e-10	1.042e-10	279.74
AM3	1e-4	6.258e-12	4.575e-12	2.339e-13	289.86
Heun	1e-4	3.066e-10	2.241e-10	1.094e-10	187.47
ABM4	1e-4	4.907e-14	3.587e-14	9.582e-15	188.69
RK4	1e-4	1.843e-14	1.347e-14	3.923e-15	72.4
ssprk3	1e-4	7.656e-13	5.597e-13	3.21e-13	122.03
ssprk54	1e-4	1.269e-10	9.274e-11	5.391e-11	113.32

errors and CPU times at $t_{\max} = 1$ using 21×21 GCL grids. Once again, RK4 is the fastest and the best accurate time scheme with $\Delta t = 10^{-4}$. BDF2, BDF3 and Trap also give acceptable accuracy 10^{-8} using $\Delta t = 10^{-3}$.

In Ref. [6], 5.9e-09 maximum error is found using 8×8 Chebyshev nodes. As is seen in Table 3, DQM with RK4 gives 5.68e-09 maximum error with 9×9 GCL grids. This result also approves the proficiency of DQM with RK4 and GCL grids.

4 Conclusion

In this study, different forms of the time schemes in two-dimensional, unsteady, convection-diffusion and convection-diffusion-reaction equations are examined. Differential quadrature method is carried out for space derivatives in these equations. A detailed analysis of DQM in collaboration with some time schemes is done in the current study.

Explicit time schemes need smaller time increments while implicit time schemes enable one to use larger time increments. Implicit methods, the second, the third order backward differentiation formulas and also trapezoidal rule, are more reliable due to their unconditionally stable nature. However, in terms of CPU times, some explicit methods can also be preferable. RK4 and spprk3 perform at all problems with very good error as well as CPU time. Another alternative is to perform predictor corrector methods for good accuracy. Having a fast convergence, ABM4 is remarkable as a competitor to RK4. These observations also point that not only spatial approximation but also temporal approximation affect the results considerably.

In addition to the problems having exact solutions, convection-diffusion-reaction problems in the absence of exact solutions may be concerned as an extension of this study.

References

1. Tian, Z.F., Ge, Y.B.: A fourth-order compact ADI method for solving two-dimensional unsteady convection-diffusion problems. J. Comput. Appl. Math. 198, 268–286 (2007). https://doi.org/10.1016/j.cam.2005.12.005.
2. Dehghan, M., Mohebbi, A.: High-order compact boundary value method for the solution of unsteady convection-diffusion problems. Math. Comput. Simulat. 79, 683–699 (2008). https://doi.org/10.1016/j.matcom.2008.04.015.
3. Mittal, R.C., Tripathi, A.: Numerical solutions of two-dimensional unsteady convection-diffusion problems using modified bi-cubic B-spline finite elements, Int. J. Comput. Math. 94, 1–21 (2017). https://doi.org/10.1080/00207160.2015.1085976.
4. Mittal, R.C, Kumar, S., Jiwari, R.: A cubic B-spline quasi-interpolation method for solving two-dimensional unsteady advection diffusion equations. Int. J. Numer. Method H. 30, 4281–4306 (2020). https://doi.org/10.1108/HFF-07-2019-0597.
5. Rashidinia, J., Khasi, M., Fasshauer, G.E.: A stable Gaussian radial basis function method for solving nonlinear unsteady convection-diffusion-reaction equations. Comput. Math. Appl. 75, 1831–1850 (2018). https://doi.org/10.1016/j.camwa.2017.12.007.
6. Hoz de la, F., Vadillo, F.: The solution of two-dimensional advection-diffusion equations via operational matrices. Appl. Numer. Math. 72, 172–187 (2013). https://doi.org/10.1016/j.apnum.2013.06.006.
7. Li, J., Feng, X., He, Y.: RBF-based meshless local Petrov Galerkin method for the multi-dimensional convection-diffusion-reaction equation. Eng. Anal. Bound. Elem. 98, 46–53 (2019). https://doi.org/10.1016/j.enganabound.2018.10.003.
8. Wang, F., Wang, C., Chen, Z.: Local knot method for 2D and 3D convection–diffusion–reaction equations in arbitrary domains. Appl. Math. Lett. 105, ID 106308, (2020). https://doi.org/10.1016/j.aml.2020.106308.
9. Houston, P., Roggendorf, S., Zee van der, G.K.: Eliminating Gibbs phenomena: A non-linear Petrov–Galerkin method for the convection-diffusion-reaction equation,. Comput. Math. Appl. 80, 851–873 (2020). https://doi.org/10.1016/j.camwa.2020.03.025.
10. Chatibi, Y., El Kinani el, H., Quhadan, A.: Lie symmetry analysis of conformable differential equations. AIMS Math. 4, 1133–1144 (2019). https://doi.org/10.3934/math.2019.4.1133.
11. Chatibi, Y., El Kinani el, H., Quhadan, A.: On the discrete symmetry analysis of some classical and fractional differential equations. Math. Meth. Appl. Sci. 44, 2868–2878 (2019). https://doi.org/10.1002/mma.6064.
12. Chatibi, Y., El Kinani el, H., Quhadan, A.: Lie symmetry analysis and conservation laws for the time fractional Black-Sholes equation. Int. J. Geom. Methods M. 17, ID 2050010, (2020). https://doi.org/10.1142/S0219887820500103.
13. Li, C., Wang, Z.: Numerical Methods for the Time Fractional Convection-Diffusion-Reaction Equation,. Numer. Func. Anal. Opt. (2021) https://doi.org/10.1080/01630563.2021.1936019.
14. Shu, C.: Differential Quadrature and Its Applications in Engineering, Springer (2000).
15. Gottlieb, S.: On high order strong stability preserving Runge-Kutta and multi step time discretizations. J. Sci. Comput. 25(112), 105–128 (2005). https://doi.org/10.1007ls10915-004-4635-5.
16. Spiteri, R., Ruuth, S.: A New Class of Optimal High-Order Strong-Stability-Preserving Time Discretization Methods. SIAM J. Numer. Anal. 40(2), 469–49 (2002). https://doi.org/10.1137/S0036142901389025.

Determinants of Inflation in Malaysia: Monetary or Real Factor?

Nurul Syafiqah Binti Mohd Shafie, Phaik Chin Tan, and Siok Kun Sek

Abstract This study examines if the inflationary of Malaysia is due to monetary or real factor. The examination is based on the Quantity Theory of Money (QTM). The data are in quarterly and range from year 1997 to 2018. The main determinants to be tested are money supply (M1, M2 and M3), real gross domestic product (RGDP) and real broad effective exchange rate (RBEER). The Autoregressive Distributed Lags (ARDL) model and the nonlinear ARDL model are employed. The empirical results detect M3 as the best proxy for money supply. The results reveal that money supply has short-run impact on inflation meanwhile the RGDP has both short-run and long-run impacts on inflation. QTM only holds partially in the short-run but invalid in the long-run. The main determinant of inflation is RGDP in the long-run. Therefore, the policymakers should strive for stable economy growth through accommodation of both fiscal and monetary policy.

Keywords Inflation · Asymmetric effect · Quantity theory of money · Real factor

1 Introduction

Inflation has become one of the monetary phenomena over a period of time. Milton Friedman proclaimed that "*inflation is always and everywhere a monetary phenomenon*" [1]. Theoretically, inflation can be defined as a measure of the rate at which the average price of a basket of selected goods and services to have persistent rises over some time in an economy. In simple terms, inflation is an increase in the cost of living. The consequences of inflation, for certain reasons, have a detrimental impact, for example, on the reduction of the real value of money and other monetary products over time, whilst the positive influence can entail alleviation of economic recessions and debt relief by reducing the true amount of debt. There are many theories formed to explain inflation. However, the validity of these theories still requires further examination.

N. S. Binti Mohd Shafie · P. C. Tan · S. K. Sek (✉)
School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia
e-mail: sksek@usm.my

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_10

In Malaysia, the historical trend of inflation was influenced by many economic factors. For instance, Cheng and Tan [2] specified that the great fall of the inflation rate in Malaysia in 1974 was owing to the decreased in gross domestic product growth. Based on Bank Negara Malaysia Annual Report 2009, one factor that triggers the inflation was the value of the ringgit dropped progressively [3]. The crude oil price also may affect the inflation rate as what had happened during 2005. In 2018, Bank Negara Malaysia reported that the decline in the inflation rate was caused by the global demand and supply factors [4].

In terms of empirical research, the inappropriate use of research methodology may lead to inaccurate results. Many earlier studies applied linear modelling to examine the relationship, but the results might be biased if the true relationship is nonlinear. The conventional approach might also subject to limited information. For instance, the results do not distinguish between short-run versus long-run effects.

This study seeks to fill the limitations that addressed above. In particular, we aim to examine the validity of the quantity theory of money and to figure out if inflation is a monetary phenomenon. We also intend to study the asymmetric effect of the monetary on inflation in Malaysia, besides identifying the main determinant of inflation in Malaysia. For this purpose, the study applies the Autoregressive-Distributed Lag (ARDL) and Non-linear Autoregressive-Distributed Lag (NARDL) models. The outlines of this paper are as follows: Sect. 2 offers the discussions on the literature; Sect. 3 describes the data; Sect. 4 discusses the methods applied; Sect. 5 summarizes the key findings of the study; Finally, Sect. 6 concludes and finalizes the study.

2 Literature Reviews

2.1 Theoretical Reviews

There are various theories describing the determinants of inflation. The two main theories are Quantity Theory of Money (QTM) and Keynesian theory which proposed monetary factor and real activity or gross domestic product (GDP) as the main determinant of inflation respectively. The other theories have proposed more factors of inflation. But overall, majority theories agreed that monetary factor and GDP are the two main factors that might affect inflation. As summarized in Table 1, these factors can be classified as monetary-based, which includes money supply, and non-monetary based.

Table 1 Theories on the determinants of inflation

No	Theory	Determinant of inflation
1	Quantity Theory of Money (QTM)	Money supply
2	Keynesian theory	GDP
3	Demand-pull inflation	GDP, expected of inflation, money supply, government spending
4	Cost-push inflation	Wage-inflation, monopoly, government regulation and tax, exchange rate
5	Mixed demand inflation	GDP, expected of inflation, money supply, government expenditure and tax, wage inflation, monopoly, exchange rate
6	Structural theory	GDP, wage inflation
7	Rational expectation theory	Expected of inflation
8	Purchasing power parity model	GDP

2.2 *Quantity Theory of Money (QTM)*

QTM states that money supply is directly proportional to the change of price, holding the velocity of money and the income constant. Equation 1 below expressed QTM's equation [5],

$$MV = PY, \tag{1}$$

where M represents the suitable measure of money supply (M1, M2 and M3), V denotes the income velocity of money obtained by $Y \times P/M$, P is the aggregate price level expressed by CPI, and Y defines the RGDP. Expressing Eq. 1 in growth form, its logarithm in lower case is denoted as:

$$\begin{aligned} \mathrm{m} + \mathrm{v} &= \mathrm{p} + \mathrm{y}, \\ \mathrm{p} &= \mathrm{m} + \mathrm{v} - \mathrm{y}. \end{aligned} \tag{2}$$

There are two propositions on QTM. Firstly, the coefficient of money is gauged to be 1 in long run estimate to show the proportionality relationship between inflation and money supply. Second, the velocity of money is orthogonal to the money stock growth rate [6]. Grauwe and Polan [6] discussed the linkage between real and nominal GDP in QTM, i.e., $Y_n = PY$, where Y_n is the nominal GDP. The simplest form of QTM is known as Cambridge equation:

$$M = VY_n.$$

After substituting $Y_n = PY$ into the Cambridge equation, the monetary relationship is obtained:

$$M = VPY,$$

where P is a deflator and Y is the real expenditure. Then, it can be expressed in the proportional rate of change by letting the dot stand for logarithmic differentiation [7],

$$\dot{m} = \dot{v} + \dot{p} + \dot{y}. \tag{3}$$

Equation 3 can be written into econometric form, i.e.,

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + \varepsilon_t, \tag{4}$$

where $Y_t,\ X_t,\ Z_t$, and ε_t are inflation rate, money growth rate, RGDP growth rate of the country, and an error term, respectively. The velocity of money, v is excluded from the estimation by assuming it is covered in the error term of the estimate [6, 8, 9]. QTM predicts the coefficient of money growth is positive 1 and the coefficient of RGDP is minus 1, i.e., $\beta_1 = +1$ and $\beta_2 = -1$.

2.3 Empirical Reviews

The factors which affect inflation can mainly be grouped into monetary (money supply) and non-monetary factors. Among these factors, the money supply is the utmost determinant that affects inflation significantly. It supports the Quantity Theory of Money (QTM), which suggests that the monetary factor is the main determinant of inflation. Most studies revealed a positive relationship between money supply and inflation, but QTM only holds partially, for instance Ndidi [10], Hashim et al. [11], Kirimi [12], Adayleh [13], Bawono [14], and Setiartiti and Hapsari [15]. By contrast, some studies reported that QTM does not hold, these include Ellahi [16], and Saxena and Singh [17].

The second common variable used is the growth of real gross domestic products (RGDP). Results are mixed. The studies that reported a negative relationship include Ellahi [16], Chaudhary and Xiumin [18], Karadzic [19], Kirimi [12], Ochieng et al. [20], and Bawono [14]. Some studies reported a positive relationship, they include Uddin et al. [21], Odusanya and Atanda [22], Gyebi and Boafo [23], Saxena and Singh [17], and Wardhono et al. [24]. For Malaysia, there was a positive relationship in the recent study by Mun et al. [25], but Hashim et al. [11] reported a negative relationship between the variables tested.

Some studies examined exchange rate as the determinant of inflation and found a negative relationship, for instance Uddin et al. [21], and Falnita and Sipos [26]. Others studies reported no significant relationship [10, 12, 20, 22]. Theoretically, the inflation rate is highly affected by an interest rate with a negative relationship. This theory was supported by Hashim et al. [11] and Adayleh [13] for the cases of Malaysia and Jordan, respectively. However, the study by Falnita and Sipos [26] revealed a significant positive correlation which contradicts the theory.

3 Data

The analysis uses the data of Malaysia on consumer price index (CPI) as the dependent variable, while independent variables include money supply (M1, M2 and M3), real gross domestic product (RGDP), and real broad effective exchange rate (RBEER). To carry out the study, we obtain the data from several databases, i.e., CPI from the Global Economy database, money supply from the Bank Negara Malaysia (BNM) database, nominal GDP and CPI growth from the CEIC database, and RBEER from Federal Reserve Economic Data (FRED) database. The study utilises the quarterly data of Malaysia from the year 1997 to 2018.

4 Methodology

This research started with the conversion of data into a log form. Prior to the estimation, all variables are checked with unit-root tests of Augmented Dickey-Fuller (ADF), Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, and breakpoint unit root. The results show that some variables are stationary at level, I(0) but all variables are stationary after first differencing, I(1). This condition satisfies the requirement of ARDL and NARLD, i.e. variables with mixture of I(0) and I(1) and no variable higher than I(1). Next, the estimations on the inflation equations are made in two forms: log level and log growth rate of price (consumer price index) based on QTM theory. Finally, diagnostic checking is performed on the log level and growth rate on ARDL and NARDL models.

4.1 ARDL Versus NARDL Models

ARDL model is a linear time series model where dependent and independent variables are related contemporaneously and across historical value. In general, if y_t is the dependent variable and $x_1, x_2, \ldots, x_k$ are k explanatory variables in a general ARDL $(p, q_1, q_2, \ldots, q_k)$ model, where the model consists of lag p on the dependent variable and lag q on independent variables, it can be written as [27]:

$$y_t = \sum_{j=1}^{p} \lambda_j y_{t-j} + \sum_{j=0}^{q} \delta_j x_{t-j} + \varepsilon_t,$$

where t is the number of periods, y_t is the dependent variable; x_t represents $k \times 1$ vector of independent variables, δ_j is $k \times 1$ coefficient vectors, λ_j is the vector of scalars, and ε_t is the disturbance term with zero mean and finite variance. The equation can also be expressed in error correction format:

$$\Delta y_t = \alpha y_{t-1} + \beta x_t + \sum_{j=1}^{p-1} \lambda_j^* \Delta y_{t-j} + \sum_{j=0}^{q-1} \delta_j^* \Delta x_{t-j} + \varepsilon_t,$$

where $\alpha = -1\left(1 - \sum_{j=1}^{p} \lambda_j\right)$; $\beta = \sum_{j=0}^{q} \delta_j$; $\lambda_j^* = \sum_{m=j+1}^{p} \lambda_m,\ j = 1, 2, ..., p-1$; $\delta_j^* = \sum_{m=j+1}^{q} \delta_m,\ j = 1, 2, ..., q-1$. The equation can be summarized as:

$$\Delta y_t = \alpha(y_{t-1} - \theta x_t) + \sum_{j=1}^{p-1} \lambda_j^* \Delta y_{t-j} + \sum_{j=0}^{q-1} \delta_j^* \Delta x_{t-j} + \varepsilon_t,$$

where $\theta = -\beta/\alpha$ is the long-run equilibrium relationship among y_t and x_t; λ_j^* is the short-run effect of dependent variable on the dependent variable; δ_j^* is the immediate effect of independent variables on dependent variable; α is speed of adjustment on measuring the speed of convergence of y_t to move to its long-run equilibrium as x_t change, and the coefficient is always negative indicates stability in long-run relationship; θ is the indicator that measure the pass-through effect of shocks from independent variables to dependent variables.

Meanwhile, the NARDL model is developed to capture the nonlinear and asymmetric relationship among the variables and also aid to distinguish the short-term and long-term effects of independent variables to dependent variable. In NARDL model, we develop positive and negative shocks for variables and decompose these shocks by white noise error [28]:

$$\begin{aligned} x_t &= x_0 + x_t^+ + x_t^- \\ x_t^+ &= \sum_{j=1}^{t} \Delta x_j^+ = \sum_{j=1}^{t} \max\left(\Delta x_j, 0\right) \\ x_t^- &= \sum_{j=1}^{t} \Delta x_j^- = \sum_{j=1}^{t} \min\left(\Delta x_j, 0\right) \end{aligned} \tag{5}$$

Generally, NARDL equation can be defined as:

$$\Delta y_t = \alpha\left(y_{t-1} - \theta^+ x_t^+ - \theta^- x_t^-\right) + \sum_{j=1}^{p-1} \lambda_j^* \Delta y_{t-j} + \sum_{j=0}^{q-1} \delta_j^{*+} \Delta x_{t-j}^+ + \sum_{j=0}^{q-1} \delta_j^{*-} \Delta x_{t-j}^- + \varepsilon_t,$$

where $\theta^{+} = -\beta^{+}/\alpha$ and $\theta^{-} = -\beta^{-}/\alpha$ are the long-run equilibrium relationship among y_t and x_t^{+} and x_t^{-}, respectively.

It is divided into two main parts in testing the nexus of CPI inflation-money supply using QTM hypothesis (Eq. 4): (I) regression based on log level variables; (II) regression based on growth rate of variables. Money supply is proxy by M1, M2 and M3 respectively. The real exchange rate (*LRBEER*) is added as an additional factor for testing the nexus. The following are the relationships to be examined:

Model 1: *LCPI = f (LM1, LRGDP, LRBEER)* Model 2: *LCPI = f (LM2, LRGDP, LRBEER)* Model 3: *LCPI = f (LM3, LRGDP, LRBEER)* Model 4: *LCPI = f (LM1POS, LM1NEG, LRGDP, LRBEER)* Model 5: *LCPI = f (LM2POS, LM2NEG, LRGDP, LRBEER)* Model 6: *LCPI = f (LM3POS, LM3NEG, LRGDP, LRBEER)*	Model 7: *DLCPI = f (DLM1, DLRGDP, DLRBEER)* Model 8: *DLCPI = f (DLM2, DLRGDP, DLRBEER)* Model 9: *DLCPI = f (DLM3, DLRGDP, DLRBEER)* Model 10: *DLCPI = f (DLM1POS, DLM1NEG, DLRGDP, DLRBEER)* Model 11: *DLCPI = f (DLM2POS, DLM2NEG, DLRGDP, DLRBEER)* Model 12: *DLCPI = f (DLM3POS, DLM3NEG, DLRGDP, DLRBEER)*

where 'L' and 'DL' indicate variables in the log form and in the log differenced (or growth rate), respectively. Model (1) to (3) and (7) to (9) are estimated with ARDL while Model (4) to (6) and (10) to (12) are estimated using NARDL. Under NARDL, the money supply variable in each model is decomposed into increasing (positive) and decreasing (negative) series based on Eq. (5) i.e., *LM1POS, LM2POS, LM3POS, LM1NEG, LM2NEG,* and *LM3NEG* for money supply in log form, whereas *DLM1POS, DLM2POS, DLM3POS, DLM1NEG, DLM2NEG,* and *DLM3NEG* for money supply in log differenced form.

5 Results

Model 1 to 12 are estimated and their performances are evaluated based on Akaike Information Criterion (AIC), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Log-likelihood Function (LOGL), and R-squared as indicators for choosing the best model. The three former indicators should have the smallest values but larger values for the two latter indicators. For the relationship in the log level form, Model 4, Model 5, and Model 6 are found to be the best models. For the growth rate relationship, the best models are Model 8, Model 9, and Model 10. The results for the estimations of the best models are summarised in Tables 2 and 3, respectively. *, ** and *** indicate the significance at 10%, 5% and 1% respectively. Due to the space constraint, only the significant short-run estimates are reported in Tables 2 and 3.

Table 2 Summary on the results of NARDL estimation for log level

Model 4: NARDL(1, 4, 0, 0, 0)		Model 5: NARDL(1, 5, 5, 1, 0)		Model 6: NARDL(6, 5, 5, 6, 2)	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Short-run estimate:					
D(LM1NEG(-2))	0.1385*	D(LM2POS(-2))	0.1543*	D(LM3NEG(-4))	−1.7845***
D(LM1NEG(-3))	−0.1716***	CointEq(-1)	−0.2884***	D(LM3POS(-1))	−0.2184*
D(LRGDP)	0.0638***			D(LM3POS(-2))	0.4338***
CointEq(-1)	−0.2178***			D(LM3POS(-3))	−0.2475**
				D(LRGDP)	0.0478*
				D(LRGDP(-5))	0.0510*
				D(LRBEER(-1))	−0.0794*
				CointEq(-1)	−0.2668***
Long-run estimate:					
LM1POS	0.3098**	LM2POS	1.8902	LM3POS	−0.3115
LM1NEG	0.021	LM2NEG	−0.0103	LM3NEG	0.0552
LRGDP	0.2931***	LRGDP	0.2998***	LRGDP	0.2929**
LRBEER	−0.0318	LRBEER	−0.0742	LRBEER	0.0009
C	1.2987	C	1.4142	C	–
@TREND	–	@TREND	–	@TREND	−0.0012
Bound (F test)	11.3767***	Bound (F test)	6.2001***	Bound (F test)	4.5450**
R-squared	0.9983	R-squared	0.9984	R-squared	0.9989
Adjusted R-squared	0.9981	Adjusted R-squared	0.998	Adjusted R-squared	0.9984
LOGL	309.9842	LOGL	310.6694	LOGL	326.8487
AIC	−7.2285	AIC	−7.1627	AIC	−7.2402
RMSE	0.008	RMSE	0.007	RMSE	0.0063
MAPE	0.1285	MAPE	0.1204	MAPE	0.1124

From Tables 2 and 3, the speed of adjustment (indicated by CointEq(-1)) is negative and significant in all models, indicating there is a convergence of inflation to the long-run equilibrium level. This parameter also indicates the existence of the long-run relationship in the model, which is consistent to our bound test. As observed, the bound (F test) is significant (reject no long-run relationship) in each model. This validates the application of ARDL and NARDL models. Table 2 summarizes the best models selected for the log level inflation equation. All best models are from NARDL estimates. The NARDL reveals extra information on the dynamic asymmetric relationship between inflation and money supply which is not found in the linear regression. Using different proxies for money supply give different results. As observed, M2 does not have significant effect on inflation in both short- and long-run.

Comparisons of results are made between log level and log growth rate inflation, with different M1, M2 and M3 proxies, between short- and long-run estimates. The results from Tables 2 and 3 are summarized in Table 4. Note that + indicates a

Table 3 Summary on the best results of ARDL and NARDL estimation for growth rate

Model 8: ARDL(2, 0, 3, 1)		Model 9: ARDL(3, 6, 3, 2)		Model 10: NARDL(1, 0, 3, 3, 1)	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Short-run estimate:					
D(DLCPI(-1))	0.2284**	D(DLCPI(-2))	−0.2199*	D(DLM1POS(-2))	0.1167**
D(DLRGDP(-2))	−0.0357*	D(DLM3(-2))	0.1961***	D(DLRGDP)	0.0723***
D(DLRBEER)	0.0683*	D(DLM3(-5))	−0.1834**	D(DLRGDP(-2))	−0.0561**
CointEq(-1)	−0.9906***	D(DLRGDP(-1))	−0.0479*	D(DLRBEER)	0.0824**
		CointEq(-1)	−0.8764***	CointEq(-1)	−0.8365***
Long-run estimate:					
DLM2	0.0616	DLM3	0.1199	DLM1POS	−0.0464
				DLM1NEG	−0.0399
DLRGDP	0.0642	DLRGDP	0.1717**	DLRGDP	0.2384***
DLRBEER	−0.0379	DLRBEER	0.0075	DLRBEER	−0.0107
C	0.0030*	C	−0.0005	C	−0.0074
Bound (F test)	11.7596***	Bound (F test)	7.3424***	Bound (F test)	12.8858***
R-squared	0.264	R-squared	0.3871	R-squared	0.2695
Adjusted R-squared	0.1745	Adjusted R-squared	0.2217	Adjusted R-squared	0.1443
LOGL	306.2975	LOGL	305.9564	LOGL	306.2928
AIC	−7.0547	AIC	−7.11	AIC	−7.0673
RMSE	0.0067	RMSE	0.0058	RMSE	0.0061
MAPE	165.812	MAPE	153.1323	MAPE	169.2914

Table 4 Summary of main findings

Model	Money supply	Best model	Inflation-M		Inflation-RGDP		Inflation-RBEER	
			Short-run	Long-run	Short-run	Long-run	Short-run	Long-run
Log level	M1	NARDL/Model 4	+ (↓)	+ (↑)	0	+	0	0
	M2	NARDL/Model 5	0	0	0	+	0	0
	M3	NARDL/Model 6	+	0	0	+	0	0
Log growth rate	M1	NARDL/Model 10	+ (↑)	0	+	+	+	0
	M2	ARDL/Model 8	0	0	0	0	0	0
	M3	ARDL/Model 9	+	0	0	+	0	0

positive relationship, 0 indicates no significant relationship, (↓) indicates decreases series of money supply while (↑) indicates increases series of money supply. For the NARDL model, the inflation-money supply (M) relationship is determined by both M increases and decreases. But the table only summarizes the significant effect (significant at 5% and 1%), the non-significance estimates are excluded. For the ARDL model, the relationship of inflation-M is determined by the net supply (symmetric) which does not distinguish between M increases and decreases. Table 4 shows that in terms of inflation-RBEER, there is no significant relationship hold in both short-and long-run (majority case). Hence real exchange rate does not influence inflation. In terms of inflation-RGDP, the positive relationship exists in the long-run but weakly exists in the short-run. The effect is large and highly significant in the long-run, indicating that real GDP might be the main determinant causing to higher inflation in Malaysia in the long-run.

In general, the relationship between inflation-M exists and is positive in the short-run (using proxy of M1 and M3) as shown in both log level and log growth rate models. However, there is no one-to-one relationship with coefficient value equal to +1 as suggested by the QTM theory, hence QTM is only valid partially in the short-run. The relationship does not hold in the long-run (except log level M1 (Model 4)). Money supply proxy by M2 does not show any impact on inflation in all cases. At the final step, diagnostic tests are performed to check the properties of residuals of estimates. All models pass the autocorrelation LM test and the heteroscedasticity test (Harvey), hence the results are reliable.

6 Conclusion

This study applies the ARDL and NARDL models to examine the validity of QTM and also to find out if inflation is a monetary phenomenon, besides to study the asymmetric effect of monetary on inflation in Malaysia. The examination on QTM are in two forms which are inflation in log level and log growth rate form, with money supply is proxy by M1, M2, and M3. The results show that QTM only holds partially in the short run and does not hold in the long run. For non-monetary factors, real effective exchange rate does not have any significant relationship with inflation in all models. However, real GDP could be the main determinant of inflation. The impact is large and highly significant in majority models. To be concluded, inflation in Malaysia is more to real factor determined in the long-run. The monetary factor only can explain the inflation in the short-run. To control and improve the current situation of the inflation rate in Malaysia, the government should focus on money supply M3 in the shorter term and target on real GDP stability in the long term. Government can manipulate the high inflation rate because of the demand-pull factor through the implementation of monetary policy by the method of credit control such as control the interest rates, reserves ratio, and securities. An effective monetary policy to target at low inflation and output stability could be a good option to maintain both stability in price and output to maintain the economic stability.

Acknowledgements We would like to acknowledge our thankful to "Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme: FRGS/1/2018/STG06/USM/02/6.

References

1. Friedman, M.: Studies in the Quantity Theory of Money. 1st edn. Chicago: University of Chicago Press (1956).
2. Cheng, M.Y., Tan, H.B.: Inflation in Malaysia. International Journal of Social Economics 29(5), 411–425 (2002).
3. Bank Negara Malaysia: Bank Negara Malaysia Annual Report 2009. Kuala Lumpur: Bank Negara Malaysia (2010)
4. Bank Negara Malaysia: Bank Negara Malaysia Annual Report 2018. Kuala Lumpur: Bank Negara Malaysia (2019)
5. Fisher, I.: The purchasing power of money. 1st edn. New York: The Macmillan Co. (1911).
6. Grauwe, P.D., Polan, M.: Is inflation always and everywhere a monetary phenomenon? Scandinavian Journal of economics 107(2), 239–259 (2005).
7. Hillinger C, Süssmuth B (2010) The Quantity Theory of Money: An assessment of its real Linchpin prediction. CESifo Working Paper Series No. 2995, 1–13
8. Moroney, J.R.: Money growth, output growth, and inflation: Estimation of a Modern Quantity Theory. Southern Economic Journal 69(2), 398–413 (2002).
9. Aslan Ö (2010) Testing quantity theory of money for the Turkish economy. https://mpra.ub.uni-muenchen.de/21704/
10. Ndidi, D.E.: Determinants of inflation in Nigeria (1970-2010). The Business & Management Review 3(2), 106–114 (2013).
11. Hashim, M.J., Osman, I., Elias, N.L.: The determinants of inflation in Malaysia. In: 3rd International Conference on Accounting, Business and Economics. 26–28 August 2014, Terengganu (2014).
12. Kirimi WN. (2014) The determinants of inflation in Kenya (1970–2013). Degree thesis, University of Nairobi
13. Adayleh RM (2018) Determinants of inflation in Jordanian economy: FMOLS approach. J Internet Banking and Commerce 23(2)
14. Bawono, A.: Factors influencing the inflation of Indonesia in Islamic perspectives. Jurnal Ilmiah Ekonomi Islam, 5(2), 81–88, (2019).
15. Setiartiti, L., Hapsari, Y.: The determinants of inflation rate in Indonesia. Jurnal Ekonomi & Studi Pembangunan, 20(1), 112–123, (2019).
16. Ellahi, N.: The determinants of inflation in Pakistan: An econometric analysis. The Romanian Economic Journal 64(1), 2–12 (2017).
17. Saxena SP, Singh A (2015) Determinants of inflation in India: An econometric analysis. Scholedge Int J Manag Dev 2(8)
18. Chaudhary, S.K., Xiumin, L.: Analysis of the determinants of inflation in Nepal. American Journal of Economics 8(5), 209-–212 (2018).
19. Karadzic, V.: A regression model of inflation in Montenegro. Montenegrin Journal of Economics 10(2), 103–110 (2014).
20. Ochieng, O., Mukras, M.S., Momanyi, G.: The determinants of inflation in the Kenyan economy. International Journal of Economics 1(1), 46–60 (2016).
21. Uddin, S., Chowdhury, N.M., Hossain, M.A.: Determinants of inflation in Bangladesh: An econometric investigation. Journal of World Economic Research 3(6), 83–94 (2014).
22. Odusanya, I.A., Atanda, A.A.: Analysis of inflation and its determinants in Nigeria. Pakistan Journal of Social Sciences 7(2), 97–100 (2010).
23. Gyebi, F., Boafo, G.K.: Macroeconomic determinants of inflation in Ghana from 1990-2009. International Journal of Business and Social Research (IJBSR) 3(6), 81–93 (2013).

24. Wardhono A, Nasir MA, Qori'ah CG, Idrawati Y (2021) Movement of inflation and New Keynesian Pillips Curve in ASEAN. Economies 9(34), 1–9
25. Mun JW, Teng LY, Yi LH, Jie TJ (2017) Effect of macroeconomic variables toward inflation in Malaysia's economy. Degree thesis, Universiti Tunku Abdul Rahman
26. Falnita E, Sipos C (2007) A multiple regression model of inflation rate in Romania in the enlarged EU. Degree thesis, West University of Timisoara
27. Pesaran, M.H., Shin, Y., Smith, R.P.: Pooled mean group estimation of dynamic heterogeneous panels. Journal of American Statistical Association 94(446), 621–624 (1999).
28. Ali, I., Khan, I., Ali, H., Baz, K., Zhang, Q., Khan, A., Huo, X.: Does cereal crops asymmetrically affect agriculture gross domestic product in Pakistan? Using NARDL model approach. Ciência Rural 50(5), 1–12 (2020).

PQ-Calculus of Fibonacci Divisors and Method of Images in Planar Hydrodynamics

Oktay K. Pashaev

Abstract By introducing the hierarchy of Fibonacci divisors and corresponding quantum derivatives, we develop the golden calculus, hierarchy of golden binomials and related exponential functions, translation operator and infinite hierarchy of Golden analytic functions. The hierarchy of Golden periodic functions, appearing in this calculus we relate with the method of images in planar hydrodynamics for incompressible and irrotational flow in bounded domain. We show that the even hierarchy of these functions determines the flow in the annular domain, bounded by concentric circles with the ratio of radiuses in powers of the Golden ratio. As an example, complex potential and velocity field for the set of point vortices with Golden proportion of images are calculated explicitly.

Keywords Fibonacci divisors · Golden calculus · Hydrodynamic images

1 Golden Ratio and Inversion in Circle

The usual definition of Golden proportion or the Golden ratio is related with division of interval $x + y$ in proportion $\frac{x+y}{x} = \frac{x}{y} \quad \Rightarrow \quad \varphi^2 = \varphi + 1$, were $\frac{x}{y} = \varphi = \frac{1+\sqrt{5}}{2} \approx$ 1.6 - Golden Ratio. Here we propose new definition of Golden Ratio, connected with reflection in circle with radius R. Let a and b are symmetric points with respect to the circle at distance R between them, satisfying equations

$$a\,b = R^2, \quad b - a = R.$$

Then, distances to these points from origin are in Golden proportion of R,

$$a = \frac{1}{\varphi}R, \quad b = \varphi R.$$

O. K. Pashaev (✉)
Department of Mathematics, Izmir Institute of Technology, 35430 Izmir, Turkey
e-mail: oktaypashaev@iyte.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_11

As well known, symmetric points with respect to circle at origin in complex plain are z and $R^2/\bar{z}$. These points correspond to position of a vortex and its image in the circle, according to method of images in hydrodynamics. Then, due to above definition, if the distance between vortex and its image is R, positions of vortex and the image are in Golden proportion. For unit circle with $R = 1$, these positions are $z = \varphi e^{i\theta}$ and $z^* = \frac{1}{\varphi} e^{i\theta}$.

If method of images is applied to problem with two circles, then an infinite set of images arises [1]. These images can be counted by q-periodic functions [2] and two circle theorem [3] in q-calculus with $q = r_2^2/r_1^2$. For annular domain with two concentric circles of radiuses r_1 and r_2, it can be reformulated in terms of PQ-calculus, with $P = r_1^2$ and $Q = r_2^2$. Then, the PQ number in this calculus

$$[n]_{PQ} = \frac{P^n - Q^n}{P - Q}$$

for $P = \varphi^k$ and $Q = \varphi'^k$ becomes Binet formula for Fibonacci divisors (1). This implies that calculus of Fibonacci divisors [4] can be applied to problem of hydrodynamic images in annular domain with two circles and the Golden ratio of images.

2 Calculus of Fibonacci Divisors

The ratio of two Fibonacci numbers F_n/F_m is not in general integer number. However, surprising fact is that F_{kn}, where $k, n \in Z$ is dividable by F_k.

Definition 1 The infinite sequence of integer numbers

$$\frac{F_{kn}}{F_k} \equiv F_n^{(k)}$$

we call Fibonacci divisors conjugate to F_k.

The Binet formula for these numbers

$$F_n^{(k)} = \frac{(\varphi^k)^n - (\varphi'^k)^n}{\varphi^k - \varphi'^k}, \tag{1}$$

leads to recursion relation $F_{n+1}^{(k)} = L_k F_n^{(k)} + (-1)^{k-1} F_{n-1}^{(k)}$, where L_k are Lucas numbers. The first few sequences of Fibonacci divisors $F_n^{(k)}$ for $k = 1, 2, 3, 4, 5$ and $n = 1, 2, 3, 4, 5, \ldots$ are

$$
\begin{aligned}
k = 1;\quad & F_n^{(1)} = F_n = 1, 1, 2, 3, 5, ...\\
k = 2;\quad & F_n^{(2)} = F_{2n} = 1, 3, 8, 21, 55, ...\\
k = 3;\quad & F_n^{(3)} = \frac{1}{2}F_{3n} = 1, 4, 17, 72, 305, ...\\
k = 4;\quad & F_n^{(4)} = \frac{1}{3}F_{4n} = 1, 7, 48, 329, 2255, ...\\
k = 5;\quad & F_n^{(5)} = \frac{1}{5}F_{5n} = 1, 11, 122, 1353, 15005, ...
\end{aligned}
$$

Definition 2 The Golden derivative operator ${}_{(k)}D_F^x$, corresponding to Fibonacci divisors conjugate to F_k, $k \in Z$ acts on arbitrary function $f(x)$ as

$$
{}_{(k)}D_F^x[f(x)] = \frac{f(\varphi^k x) - f(\varphi'^k x)}{\left(\varphi^k - \varphi'^k\right) x}. \tag{2}
$$

For even k in the limit $k \to 0$ it gives usual derivative $\lim_{k\to 0}\, {}_{(k)}D_F^x f(x) = f'(x)$ and ${}_{(k)}D_F^x\, x^n = F_n^{(k)} x^{n-1}$.

Definition 3 The product of Fibonacci divisors,

$$
F_1^{(k)} F_2^{(k)} \dots F_n^{(k)} = \prod_{i=1}^{n} F_i^{(k)} \equiv F_n^{(k)}! \tag{3}
$$

—the Fibonacci divisors factorial, can be considered as k-th Fibonorial or generalized Fibonorial. The Fibonomial coefficients for Fibonacci divisors are

$$
{}_{(k)}\begin{bmatrix} n \\ m \end{bmatrix}_F = \frac{F_1^{(k)} F_2^{(k)} \dots F_{n-m+1}^{(k)}}{F_1^{(k)} F_2^{(k)} \dots F_m^{(k)}} = \frac{F_n^{(k)}!}{F_m^{(k)}!\,F_{n-m}^{(k)}!}.
$$

Hierarchy of Golden Binomials

Definition 4 The k-th Golden binomial is defined by polynomial

$$
{}_{(k)}\,(x - a)_F^n = \prod_{s=1}^{n} \left(x - \varphi^{k(n-s)} \varphi'^{k(s-1)} a\right).
$$

It can be expanded in powers of x:

$$
{}_{(k)}\,(x + y)_F^n = \sum_{m=0}^{n} {}_{(k)}\begin{bmatrix} n \\ m \end{bmatrix}_F (-1)^{k\frac{m(m-1)}{2}} x^{n-m} y^m.
$$

The k-th Golden derivative acts on this binomial as

$$
\begin{aligned}
{}_{(k)}D_F^x \ {}_{(k)}(x+y)_F^n &= F_n^{(k)} \ {}_{(k)}(x+y)_F^{n-1}, \\
{}_{(k)}D_F^y \ {}_{(k)}(x+y)_F^n &= F_n^{(k)} \ {}_{(k)}\left(x+(-1)^k y\right)_F^{n-1}, \\
{}_{(k)}D_F^y \ {}_{(k)}(x-y)_F^n &= -F_n^{(k)} \ {}_{(k)}\left(x-(-1)^k y\right)_F^{n-1}.
\end{aligned}
$$

Proposition 1 *Let, entire complex valued function of complex variable z is*

$$f(z) = \sum_{n=0}^{\infty} a_n \frac{z^n}{n!}. \tag{4}$$

Then, for any integer k exists entire complex function

$$ {}_{(k)}\ f_F(z) = \sum_{n=0}^{\infty} a_n \frac{z^n}{F_n^{(k)}!}. \tag{5}$$

Definition 5 We introduce entire exponential functions

$$ {}_{(k)}e_F^x \equiv \sum_{n=0}^{\infty} \frac{x^n}{F_n^{(k)}!}, \qquad {}_{(k)}E_F^x \equiv \sum_{n=0}^{\infty} (-1)^{k\frac{n(n-1)}{2}} \frac{x^n}{F_n^{(k)}!}. \tag{6}$$

The kth Golden derivative acts on these functions as

$$ {}_{(k)}D_F^x \left({}_{(k)}e_F^{\lambda x}\right) = \lambda \ {}_{(k)}e_F^{\lambda x}, \qquad {}_{(k)}D_F^x \left({}_{(k)}E_F^{\lambda x}\right) = \lambda \ {}_{(k)}E_F^{(-1)^k \lambda x}. \tag{7}$$

Two exponential functions are related by formula

$$ {}_{(k)}E_F^x = {}_{(-k)}\ e_F^x. \tag{8}$$

The product of the exponentials is represented by series in powers of kth Golden binomial

$$ {}_{(k)}E_F^x \cdot {}_{(k)}\ e_F^y = \sum_{n=0}^{\infty} \frac{{}_{(k)}(x+y)_F^n}{F_n^{(k)}!} \equiv {}_{(k)}\ e_F^{{}_{(k)}(x+y)_F}. \tag{9}$$

Translation operator ${}_{(k)}E_F^{y_{(k)}D_F^x}$ generates these binomials and kth Golden functions as follows

$$ {}_{(k)}E_F^{y_{(k)}D_F^x} x^n = {}_{(k)}\ (x+y)_F^n, \tag{10}$$

$$ {}_{(k)}E_F^{y_{(k)}D_F^x} f(x) = {}_{(k)}\ E_F^{y_{(k)}D_F^x} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n \cdot {}_{(k)}(x+y)_F^n. \tag{11}$$

Hierarchy of Golden Analytic Functions

Definition 6 By translation operator we introduce complex valued kth Golden analytic binomials

$$ {}_{(k)}E_F^{iy_{(k)}D_F^x} x^n ={}_{(k)} (x+iy)_F^n \tag{12}$$

and the hierarchy of kth Golden analytic functions

$$ {}_{(k)}E_F^{iy_{(k)}D_F^x} f(x) = \sum_{n=0}^{\infty} a_n \cdot {}_{(k)}(x+iy)_F^n \equiv f\left({}_{(k)}(x+iy)_F\right). \tag{13}$$

For every integer k it satisfies the $\bar{\partial}$-equation

$$ \frac{1}{2}\left({}_{(k)}D_F^x + i\,{}_{(-k)}D_F^y\right) f\left({}_{(k)}(x+iy)_F\right) = 0. \tag{14}$$

For the real and imaginary parts of these functions

$$ u(x,y) ={}_{(-k)} \cos_F\left(y_{(k)}D_F^x\right) f(x), \quad v(x,y) ={}_{(-k)} \sin_F\left(y_{(k)}D_F^x\right) f(x), \tag{15}$$

we have the Cauchy-Riemann equations

$$ {}_{(k)}D_F^x u(x,y) ={}_{(-k)} D_F^y v(x,y), \quad {}_{(-k)}D_F^y u(x,y) = -{}_{(k)}D_F^x v(x,y), \tag{16}$$

and functions are solutions of the hierarchy of Golden Laplace equations

$$ \left({}_{(k)}D_F^x\right)^2 \phi(x,y) + \left({}_{(-k)}D_F^y\right)^2 \phi(x,y) = 0. \tag{17}$$

Golden periodic functions The set of Golden derivatives determines hierarchy of Golden periodic functions for every natural k. If function $f(x)$ is Golden periodic ($k=1$),

$$ D_F^x(f(x)) = 0 \Longleftrightarrow f(\varphi x) = f(\varphi' x), \tag{18}$$

then, it is also periodic for arbitrary kth order Golden derivative,

$$ \begin{aligned} D_F^x(f(x)) = 0 \quad &\Rightarrow \quad {}_{(k)}D_F^x(f(x)) = 0, \\ f(\varphi x) = f(\varphi' x) \quad &\Rightarrow \quad f(\varphi^k x) = f(\varphi'^k x), \end{aligned} $$

for $k = 2, 3, \ldots$ But the opposite is not in general true. Indeed, function

$$ f(x) = \sin\left(\frac{\pi}{\ln \varphi^2} \ln|x|\right) $$

is Golden periodic function with $k=2$, but it is not Golden periodic ($k=1$).

3 Hydrodynamic Images and Golden Periodic Functions

3.1 Two Dimensional Stationary Flow

We consider incompressible and irrotational planar flow,

$$div\, \mathbf{u} = 0 \Rightarrow u_1 = \psi_y,\ u_2 = -\psi_x, \quad rot\, \mathbf{u} = 0 \Rightarrow u_1 = \varphi_x,\ u_2 = \varphi_y,$$

where real functions $\varphi(x, y)$ and $\psi(x, y)$ are velocity potential and stream function, correspondingly. These functions are harmonically conjugate and satisfy Cauchy-Riemann equations, $\varphi_x = \psi_y$, $\varphi_y = -\psi_x$. Combined together, they determine complex potential $f(z) = \varphi + i\psi$, as analytic function $\frac{\partial}{\partial \bar{z}} f(z) = 0$, of $z = x + iy$. Corresponding complex velocity $V(\bar{z})$, where $\bar{V}(z) = \frac{\partial}{\partial z} f(z)$, is anti-analytic function of z. For hydrodynamic flow in bounded domain, the problem is for given C-the boundary curve, find analytic function (complex potential) $F(z)$, with boundary condition $\Im F|_C = \psi|_C = 0$. This equation determines the stream lines of the flow, such that normal velocity to the curve $v_n|_C = 0$.

Two Circle Theorem. Applying two circles theorem [3] for flow $f(z)$, restricted to annular domain: $1 < |z| < \sqrt{\varphi}$ between two concentric circles $C_1 : |z| = 1$, $C_2 : |z| = \sqrt{\varphi}$, we get complex potential

$$F_\varphi(z) = f_\varphi(z) + \bar{f}_\varphi\left(\frac{1}{z}\right),$$

where $q = \frac{r_2^2}{r_1^2} = \varphi$, flow in even annulus and flow in odd annulus are correspondingly

$$f_\varphi(z) = \sum_{n=-\infty}^{\infty} f(\varphi^n z), \qquad \bar{f}_\varphi\left(\frac{1}{z}\right) = \sum_{n=-\infty}^{\infty} \bar{f}\left(\varphi^n \frac{1}{z}\right).$$

Golden φ-periodicity of flow The Golden periodicity $f_\varphi(\varphi z) = f_\varphi(z) \Rightarrow F_\varphi(\varphi z) = F_\varphi(z)$ implies that complex potential of the flow is invariant under Golden Ratio rescaling and as follows, it is Golden φ-periodic function,

$$D_z f_\varphi(z) = \frac{f(\varphi z) - f(z)}{(\varphi - 1)z} = 0.$$

Corresponding complex velocity

$$\bar{V}(z) = \sum_{n=-\infty}^{\infty} \varphi^n \bar{v}(\varphi^n z) - \frac{1}{z^2} \sum_{n=-\infty}^{\infty} \varphi^n v\left(\varphi^n \frac{1}{z}\right)$$

is Golden φ-scale invariant function $\bar{V}(\varphi z) = \varphi^{-1} \bar{V}(z)$.

Golden φ scale-invariant analytic fractal. The scale invariant function $f(\varphi z) = \varphi^d f(z)$ is subject to the φ-difference equation

$$zD_z f(z) = [d]_\varphi f(z).$$

Solution of this equation is representable as $f(z) = z^d A_\varphi(z)$, where $A_\varphi(\varphi z) = A_\varphi(z)$ is an arbitrary φ-periodic function, playing the role of φ-periodic modulation.

Golden Weierstrass-Mandelbrot fractal. As an example we consider

$$W(t) = \sum_{n=-\infty}^{\infty} \frac{1 - \cos \varphi^n t}{\varphi^{nd}}, \quad 0 < d < 1,\ \varphi > 1,$$

—continuous but nowhere differentiable function, representing fractal with dimension $2 - d$. It is Golden self-similar function $W(\varphi t) = \varphi^d W(t)$, satisfying φ-difference equation

$$tD_t W(t) = [d]_\varphi W(t).$$

By decomposing it as $W(t) = t^d A_\varphi(t)$ we extract the Golden φ-scale periodic part $A_\varphi(\varphi t) = A_\varphi(t)$, where in terms of Fibonacci numbers

$$A_\varphi(t) = \sum_{n=-\infty}^{\infty} \frac{1 - e^{i\varphi^n t}}{(\varphi^d)^n t^d} = \sum_{n=-\infty}^{\infty} \frac{1 - \cos(\varphi F_n + F_{n-1})t - i \sin(\varphi F_n + F_{n-1})t}{(\varphi^d)^n t^d}.$$

Elliptic Function Form. Let complex potential is Golden periodic analytic function $F(\varphi z) = F(z)$. The Golden ratio can be represented

$$\varphi = e^{2\pi i \frac{\omega'}{\omega}}$$

by arbitrary real ω and pure imaginary $\omega' = -i \frac{\omega}{2\pi} \ln \varphi$. Function $F(z) \equiv \Phi\left(\frac{\omega}{i\pi} \ln z\right) = \Phi(u)$ is double periodic function: $\Phi(u + 2\omega') = \Phi(u)$, $\Phi(u + 2\omega) = \Phi(u)$. It is an elliptic function on Golden torus, determined by its singular points.

Golden φ periodic flow. Simplest example of Golden φ periodic function (as principal branch) is

$$F(z) = z^{\frac{2\pi i}{\ln \varphi}} = e^{\frac{2\pi i}{\ln \varphi} \ln z} = F(\varphi z).$$

Rewritten in polar coordinates $z = re^{i\theta}$,

$$F(z) = e^{-\frac{2\pi}{\ln \varphi}\theta} \left(\cos(2\pi \log_\varphi r) + i \sin(2\pi \log_\varphi r)\right)$$

it gives stream function $\psi(r, \theta) = e^{-\frac{2\pi}{\ln \varphi}\theta} \sin(2\pi \log_\varphi r)$ and complex velocity

$$\bar{V}(z) = \frac{dF}{dz} = \frac{2\pi i}{\ln\varphi}\frac{1}{z}e^{\frac{2\pi i}{\ln\varphi}\log_\varphi z} = \frac{\Gamma}{2\pi i}\frac{1}{z}A_\varphi(z).$$

In the last form it represents Golden modulated point vortex at origin, with strength $\Gamma = -\frac{4\pi^2}{\ln\varphi}$ and stream lines $\psi|_C = 0$ at $\sin(2\pi\log_\varphi r) = 0$, or $2\pi\log_\varphi r = \pi n$, $n = 0, \pm1, \pm2, \ldots$. These lines represent an infinite set of concentric circles with radiuses $r_n = \varphi^{\frac{n}{2}}$. The squared ratio of two successive radiuses is the Golden Ratio $q = \frac{r_{n+1}^2}{r_n^2} = \varphi$. For the flow in Golden annulus $r_0 = 1$ and $r_1 = \sqrt{\varphi}$ we have $D_\varphi F(z) = 0$, and in k-th Golden annulus $r_0 = 1$ and $r_k = \varphi^{\frac{k}{2}}$ it gives $D_{\varphi^k}F(z) = 0$. Superposition

$$F_k(z) = \sum_{N=-\infty}^{+\infty} a_N z^{\frac{2\pi i}{k\ln\varphi}N}$$

describes the flow in circular annulus with radius $r = 1$ and $R = \varphi^{\frac{k}{2}}$, so that $F_k(\varphi^k z) = F_k(z)$, and the flow is φ^k-periodic.

3.2 *Vortex in Golden Annular Domain*

For point vortex at position z_0 in Golden annular domain, $1 < |z_0| < \varphi^{\frac{k}{2}}$, by Two Circle Theorem

$$F_k(z) = \frac{\Gamma}{2\pi i}\sum_{n=-\infty}^{\infty}\ln\frac{z - z_0\varphi^{kn}}{z - \frac{1}{\bar{z}_0}\varphi^{kn}}$$

and

$$\bar{V}(z) = \frac{\Gamma}{2\pi i}\sum_{n=-\infty}^{\infty}\left[\frac{1}{z - z_0\varphi^{kn}} - \frac{1}{z - \frac{1}{\bar{z}_0}\varphi^{kn}}\right].$$

The flow is Golden φ^k periodic $F_k(\varphi^k z) = F_k(z)$, with self-similar complex velocity $\bar{V}_k(\varphi^k z) = \frac{1}{\varphi^k}\bar{V}(z)$. It represents modulation of point vortex by Golden periodic function

$$\bar{V}(z) = \frac{\Gamma}{2\pi i\, z}A_k(z).$$

Golden Ratio of pole singularities Pole singularities are located at positions $z_n = z_0\varphi^{kn}$ and at symmetric points $z_n^* = \frac{1}{\bar{z}_0}\varphi^{kn}$, where $n = 0, \pm1, \pm2, \ldots \pm\infty$. The ratio of two image positions is power of Golden ratio $\frac{|z_{n+1}|}{|z_n|} = \varphi^k$. In addition, the distance between symmetric points is growing in geometric progression

$$|z_n - z_n^*| = |z_0 - z_0^*|\left(\varphi^k\right)^n.$$

Hierarchy of Golden Logarithmic Functions. The set of vortex images is determined completely by singularities of the φ-Logarithmic function,

$$Ln_{\varphi}(1-z) \equiv -\sum_{n=1}^{\infty} \frac{z^n}{[n]_{\varphi}}.$$

It converges for $|z| < \varphi$, where φ - number

$$[n]_{\varphi} \equiv 1 + \varphi + \varphi^2 + \cdots + \varphi^{n-1} = \frac{\varphi^n - 1}{\varphi - 1}$$

expressed by Fibonacci numbers is $[n]_{\varphi} = (F_{n+1} - 1)\varphi + F_n$. More general function, φ^k-logarithm $(0 < |z| < \varphi^k)$:

$$Ln_{\varphi^k}(1+z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} z^n}{[n]_{\varphi^k}} = \frac{1}{\varphi^k} \sum_{n=1}^{\infty} \frac{z}{\varphi^{kn} + z} \tag{19}$$

is expressible by Fibonacci divisors $\varphi^{kn} = \varphi^k F_n^{(k)} + (-1)^{k+1} F_{n-1}^{(k)}$ and numbers

$$[n]_{\varphi^k} = \frac{\varphi^k F_n^{(k)} + (-1)^{k+1} F_{n-1}^{(k)} - 1}{\varphi^k - 1}.$$

This function possess an infinite number of simple pole singularities at $z_n = -\varphi^{kn}$.

The logarithm function is related to entire exponential functions

$$e_{\varphi}(z) = \sum_{n=0}^{\infty} \frac{z^n}{[n]_{\varphi}!}, \quad E_{\varphi}(z) = \sum_{n=0}^{\infty} \varphi^{n(n-1)/2} \frac{z^n}{[n]_{\varphi}!},$$

which by Euler identities for φ-binomial can be written as infinite product

$$e_{\varphi}(z) = E_{\frac{1}{\varphi}}(z) = \prod_{n=0}^{\infty} \left(1 + \frac{z}{\varphi^{n+2}}\right).$$

Zeroes of φ - exp function, due to $\varphi Ln_{\varphi}(1 - \alpha z) = z \frac{d}{dz} \ln e_{\varphi}(-\varphi \alpha z)$, contribute to complex potential

$$F(z) = \sum_{s=1}^{N} i\kappa_s \ln(z - z_s) + \sum_{s=1}^{N} i\kappa_s \ln \frac{e_{\varphi}\left(-\varphi \frac{z}{z_s}\right) e_{\varphi}\left(-\varphi \frac{z_s}{z}\right)}{e_{\varphi}(-\varphi z \bar{z}_s)\, e_{\varphi}\left(-\frac{\varphi^2}{z \bar{z}_s}\right)},$$

so that all images in the second sum are determined by zeros of these functions.

3.3 Hydrodynamic Images and k-th Golden Derivatives

For even $k = 2l$, the Fibonacci divisor derivative is determined by finite difference

$$z\,_{(k)}D_F^z[f(z)] = \frac{f(\varphi^k z) - f(\frac{1}{\varphi^k}z)}{\left(\varphi^k - \frac{1}{\varphi^k}\right)},$$

vanishing for Golden periodic function $F(z)$: $_{(k)}D_F^z F(z) = 0$. In annular domain bounded by circles $1 < |z| < \varphi^{\frac{k}{2}}$, the flow is k-th Golden periodic, $F_k(\varphi^k z) = F_k(z)$, so that $_{(k)}D_F^z F_k(z) = 0$.

3.4 Single Vortex Motion

For single vortex motion, subject to equation $\dot{z}_0 = i\omega z_0$, the solution describes uniform rotation $z_0(t) = z_0(0)e^{i\omega t}$, with angular velocity

$$\omega = \frac{\varphi\kappa}{|z_0|^2}\left(Ln_\varphi\left(1 - |z_0|^2\right) - Ln_\varphi\left(1 - \frac{\varphi}{|z_0|^2}\right)\right).$$

The vortex is stationary, $\omega = 0$, at geometric mean distance $|z_0| = \varphi^{\frac{1}{4}}$ and ratio of frequencies at boundary circles is the Golden ratio: $\frac{|\omega_1|}{|\omega_2|} = \varphi$.

Semiclassical quantization of vortex motion The Bohr-Zommerfeld quantization of single vortex motion gives discrete energy spectrum

$$E_n = \frac{\Gamma^2}{4\pi}\ln\left|e_\varphi\left(-\varphi\left(n + \frac{1}{2}\right)\right)e_\varphi\left(\frac{-\varphi^2}{(n + \frac{1}{2})}\right)\right|.$$

This expression never vanishes, since zeros of exponential functions in r.h.s. should satisfy following equations, $n + \frac{1}{2} = \varphi^{k+1}$ or $n + \frac{1}{2} = \varphi^{-k}$. But in both equations the l.h.s is rational number, while the r.h.s. is irrational.

3.5 N Vortex Dynamics

For N—point vortices with circulations $\Gamma_1, \ldots, \Gamma_N$, at positions $z_1, \ldots, z_N$, equations of motion are

$$\dot{\bar{z}}_n = \frac{1}{2\pi i}\sum_{j=1(j\neq n)}^{N}\frac{\Gamma_j}{z_n - z_j} + \frac{1}{2\pi i}\sum_{j=1}^{N}\sum_{n=\pm 1}^{\pm\infty}\frac{\Gamma_j}{z_n - z_j\varphi^n} - \frac{1}{2\pi i}\sum_{j=1}^{N}\sum_{n=-\infty}^{\infty}\frac{\Gamma_j}{z_n - \frac{1}{\bar{z}_n}\varphi^n}.$$

This is Hamiltonian system with Hamiltonian function

$$H = -\frac{1}{4\pi}\sum_{i,j=1(i\neq j)}^{N} \Gamma_i\Gamma_j \ln|z_i - z_j| - \frac{1}{4\pi}\sum_{i,j=1}^{N}\Gamma_i\Gamma_j \ln\left|\frac{e_\varphi\left(-\varphi\frac{z_i}{z_j}\right)e_\varphi\left(-\varphi\frac{z_j}{z_i}\right)}{e_\varphi\left(-\varphi z_i\bar{z}_j\right)e_\varphi\left(-\frac{\varphi^2}{z_i\bar{z}_j}\right)}\right|,$$

where the second sum describes an infinite set of images with Golden proportion of positions. The Green function of the problem

$$G_I = -\frac{1}{2\pi}\ln|z - z_l| - \frac{1}{2\pi}\ln\left|\frac{e_\varphi\left(-\varphi\frac{z}{z_l}\right)e_\varphi\left(-\varphi\frac{z_l}{z}\right)}{e_\varphi\left(-\varphi z\bar{z}_l\right)e_\varphi\left(-\frac{\varphi^2}{z\bar{z}_l}\right)}\right| + \frac{1}{4\pi}\ln\varphi$$

satisfies following conditions:1. symmetry $G_I(z, z_l) = G_I(z_l, z)$; 2. boundary values, $G_I(z, z_l)|_{C_2} = 0$ - at the outer circle, $G_I(z, z_l)|_{C_1} = \frac{1}{2\pi}\ln\left|\frac{\sqrt{\varphi}}{z_l}\right|$ - at the inner circle.

Exact solution for N identical vortices $\Gamma_l = \Gamma$, $l = 1,\ldots,N$, located at the same distance $1 < r < \sqrt{\varphi}$ is $z_l(t) = re^{i\omega t + i\frac{2\pi}{N}l}$, where rotation frequency

$$\omega = \frac{\Gamma}{2\pi r^2}\left(\frac{N-1}{2} + \varphi\sum_{j=1}^{N}\left[Ln_\varphi\left(1 - \frac{\varphi}{r^2}e^{i\frac{2\pi}{N}j}\right) - Ln_\varphi\left(1 - r^2 e^{-i\frac{2\pi}{N}j}\right)\right]\right).$$

At geometrical mean distance $r = \varphi^{1/4}$ the rotation frequency is $\omega = \frac{\Gamma(N-1)}{4\pi\sqrt{\varphi}}$.
Notes and Comments. First discussed by Poincaré [5], the problem of vortex motion around two cylinders by method of images in annular domain, was exactly formulated and solved in terms of non-symmetric q-calculus in [1]. For one vortex problem, the solution in elliptic functions was given in [6]. In paper [3] it was reformulated in terms of q-periodic functions, as the two circle theorem and then applied to several bounded domains in [7]. In present work, for annular domain with Golden ratio of circles, we solved problem by "post-quantum", PQ calculus for Fibonacci divisors. The approach shows clear intuitive picture of vortex images and converges faster than the elliptic function case. It provides convergent and compact form for infinite sums of images in terms of q-elementary functions, and illuminates a new type of hidden symmetry related with quantum groups and quantum symmetry.

References

1. Pashaev, O.K., Yilmaz O.: Vortex images and q-elementary functions. J. Phys. A: Math. Theor. 41, 135207 (2008). https://doi.org/10.1088/1751-8113/41/13/135207.
2. Pashaev, O.K., Nalci, S.: Golden quantum oscillator and Binet-Fibonacci calculus. J. Phys. A: Math. Theor. 45, 015303 (2012). https://doi.org/10.1088/1751-8113/45/1/015303.

3. Pashaev, O.K.: Two-circles theorem, q-periodic functions and entangled qubit states. J. Phys. Conf. Ser. 482, 012033 (2014). https://doi.org/10.1088/1742-6596/482/1/012033.
4. Pashaev, O.K.: Quantum calculus of Fibonacci divisors and infinite hierarchy of Bosonic-Fermionic quantum oscillators. Int. J. Geom. Methods in Modern Physics 18, 2150075 (2021). https://doi.org/10.1142/S0219887821500754.
5. Poincaré, H. : Theorie des Tourbillions. Georges Carre, Paris (1893).
6. Johnson, E.R., McDonald, N.R. : The motion of a vortex near two circular cylinders. Proc. Roy. Soc. London A460, 939-954 (2004). https://doi.org/10.1098/rspa.2003.1193.
7. Pashaev, O.K. : Variations on a theme of q-oscillator. Phys. Scr. 90, 074010 (2015). https://doi.org/10.1088/0031-8949/90/7/074010.

Generalized Riesz Potential Operator in the Modified Morrey Spaces

Abdulhamit Kucukaslan

Abstract In this paper, we prove the boundedness of generalized fractional maximal operator M_ρ and generalized Riesz potential operator I_ρ in the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$. We show that the sufficient conditions for the boundedness of the operator M_ρ and the operator I_ρ from the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to another one $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $1 < p < q < \infty$ and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to the weak modified Morrey spaces $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $p = 1$, $1 < q < \infty$. We get the boundedness of our two-operators I_ρ and M_ρ in the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ using the local estimate given in the Lemma 2.

Keywords Generalized Riesz potential operator · Generalized fractional maximal operator · Modified Morrey spaces

1 Introduction

Morrey spaces $L_{p,\lambda}(\mathbb{R}^n)$ were introduced by Morrey in [1] and defined as following: For $1 \le p \le \infty$, $0 \le \lambda < n$, $f \in L_{p,\lambda}(\mathbb{R}^n)$ if $f \in L_p^{loc}(\mathbb{R}^n)$ and

$$\|f\|_{L_{p,\lambda}(\mathbb{R}^n)} := \sup_{x\in\mathbb{R}^n, r>0} r^{-\frac{\lambda}{p}} \|f\|_{L_p(B(x,r))} < \infty \tag{1}$$

holds. These spaces appeared to be useful in the study of local behavior properties of the solutions of second order elliptic PDEs. Morrey spaces found important applications to potential theory [2] and [3], elliptic equations with discountinuous coefficients [4], Navier-Stokes equations [5] and Shrödinger equations [6]. For more information about the Morrey spaces can be seen in the book [7].

A. Kucukaslan (✉)
Faculty of Aerospace Sciences, Department of Aerospace Engineering, Ankara Yildirim Beyazit University, Etlik, Ankara, Turkey
e-mail: aakucukaslan@gmail.com

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_12

In recent years, many authors have been working on the boundedness of classical operators of harmonic analysis in the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$. It can be seen some examples of these works in [8–12]. The boundednesses of fractional maximal operator M_α and Riesz potential operator in the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ were investigated by Guliyev et al. in [13]. But, the generalization of these two-operators which are the generalized fractional maximal operator M_ρ and the generalized fractional integral operator I_ρ, respectively, in the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ have not been studied, yet.

Let $f \in L^1_{loc}(\mathbb{R}^n)$ which is the set of locally integrable functions. For a measurable function $\rho : (0,\infty) \to (0,\infty)$ the generalized fractional integral operator I_ρ and the generalized fractional maximal operator M_ρ are defined by

$$I_\rho f(x) := \int\limits_{\mathbb{R}^n} \frac{\rho(|x-y|)}{|x-y|^n} f(y)dy, \quad M_\rho f(x) := \sup_{r>0} \frac{\rho(r)}{r^n} \int\limits_{B(x,r)} |f(y)|dy \tag{2}$$

for any suitable function f on $\mathbb{R}^n$, respectively. If $\rho(r) \equiv r^\alpha$, then $I_\rho \equiv I_\alpha$ is Riesz potential and $M_\rho \equiv M_\alpha$ is the fractional maximal operator, respectively. If $\rho(r) \equiv 1$, then $M_\rho \equiv M$ is the Hardy-Littlewood maximal operator. The operators I_α and M_α play important role in real and harmonic analysis (see, for example [5, 14–17]).

There is a close relation between the operators M_ρ and I_ρ (see [18], pp. 78), such that

$$M_\rho f(x) \le C\, I_\rho(|f|)(x). \tag{3}$$

The generalized Riesz potential operator, so-called generalized fractional integral operator I_ρ was initilally investigated in [19]. Nakai [20] proved the boundedness of I_ρ in the generalized Morrey-type spaces. Nowadays many authors have been culminating important observations about I_ρ and M_ρ especially in connection with Morrey-type spaces. Some of these studies are investigated by Guliyev et al. [21]; Eridani et al. [22–24]; Kucukaslan et al. [18, 25–29].

In the present work, we prove sufficient conditions for the boundedness of the generalized fractional maximal operator M_ρ and the generalized Riesz potential operator I_ρ from the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $1 < p < q < \infty$ and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to the weak modified Morrey spaces $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $p = 1, 1 < q < \infty$.

Throughout the paper we use the letter C for a positive constant, independent of appropriate parameters and not necessarily the same at each occurrence.

2 Preliminaries

$B(x,t)$ denotes the open ball centered at x of radius t for $x \in \mathbb{R}^n$ and $t > 0$. $|B(x,t)| = \omega_n t^n$ and ω_n denotes the volume of the unit ball in the Euclidean space $\mathbb{R}^n$.

Now we recall the definitions of Morrey spaces and modified Morrey spaces.

Definition 1 ([30]) Let $1 \le p < \infty, 0 \le \lambda \le n$. We denote by $L_{p,\lambda}(\mathbb{R}^n)$ the Morrey space, and by $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ the modified Morrey space, as the set of locally integrable functions $f(x)$, $x \in \mathbb{R}^n$, with the finite norms

$$\|f\|_{L_{p,\lambda}(\mathbb{R}^n)} := \sup_{x\in\mathbb{R}^n, t>0} t^{-\frac{\lambda}{p}} \|f\|_{L_p(B(x,t))}, \tag{4}$$

$$\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)} := \sup_{x\in\mathbb{R}^n, t>0} [\min\{1,t\}]^{-\frac{\lambda}{p}} \|f\|_{L_p(B(x,t))}, \tag{5}$$

respectively.

Note that

$$\widetilde{L}_{p,0}(\mathbb{R}^n) = L_{p,0}(\mathbb{R}^n) = L_p(\mathbb{R}^n), \tag{6}$$

$$\widetilde{L}_{p,\lambda}(\mathbb{R}^n) \hookrightarrow L_{p,\lambda}(\mathbb{R}^n) \tag{7}$$

and

$$\max\{\|f\|_{L_{p,\lambda}(\mathbb{R}^n)}, \|f\|_{L_p(\mathbb{R}^n)}\} \le \|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)} \tag{8}$$

and if $\lambda < 0$ or $\lambda > n$, then $L_{p,\lambda}(\mathbb{R}^n) \equiv \widetilde{L}_{p,\lambda} \equiv \Theta$, where Θ is the set of all functions equivalent to 0 on $\mathbb{R}^n$.

Definition 2 ([30]) Let $1 \le p < \infty, 0 \le \lambda \le n$. We denote by $WL_{p,\lambda}(\mathbb{R}^n)$ the weak Morrey space and by $W\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ the modified weak Morrey space as the set of all locally integrable functions $f(x)$, $x \in \mathbb{R}^n$ with finite norms

$$\|f\|_{WL_{p,\lambda}(\mathbb{R}^n)} := \sup_{x\in\mathbb{R}^n, t>0} t^{-\frac{\lambda}{p}} \|f\|_{WL_p(B(x,t))}, \tag{9}$$

$$\|f\|_{W\widetilde{L}_{p,\lambda}(\mathbb{R}^n)} := \sup_{x\in\mathbb{R}^n, t>0} [\min\{1,t\}]^{-\frac{\lambda}{p}} \|f\|_{WL_p(B(x,t))} \tag{10}$$

respectively.

Note that

$$WL_p(\mathbb{R}^n) = WL_{p,0}(\mathbb{R}^n) = W\widetilde{L}_{p,0}(\mathbb{R}^n), \tag{11}$$

$$L_{p,\lambda}(\mathbb{R}^n) \subset WL_{p,\lambda}(\mathbb{R}^n) \textit{ and } \|f\|_{WL_{p,\lambda}(\mathbb{R}^n)} \le \|f\|_{L_{p,\lambda}(\mathbb{R}^n)} \tag{12}$$

$$\widetilde{L}_{p,\lambda}(\mathbb{R}^n) \subset W\widetilde{L}_{p,\lambda}(\mathbb{R}^n) \textit{ and } \|f\|_{W\widetilde{L}_{p,\lambda}(\mathbb{R}^n)} \le \|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}. \tag{13}$$

The following lemma was proved in [22].

Lemma 1 *([22])* (i) *Let* $1 < p < q < \infty$. *Then the operator* I_ρ *is bounded from* $L_p(\mathbb{R}^n)$ *to* $L_q(\mathbb{R}^n)$ *if and only if there exists* $C > 0$ *such that for all* $r > 0$

$$\rho(r) \leq Cr^{\frac{n}{p}-\frac{n}{q}}. \tag{14}$$

(ii) *Let* $p = 1, 1 < q < \infty$. *Then the operator* I_ρ *is bounded from* $L_1(\mathbb{R}^n)$ *to* $WL_q(\mathbb{R}^n)$ *if and only if there exists* $C > 0$ *such that for all* $r > 0$

$$\rho(r) \leq Cr^{n-\frac{n}{q}}. \tag{15}$$

The following lemma was proved in [21] which is our main tools to obtain the boundedness of generalized Riesz potential operator I_ρ in the modified Morrey spaces.

Lemma 2 *([21]) Let* $1 \leq p < \infty$ *and* $f \in L_p^{loc}(\mathbb{R}^n)$. *Then the following inequality*

$$|I_\rho f(x)| \leq C\left(\rho(t)Mf(x) + \int_t^\infty \|f\|_{L_p(B(x,r))}\frac{\rho(r)}{r^{\frac{n}{p}+1}}dr\right), \tag{16}$$

satisfies for the generalized Riesz potential operator I_ρ *and for all* $y \in B(x,r), t > 0$, *where* C *is a positive constant and* Mf *is the Hardy-Littlewood maximal function.*

3 The boundedness of I_ρ and M_ρ in the spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$

When we consider our two-operators I_ρ and M_ρ, we will always assume that ρ satisfies *the Dini condition*:

$$\int_0^1 \frac{\rho(s)}{s}ds < \infty, \qquad \sup_{1\leq t<\infty}\frac{\rho(t)}{t^n} < \infty, \tag{17}$$

respectively, so that I_ρ and M_ρ are well defined, at least for characteristic functions $1/|x|^{2n}$ of complementary balls:

$$f(x) = \frac{\chi_{\mathbb{R}^n\setminus B(0,1)}(x)}{|x|^{2n}}. \tag{18}$$

Also ρ satisfies *the growth condition*: there exist constants $C > 0$ and $0 < 2k_1 < k_2 < \infty$ such that

$$\sup_{\frac{r}{2}<s\leq\frac{3r}{2}}\frac{\rho(s)}{s^n} \leq C\int_{k_1 r}^{k_2 r}\frac{\rho(t)}{t^{n+1}}dt, r > 0; \sup_{r<s\leq 2r}\frac{\rho(s)}{s^n} \leq C\sup_{k_1 r<t<k_2 r}\frac{\rho(t)}{t^n}, r > 0 \tag{19}$$

for the operators I_ρ and M_ρ, respectively. Also we will put the following conditions on ρ (see, [20]):

$$\int_0^r \frac{\rho(t)}{t}dt \leq C\rho(r), \quad r, t > 0 \tag{20}$$

and

$$\frac{\rho(r)}{r^n} \leq C\frac{\rho(s)}{s^n}, \quad s \leq r \tag{21}$$

so that the sufficient conditions for the boundedness of generalized fractional integral operator I_ρ and generalized fractional maximal operator M_ρ are satisfied on the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$.

In the following theorem which is the one of main results of our paper, we give the sufficient conditions for the boundedness of the generalized Riesz potential operator in the modified Morrey spaces. We prove Theorem 1 using the local estimate given in Lemma 2.

Theorem 1 *Let $0 \leq \lambda < n$, $1 \leq p < q < \infty$, the function ρ be a positive, measurable function and the conditions (20)–(21) be satisfied, and $f \in \widetilde{L}_{p,\lambda}(\mathbb{R}^n)$.*

(i) If $1 < p < q < \infty$ and ρ satisfies the condition (14), then the generalized Riesz potential operator I_ρ is bounded from $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ and the following norm inequality satisfies, i.e.,

$$\|I_\rho f\|_{\widetilde{L}_{q,\lambda}(\mathbb{R}^n)} \leq C\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}. \tag{22}$$

(ii) If $p = 1$, $1 < q < \infty$ and ρ satisfies the condition (15), then the generalized Riesz potential operator I_ρ is bounded from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ and the following norm inequality satisfies, i.e.,

$$\|I_\rho f\|_{W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)} \leq C\|f\|_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}. \tag{23}$$

Proof (i) Let $0 \leq \lambda < n$, $1 \leq p < q < \infty$, the function ρ be a positive, measurable function and the conditions (20)–(21) be satisfied, and $f \in \widetilde{L}_{p,\lambda}(\mathbb{R}^n)$. From the Lemma 2 we get

$$\|I_\rho f\|_{\widetilde{L}_{q,\lambda}(\mathbb{R}^n)} = \sup_{x\in\mathbb{R}^n, t>0} [\min\{1,t\}]^{-\frac{\lambda}{q}} \|I_\rho f\|_{L_q(B(x,t))} \tag{24}$$

$$= \sup_{x\in\mathbb{R}^n, t>0} \left([\min\{1,t\}]^{-\lambda} \int_{B(x,t)} |I_\rho f(y)|^q dy\right)^{\frac{1}{q}} \tag{25}$$

$$\leq C \sup_{x\in\mathbb{R}^n, t>0} [\min\{1,t\}]^{-\frac{\lambda}{q}} \tag{26}$$

$$\times\left(\int\limits_{B(x,t)}\left(\rho(r)Mf(y)+\int\limits_{r}^{\infty}\|f\|_{L_p(B(x,\tau))}\frac{\rho(\tau)}{\tau^{\frac{n}{p}+1}}d\tau\right)^{q}dy\right)^{1/q} \tag{27}$$

$$\leq C\sup_{x\in\mathbb{R}^n,t>0}[\min\{1,t\}]^{-\frac{\lambda}{q}} \tag{28}$$

$$\times\left(\int\limits_{B(x,t)}\left(\rho(r)Mf(y)+\|f\|_{\widetilde{L}_{p,\lambda}}\min\left\{\int\limits_{r}^{\infty}\frac{\rho(\tau)}{\tau^{\frac{n}{p}+1}}d\tau,\int_{r}^{\infty}\frac{\rho(\tau)}{\tau^{\frac{n-\lambda}{p}+1}}d\tau\right\}\right)^{q}dy\right)^{1/q} \tag{29}$$

$$=\sup_{x\in\mathbb{R}^n,t>0}[\min\{1,t\}]^{-\frac{\lambda}{q}} \tag{30}$$

$$\times\left(\int\limits_{B(x,t)}\left(\rho(r)Mf(y)+\|f\|_{\widetilde{L}_{p,\lambda}}\min\left\{\frac{\rho(r)}{r^{\frac{n}{p}}},\frac{\rho(r)}{r^{\frac{n-\lambda}{p}}}\right\}\right)^{q}dy\right)^{1/q}. \tag{31}$$

Thus choosing $\rho(r)=\left(\frac{\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}}{Mf(y)}\right)^{\frac{q-p}{q}}$ for all $y\in B(x,t)$ we obtain

$$\|I_\rho f\|_{\widetilde{L}_{q,\lambda}(\mathbb{R}^n)}\leq C\sup_{x\in\mathbb{R}^n,t>0}[\min\{1,t\}]^{-\frac{\lambda}{q}}\|f\|_{\widetilde{L}_{p,\lambda}}^{1-\frac{p}{q}}\|Mf\|_{L_p(B(x,t))}^{\frac{p}{q}}. \tag{32}$$

Hence from the boundedness of Hardy-Littlewood maximal operator M in the spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ (see [13], pp. 493) we get

$$\|I_\rho f\|_{\widetilde{L}_{q,\lambda}(\mathbb{R}^n)}\leq C\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}^{1-\frac{p}{q}}\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}^{\frac{p}{q}}=\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}, \tag{33}$$

which completes the boundedness of the generalized Riesz potential operator I_ρ from $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$.

(ii) Let $p=1$ and $1<q<\infty$. From the Theorem 2 we have

$$|I_\rho f(x)|\leq C\left(\rho(t)Mf(x)+\int\limits_{t}^{\infty}\|f\|_{L_1(B(x,r))}\frac{\rho(r)}{r^{n+1}}dr\right) \tag{34}$$

$$\leq C\left(\rho(t)Mf(x)+\|f\|_{\widetilde{L}_{1,\lambda}}\min\left\{\frac{\rho(t)}{t^n},\frac{\rho(t)}{t^{n-\lambda}}\right\}\right). \tag{35}$$

Thus choosing $\rho(r) = \left(\frac{\|f\|_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}}{Mf(y)}\right)^{\frac{q-1}{q}}$ for all $y \in B(x,t)$ we obtain

$$|I_\rho f(x)| \le C(Mf(x))^{1/q}\|f\|^{1-1/q}_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}. \tag{36}$$

From the inequality (36) and Theorem 1 (*ii*) in ([13] pp. 494) we get

$$\|I_\rho f\|^q_{W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)} = \sup_{x\in\mathbb{R}^n, t>0}[\min\{1,t\}]^{-\lambda}\|I_\rho f\|^q_{WL_q(B(x,t))} \tag{37}$$

$$= \sup_{r>0} r^q \sup_{x\in\mathbb{R}^n, t>0}[\min\{1,t\}]^{-\lambda}|\{y\in B(x,t) : |I_\rho f(y)| > r\}| \tag{38}$$

$$\le C\sup_{r>0} r^q \sup_{x\in\mathbb{R}^n, t>0}[\min\{1,t\}]^{-\lambda}|\{y\in B(x,t) : (Mf(y))^{1/q}\|f\|^{1-1/q}_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)} > r\}| \tag{39}$$

$$= \sup_{r>0} r^q \sup_{x\in\mathbb{R}^n, t>0}[\min\{1,t\}]^{-\lambda}\left|\left\{y\in B(x,t) : Mf(y) > \left(\frac{r}{\|f\|^{1-1/q}_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}}\right)^q\right\}\right| \tag{40}$$

$$\le C\sup_{r>0} r^q\left(\frac{\|f\|^{1-\frac{1}{q}}_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}}{r}\right)^q\|f\|_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)} \tag{41}$$

$$= \|f\|^q_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}, \tag{42}$$

which completes the boundedness of generalized Riesz potential operator I_ρ from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to the weak space $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$.

Corollary 1 *In the Theorem 1, in the special case if we choose $\rho(t) = t^\alpha$ then we get sufficient conditions for the boundedness of Riesz potential operator I_α from $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ for $1 < p < q < \infty$, and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ for $p = 1$, $1 < q < \infty$, (see Theorem 2, in [13]).*

Theorem 2 *Let $0 \le \lambda < n$, $1 \le p < q < \infty$, the function ρ be a positive, measurable function and the conditions (20)–(21) be satisfied, and $f \in \widetilde{L}_{p,\lambda}(\mathbb{R}^n)$.*

(i) If $1 < p < q < \infty$ and ρ satisfies the condition (14), then the generalized fractional maximal operator M_ρ is bounded from $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ and the following norm inequality satisfies, i.e.,

$$\|M_\rho f\|_{\widetilde{L}_{q,\lambda}(\mathbb{R}^n)} \le C\|f\|_{\widetilde{L}_{p,\lambda}(\mathbb{R}^n)}. \tag{43}$$

(ii) If $p = 1$, $1 < q < \infty$ and ρ satisfies the condition (15), then the generalized fractional maximal operator M_ρ is bounded from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ and the following norm inequality satisfies, i.e.,

$$\|M_\rho f\|_{W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)} \leq C\|f\|_{\widetilde{L}_{1,\lambda}(\mathbb{R}^n)}. \tag{44}$$

Proof The inequality (3) implies that the generalized fractional maximal operator M_ρ is dominated by the generalized Riesz potential operator I_ρ. Hence the proof of Theorem 2 is step by step the same as in the proof of Theorem 1.

Corollary 2 *In the Theorem 2, in the special case if we choose $\rho(t) = t^\alpha$ then we get sufficient conditions for the boundedness of fractional maximal operator M_α from $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ for $1 < p < q < \infty$, and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ for $p = 1, 1 < q < \infty$, (see Corollary 1, in [13]).*

Corollary 3 *In the Theorem 2, in the special case if we choose $\rho(t) \equiv 1$ then we get sufficient conditions for the boundedness of Hardy-Littlewood maximal operator M from $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ for $1 < p < \infty$, and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$ for $p = 1$, (see Theorem 1, in [13]).*

4 Conclusions

This work is a generalization of the main results of the paper [13]. Our main results are Theorems 1 and 2. In the Theorem 1, we prove sufficient conditions for the generalized Riesz potential operator I_ρ from the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $1 < p < q < \infty$ and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to the weak modified Morrey spaces $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $p = 1, 1 < q < \infty$.

It is well-known that (see [18], pp. 78) the inequality (3) is valid for our two-operators I_ρ and M_ρ. Since the generalized fractional maximal operator M_ρ is dominated by the generalized Riesz potential operator I_ρ then we get the Theorem 2, in which we prove the boundedness of the generalized fractional maximal operator M_ρ from the modified Morrey spaces $\widetilde{L}_{p,\lambda}(\mathbb{R}^n)$ to $\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $1 < p < q < \infty$ and from $\widetilde{L}_{1,\lambda}(\mathbb{R}^n)$ to the weak modified Morrey spaces $W\widetilde{L}_{q,\lambda}(\mathbb{R}^n)$, for $p = 1, 1 < q < \infty$.

References

1. Morrey, C.B.: On the solutions of quasi-linear elliptic partial differential equations, Trans. Amer. Math. Soc. 43, 126–166 (1938).
2. Adams, D.R.: A note on Riesz potentials, Duke Math. 42(4), 765–778 (1975).
3. Adams, D.R.: Choquet integrals in potential theory, Publ. Mat. 42, 3–66 (1998).
4. Di Fazio, G., Palagachev, D.K., Ragusa, M.A.: Global Morrey ragularity of strong solutions to the Dirichlet problem for elliptic equations with discontinuous coefficients, J. Funct. Anal. 166(2), 179–196 (1999).
5. Taylor, M.E.: Analysis on Morrey spaces and applications to Navier-Stokes and other evolution equations, Comm. Partial Differential Equations. 17, 1407–1456 (1992).

6. Ruiz, A., Vega, L.: Unique continuation for Schrödinger operators with potentials in the Morrey class, Publ. Math. 35(2), 291–298 (1991), Conference of Mathematical Analysis (El Escorial, 1989).
7. Kufner, A., John, O., Fucik, S.: Function spaces, Noordhoff, Leyden and Academia, Prague, (1977).
8. Guliyev, E.V., Mammadov, Y.Y.: Some embeddings into the Morrey and modified Morrey spaces sssociated with the Dunkl operator, Abstr. Appl. Anal. Volume 2010, Article ID 291345, 10 pages.
9. Guliyev, V.S., Mammadov, Y.Y.: Riesz potential on the Heisenberg group and modified Morrey spaces, An. St. Univ. Ovidius Constanta. 20(1), 189–212, (2012).
10. Guliyev, V.S. Rahimova, K.: Parabolic fractional maximal operator and modified parabolic Morrey spaces, Journal of Function Spaces and Applications. 2012, Article ID 543475, 20 pages, (2012).
11. Guliyev, V.S., Rahimova, K.: Parabolic fractional integral operator in modified parabolic Morrey spaces, Proc. Razmadze Mathematical Institute. 163, 85–106, (2013).
12. Guliyev, V.S., Omarova, M.N., Ragusa, M.A., Scapellato, A.: Regularity of solutions of elliptic equations in divergence form in modified local generalized Morrey spaces, Analysis and Mathematical Physics. 11:13. https://doi.org/10.1007/s13324-020-00433-9, (2021)
13. Guliyev, V.S., Hasanov, J., Zeren, Y.: Necessary and sufficient conditions for the boundedness of Riesz potential in modified Morrey spaces, Journal of Mathematical Inequalities. 5(4), 491–506, (2011).
14. Burenkov, V.I., , Guliyev, H.V.: Necessary and sufficient conditions for boundedness of the maximal operator in the local Morrey-type spaces, Studia Mathematica. 163(2), 157–176, (2004).
15. Burenkov, V.I., , Guliyev, V.S.: Necessary and sufficient conditions for the boundedness of the Riesz operator in local Morrey-type spaces, Potential Analysis. 30(3), 211–249, (2009).
16. Stein, E.M., Weiss, G.: Introduction to Fourier Analysis on Euclidean Spaces, Princeton Univ. Press, (1971).
17. Stein, E.M.: Singular integrals and differentiability properties of functions, Princeton Math. Ser. 30. Princeton University Press, Princeton, (1971).
18. Kucukaslan, A., Guliyev, V.S., Serbetci, A.: Generalized fractional maximal operators on generalized local Morrey spaces, Commun. Fac. Sci. Univ. Ank. Ser. A1. Math. Stat. 69(1), 73–87, (2020).
19. Gadjiev, A.D.: On generalized potential-type integral operators, Dedicated to Roman Taberski on the occasion of his 70th birthday. Funct. Approx. Comment. Math. 25, 37–44, (1997).
20. Nakai, E.: On generalized fractional integrals, Taiwanese J. Math. 5, 587–602, (2001).
21. Guliyev, V.S., Ismayilova, A.F., Kucukaslan, A., Serbetci, A.: Generalized fractional integral operators on generalized local Morrey spaces, J. Funct. Spaces. Article ID 594323, 8 pages, (2015).
22. Eridani, A., Gunawan, H., Nakai, E., Sawano, Y.: Characterizations for the generalized fractional integral operators on Morrey spaces, Math. Inequal. Appl. 17(2), 761–777, (2014).
23. Eridani, A., Gunawan, H., Nakai, E.: On generalized fractional integral operators, Sci. Math. Jpn. 60(3), 539–550, (2004).
24. Eridani, A.: On the boundedness of a generalized fractional integral on generalized Morrey spaces, Tamkang J. Math. 33(4), 335–340, (2002).
25. Kucukaslan, A., Hasanov, S.G., Aykol, C.: Generalized fractional integral operators on vanishing generalized local Morrey spaces, Int. J. of Math. Anal. 11(6), 277–291, (2017).
26. Kucukaslan, A.: Equivalence of norms of the generalized fractional fractional integral operator and the generalized fractional maximal operator on generalized weighted Morrey spaces, Ann. Funct. Anal. 11, 1007–1026, (2020).
27. Kucukaslan, A.: Two-type estimates for the boundedness of generalized fractional maximal operators on generalized weighted local Morrey spaces, Turk. J. Math. Comput. Sci. 12(1), 57–66, (2020).

28. Kucukaslan, A.: Two-type estimates for the boundedness of generalized Riesz potential operators on generalized weighted local Morrey spaces, Tbilisi Math. Journal (Accepted in 2021, not published yet).
29. Nakai, E.: Generalized fractional integrals on generalized Morrey spaces, Math. Nachr. 287(2-3), 339–351, (2014).
30. Mammadov, Y.Y.: Some embeddings into the Morrey spaces associated with the Dunkl operator on R, in Proceedings of NASA, Embedding Theorems. Harmonic Analysis, pp. 258–269, (2007).

Do Energy and Economic Growth Contribute to Environmental Degradation? Empirical Evidence From Selected European Countries

Sayed Kushairi Sayed Nordin and Siok Kun Sek

Abstract Energy as one of the factors contributing to environmental degradation, as found in literature, provides a necessary input for economic growth in most countries. This study investigates relationships of energy consumption and economic growth to environmental degradation for a panel of 25 selected European (EU) countries from 2000 to 2019. Carbon dioxide (CO_2) emission, gross domestic product (GDP) and energy consumption are used as proxies in the analysis. Besides ordinary least square regression (OLS), we employ a spatial model to measure the spatial dependence effect in the region. The Lagrange Multiplier (LM) test shows that the Spatial Durbin Model (SDM) is most appropriate for modeling the relationship. The estimated results indicate that there is a spatial effect among the variables. Thus, this study provides a better understanding of the inter-relationship among the variables in developed countries like the EU to attain sustainable development.

Keywords CO_2 emissions · Energy · Economic growth · Spatial models

1 Introduction

Pollution is one of the most pressing issues responsible for climate change and environmental problems around the globe. Among the causal to pollution, emissions of greenhouse gases due to energy consumption is believed to be a significant problem. Energy resources, on the one hand, provide the necessary resources in production to boost economic progress. Still, on the other hand, they might produce greenhouse

S. K. S. Nordin (✉) · S. K. Sek
School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia
e-mail: sayedkushairi@utem.edu.my

S. K. Sek
e-mail: sksek@usm.my

S. K. S. Nordin
Fakulti Teknologi Kejuruteraan Mekanikal dan Pembuatan, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Durian Tunggal, Melaka, Malaysia

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_13

gases that might harm the environment. Many studies have been conducted to study the nexus between energy consumption, economic growth, and CO_2 emissions. One of the famous theories, the Environmental Kuznets Curve (EKC), has established a link between economic growth and environmental quality, depicted in an inverted-U curve. According to the EKC theory, the more substantial economic growth initially leads to higher CO_2 emissions before reversing after the economy has reached its optimal level.

To reduce the emissions of greenhouse gases, coal and fossil fuel-based energy should be regulated. However, environmental degradation is unavoidable in the progress of economic development. This is the cost of industrialization, which may abrupt the progress of economic growth [1]. Such trade-off condition also brings challenges to balancing the outcome to achieve the mutually benefits [2]. Unsustainable development has the potential to harm ecosystems and human health. Global warming and climate change have resulted in rising sea levels, melting ice caps and glaciers, and causing extreme weather such as droughts, major floods, and tornadoes.

Meanwhile, the National Energy Consumption Technology Laboratory reports that there are two significant sources of CO_2 emissions: anthropogenic and natural CO_2. Human activities such as worldwide trade, travel industries, power generation, transportation, industrial sources, chemical manufacture, fossil fuel combustion, and agricultural production release CO_2 from anthropogenic sources. According to [3], the relationship between economic growth and the environment can be explained in three ways: win-win, win-lose, and lose-lose.

As one of the world's major oil-importing countries, Germany aims to reduce CO_2 emissions by 40% by 2020 and 80% to 95% by 2050 [4]. In 2020 and 2050, primary energy consumption is expected to be lowered by 20% and 50%, respectively, compared to 2008. According to the International Energy Agency, other European countries, such as Italy, France, and Poland, were among the top 20 emitters of CO_2 in 2015. As a result of missing the Paris climate agreement's 2016 target, France has agreed to reconsider its CO_2 emissions target.

In [5], the empirical evidence confirms a long-term positive association between economic growth and renewable energy consumption in the EU. Even though the relationship between economic growth, energy consumption, and environmental degradation has been extensively studied, there is no consensus on the spatial connection. These three variables have the potential to be geographically connected between neighbouring countries in this light. As a result, this study applies spatial panel regression to avoid the potential coefficient bias that non-spatial models disregard.

This study contributes to the literature of energy consumption-growth-environmental quality in three ways: first, to the best of our knowledge, the first study uses a spatial panel model to predict the link between these three variables in EU countries. Second, most of the previous works focused on examining spatial effects based on the national level. Still, this study is one of the limited spatial studies that focused on the regional level. Third, we compare the results from different spatial weighting matrices in examining the link between these variables. We use three distinct models to assess the spatial interaction effects in different countries.

The remainder of this research is organized as follows: Section 2 reviews the literature, Section 3 is an overview of the data source and methods, and Section 4 summarizes the empirical findings. Section 5 concludes with the findings and suggestions.

2 Literature Review

The study on the nexus between energy consumption, economic growth and environmental quality is broad. The topic has attracted experts from different fields, including economics, ecologists, industrial players, and legislators. The inverted U-shaped EKC illustrates how pollution rises in tandem with economic expansion before reversing its direction after a country reaches its optimal state of economic growth. This research, however, has yet to provide consistent outcomes as results might vary using different methodologies, data and periods.

OLS regression, a non-spatial panel approach, had been widely utilized by numerous academics before panel regression became popular. One of the prominent researchers, [6], introduced the neoclassical growth model and concluded that a country's economy would enter a stable phase when no new investment is necessary. The study also stated that technological advancement is necessary for long-term economic success. On the other hand, the economic expansion comes with the cost of development, which leads to the degradation quality of the environment. The emissions of greenhouse gases increase as industrial growth accelerates. According to [7], CO_2 emissions have a detrimental impact on economic growth in Turkey, whereas energy consumption has a favourable impact. Authors [3, 8] and [9] found a positive link between CO_2 emissions and economic growth. Yassin and Aralas [10], on the other hand, applied panel data from 1971 to 2005 and found no strong link between the variables. Yassin and Aralas [10] examined the regional level of selected Association of Southeast Asian Nations (ASEAN) countries using homogeneous and heterogeneous estimators to investigate pollution in the area from 1990 to 2016. The study concluded that the CO_2 emissions of ASEAN could rise as carbon-intensive activities and industrialization continue to grow.

The effect of spatial interaction in neighbouring areas is not considered by classic panel methods, including EKC's estimation. As a result, these strategies are highly susceptible to estimation bias. Countries are frequently assumed to be homogeneous and spatially independent. When there exists a dependency among countries, there is always a risk of spatial dependence on the data. Zhang et al. [11] examined 30 Chinese provinces from 2005 to 2012 and discovered a sizeable spatial influence of CO_2 emissions. Regarding understanding the components that contribute to CO_2 emissions, the spatial autoregressive with autoregressive disturbance model (SAC) outperformed other models. Hao et al. [12] estimated the Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and SDM using data from 29 Chinese provinces from 1995 to 2012. The findings backed up the existence of an EKC for China's per capita coal use. They encouraged the government to find a way to minimize energy consumption and greenhouse gas emissions. Kang et al. [13] also investigated how energy-related factors affect CO_2 emissions in China. From 1997 to 2012, the SDM illustrated that GDP is the primary cause of rising CO_2 emissions. Another study was conducted by [14] in 30 provinces in China. Economic growth is expected to impact sustainable development in the long run positively. The Moran's I result showing that a positive clustering among the provinces.

Rios and Gianmoena [15] applied the data from 1970 to 2014 to conduct the dynamic geographic panel research in assessing the evolution of cross-country CO_2 emissions. According to the findings, there exists an association between the initial CO_2 and growth rates. Maddison [16], on the other hand, focused the examination on various contaminants (SO_2, NOx, VOCs, and CO_2). The findings contradicted the notion that environmental degradation can be rectified by economic growth on its own. From 2001 to 2009, [17] examined the spatial dependence on CO_2 emissions at the state level in the United States. The SAR with fixed effects revealed that economic spillovers were significantly positive while state-level emissions were significantly negative. Authors [18, 19] and [20] employed spatial panel regression to look at data from Chinese provinces.

3 Methodology

This study investigates the relationship between economic growth and energy consumption on environmental degradation from a spatial perspective. An annual panel data of 25 selected EU countries (N=25) from the year 2000 to the year 2019 ($T = 20$) are obtained from The World Bank and BP Statistical Review of World Energy websites. The shapefile for spatial analysis is generated from www.diva-gis.org/gdata. The sampled countries are Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Switzerland, Turkey and Ukraine. The data are converted into logarithm form for consistency purposes. The details of the variables used are given below (Table 1).

The analysis involves several steps. First of all, all variables are checked with their stationarity using the Levin-Lin-Chu panel unit root test. The rejection of the null hypothesis reveals the stationary of the series. The next step is to construct the spatial weight matrix to capture the spatial relations among the spatial objects, diagnose the spatial dependence in OLS regression and comparing the results between spatial regression models.

A weight matrix is essential in spatial modelling. The use of a weight matrix makes spatial models different from non-spatial models. The purpose of the spatial weight matrix is to describe the spatial relations between N spatial objects, the countries. For every spatial weight, W_{ij}, indicates the "spatial influence" of object j on object i [20].

Table 1 Description of variables

Variables	Definition	Unit of Measurement
CO_2 emission	The carbon emissions reflect only those through consumption of oil, gas and coal for combustion-related activities and are based on 'Default CO_2 Emissions Factors for Combustion' listed by the IPCC in its Guidelines for National Greenhouse Gas Inventories (2006)	Million tonnes
GDP	Gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products	Per capita (constant 2010 US$)
Energy consumption	Commercially traded fuels, such as modern renewables used to create power, are classified as primary energy. Energy from all sources of non-fossil power generation is used on an input-equivalent basis	Gigajoule per capita

$$W_{ij} = \begin{cases} 1, & if\ share\ a\ common\ border\ or\ or\ a\ single\ common\ point \\ 0, & elsewhere \end{cases}$$

In the analysis, we apply a contiguity matrix to indicate whether spatial units share a boundary or not. The higher order of weight matrix (second-order) is used to capture the effects from neighbours of the neighbours of spatial objects. Figure 1 exhibits the connectivity maps for both orders. The number of countries connected as neighbours for the second-order weight matrix is remarkably higher than the first order.

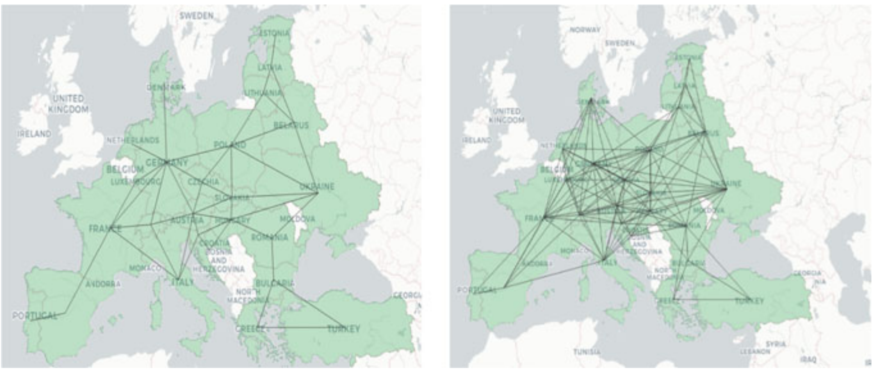


Fig. 1 Connectivity map of the first-order (left) and second-order (right) spatial weight matrix

Three spatial models, namely Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM) and Spatial Durbin Model (SDM), are used in the modelling part. The SAR is useful to investigate the direct impact of the dependent variable, Y_i on its neighbouring Y_j and vice versa, and it uses the spatial lag effect of the dependent variable. The SEM refers to a situation in which the unobserved shock to country i is influenced by unobserved shocks in adjacent countries. Meanwhile, the SDM assumes the Y_i is spatially dependent on the independent and dependent variables of other adjacent countries. The empirical models are specified as:

SAR:

$$lnCO_{2it} = \rho \sum_{j=1}^{n} W_{ij} lnCO_{2jt} + \beta_1 lnGDP_{it} + \beta_2 lnENE_{it} + \mu_i + \lambda_i + \varepsilon_{it} \quad (1)$$

SEM:

$$lnCO_{2it} = \beta_o + \beta_1 lnGDP_{it} + \beta_2 lnENE_{it} + \mu_i + \lambda_i + \varphi_{it} \quad (2)$$

$$\varphi_{it} = \delta \sum_{j=1}^{n} W_{ij} \varphi_{jt} + \varepsilon_{it} \quad (3)$$

SDM:

$$\begin{aligned} lnCO_{2it} &= \rho \sum_{j=1}^{n} W_{ij} lnCO_{2jt} + \beta_1 lnGDP_{it} + \beta_2 lnENE_{it} \\ &\quad + \theta_1 \sum_{j=1}^{n} W_{ij} ln(GDP)_{jt} + \theta_2 \sum_{j=1}^{n} W_{ij} ln(ENE)_{jt} + \mu_i + \lambda_i + \varepsilon_{it} \end{aligned} \quad (4)$$

where i and j represent countries i and j respectively ($i \neq j$) and t denotes the period in the year. $\beta's$ are parameters and ρ is the spatial parameter indicating the dependency among the dataset. W_{ij} is an element in the square $N \times N$ spatial weight matrix. The μ_i represents the time-fixed effect of spatial units. The λ_i denotes spatially fixed effects and ε_{it} is the random disturbance term. φ_{it} in SEM is the spatial correlation error term. δ denotes the spatial autoregressive coefficient, which shows the effects of residuals of the local area, and ε_{it} is an i.i.d (independent and identically distributed) residual. θ in SDM denotes the spatial autocorrelation coefficient of independent variables.

LM tests are applied to test whether the spatial dependence in OLS regression. If a dependency exists, the OLS regression is inefficient and biased to regress the data. For spatial error, the H_o of the LM test is that the error has no spatial autocorrelation versus the H_a of the error has spatial autocorrelation. For spatial lag, the H_o for the

LM test is that the spatial lagged dependent variable has no spatial autocorrelation versus the H_a of spatial autocorrelation.

4 Empirical Results

The panel unit-root test of Levin-Lin-Chu is conducted to check the stationarity of each variable. From Table 2, the null hypothesis of unit root is rejected, implying that all panel series are stationary (no unit root). Next, the data is estimated using OLS before proceeding to spatial regression analysis. The result of OLS is presented in Table 3. The coefficients are not significant for both independent variables. A low value of R-square (0.2%) indicates that the OLS model is not well performed to explain the data.

Table 4 reports the results of diagnostic tests. Based on LM and Robust LM results, most of the statistics are significant across different orders of weight matrices. The null hypothesis of the error has no spatial autocorrelation, and the spatial lagged dependent variable that has no spatial autocorrelation will be rejected. Therefore, the SDM that incorporates both features of SAR and SEM is the most suitable model to deal with spatial dependence [14]. The results also reveal that OLS regression is not appropriate to model the data because it is inefficient and biased.

Table 5 compares the maximum likelihood (ML) estimation of the three models using different weight matrices. In general, all the spatial models fit the data very well, with high values of R-square greater than 95%. As we can see, the coefficients

Table 2 Test for unit root

	Level	
	Without trend	Trend
$LNCO_2$	0.219	−2.545***
LNGDP	−3.576***	−3.836***
LNENE	0.018***	−2.264***

*significant at 10%, ** significant at 5%, *** significant at 1%

Table 3 OLS regression results

Variables	Coefficient
LNGDP	0.002 (0.088)
LNENE	0.135 (0.198)
Constant	3.610 (0.743)
R-square	0.002

Standard errors are given in the bracket

Table 4 Diagnostic tests for spatial dependence in OLS regression

	Spatial contiguity weight matrix	
	Order 1	Order 2
Spatial error:		
LM	13.014***	1.035
Robust LM	85.660***	34.106***
Spatial Lag:		
LM	14.688***	0.641
Robust LM	87.334***	33.712***
Suggested model	SDM	SDM

*significant at 10%, ** significant at 5%, *** significant at 1%

Table 5 Spatial regression results

	Spatial contiguity weight matrix					
	Order 1			Order 2		
	SDM	SAR	SEM	SDM	SAR	SEM
LNGDP	−0.903*** (0.110)	−0.092 (0.089)	−0.199* (0.109)	−0.534*** (0.092)	0.024 (0.091)	0.092 (0.104)
LNENE	−0.084 (0.179)	0.048 (0.195)	−0.079 (0.193)	−0.225 (0.176)	0.116 (0.199)	0.073 (0.201)
Constant	−3.740*** (−3.740)	4.096*** (0.738)	6.528*** (1.1271)	−1.278 (1.118)	3.817*** (0.7814)	3.036*** (0.785)
Rho	0.055*** (0.0498)	0.188*** (0.049)		−0.382*** (0.092)	−0.075 (0.091)	
Sigma	1.109*** (0.035)	1.235*** (0.039)	1.224*** (0.039)	1.094*** (0.034)	1.257*** (0.039)	1.254*** (0.397)
Lambda			0.252*** (0.057)			−0.1628 (0.117)
W* LNGDP	0.951*** (0.167)			1.963*** (0.193)		
W* LNENE	1.609*** (0.179)			−1.184*** (0.433)		
Global Moran I	−0.264***	0.271***	0.244***	0.744***	0.395***	0.458***
R-square	0.979	0.992	0.989	0.957	0.992	0.992
AIC	0.033	0.012	0.016	0.068	0.011	0.012

*significant at 10%, ** significant at 5%, *** significant at 1%

are not consistent across the models. However, to explain the model, SDM is the most appropriate model for both orders of weight matrix as suggested in previous diagnostic tests (Table 4). Therefore, we use SDM for interpreting factors affecting CO_2 emissions in the EU.

Based on the chosen model, all coefficients for independent variables are consistent for both order weight matrices. GDP and energy have a significant and negative effect on CO_2 emissions. Moving to spatial analysis, it is worth mentioning that the spatial correlation parameter, Rho of the estimation, is significant with the opposite direction of the effect when using order one and order two weight matrix. The results indicate that CO_2 emissions in adjacent countries affect each other. The estimate of 0.055 suggests that a 1% increase in the CO_2 emissions in the adjacent countries should result in about a 0.055% increase in local CO_2 emissions. Meanwhile, the estimate of −0.382 indicates that a 1% increase in the CO_2 emissions in the adjacent countries should result in a 0.382% decrease in local CO_2 emissions. The product terms of independent variables and spatial weights matrix W in the SDM indicate how independent variables in adjacent countries affect other country's dependent variable. The spatial spillover effects of adjacent GDP and energy consumption areas were highly significant to CO_2 emissions. All the products terms were positive except W*LNENE for order two weight matrix, which implies that an increase in energy consumption in adjacent countries should decrease CO_2 emissions in a country. In summary, the results indicate that regardless of the order of weight matrices in spatial analysis, economic growth and energy consumption significantly affect CO_2 emissions in adjacent countries. Comparing between SDM model in order one and order two, order 1 is preferable as it shows lower AIC (0.033) and higher R-square (97.9%).

5 Conclusion

In this paper, we examine the relationship between economic growth, energy consumption and environmental degradation. We compare the results of spatial models with the non-spatial model (OLS regression) using annual panel data from 2000–2019. Our main objective is to investigate the spatial effects of independent variables (GDP and energy) on the dependent variable (CO_2 emissions). The panel unit root tests show that all the variables are stationary. The diagnostics tests suggest that the OLS regression is miss-specified as the dataset has a spatial dependence issue. Thus, we can conclude that the spatial model is more appropriate to model the relationship. Based on LM tests, SDM was favourable than SEM and SAR. The spatial models show that spatial interactions exist among the variables. The results imply that GDP and energy significantly affect each other among adjacent countries in the EU region. The empirical results are consistent with the past studies, in which an increase in economic growth and energy consumption leads to an increase in CO_2. Overall, it is possible to conclude that all the variables are related to sampled countries of the EU. The results offer a better understanding of the spillover effects of economic growth and energy consumption on environmental quality. Indirectly, it proves that support from neighbouring countries is necessary to achieve sustainability both locally and globally.

Acknowledgement This work is supported by the Research University Grant (1001/PMATHS/8016115), Universiti Sains Malaysia.

References

1. Dasgupta, P., Maler, K: Net National Product, Wealth, and Social Well- Being. Environment and Development Economics 5, 69-93 (2000)
2. Apergis, N., Payne, J.E.: CO_2 emissions, energy usage, and output in Central America. Energy Policy 37, 3282–3286 (2009).
3. Karakas, A.: Economic Growth-CO_2 Emission Relationship in OECD and Non-OECD Countries: A Panel Data Analysis for the Period between 1990-2011. The International Journal of Humanities & Social Studies 2, 57-62 (2014).
4. Each Country's Share of CO_2 Emissions, https://www.ucsusa.org/global-warming/science-and-impacts/science/each-countrys-share-of-co2.html, last accessed 2021/06/20.
5. Cerovi, J., Muhadinovi, M., Radonji, M.: How does renewable energy consumption affect economic growth in the traditional and new member states of the European Union. Energy Reports 6, 505–513 (2020).
6. Solow, R.M.: A Contribution to the Theory of Economic Growth. Quarterly Journal of Economics 70, 65-94 (1956).
7. Wang, Y., Kang, L., Wu, X., Xiao, Y.: Estimating the environmental Kuznets curve for ecological footprint at the global level: A spatial econometric approach. Ecological Indicators. 34, 15-21 (2013).
8. Yin, J., Zheng, M., Chen, J.: The effects of environmental regulation and technical progress on CO_2 Kuznets curve: An evidence from China. Energy Policy 77, 97-108 (2015).
9. Ozturk, I.: A literature survey on energy-growth nexus. Energy Policy 38, 340-349. (2010).
10. Yassin, J., Aralas, S., B.: Economic Structural Transformation and CO_2 Emissions in Asian Countries: Homogeneous Vs. Heterogeneous Estimators. SSRN Electronic Journal, 4(13), 57–68 (2019).
11. Zhang, Q., Yang, J., Sun, Z., Wu, F.: Analyzing the impact factors of energy-related CO_2 emissions in China: What can spatial panel regressions tell us? Journal of Cleaner Production 161, 1085-1093 (2017).
12. Hao, Y., Liu, Y., Weng, J., Gao, Y.: Does the Environmental Kuznets Curve for coal consumption in China exist? New evidence from spatial econometric analysis. Energy 114, 1214–1223 (2016).
13. Kang, Y., Zhao, T., Wu, P.: Impacts of energy-related CO2 emissions in China: a spatial panel data technique. Natural Hazards 81, 405–421 (2016).
14. Li, J., Luo, Y., & Wang, S. Spatial effects of economic performance on the carbon intensity of human well-being: The environmental Kuznets curve in Chinese provinces. Journal of Cleaner Production, 233, 681–694 (2019).
15. Rios, V., Gianmoena, L.: Convergence in CO_2 emissions: A spatial economic analysis with cross-country interactions. Energy Economics 75, 222–238 (2018).
16. Maddison, D.: Environmental Kuznets curves: A spatial econometric approach. Journal of Environmental Economics and Management 51, 218–230 (2006).
17. Burnett, J.W., Bergstrom, J.C., Dorfman, J. H.: A spatial panel data approach to estimating U.S. state-level energy emissions. Energy Economics 40, 396–404 (2013).
18. Li, M., Li, C., Zhang, M.: Exploring the spatial spillover effects of industrialization and urbanization factors on pollutants emissions in China's s Huang-Huai-Hai region. Journal of Cleaner Production 195, 154–162 (2018).
19. Li, K., Fang, L., He, L.: How urbanization affects China's energy efficiency: A spatial econometric analysis Journal of Cleaner Production 200, 1130–1141 (2018).

20. Que, W., Zhang, Y., Liu, S., Yang, C.: The spatial effect of fiscal decentralization and factor market segmentation on environmental pollution. Journal of Cleaner Production 184, 402–413 (2018).

Some Characterizations for Harmonic Complex Fibonacci Sequences

Emel Karaca and Fatih Yilmaz

Abstract In this study, we define and construct a new number system, called the harmonic complex Fibonacci sequences (HCF), which is inspred by the well-known harmonic and complex numbers in literature. Some algebraic properties are examined in detail. Furthermore, by using generating fuction, Binet formula and Cassini identity are shown.

Keywords Complex number · Harmonic number · Fibonacci sequence

1 Introduction

In mathematics, there are several number systems which are studied by many mathematicians actively, such as harmonic, algebraic, complex, hybrid etc. One of the well-known number system is harmonic number system.

The nth harmonic number, denoted by H_n, is defined by

$$H_n = \sum_{k=1}^{n} \frac{1}{k},$$

where $H_0 = 0$. These numbers are required in many areas of science such as in calculations of high energy physics, in computer science, in the efficiency analysis of algorithms, see for details in [1, 2].

Especially, in algebra and number theory, there are a lot of research articles about special sequences such as Lucas, Pell, Horadam, Fibonacci, etc., see [3–9]. Among these sequences, Fibonacci sequence has attracted many mathematicians. The definition of Fibonacci sequence is given as follows:

E. Karaca (✉) · F. Yilmaz
Department of Mathematics, Ankara Haci Bayram Veli University, Ankara, Turkey
e-mail: emel.karaca@hbv.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_14

The Fibonacci sequence is defined by the following recurrence relation, for $n \geq 0$:

$$F_{n+2} = F_{n+1} + F_n$$

with $F_0 = 0$, $F_1 = 1$.

There is a fact that the complex numbers are quite easy to describe in terms of real numbers. In other words, every complex number has the form $x + yi$ where x and y are real numbers. If a complex number is at the denominator of a fraction, it can be expressed as below:

$$\frac{1}{x+iy} = \frac{x-iy}{x^2+y^2} = \frac{x}{x^2+y^2} - \frac{y}{x^2+y^2}i.$$

At this study, we give a new persoective to generalization of the complex numbers. Motivated by such researchs, we defined harmonic complex Fibonacci sequence (HCF). Then we obtain some significiant algebraic properties in detail. Furthermore, by exploitting generating fuction, Binet formula and Cassini identity has been examined.

This study is organized as follows: In Sect. 2, complex Fibonacci numbers are defined. Some fundamental properties of these numbers are represented. Moreover, generating function, Binet formula and Cassini identity are proved in detail. In Sect. 3, obtained results are discussed.

2 On Harmonic Complex Fibonacci Sequences

In this section, we define a new approach to complex numbers by inspring harmonic numbers. In this approach, we combined harmonic numbers with complex numbers. We consider harmonic complex Fibonacci numbers as

$$H_n^C = \sum_{k=1}^{n} \frac{1}{F_k + iF_{k+1}}. \tag{1}$$

Let us denote the set of the harmonic hybrid Fibonacci numbers as follows:

$$K_1 = \{\sum_{k=1}^{n} \frac{1}{F_k + iF_{k+1}} : F_k \; is \; k-th \; Fibonacci \; number, \; i^2 = -1, \; 1 < n < \propto\}. \tag{2}$$

For all $H_n^C = \sum_{k=1}^{n} \frac{1}{F_k+iF_{k+1}}$, $H_m^C = \sum_{k=1}^{m} \frac{1}{F_k+iF_{k+1}} \in K_1$, the fundamental operators are defined as below:

(a) Addition:

(i) If $m = n$, $H_n^C + H_m^C = 2\sum_{k=1}^{n} \frac{1}{F_k+iF_{k+1}}$.

(ii) If $m < n$, $H_n^C + H_m^C = 2\sum_{k=1}^{m} \frac{1}{F_k+iF_{k+1}} + \sum_{k=1}^{m+1} \frac{1}{F_k+iF_{k+1}}$.

(ii) If $n < m$, $H_n^C + H_m^C = 2\sum_{k=1}^{n} \frac{1}{F_k+iF_{k+1}} + \sum_{k=n+1}^{m} \frac{1}{F_k+iF_{k+1}}$.
(b) Multiplication:

$$H_n^C \cdot H_m^C = \left(\sum_{k=1}^{n} \frac{1}{F_k + iF_{k+1}}\right) \cdot \left(\sum_{k=1}^{m} \frac{1}{F_k + iF_{k+1}}\right)$$

is calculated as distributing terms on the right by exploitting the each product of unit.
(c) Complex conjugate:

$$\overline{H_n^C} = \sum_{k=1}^{n} \frac{F_k + iF_{k+1}}{F_k^2 + F_{k+1}^2}.$$

From the definition of complex conjugate, the norm is calculated as

$$N(H_n^C) = \| H_n^C \| = \sqrt{H_n^C \overline{H_n^C}}. \tag{3}$$

Now, let us consider some algebraic properties and significiant theorems of HCF. There is an isomorphism between the map $\phi : K \longrightarrow M_{2\times 2}$, where K is a ring. Considering this isomorphism, let us denote the matrix representation of H_n^C :

$$M = \begin{pmatrix} \sum_{k=1}^{n} \frac{F_k}{F_k^2+F_{k+1}^2} & \sum_{k=1}^{n} \frac{-F_{k+1}}{F_k^2+F_{k+1}^2} \\ \sum_{k=1}^{n} \frac{F_{k+1}}{F_k^2+F_{k+1}^2} & \sum_{k=1}^{n} \frac{F_k}{F_k^2+F_{k+1}^2} \end{pmatrix},$$

where the matrix representations of units are $1 \leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $i \leftrightarrow \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.
Let M be the matrix representation of $\phi(H_n^C)$ is called harmonic complex Fibonacci matrix corresponding to H_n^C, where $\phi(H_n^C) = M$ and ϕ is a ring isomorphism.

Definition 1 Assume that M is a 2 by 2 real matrix corresponding to H_n^C. The determinant of M is defined as

$$\begin{aligned} det(M) &= \left(\sum_{k=1}^{n} \frac{F_k}{F_k^2 + F_{k+1}^2}\right)^2 det(i) + \left(\sum_{k=1}^{n} \frac{F_{k+1}}{F_k^2 + F_{k+1}^2}\right)^2 det(1), \\ &= \sum_{k=1}^{n} \frac{1}{F_k^2 + F_{k+1}^2}. \end{aligned}$$

Theorem 1 *Let M and N be the matrix representations of any non-zero H_n^C and H_m^C, respectively. The following equalities are satisfied:*
(i) $det(M) = det(M^T) = det(\bar{M})$,
(ii) For any $\lambda \in C$, $det(\lambda M) = \lambda^2 det(M)$,
(iii) $det(MN) = detM.detN$.

Proof Exploiting the definitions of conjugate and transpose for matrices given above and properties of determinant, the proof can be shown easily.

Corollary 1 *Let M be the matrix representation of any non-zero H_n^C. Then, the following equation is satisfied:*

$$det(M)^2 = det(M^T \bar{M}). \tag{4}$$

Example 1 For $n = 1$ and $m = 2$, let us consider two matrices M and N corresponding to H_n^C and H_m^C:

$$M = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

and

$$N = \begin{pmatrix} \frac{7}{10} & -\frac{9}{10} \\ \frac{9}{10} & \frac{7}{10} \end{pmatrix}.$$

We can simply calculate that $det(MN) = det M.det N$. Moreover, the transpose of M is

$$M^T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

The conjugate of M is equal to M. Therefore, it is seen that $det(M)^2 = det(M^T \bar{M})$. Additionally, we verify the condition (iii) given Theorem 1 as calculating

$$det(MN) = det(M).det(N) = 260.$$

Theorem 2 *Let H_n^C and H_m^C be any non-zero harmonic complex Fibonacci sequences. The following properties are satisfied:*
(i) $\overline{H_n^C}^T = H_n^C$.
(ii) If H_n^C is invertible, $\overline{H_n^C}^{-1} = \overline{((H_n^C))^{-1}}$.
(iii) If H_n^C is invertible, $((H_n^C)^T)^{-1} = ((H_n^C)^{-1})^T$.
(iv) If H_n^C and H_m^C are invertible, $(H_n^C H_m^C)^{-1} = (H_m^C)^{-1}(H_n^C)^{-1}$.

Proof (i)–(iii) and (iv) may be proved by considering the properties of transpose, inverse and complex conjugate of H_n^C. We will only prove (ii).

(ii) Let (H_n^C) be invertible. Hence, we give

$$\overline{(H_n^C)}^{-1} = \sum_{k=1}^{n} (F_k + i F_{k+1}) \tag{5}$$

and

$$\overline{(H_n^C)^{-1}} = \frac{\sum_{k=1}^{n} \frac{F_k + iF_{k+1}}{F_k^2 + F_{k+1}^2}}{\sum_{k=1}^{n} \frac{1}{F_k^2 + F_{k+1}^2}}$$
$$= \sum_{k=1}^{n} (F_k + iF_{k+1}).$$

So, we obtain the desired result.

Example 2 For $n = 1$, let us consider H_1^C, and denote its matrix representation with

$$M = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

The inverse of M is

$$M^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

It is very easy to see that all the conditions of Theorem 2 are satisfied.

Theorem 3 *(Generating Function): For $n \geq 0$, the generating function of H_n^C is*

$$G(t) = \frac{H_0^C + (H_1^C - H_0^C)t}{1 - t - t^2}.$$

Proof By using the definition of generating function, we have

$$G(t) = \sum_{k=1}^{\propto} H_n^C t^n.$$

By writing the following equalities, i.e.:

$$G(t) = H_0^C + H_1^C t + H_2^C t^2 + \cdots + H_n^C t^n + \cdots \tag{6}$$
$$tG(t) = H_0^C t + H_1^C t^2 + H_2^C t^3 + \cdots + H_n^C t^{n+1} + \cdots \tag{7}$$
$$t^2 G(t) = H_0^C t^2 + H_1^C t^3 + H_2^C t^4 + \cdots + H_n^C t^{n+2} + \cdots \tag{8}$$

So, we obtain

$$G(t)(1 - t - t^2) = H_0^C + (H_1^C - H_0^C)t.$$

Then, we get

$$G(t) = \frac{H_0^C + (H_1^C - H_0^C)t}{1 - t - t^2}.$$

Theorem 4 *(Binet Formula) For $n \geq 1$, the Binet formula is*

$$H_n^C = \sum_{k=1}^{n} \frac{\alpha - \beta}{(A_*\alpha^k - B_*\beta^k)} \tag{9}$$

where $A = 1 + i\alpha$, $B = 1 - i\beta$ and $\alpha = \frac{1+\sqrt{5}}{2}$, $\beta = \frac{1-\sqrt{5}}{2}$.

Proof Using the definition of the harmonic complex Fibonacci numbers and the Binet's formula of the Fibonacci numbers, it can be seen easily.

Theorem 5 *(Cassini Identity) Let H_n^C be the sequence of harmonic complex Fibonacci sequence. Then for $n \geq 1$:*

$$H_{n+1}^C H_{n-1}^C - (H_n^C)^2 = 5(-1)^{n-1} AB, \tag{10}$$

where $A = \frac{-2\sqrt{5}+7-i(2\sqrt{5}+3)}{10\sqrt{5}}$ and $B = \frac{-2\sqrt{5}-7-i(2\sqrt{5}-3)}{10\sqrt{5}}$.

Proof This proof employs Binet formula for harmonic complex Fibonacci numbers:

$$H_n^C = \sum_{k=1}^{n} (A\alpha^k + B\beta^k), \tag{11}$$

where $\alpha = \frac{1+\sqrt{5}}{2}$ and $\beta = \frac{1-\sqrt{5}}{2}$. Moreover, we have $\alpha.\beta = -1$ and $\alpha - \beta = \sqrt{5}$.

$$\begin{aligned}
H_{n+1}^C H_{n-1}^C - (H_n^C)^2 &= (A\alpha^{n+1} + B\beta^{n+1})(A\alpha^{n-1} + B\beta^{n-1}) - (A\alpha^n + B\beta^n)^2 \\
&= AB\alpha^{n-1}\beta^{n-1}(\alpha^2 - 2\alpha\beta + \beta^2) \\
&= AB\alpha^{n-1}\beta^{n-1}(\alpha - \beta)^2 \\
&= 5AB(-1)^{n-1}.
\end{aligned}$$

As we put A and B into the equation given above, we find

$$H_{n+1}^C H_{n-1}^C - (H_n^C)^2 = 5(-1)^{n-1}\left(\frac{-2\sqrt{5}+7-i(2\sqrt{5}+3)}{10\sqrt{5}}\right)\left(\frac{-2\sqrt{5}-7-i(2\sqrt{5}-3)}{10\sqrt{5}}\right). \tag{12}$$

3 Conclusion

In this study, we define a new number system wich is called complex Fibonacci number. Some significiant algebraic properties are examined in detail. Furthermore, several fundamental theorems are proved and some exampes are given to support the main results.

References

1. Gromov, N.A.: Possible quantum kinematics. II. Nonminimal case. Journal of Mathematical Physics 51, (2010).
2. Gromov, N.A.: Possible quantum kinematics. Journal of Mathematical Physics 47, (2006).
3. Halici, S.: On Fibonacci quaternions. Adv. Appl. Clifford Algebras 22, 321–327 (2012).
4. Ates, F., Gök, I. and Ekmekci, N.: Algebraic properties of bi-periodic dual Fibonacci quaternions. Kragujevac Journal of Mathematics 43, 99–107 (2017).
5. Liana, A.S., Wloch, I.: Jacobsthal and Jacobsthal Lucas hybrid numbers. Annales Mathematiccae Silesianae 33, 276–283 (2019).
6. Catarino, P.: On k-Pell hybrid numbers. Journal of Discrete Mathematical Sciences and Cryptography 22, 83–89 (2019).
7. Szynal-Liana, A., Wloch, I.: On Pell and Pell-Pucas hybrid numbers. Commentat Mathematici 58, 11–17 (2018).
8. Szynal-Liana, A.: The horadam hybrid numbers. Discussiones Mathematicae General Algebra and Applications 38, 91–98 (2018).
9. KIzIlateş, C.: A new generalization of Fibonacci hybrid and Lucas hybrid numbers. Chaos, Solitons and Fractals 130, 1–5 (2020).

Topological Rings and Annihilator Conditions

Ebru Bitkin and Yeliz Kara

Abstract We introduce the class of rings which is defined by certain annihilator conditions on projection invariant ideals. We obtain connections between the former class of rings and the class of dual rings. Examples illustrate our results.

Keywords Annihilator conditions · Dual rings · Projection invariant ideals

1 Introduction

All rings are associative with unity and $\mathcal{R}$ will denote such a ring in this paper.

Kaplansky [1] introduced the concept of dual rings by defining certain annihilator conditions on closed ideals of a topological ring. Following this idea, Hajarnavis and Norton [2] studied the variation of dual ring condition in algebraic sense. Several authors have investigated different annihilator conditions on rings and modules.

Motivated by the study in [2], we introduce the class of π-dual rings defined by certain annihilator conditions on projection invariant one-sided ideals. We obtain some results to make connections between π-dual rings and related concepts. Some examples are indicated to demonstrate our results.

In Sect. 2, we present preliminary results for topological rings. Moreover, the motivation of the study on Baer rings and some current research on this field are mentioned. In Sect. 3, we introduce and explore the class of π-dual rings. Since there is a link between the theory of Baer rings and extending rings (see [3]), we also investigate some connections between π-dual rings and π-extending rings. We present examples to illustrate our results.

We use $r(\mathcal{X})$, $l(\mathcal{X})$ and $Mat_n(\mathcal{R})$, for the right annihilator of $\mathcal{X}$ in $\mathcal{R}$, the left annihilator of $\mathcal{X}$ in $\mathcal{R}$ and the $n \times n$ full matrix ring over $\mathcal{R}$, respectively. Recall that a ring $\mathcal{R}$ is called *Abelian*, if $ea = ae$ for all $a \in \mathcal{R}$ and $e = e^2 \in \mathcal{R}$. An idempotent $e \in \mathcal{R}$ is called *left semicentral* if $xe = exe$, for all $x \in \mathcal{R}$. The set of left semicentral

E. Bitkin (✉) · Y. Kara
Department of Mathematics, Bursa Uludağ University, 16059 Bursa, Turkey
e-mail: yelizkara@uludag.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_15

idempotents of R is denoted by $S_e(\mathcal{R})$. A ring $\mathcal{R}$ is a *duo ring*, if every one-sided ideal is two-sided. For undefined terminology and notions, we refer to [1, 3–5].

2 Background

In this section, we give some preliminaries for the concept of topological rings and the class of rings defined by certain annihilator conditions. Some examples are also indicated.

Definition 1 [6] A *topological ring* is a ring $\mathcal{R}$ that is also a topological space such that both the addition and the multiplication are continuous.

All Banach algebras, rings of all bounded real-valued functions on a topological space $\mathcal{X}$, rings of continuous linear operators on a normed vector space $\mathcal{X}$, the p-adic numbers with standard topology are the examples of topological rings (see [6]).

In 1948, Kaplansky [1] introduced *dual rings* on a topological ring with respect to the annihilator conditions.

Definition 2 [1] A topological ring $\mathcal{R}$ is called a *dual ring*, if for every closed right ideal $\mathcal{A}$, $r(l(\mathcal{A})) = \mathcal{A}$, and for every closed left ideal $\mathcal{B}$, $l(r(\mathcal{B})) = \mathcal{B}$.

It is evident that $l(\mathcal{B})$ is a closed left ideal and $r(l(\mathcal{B}))$ is a closed right ideal. Then, for any $\mathcal{X} \subseteq \mathcal{R}$, $r(l(\mathcal{X}))$ is the smallest closed right ideal. In particular, if $\mathcal{X}$ is a right ideal, $r(l(\mathcal{X}))$ is the closure of $\mathcal{X}$.

For many examples and detailed information about dual rings, we refer to [1].

Motivated by Kaplansky's idea on topological spaces, Hajarnavis and Norton [2] studied the following class of rings.

Definition 3 [2] A ring $\mathcal{R}$ is called a *dual ring*, if for all right ideal $\mathcal{A}$ of $\mathcal{R}$, $\mathcal{A} = r(l(\mathcal{A}))$ and for all left ideal $\mathcal{B}$ of $\mathcal{R}$, $\mathcal{B} = l(r(\mathcal{B}))$.

Notice that $\mathcal{Y}$ is a maximal right ideal of a dual ring provided that $l(\mathcal{Y})$ is a minimal left ideal. For a dual ring $\mathcal{R}$, it is proved that $\mathcal{R}_{\mathcal{R}}$ has finite uniform dimension and $\mathcal{R}/\mathcal{J}$ is semisimple Artinian, where $\mathcal{J}$ is the Jacobson radical of $\mathcal{R}$ (see [2, 3.3 Lemma] and [2, 3.4 Theorem]).

Kaplansky [5] defined the notion of *Baer* rings which dates back to the study of Operator Theory. Clark [7] generalized the idea of Baer ring and named *quasi-Baer*. The importance of such rings are the annihilators generated by idempotent elements.

Definition 4 [5, 7] A ring $\mathcal{R}$ is *Baer* (resp., *quasi-Baer*) if the right annihilator of each nonempty set (resp., ideal) is generated by an idempotent element, as a right ideal of $\mathcal{R}$.

Realize that Baer and quasi-Baer ring conditions are right-left symmetric. Clearly, every Baer ring is quasi-Baer. Since a Baer ring is right nonsingular, any prime ring which is not right nonsingular is quasi-Baer. However it is not Baer. In the following example, some examples of Baer and quasi-Baer rings are given.

Example 1 (1) Any domain is a Baer ring.
(2) The endomorphism ring of $\mathcal{V}_{\mathcal{F}}$ is Baer, where $\mathcal{V}_{\mathcal{F}}$ is a vector space over the field $\mathcal{F}$.
(3) Von Neumann algebras satisfy Baer property.
(4) The full (resp., upper triangular) matrix ring over a quasi-Baer ring is quasi-Baer.
(5) The polynomial ring of a quasi-Baer ring is quasi-Baer.

As opposed to quasi-Baer, Baer condition does not transfer to the matrix (resp., polynomial) rings.

Example 2 [3, Example 3.1.28] Consider $\mathcal{T} = Mat_2(\mathbb{Z}[x])$. Although $\mathbb{Z}[x]$ is Baer, $\mathcal{T}$ is not. However, the ring $Mat_2(\mathbb{Z})$ is Baer, by [5, Example 3]. Since $\mathcal{T} = Mat_2(\mathbb{Z}[x]) = Mat_2(\mathbb{Z})[x]$, $\mathcal{T}$ is not Baer.

Since quasi-Baer property behaves better than Baer property with respect to ring extensions, it can be discussed on a property between the Baer and quasi-Baer properties. Recently, the authors in [8] introduced the notion of π-Baer rings using the class of *projection invariant* ideals.

Definition 5 [4] A left (resp., right) ideal $\mathcal{P}$ of $\mathcal{R}$ is called *projection invariant*, if $\mathcal{P}g \subseteq \mathcal{P}$ (resp., $g\mathcal{P} \subseteq \mathcal{P}$) for each $g^2 = g \in \mathcal{R}$, denoted by ${}_{\mathcal{R}}\mathcal{P} \trianglelefteq_p {}_{\mathcal{R}}\mathcal{R}$ (resp., $\mathcal{P}_{\mathcal{R}} \trianglelefteq_p \mathcal{R}_{\mathcal{R}}$).

Obviously, every two sided ideal is projection invariant ideal. It can be seen that every one sided ideal of an Abelian ring is projection invariant. Various examples for the projection invariance concept can be constructed.

Lemma 1 (i) *Any intersection and sum of projection invariant right (resp., left) ideals is a projection invariant right (resp., left) ideal.*
(ii) $r(\mathcal{P})$ *is a projection invariant right ideal, for a projection invariant left ideal* $\mathcal{P}$ *of* $\mathcal{R}$*.*
(iii) $l(\mathcal{G})$ *is a projection invariant left ideal, for a projection invariant right ideal* $\mathcal{G}$ *of* $\mathcal{R}$*.*

Proof It is clear from [4, p. 50] and [8, Lemma 2.1].

Definition 6 [8] A ring $\mathcal{R}$ is π*-Baer*, if for each projection invariant left ideal $\mathcal{P}$, $r(\mathcal{P}) = c\mathcal{R}$ for some $c = c^2 \in \mathcal{R}$.

In the following result, the connections between π-Baer rings and related notions are mentioned.

Proposition 1 *[8, Theorem 2.1.] Consider the following assertions:*

(a) $\mathcal{R}$ *is Baer.*
(b) $\mathcal{R}$ *is* π*-Baer.*
(c) $\mathcal{R}$ *is quasi-Baer.*
Thence $(a) \Rightarrow (b) \Rightarrow (c)$*, however these arrows are irreversible, in general.*

3 Main Results

We introduce and investigate the class of π-dual rings in this section. The connections between π-dual rings and related notions are studied.

Definition 7 A ring $\mathcal{R}$ is called *right projection invariant dual ring* (denoted by, π-dual) if for each projection invariant right ideal $\mathcal{P}$ of $\mathcal{R}$, $\mathcal{P} = r(l(\mathcal{P}))$.

The left projection invariant π-dual concept can be defined similarly. $\mathcal{R}$ is said to be a *π-dual ring*, if $\mathcal{R}$ is both right and left π-dual. Clearly $\mathcal{P} \subseteq r(l(\mathcal{P}))$, for all $\mathcal{P}_{\mathcal{R}} \trianglelefteq_p \mathcal{R}_{\mathcal{R}}$.

Proposition 2 *Suppose* $\mathcal{R}$ *is a* π*-dual ring and* $\{\mathcal{P}_i\}_{i\in I}$ *a collection of projection invariant right (resp., left) ideal of* $\mathcal{R}$*. Hence*

$$l(\bigcap_{i\in I} \mathcal{P}_i) = \sum_{i\in I} l(\mathcal{P}_i) \quad \textit{(resp.,}\ \ r(\bigcap_{i\in I} \mathcal{P}_i) = \sum_{i\in I} r(\mathcal{P}_i)\textit{).}$$

Proof Suppose $\mathcal{R}$ is a π-dual ring and $\{\mathcal{P}_i\}_{i\in I}$ is a collection of projection invariant right ideal of $\mathcal{R}$. Then $\mathcal{P}_i = r(l(\mathcal{P}_i))$. Thence $l(\bigcap_{i\in I} \mathcal{P}_i) = l(\bigcap r(l(\mathcal{P}_i))) = lr\sum_{i\in I} l(\mathcal{P}_i)$. Observe that ${}_R l(\mathcal{P}_i) \trianglelefteq_p {}_{\mathcal{R}}\mathcal{R}$ by Lemma 1. It is clear from Lemma 1 that $\sum l(\mathcal{P}_i)$ is projection invariant left ideal of R. $\sum l(\mathcal{P}_i) = lr(\sum l(\mathcal{P}_i))$, as $\mathcal{R}$ is π-dual. Therefore $l(\bigcap_{i\in I} \mathcal{P}_i) = \sum_{i\in I} l(\mathcal{P}_i)$. The equality for left ideals is obtained similarly.

Lemma 2 *If any of the following assertions hold, then* $\mathcal{R}$ *is a dual ring* $\Leftrightarrow$ $\mathcal{R}$ *is a* π*-dual ring.*

(a) $\mathcal{R}$ *is indecomposable,*
(b) $\mathcal{R}$ *is Abelian,*
(c) $\mathcal{R}$ *is right duo.*

Proof Since every right ideal of an indecomposable (resp., Abelian, right duo) ring is projection invariant, the proof is straightforward.

Recall that a ring $\mathcal{R}$ is *right* π *-extending* [9], if every projection invariant right ideal $\mathcal{X}$ of $\mathcal{R}$ is essential in a direct summand of $\mathcal{R}$. The connections between the π-Baer, π-extending and π-dual conditions are obtained in the next result.

Proposition 3 (i) *Assume $\mathcal{R}$ is a π-Baer ring. Then $\mathcal{R}$ is right π-dual ring $\Leftrightarrow$ For each $\mathcal{A}_{\mathcal{R}} \trianglelefteq_p \mathcal{R}_{\mathcal{R}}$, $\mathcal{A} = e\mathcal{R}$, where $e = e^2 \in S_l(\mathcal{R})$.*
(ii) *If $\mathcal{R}$ is a π-Baer and right π-dual ring, then $\mathcal{R}$ is right π-extending.*

Proof (i) Suppose $\mathcal{R}$ is a π-Baer and right π-dual ring. Let $\mathcal{A}_{\mathcal{R}} \trianglelefteq_p \mathcal{R}_{\mathcal{R}}$. Hence $l(\mathcal{A}) = \mathcal{R}h$ for some $h = h^2 \in \mathcal{R}$, as $\mathcal{R}$ is π-Baer. Thus $r(l(\mathcal{A})) = (1-h)\mathcal{R}$. Since $\mathcal{R}$ is right π-dual, $\mathcal{A} = r(l(\mathcal{A})) = (1-h)\mathcal{R} = e\mathcal{R}$, where $g = 1 - h$. Notice from Definition 6, $g \in S_l(\mathcal{R})$. The converse is clear.

(ii) It is a result of part (i).

It is clear from [8, Proposition 2.4] that every nonsingular π-Baer ring is π-extending. Thus, it might be expected π-dual property implies nonsingularity. However, we eliminate this possibility by the next example.

Example 3 (i) $\mathcal{R} = \mathbb{Z}$ is nonsingular, but clearly $\mathcal{R}$ is not a π-dual ring.

(ii) Let $\mathcal{F}$ be a field and $\mathcal{V}$ a vector space over $\mathcal{F}$ such that $dim(\mathcal{V}_{\mathcal{F}}) = 1$. Consider the following ring

$$\mathcal{R} = \begin{bmatrix} \mathcal{F} & \mathcal{V} \\ 0 & \mathcal{F} \end{bmatrix} = \left\{ \begin{bmatrix} a & x \\ 0 & a \end{bmatrix} : a \in \mathcal{F}, x \in \mathcal{V} \right\}.$$

Then $\mathcal{R}$ is a commutative indecomposable ring. Then the singular right ideal is $\begin{bmatrix} 0 & \mathcal{V} \\ 0 & 0 \end{bmatrix} \neq 0$. However, $\mathcal{R}$ is a π-dual ring.

(iii) The ring in (ii) also illustrates that a π-extending and π-dual ring need not to be a π-Baer ring.

4 Conclusions

In this study, we explore the class of π-dual rings. Some connections between π-dual rings and related concept are obtained. As a future work, a topology can be constructed on the π-dual rings. Moreover, it will be worthwhile to investigate the correspondence of π-dual ring properties in modules.

References

1. Kaplansky, I.: Dual rings. Annals of mathematics, 689–701, (1948).
2. Hajarnavis, C.R., Norton, N.C.: On dual rings and their modules. Journal of Algebra, 93(2), 253–266, (1985).
3. Birkenmeier, G.F., Park, J.K., Rizvi, S.T.: Extensions of rings and modules. New York: Birkhäuser, (2013).

4. Fuchs, L.: Infinite Abelian Groups I, Academic Press, New York.
5. Kaplansky, I.: Rings of Operators, Benjamin, New York, (1965).
6. Warner, S.: Topological rings, Elsevier, (1993).
7. Clark, W.E.: Twisted matrix units semigroup algebras. Duke Mathematical Journal, 417–423, (1967).
8. Birkenmeier, G.F. Kara,Y. Tercan, A.: π-Baer rings. Journal of Algebra and Its Applications, 1719(2), 1850029, (2018).
9. Birkenmeier, G.F., Tercan, A., Yücel, C.C.: The extending condition relative to sets of submodules. Communications in Algebra, 42(2), 764–778, (2014).

On a Generalization of FI-Extending Modules

Yeliz Kara

Abstract In this paper, we introduce modules with the property that every f-closed submodule has a complement which is a direct summand. We provide some structural properties related to the class of generalization of extending modules.

Keywords Complement submodule · Extending module · Fully invariant submodule

1 Introduction

Throughout this paper, R will denote an associative ring with unity and M will denote a unital right R-module, respectively.

Recall that a module is *extending* [3], if every submodule is essential in a direct summand, equivalently every complement submodule is a direct summand. Various generalizations of extending modules have been studied by many authors [1, 6]. In particular, M is called a C_{11}-*module* [6], if every submodule has a complement which is a direct summand of M. Furthermore, a submodule A of M is called *fully invariant* [4], provided $\phi(A)$ is contained in A for all endomorphism ϕ of M. Many examples of fully invariance concept can be constructed in different algebraic structures. By using fully invariant submodules, a module M is said to be *FI-extending* [1], if every fully invariant submodule is essential in a direct summand of M. Module theoretical properties of FI-extending modules are investigated in [1]. It is shown in [2] that extending condition implies C_{11} condition and C_{11} condition implies FI-extending condition.

Recall from [5] that a submodule N of M is said to be *z-closed*, if M/N is nonsingular. Motivated by the idea of z-closed submodules, we introduce the notion of f-close in this study. Incidentally, we call a submodule N of M is *f-closed*, if N is fully invariant submodule of M such that M/N is nonsingular. We provide basic

Y. Kara (✉)
Department of Mathematics, Bursa Uludağ University, 16059 Bursa, Turkey
e-mail: yelizkara@uludag.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_16

properties of f-closed submodules. Further, we discuss modules with the condition that every f-closed submodule has a complement which is a direct summand. To this end, we obtain some connections between the former class of modules and generalizations of extending modules. Moreover, we explore when the aforementioned module property inherits by its submodules.

For notation $X \leq M$, $X \leq_e M$, $X \leq_c M$, and $X \trianglelefteq M$, we mean that X is a right R-submodule of M, X is an essential submodule of M, X is complement in M, and X is a fully invariant submodule of M, respectively. For undefined terminology and notions, we refer to [1, 3–5].

2 Main Results

In this section, we present the fundamental connections between the former class of modules and related notions. For the following well-known result about fully invariant submodules, we refer to [4, p. 50].

Lemma 1 *[4, p. 50] Let M be a module.*

(i) Assume $\{X_i \mid i \in I\}$ is the family of fully invariant submodules of M. Then $\bigcap_{i\in I} X_i$ and $\sum_{i\in I} X_i$ are fully invariant submodule of M.

(ii) Let $X_1 \leq X_2 \leq M$ such that X_1 is fully invariant in X_2 and X_2 is fully invariant in M. Then X_1 is fully invariant in M.

(iii) Assume $M = \bigoplus_{i\in I} M_i$ and X is fully invariant in M. Then $X = \bigoplus_{i\in I}(X \cap M_i)$, where $X \cap M_i$ is fully invariant in M_i for each $i \in I$.

Definition 1 We call a submodule N of M is *f-closed* provided that N is a fully invariant submodule of M and M/N is nonsingular.

In the following result, we obtain some useful properties of f-closed submodules.

Lemma 2 *(i) Any intersection of f-closed submodules of M is an f-closed submodule of M.*

(ii) Let $A_1, A_2 \leq M$ such that $A_1 \leq A_2$. If A_1 is an f-closed submodule of A_2 and A_2 is an f-closed submodule of M, then A_1 is an f-closed submodule of M.

(iii) Every f-closed submodule of M is a complement in M.

Proof (i) Let X_1 be an f-closed submodule of M and X_2 be an f-closed submodule of M. Then $X_1 \trianglelefteq M$ and $X_2 \trianglelefteq M$ such that $Z(M/X_1) = 0$ and $Z(M/X_2) = 0$. Notice from Lemma 1, $X_1 \cap X_2 \trianglelefteq M$. Define the homomorphism $\theta : M \to M/(X_1 \cap X_2)$, by $\theta(m) = (m + X_1, m + X_2)$. Then $M/(X_1 \cap X_2) \cong \theta(M) \leq (M/X_1) \oplus (M/X_2)$. Since $Z(M/X_1) = 0$ and $Z(M/X_2) = 0$, $Z(\theta(M)) = 0$. It follows that $Z(M/(X_1 \cap X_2)) = 0$. Therefore $X_1 \cap X_2$ is an f-closed submodule of M.

(ii) Suppose A_1 is an f-closed submodule of A_2 and A_2 is an f-closed submodule of M. It follows from Lemma 1 that $A_1 \trianglelefteq M$. Since $(M/A_1)/(A_2/A_1) \cong M/A_2$, it can be checked that $Z(M/A_1) = 0$. Hence A_1 is an f-closed submodule of M.

(iii) Let X be an f-closed submodule of M. Then $X \trianglelefteq M$ and $Z(M/X) = 0$. Assume that there exists a submodule T of M such that $X \leq_e T \leq_c M$. Then T/X is singular, hence $T/X \subseteq Z(M/X)$. Since $Z(M/X) = 0$, $T = X$. Thus X has no proper essential extension, so X is complement in M.

Definition 2 We call a module M is a C_{11}^f*-module*, if every f-closed submodule of M has a complement which is a direct summand of M.

The next result which provides a useful characterization of the C_{11}^f-modules will be used repeatedly.

Lemma 3 *A module M is a C_{11}^f-module if and only if for each f-closed submodule L of M, there exists a direct summand P of M such that $L \cap P = 0$ and $P \oplus L \leq_e M$.*

Proof It is straightforward.

We present connections between C_{11}^f-modules and generalization of extending modules in the next result.

Proposition 1 *Assume the following assertions for a module M:*

(i) M is an extending module,
(ii) M is a C_{11}-module,
(iii) M is a FI-extending module,
(iv) M is a C_{11}^f-module.

Hence $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv)$. However, these arrows are irreversible, in general.

Proof $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv)$ It can be seen that these implications hold from definitions.

$(iii) \not\Rightarrow (ii) \not\Rightarrow (i)$ It follows from [2, Proposition 1.2].

$(iv) \not\Rightarrow (iii)$ Let $T = \begin{bmatrix} \mathcal{K} & \mathcal{V} \\ 0 & \mathcal{K} \end{bmatrix} = \left\{ \begin{bmatrix} \kappa & \omega \\ 0 & \kappa \end{bmatrix} : \kappa \in \mathcal{K}, \omega \in \mathcal{V} \right\}$ be the ring, where $\mathcal{K}$ is a field and $\mathcal{V}$ is a vector space over $\mathcal{K}$ with dimension ≥ 2. Note that T is a commutative indecomposable ring, so all T-submodules of T is fully invariant. Thus, T has no proper f-closed submodule. Hence, T_T is a C_{11}^f-module. Nevertheless, T_T is not FI-extending.

Lemma 4 *Let M be a nonsingular module. Then M is a C_{11}^f-module if and only if M is an FI-extending.*

Proof Suppose M is a nonsingular C_{11}^{f}-module and $L \trianglelefteq M$. Thus, there is a complement submodule T in M such that $L \leq_e T$. Since M is nonsingular, $T \trianglelefteq M$ by [7, Proposition 4.101]. It follows from [7, Lemma 5.58 (ii)] that M/T is nonsingular. So, T is an f-closed submodule of M. Then, there exists a direct summand P of M such that $T \cap P = 0$ and $T \oplus P \leq_e M$ by Lemma 3. Then, $L \cap P = 0$ and $L \oplus P \leq_e M$. By [2, Lemma 1.1], M is FI-extending. Proposition 1 yields the converse.

The C_{11}^{f}-module property may not transfer to the submodules, in general. Let $R = \begin{bmatrix} S & S \oplus S \\ 0 & S \end{bmatrix}$, where S is a simple domain which is not a division ring. Then R is a right nonsingular ring. By [1, Example 4.11], R_R is not FI-extending. Therefore it is not a C_{11}^{f}-module by Lemma 4. However, the injective hull of R_R fulfills the C_{11}^{f}-module property. In the next result, we focus on when the C_{11}^{f}-module property inherits by submodules.

Proposition 2 *Assume M is a C_{11}^{f}-module. Then every f-closed submodule of M fulfills C_{11}^{f}-module property.*

Proof Suppose L is an f-closed submodule of M and L' is an f-closed submodule of L. By Lemma 2, L' is an f-closed submodule of M. Thereby, there exists a direct summand P of M such that $L' \cap P = 0$ and $L' \oplus P \leq_e M$. Accordingly, $M = P \oplus P'$ for some submodule P' of M. Since $L \trianglelefteq M$, Lemma 1 gives that $L = (L \cap P) \oplus (L \cap P')$. Observe that $L' \cap (L \cap P) = 0$ and $L \cap (L' \oplus P) = L' \oplus (L \cap P) \leq_e L$. Therefore, L is a C_{11}^{f}-module by Lemma 3.

3 Conclusion

In the current research, we investigate the class of C_{11}^{f}-modules by asking that every f-closed submodule has a complement which is a direct summand rather than every submodule. We explore some module theoretical properties for the former class.

Finally, decomposition results and applications on Abelian groups of C_{11}^{f}-module property can be discussed as a future work.

References

1. Birkenmeier, G. F., Müller, B. J., Rizvi, S. T., (2002). Modules in which every fully invariant submodules essential in a direct summand. Comm. Algebra 30(3), 1395–1415.
2. Birkenmeier, G. F., Tercan, A., (2007). When some complement of a submodule is a summand. Comm. Algebra, 35(2), 597–611.
3. Dung, N. V., Huynh, D. V., Smith, P. F., Wisbauer, R., (1994). Extending Modules, Longman, Harlow.

4. Fuchs, L., (1970). Infinite Abelian Groups I, Academic Press, New York.
5. Goodearl, K. R., (1976). Ring Theory: Nonsingular Rings and Modules, Dekker, New York.
6. Smith, P. F., Tercan A., (1993). Generalizations of CS-modules. Comm. Algebra 21(6), 1809–1847.
7. Tercan, A., Yücel, C. C. (2016). Module theory. Extending Modules and Generalizations, Bassel: Birkhäuser-Springer.

A Solution of Fractional Bio-Chemical Reaction Model by Adomian Decomposition Method

Gunvant A. Birajdar

Abstract This paper focuses on the modeling of the bio-chemical reaction viz anaerobic digestion which is biochemical process of producing biogas which is the biological degradation of biomass. This chemical phenomenon forms as an system of fractional differential equations. Therefore, the attempt has been made to model this bio-medical process and to find its solution by using powerful Adomian decomposition method. For this Caputo fractional operator is used to represent the fractional derivative.

Keywords Anaerobic · System of fractional equations · Caputo fractional derivative

1 Introduction

Nowadays, the application of fractional calculus are found in every branch like physics, bio-chemistry, viscoelasticity and engineering etc. Along with this many more real world problems and its applications exhibits by fractional differential equations. Many researchers have discussed such problems and its application [1, 2]. Because of this number of researchers & scientists attracted towards this field. The importance of such important biological process energised the researchers to build simple, coherent and high accurate methods for solving such mathematical models which governs differential equation or system of differential equations of fractional order. Djouad et al. [3] gave a light on chemical kinetics for gas phase. It is observed that Jajarmi et al. [4] formulated a model for the dengue fever which is in the forms of system of differential equation of fractional order. Baleanu et al. [5] developed a fractional model by using nonsingular derivative operator. In biochemical process Anaerobic digestion is one of the important process which are having many applications. This is an activity of generating biogas which is the biological

G. A. Birajdar (✉)
School of Rural Development, Tata Institute of Social Sciences, 413601 Tuljapur, Osmanabad, MS, India

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*, Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_17

decomposition of biomass [6–8]. In this biochemical activity involving multiple and complex stages of metabolic interactions, presented by a group of microbial populations in the absent of oxygen. Further, this can be divided into four stages of biodegration viz mehtanogenesis , acidogenisis, hydrolysis and acitogenesis [9–12]. The mathematical formation of biochemical activities is obtained by the number of chemical reactions carry out in every phase of the process. Such process governs as set of coupled nonlinear fractional differential equations. The solution of such system of equations permits exactly to predict the concentration of chemical species at random time using the initial conditions. So, it becomes challenging task to obtain the solution of such initial value problem. There are the known and popular classical technics are available in literature viz iterative method [13], numerical methods [14–19],and Adomian decomposition method (ADM) [17, 20–23]. In 1980 Gorge Adomian introduced the method called ADM [20, 21] which is useful for obtaining the solution of nonlinear equations. Wazwaz [24] discussed the application of this method by solving varieties of nonlinear differential equations. Dhaigude and Birajdar [25] developed the discrete ADM to obtain the solution of system of fractional partial differential equations together with initial condition.

1.1 Preliminaries

This section is devoted for the main definitions and its important properties of Riemann-Liouville (R-L) fractional integral and derivatives as well as the its relation with Caputo fractional derivative.

Definition 1 [1] A real valued function $g(y)$, $y > 0$ is said to be in space C_α, $\alpha \in \Re$ if there exists a real number $p > \beta$ such that $g(y) = y^p g_1(y)$ where $g_1(y) \in C[0, \infty)$.

Definition 2 [1] A function $g(y)$, $y > 0$ is called to be in space C^m_β, $m \in N \bigcup \{0\}$ if $g^m \in C_\beta$.

Definition 3 [1] Consider $g \in C_\beta$ and $\beta \geq -1$, then R-L fractional integral of $g(y)$ with respect to y of order β is denoted by $J^\beta g(y)$ and is defined as

$$J^\beta g(y) = \frac{1}{\Gamma(\beta)} \int_0^y (y-\tau)^{\beta-1} f(\tau) d\tau, \quad t > 0,\ \beta > 0.$$

The well known property [1]of the Riemann-Liouville operator J^β is

$$J^\beta y^\gamma = \frac{\Gamma(\gamma+1) y^{\gamma+\beta}}{\Gamma(\gamma+\beta+1)}.$$

Definition 4 [26] For m to be the smallest integer that exceeds $\beta > 0$, the Caputo fractional derivative of $g(y)$ with respect to y of order $\beta > 0$ is defined as

$$D_y^{\beta} g(y) = \begin{cases} \frac{1}{\Gamma(m-\beta)} \int\limits_0^y (y-\tau)^{m-\beta-1}\, g^m d\tau, & \text{for } m-1 < \beta < m; \\ g^m(y), & \text{for} \beta = m \in N. \end{cases}$$

Note that the relation between Riemann-Liouville operator and Caputo fractional differential operator is given as follows

$$J^{\beta}(D_y^{\beta} g(y)) = J^{\beta}(J^{m-\beta} g^{(m)}(y)) = J^m f^{(m)}(y) = g(y) - \sum_{k=0}^{m-1} g^{(k)}(0) \frac{g^k}{k!}.$$

1.2 Modeling of Metabolic Process

The activities of the anaerobic digestion are:
Hydrolysis: The initial phase of degradation, in which composite organic molecules viz fats, carbohydrates and proteins decompose into soluble monomers form. The enzymes secreted during the reaction are from fermentative bacterias and hydrolytics like lipase, cellulase and protease. The organic raw is braked into a sugar (glucose) during hydrolysis reaction which can exhibit by following equation

$$C_6H_{10}O_5 + H_2O = C_6H_{12}O_6. \tag{1}$$

Acidogenesis: In the process of Acidogenesis sugars are fermented in to simple organic compounds especially ketones (e.g.glycerol, acetone), alcohols (e.g. ethanol, methanol), and acids (e.g. propionic, formic, lactic, butyric, or succinic acids). We can see in the following an example of product gained on acidogenesis activity as well as its corresponding value of ΔG

$$C_6H_{12}O_6 \rightarrowtail C_4H_8O_2 + 2CO_2 + 2H_2, \qquad \Delta G = -264.19 \text{ KJ/mol}. \tag{2}$$

Acetogenesis: The third steps in this process is Acetogenesis resulted in a mixture of hydogen (H_2), acetate, and carbon dioxide (CO_2) during the fermentation. From the lipid hydrolysis the long chain fatty acids are oxidized to propionate or acetate and formes gaseous hydrogen which represents the following reaction equation.

$$2C_4H_8O_2 + 2H_2O + CO_2 \rightarrowtail 4C_2H_4O_2 + CH_4, \qquad \Delta G = -35 \text{ KJ/mol}. \tag{3}$$

Methanogenesis: The final stage in this process is Methanogensis carried out by methanogen micooraganisms where CO_2 and methane are produced. In this activity the methanogenic archaea mainly converts CO_2, acetic acid and hydrogen into methane. It is classified into two groups [12].

Acetoclastic methanogenesis they generate Methane from methanol or acetic acid. These are the prime and important microorganisms in anaerobic digestion, which produces methane approximately 60–70%. Hydrogenotrophic methanogenesis process they uses hydrogen as reducing agent and produce methane from hydrogen and CO_2 using CO_2 as origin of carbon. The chemical reaction is shown in Table 1 with its ΔG value of each reaction. The theoretical yield of CH_4 and CO_2 can be obtained if the substrate composition is known and is given by

$$\mathrm{C}_n\mathrm{H}_p O_q + \left[n - \frac{p}{4} - \frac{q}{2}\right]\mathrm{H}_2 O \rightarrowtail \left[n - \frac{p}{4} - \frac{q}{2}\right]\mathrm{H}_2. \tag{4}$$

where $\mathrm{C}_n + \mathrm{H}_p O_q$ is organic matter and n, p and q are dimensionless coefficients.

2 Mathematical Formulation of the Bio-chemical Process

For the formulation of any chemical reaction the stoichiometric equation, we have

$$\sum_{i=1}^{M_a} u_i v_i = 0, \tag{5}$$

where v_i is the stoichiometric coefficient of i-th species u_i and M_s is the number of species. In general the constant v_j are $-$ve and +ve for reagents and products respectively by convention. Applying the law of mass action the rate of reaction, we have

$$s_j = l_j \prod_{i}^{M_a} u_i^{w_{ji}}, \tag{6}$$

where u_j is the molar concentration of species i and l_j the rate coefficients that can be calculated using the Gibbs free energy (ΔG) of each reaction

$$l_j = \exp\left(\frac{-\Delta G}{RT_j}\right), \tag{7}$$

Table 1 Set of chemical reactions of the anaerobic digestion process

Phases	Reaction	Rates
I	$C_6H_{10}O_5 + H_2O = C_6H_{12}O_6$	$s_0 = l_0[C_6H_{10}O_5][H_2O]$
II	$C_6H_{12}O_6 = C_4H_8O_2 + 2CO_2 + 2H_2$	$s_1 = l_1[C_6H_{10}O_6]$
III	$C_4H_8O_2 + H_2O + \frac{1}{2}CO_2 = 2C_2H_4O_2 + \frac{1}{2}CH_4$	$s_2 = l_2[C_4H_8O_2][H_2O][CO_2]^{\frac{1}{2}}$
IV	$\frac{1}{2}CO_2 + 2H_2 = \frac{1}{2}CH_4 + H_2O$	$s_3 = [CO_2]^{\frac{1}{2}}[H_2]^2$
V	$2C_2H_4O_2 = 2CH_4 + 2CO_2$	$s_4 = l_4 = [C_2H_4O_2]^2$

where $R = 8.3144$ J/Kmol is the universal gas constant and T_j is the absolute temp (Kelvins). We obtain the fractional system of differential equations (FSDEs) is written as

$$D_t^\alpha u_j(t) = \sum_{j}^{M_b} v_{ji} s_j,\ j = 1, 2, 3, \cdots, M_a. \tag{8}$$

We can observe that the system (8) is nonlinear in general. Every species take part in reaction with corresponding production rate. The above fractional system having the different reaction stages: (A) Hydrolysis, (B) Acidogenesis, (C) Acetogenesis, (D) Hydrogenotrophic Methanogenesis, and (E) Acetoclastic Methanogenesis. The above Table 1 provides the different phases of reaction, viz (A), (B), (C) and (D) which can be written in system.

Table 2 Chemical compounds, chemical formulas and abbreviations

	Chemical compounds	Chemical formulas	Abbreviations (u_i)
1	Cellulose	$C_6H_{10}O_5$	u_1
2	Glucose	$C_6H_{12}O_6$	u_2
3	Butyric acid	$C_4H_8O_2$	u_3
4	Acetic acid	$C_2H_4O_2$	u_4
5	Methane	CH_4	u_5
6	Carbon dioxide	CO_2	u_6
7	Hydrogen	H_2	u_7
8	Water	H_2O	u_8

In the Table 2, we can observe the chemical compounds are the parts of anaerobic digestion process and which abbreviated in chemical formula.

The concentration variations u_i(i = 1, 2, 3, ..., 8) are based on system (8). So the system of fractional differential equations is composed of eight fractional FDEs together with initial conditions as fallow:

$$\begin{cases} D_t^{\alpha} u_1(t) = -s_0 u_1 u_8, & u_1(0) = 1, \\ D_t^{\alpha} u_2(t) = -s_0 u_1 u_8 - s_1 u_2, & u_2(0) = 0, \\ D_t^{\alpha} u_3(t) = s_1 u_2 - s_2 u_3 u_8 u_6^{\frac{1}{2}}, & u_3(0) = 0, \\ D_t^{\alpha} u_3(t) = s_1 u_2 - s_2 u_3 u_8 u_6^{\frac{1}{2}}, & u_3(0) = 0, \\ D_t^{\alpha} u_4(t) = 2 s_2 u_3 u_8 u_6^{\frac{1}{2}} - 2 s_4 u_4^2, & u_4(0) = 0, \\ D_t^{\alpha} u_5(t) = s_2 u_3 u_8 u_6^{\frac{1}{2}} + \frac{1}{2} s_3 u_6^{\frac{1}{2}} u_7^2 + 2 s_4 u_4^2, & u_5(0) = 0, \\ D_t^{\alpha} u_6(t) = 2 s_1 u_2 - s_2 u_3 u_8 u_6^{\frac{1}{2}} - \frac{1}{2} s_3 u_6^{\frac{1}{2}} u_7^2 + 2 s_4 u_4^2 & u_6(0) = 0 \\ D_t^{\alpha} u_7(t) = 2 s_1 u_2 - 2 s_3 u_6^{\frac{1}{2}} u_7^2, & u_7(0) = 1 \\ D_t^{\alpha} u_8(t) = -s_0 u_1 u_8 - s_2 u_3 u_8 u_6^{\frac{1}{2}} + s_3 u_6^{\frac{1}{2}} u_7^2, & u_8(0) = 1. \end{cases} \tag{9}$$

3 Adomian Decomposition Method

Consider the system of fractional differential equations with initial conditions of the following form

$$\begin{cases} D_t^{\alpha} u_1(t) = N_1(t, u_1, u_2, ..., u_n) = g_1(t), & u_1(0) = u_{1,0}, \\ D_t^{\alpha} u_2(t) = N_2(t, u_1, u_2, ..., u_n) = g_2(t), & u_2(0) = u_{2,0}, \\ ..., & ... \\ ..., & ... \\ D_t^{\alpha} u_n(t) = N_n(t, u_1, u_2, ..., u_n) = g_n(t), & u_n(0) = u_{n,0}, \end{cases} \tag{10}$$

where $N_k(t, u_1, u_2, ..., u_n)$, $k = 1, 2, 3, ...n$ are linear and nonlinear function. Applying the operator J^{α} (10) we get,

$$\begin{cases} u_1(t) = u_1(0) - J^{\alpha}\{N_1(t, u_1, u_2, ..., u_n) + g_1(t)\}, \\ u_2(t) = u_2(0) - J^{\alpha}\{N_2(t, u_1, u_2, ..., u_n) + g_2(t)\}, \\ ..., \\ ..., \\ u_n(t) = u_n(0) - J^{\alpha}\{N_n(t, u_1, u_2, ..., u_n) + g_n(t)\}. \end{cases} \tag{11}$$

As per the procedure of the Adomian decomposition method the linear terms u_n and nonlinear terms N_n can be decomposed by an infinite series of components such as

$$u_1(t) = \sum_{n=0}^{\infty} u_{1,n}(t), u_2(t) = \sum_{n=0}^{\infty} u_{2,n}(t), ..., u_n(t) = \sum_{n=0}^{\infty} u_{n,n}(t), \tag{12}$$

and the nonlinear functions

$$\begin{cases} N_1(t, u_1, u_2, ..., u_n) & = \sum_{n=0}^{\infty} A_n, \\ N_2(t, u_1, u_2, ..., u_n) & = \sum_{n=0}^{\infty} B_n, \\ ..., \\ ..., \\ N_n(t, u_1, u_2, ..., u_n) & = \sum_{n=0}^{\infty} Z_n, \end{cases} \tag{13}$$

respectively. Note that $u_1(t), u_2(t), ..., u_n(t)(n \geq 0)$ are the approximations of $u_1(t), u_2(t), ..., u_n(t)$, where $A_n, B_n, ..., Z_n$ are the Adomian polynomial generated according to nonlinearity. In general the Adomian polynomial is defined as

$$P_n = \frac{1}{n!}\left[\frac{d^n}{d^n} g\left(\sum_{k=0}^{\infty} \lambda^k v_k(t)\right)\right]_{\lambda=0}, n \geq 0. \tag{14}$$

Substituting equation (13) in equation (12), we have

$$\begin{cases} u_1(t) = & u_1(0) + J^{\alpha} g_1(t) - J^{\alpha}\left(\sum_{n=0}^{\infty} A_n\right), \\ u_2(t) = & u_2(0) + J^{\alpha} g_2(t) - J^{\alpha}\left(\sum_{n=0}^{\infty} B_n\right), \\ ..., \\ ..., \\ u_n(t) = & u_1(0) + J^{\alpha} g_n(t) - J^{\alpha}\left(\sum_{n=0}^{\infty} Z_n\right). \end{cases} \tag{15}$$

On simplifying above equation (13), we get the following recursive relations as follows

$$\begin{cases} u_{1,0}(t) = & u_1(0) + J^{\alpha} g_1(t) \quad & u_{1,n+1}(t) = -J^{\alpha}(A_n), n \geq 0, \\ u_{2,0}(t) = & u_1(0) + J^{\alpha} g_1(t) \quad & u_{2,n+1}(t) = -J^{\alpha}(B_n), n \geq 0, \\ ..., \\ ..., \\ u_{n,0}(t) = & u_1(0) + J^{\alpha} g_1(t) \quad & u_{n,n+1}(t) = -J^{\alpha}(Z_n), n \geq 0. \end{cases} \tag{16}$$

Therefore, the solution for fractional system of equations is given by

$$\begin{cases} u_1(t) = & u_{1,1}(t) + u_{1,2}(t) + ... + u_{1,n}(t), \\ u_2(t) = & u_{2,1}(t) + u_{2,2}(t) + ... + u_{2,n}(t), \\ ..., \\ ..., \\ u_n(t) = & u_{n,1}(t) + u_{n,2}(t) + ... + u_{n,n}(t). \end{cases} \tag{17}$$

Solution of the Problem

In this section we are giving the solution of above fractional system of equations up to two terms

Equations 1

$$\begin{aligned} A_0 =& - k_0 u_{1,0} u_{8,0}, \\ u_{1,1} =& J^{\alpha}(A_0 dt) = A_0 \frac{t^{\alpha}}{\Gamma(\alpha + 1)}. \end{aligned}$$

Equations 2

$$\begin{aligned} B_0 =& - s_0 u_{1,0} u_{8,0}, \\ u_{2,1} =& J^{\alpha}(B_0 dt) - J^{\alpha}(s_1 u_{2,0}) = (B_0 - s_1 u_{2,0}) \frac{t^{\alpha}}{\Gamma(\alpha + 1)}. \end{aligned}$$

Equations 3

$$\begin{aligned} C_0 =& - s_2 u_{3,0} u_{8,0} \sqrt{u_{6,0}}, \\ u_{3,1} =& J^{\alpha}(C_0 dt) - J^{\alpha}(s_1 u_{3,0}) = (C_0 + s_1 u_{3,0}) \frac{t^{\alpha}}{\Gamma(\alpha + 1)}. \end{aligned}$$

Equations 4

$$\begin{aligned} D_0 =& 2 s_2 u_{3,0} u_{8,0} \sqrt{u_{6,0}} - 2 s_4 u_{4,0}^2, \\ u_{4,1} =& J^{\alpha}(C_0 dt) - J^{\alpha}(s_1 u_{3,0}) = (C_0 + s_1 u_{3,0}) \frac{t^{\alpha}}{\Gamma(\alpha + 1)}. \end{aligned}$$

Equations 5

$$\begin{aligned} E_0 =& \frac{1}{2} s_2 u_{3,0} u_{8,0} \sqrt{u_{6,0}} - \frac{1}{2} s_3 \sqrt{u_{6,0}} u_{7,0}^2 + 2 s_4 u_{4,0}^2, \\ u_{5,1} =& J^{\alpha}(E_0 dt) = (E_0) \frac{t^{\alpha}}{\Gamma(\alpha + 1)}. \end{aligned}$$

Equations 6

$$F_0 = \frac{-1}{2} s_2 u_{3,0} u_{8,0} \sqrt{u_{6,0}} - \frac{1}{2} s_3 \sqrt{u_{6,0}} u_{7,0}^2 + 2 s_4 u_{4,0}^2,$$
$$u_{6,1} = J^{\alpha}\{(F_0 + 2s_1 u_{2,0} dt)\} = (F_0 + 2s_1 u_{2,0}) \frac{t^{\alpha}}{\Gamma(\alpha+1)}.$$

Equations 7

$$G_0 = -2 s_3 \sqrt{u_{6,0}} u_{7,0}^2,$$
$$u_{7,1} = J^{\alpha}\{(G_0 + 2s_1 u_{2,0} dt)\} = (G_0 + 2s_1 u_{2,0}) \frac{t^{\alpha}}{\Gamma(\alpha+1)}.$$

Equations 8

$$H_0 = - s_0 u_{1,0} u_{8,0} - s_2 u_{3,0} u_{8,0} \sqrt{u_{6,0}} + s_3 \sqrt{u_{6,0}} u_{7,0}^2,$$
$$u_{8,1} = J^{\alpha}\{H_0 dt\} = (H_0) \frac{t^{\alpha}}{\Gamma(\alpha+1)}.$$

Then the solution of system by the ADM using initial iterations, we get

For two term	For three terms
$U1 = u_{1,0} + u_{1,1}$	$U1 = u_{1,0} + u_{1,1} + u_{1,2}$
$U2 = u_{2,0} + u_{2,1}$	$U2 = u_{2,0} + u_{2,1} + u_{0,2}$
$U3 = u_{3,0} + u_{3,1}$	$U3 = u_{3,0} + u_{3,1} + u_{3,2}$
$U4 = u_{4,0} + u_{4,1}$	$U4 = u_{4,0} + u_{4,1} + u_{4,2}$
$U5 = u_{5,0} + u_{5,1}$	$U5 = u_{5,0} + u_{5,1} + u_{5,2}$
$U6 = u_{6,0} + u_{6,1}$	$U6 = u_{6,0} + u_{6,1} + u_{6,2}$
$U7 = u_{7,0} + u_{7,1}$	$U7 = u_{7,0} + u_{7,1} + u_{7,2}$
$U8 = u_{8,0} + u_{8,1}$	$U8 = u_{8,0} + u_{8,1} + u_{8,2}.$

Conclusion

The paper is devoted to model the biochemical process where the biogas production process was presented as cellule as a substrate.This is done using the Gibbs free energy value. For obtaing the solution an efficient and Powerful Adomian decomposition method is used.

References

1. Podlubny I.: Fractional Differential Equations, Academic Press, San Diego, (1999).
2. Samko S. G., Kilbas A. A., Marichev O. I.: Fractional Integral and Derivatives: Theory and Applications, Gordon and Breach, Yverdon, (1993).
3. Djouad R., Sportisse B., Audiffren N.: Numerical simulation of aqueous-phase atmospheric models: use of a non-autonomous Rosenbrock method. Atmos. Environ. 36(5), 873–879 (2002).
4. Jajarmi A., Arshad S., Baleanu D.: A new fractional modelling and control strategy for the outbreak of dengue fever. Physica A Stat. Mech. Appl. 535(1), 1–14 (2019).
5. Baleanu D., Jajarmi A., Sajjadi S. S., Mozyrska D.: A new fractional model and optimal control of a tumor-immune surveillance with non-singular derivative operator. Chaos Interdiscip. J. Nonlinear Sci. 29(8), 1–15 (2019).
6. Jajarmi A., Baleanu D., Sajjadi S.S., Asad J. H.: A new feature of the fractional Euler-Lagrange equations for a coupled oscillator using a nonsingular operator approach. Front. Phys. 7(196), 1–9 (2019).
7. McKendry P.: Energy production from biomass (part 1): overview of biomass. Bioresour. Technol. 83(1), 37–46 (2002).
8. Twidell J., Weir T.: Renewable energy resources, 2nd edn. Taylor and Francis, New York (2006)
9. Prokopová Z., Prokop R., Modelling and simulation of dry anaerobic fermentation. In: European Conference on Modelling and Simulation, pp. 200–205 (2010).
10. Silva M. I., De Bortoli A. L.:Sensitivity analysis for verification of an anaerobic digestion model. Int. J. Appl. Comput. Math. 6(38), 1–12 (2020).
11. Silva M. I., Bortoli A. L.De :Development of a Model for the Process of Anaerobic Digestion and Its Solution by the Modified Adomian Decomposition Method. Int. J. Appl. Comput. Math. 7(5), 1–14 (2021).
12. Zieminski K., Frac M.: Methane fermentation process as anaerobic digestion of biomass: Transformations, stages and microorganisms. Afr. J. Biotechnol. 11(18), 4127–4139 (2012).
13. Daftardar-Gejji V., Bhalekar S: Solving multi-term linear and nonlinear diffusion-wave equations of fractional order by Adomian decomposition. Appl. Math. Comput. 202, 113–120 (2008).
14. Birajdar G. A.: Numerical solution of time fractional Navier-Stokes equation by discrete Adomian decomposition method. Nonlinear Eng. 3(1), 1–6 (2014).
15. Birajdar G. A.: An implicit numerical method for semilinear space-time fractional diffusion equation, Walailak J. Sci. & Tech. 12(11), 983–994 (2015).
16. Birajdar G. A.: Stability of nonlinear fractional diffusion equation. Libertas Mathematica (New Series), 36(1), 1–12 (2016).
17. Birajdar G. A.: An implicit finite difference method for semilinear time-fractional diffusion equation. Dyn. Cont. Dis. Impul. Syst. Series A: Math. Anal. 27(3a), 209–217 (2020).
18. Dhaigude D. B., Birajdar G. A.: Numerical solution of system of fractional partial differential equations by discrete Adomian decomposition method. J. Frac. Calc. Appl. 3(12), 1–11 (2012).
19. Dhaigude D. B., Birajdar G. A.: Numerical solution of fractional partial differential equations by discrete Adomian decomposition method. Adv. Appl. Math. Mech. 6(1), 107–119 (2014)
20. Adomian G.: Solving Frontier Problems of Physics: The Decomposition Method. Kluwer, Boston, (1994).
21. Adomian G. :A review of the decomposition method in applied mathematics. J. Math. Anal. Appl. 135, 501–544 (1988).
22. Birajdar G. A.: A New Approach for Nonlinear Fractional Heat Transfer Model by Adomian Decomposition Method, Springer Proceedings in Mathematics & Statistics, vol. 344, 674-678, Springer, Singapore, https://doi.org/10.1007/978-981-33-4646-8_28.
23. Daftardar-Gejji V., Jafari H.: Adomian decomposition: a tool for solving a system of fractional differential equations. J. Math. Anal. Appl. 301, 508–518 (2005).
24. Wazwaz A. M.: Partial Differential Equations and Solitary Waves Theory, Higher Education Press and Springer-Verlag, Beijing and Berlin, (2009).

25. Dhaigude D. B., Birajdar G. A., Nikam V. R.:Adomain decomposition method for fractional Benjamin-Bona-Mahony-Burger's equations. Int. J. Appl. Math. Mech. 8(12), 42–51(2012).
26. Caputo M.: Linear models of dissipition whose Q is almost independent, II, Geo-phys. J. Roy. Astron. 13, 529–5397 (1967).

On Fixed Point Results for Mixed Nonexpansive Mappings

Isa Yildirim

Abstract In this presented paper, we consider the S-iteration process for two mappings which include mappings such as satisfying condition (C) and α-nonexpansive mapping in the framework of convex metric spaces. We also prove the some fixed point theorems on strong convergence and Δ-convergence of the proposed algorithm.

Keywords Iterative process · Convex metric space · Condition (C) · α-nonexpansive mapping.

2000 Mathematics Subject Classification: Primary 47H09 · Secondary 65J15

1 Introduction and Preliminaries

Let (Z, ρ) be a metric space and let V be a nonempty subset of Z. Assume that $P : V \to V$ is a mapping. The set $F(T)$ shows that the set of fixed points of P, that is, $F(P) = \{v \in V : Pv = v\}$. We also suppose that P is a mapping on V.

(i) If $\rho(Pv, Pz) \leq \rho(v, z)$ for all $v, z \in V$, the mapping P is called nonexpansive.

(ii) If $\rho(Pv, Pz) \leq \rho(v, z)$ for all $v \in V$ and $z \in F(P)$, the mapping P is called quasi-nonexpansive.

(iii) If $\rho(v, Pv) \leq \rho(v, z)$ implies $\rho(Pv, Pz) \leq \rho(v, z)$ for all $v, z \in V$, the mapping P said to satisfy condition (C).

(iv) If $\rho(Pv, Pz)^2 \leq \alpha\rho(Pv, z)^2 + \alpha\rho(v, Pz)^2 + (1 - 2\alpha)\rho(v, z)^2$ for all $v, z \in V$ and for some $\alpha < 1$, the mapping P is called α-nonexpansive.

From the above definitions, we know that the condition (C) is stronger than quasi-nonexpansiveness but is weaker than nonexpansiveness [13].

I. Yildirim (✉)
Department of Mathematics, Faculty of Science, Atatürk University, 25240 Erzurum, Turkey
e-mail: isayildirim@atauni.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_18

Definition 1.1 [15] Let (Z, ρ) be a metric space and let $\Omega : Z \times Z \times [0,1] \to Z$ be a mapping. Assume that the mapping satisfies the following inequality

$$\rho\left(v, \Omega\left(z, y; \mu\right)\right) \leq \mu\rho\left(v, z\right) + \left(1-\mu\right)\rho\left(v, y\right).$$

for all $v, z, y \in Z$ and all $\mu \in [0, 1]$. Then the metric space (Z, ρ) is called a convex metric space.

The mapping Ω doesn't need to continuous. But, if the following inequality

$$\rho(\Omega(v, z, \mu), \Omega(v, y, \mu) \leq (1-\mu)\rho(z, y)$$

holds in the convex metric space Z, then the mapping Ω becomes continuous.

Definition 1.2 [12] Let (Z, ρ) be a convex metric space. This space Z is called uniformly convex if for all $v, z, y \in Z$, $\delta > 0$ and $\eta \in (0, 2]$, there exists a $\theta > 0$ such that $\rho\left(\Omega(z, y, \frac{1}{2}), v\right) \leq (1-\theta)\,\delta < \delta$ whenever $\rho(z, v) \leq \delta$, $d(y, v) \leq \delta$ and $d(z, y) \geq \delta\eta$.

We suppose that $\{v_n\}$ is a bounded sequence in Z. The mapping $\zeta(., \{v_n\})$ is defined on Z by

$$\zeta(v, \{v_n\}) = \limsup_{n\to\infty} \rho(v, v_n),\ \ v \in Z.$$

The number $\zeta_V(\{v_n\})$ is the asymptotic radius of the sequence $\{v_n\}$ with respect to $V \subseteq Z$ and it is defined as

$$\zeta_V(\{v_n\}) = \inf_{v\in V} \zeta(v, \{v_n\})$$

and the asymptotic center $C_V(\{v_n\})$ of $\{v_n\}$ with respect to V is the set

$$C_V(\{v_n\}) = \{z \in V : \zeta(z, \{v_n\}) = \zeta_V(\{u_n\})\}.$$

Let $\{v_n\}$ be a sequence in (Z, ρ). If $\rho(v_{n+1}, v) \leq \rho(v_n, v)$ for $v \in V$, the sequence $\{v_n\}$ is Fejer monotone with respect to a subset V of Z. If v is the unique asymptotic center for every subsequence $\{z_n\}$ of $\{v_n\}$, then the sequence $\{v_n\}$ $\triangle$-converges to $v \in Z$. In this case, we write $\triangle - lim_n v_n = v$.

Recently, Takahashi et al. [14] obtained some fixed point results for two nonexpansive mappings such as P and R in Banach space using the following iteration

$$\begin{cases} v_{n+1} = (1-\gamma_n)\,v_n + \gamma_n P z_n \\ \ \ z_n = (1-\delta_n)\,v_n + \delta_n R v_n \end{cases} \tag{1}$$

where $0 < \gamma_n,\ \delta_n < 1$. After then, Dhompongsa et al. [4] used the algorithm (1) to obtain some convergence results for a nonspreading mapping and a mapping satisfying condition (C) in Hilbert spaces (see also [2, 3, 8–11]). Moreover, Wattanawitoon et al. [16] proved some convergence theorems for an α-nonexpansive mapping and a mapping satisfying the condition (C) using the algorithm (1) in Hilbert spaces.

In this paper, we consider S-iteration process for two mappings such as an α-nonexpansive mapping and a mapping satisfying condition (C) in the framework of convex metric spaces. We also show that some convergence theorems using this iteration process in such spaces. Our algorithm is as under

$$\begin{cases} v_{n+1} = \Omega(Pv_n, Rz_n, \gamma_n) \\ z_n = \Omega(v_n, Pv_n, \delta_n) \end{cases} \tag{2}$$

where $0 < \gamma_n, \delta_n < 1$.

If we take a normed space instead of convex metric space, the algorithm (2) reduces to the following algorithm

$$\begin{cases} v_{n+1} = \gamma_n Pv_n + (1-\gamma_n)\, Rz_n \\ z_n = \delta_n v_n + (1-\delta_n)\, Pv_n, \end{cases} \tag{3}$$

where $0 < \gamma_n, \delta_n < 1$.

This iteration process (3) is independent of the iteration process (1), but the process (3) is faster than the Mann, Ishikawa and Noor iteration processes. If we take $P = R$ in (3), we obtain the S-iteration process [1].

Now, we will give some important lemmas that we will use to prove the main results. Throughout the paper, we will take $F = F(P) \cap F(R)$.

Lemma 1.1 *[6] Let (Z, ρ) be a complete and uniformly convex metric space and let V be a nonempty, closed and convex subset Z. Then every bounded sequence $\{v_n\}$ in Z has a unique asymptotic center with respect to V.*

Lemma 1.2 *[7] Let (Z, ρ) be a uniformly convex metric space with continuous convex structure Ω. Assume that $v \in Z$ and $\{a_n\}$ be a sequence in $[t, r]$ for some $t, r \in (0, 1)$. If $\{v_n\}$ and $\{z_n\}$ are sequences in Z such that $limsup_{n\to\infty} d(v_n, v) \leq \delta$, $limsup_{n\to\infty} d(z_n, v) \leq \delta$ and $lim_{n\to\infty} \rho(\Omega(v_n, z_n, a_n), v) = \delta$ for some $\delta \geq 0$, then $lim_{n\to\infty} \rho(v_n, z_n) = 0$.*

Lemma 1.3 *[5] Let (Z, ρ) be a metric space and let V be a subset of a metric space Z. Let P be an α-nonexpansive mapping on V and let R be a mapping on V satisfying condition (C) with $F \neq \emptyset$. Then P and R are quasi-nonexpansive.*

Lemma 1.4 *[5] Let (Z, ρ) be a complete and uniformly convex metric space and let V be a nonempty, closed and convex subset of Z with continuous convex structure Ω. Assume that P is an α-nonexpansive mapping on V and R is a mapping on V satisfying condition (C) such that $F \neq \emptyset$. If $\{z_n\}$ is any bounded sequence in V with $C(\{z_n\}) = \{z\}$ and*

$$\lim_{n\to\infty} d(z_n, Sz_n) = \lim_{n\to\infty} d(z_n, Tz_n) = 0,$$

then $z \in F$.

2 Main Results

Lemma 2.1 *Let (Z, ρ) be a convex metric space and let V be a nonempty, closed and convex subset of Z. Let $P : V \to V$ be an α-nonexpansive mapping and let $R : V \to V$ be a mapping satisfying condition (C) such that $F \neq \emptyset$. Then for the sequence $\{v_n\}$ in (2), we get the followings:*
(1) $\{v_n\}$ is a Fejer monotone sequence with respect to F,
(2) $\lim_{n\to\infty} \rho(v_n, q)$ exists for all $q \in F$,
(3) $\lim_{n\to\infty} \rho(v_n, F)$ exists.

Proof From convex structure of metric space and the iteration process (2), for any $q \in F$, we have

$$\begin{aligned} \rho(z_n, q) &= \rho(\Omega(v_n, Pv_n, \delta_n), q) \\ &\leq \delta_n \rho(v_n, q) + (1-\delta_n)\rho(Pv_n, q) \\ &\leq \delta_n \rho(v_n, q) + (1-\delta_n)\rho(v_n, q) \\ &= \rho(v_n, q) \end{aligned} \tag{4}$$

and

$$\begin{aligned} \rho(u_{n+1}, q) &= \rho(\Omega(Pv_n, Rz_n, \gamma_n), q) \\ &\leq \gamma_n \rho(Pv_n, q) + (1-\gamma_n)\rho(Rz_n, q) \\ &\leq \gamma_n \rho(v_n, q) + (1-\gamma_n)\rho(z_n, q). \end{aligned} \tag{5}$$

If we combine the inequalities (4) and (5), we get

$$\rho(v_{n+1}, q) \leq \gamma_n \rho(v_n, q) + (1-\gamma_n)\rho(z_n, q) = d(v_n, q). \tag{6}$$

By using (6), we obtain that (1) $\{v_n\}$ is a Fejer monotone sequence with respect to F and (2) $\lim_{n\to\infty} \rho(v_n, q)$ exists for each $q \in F$. If we take inf of both sides in (6), we get

$$\inf_{q\in F} \rho(v_{n+1}, q) \leq \inf_{q\in F} \rho(v_n, q)$$

which implies that (3) $\lim_{n\to\infty} \rho(v_n, F)$ exists.

Lemma 2.2 *Let (Z, ρ) be a convex metric space and let V be a nonempty, closed and convex subset of Z. Let $P : V \to V$ be an α-nonexpansive mapping and let $R : V \to V$ be a mapping satisfying condition (C) such that $F \neq \emptyset$. Then for the sequence $\{v_n\}$ in (2), we have the following limits*

$$\lim_{n\to\infty} \rho(v_n, Pv_n) = \lim_{n\to\infty} \rho(v_n, Rv_n) = 0.$$

Proof From (ii) in Lemma 2.1, we know $\lim_{n\to\infty} \rho(v_n, q)$ for $q \in F$. Let's say σ. If $\sigma = 0$, the proof is clear. We suppose that $\sigma > 0$. Since the mapping P satisfies condition (C), the mapping P is quasi-nonexpansive mapping (see Lemma 1.3). That's why we write $\rho(Pz_n, q) \le \rho(z_n, q)$. If we take lim sup of this inequality, we obtain

$$\limsup_{n\to\infty} \rho(Pz_n, q) \le \limsup_{n\to\infty} \rho(z_n, q).$$

Using Lemma 2.1, we have $\rho(z_n, q) \le \rho(v_n, q)$. If we take again lim sup of this inequality, we get

$$\limsup_{n\to\infty} \rho(z_n, q) \le \limsup_{n\to\infty} \rho(v_n, q) = \sigma$$

which implies $\limsup_{n\to\infty} \rho(z_n, q) \le \sigma$. Also,

$$\begin{aligned} \limsup_{n\to\infty} \rho(Pz_n, q) &\le \limsup_{n\to\infty} \rho(z_n, q) \le \sigma \\ &\Rightarrow \limsup_{n\to\infty} \rho(Pz_n, q) \le \sigma. \end{aligned}$$

Since R is an α-nonexpansive, we have

$$\begin{aligned} \rho(Rv_n, q) &\le \rho(v_n, q) \\ &\Rightarrow \limsup_{n\to\infty} \rho(Rv_n, q) \le \sigma. \end{aligned} \tag{7}$$

Also

$$\sigma = \lim_{n\to\infty} \rho(v_{n+1}, q) = \lim_{n\to\infty} \rho(\Omega(Pv_n, Rz_n, \gamma_n), q)$$

and from Lemma 1.2

$$\lim_{n\to\infty} \rho(Pv_n, Rz_n) = 0. \tag{8}$$

Now

$$\begin{aligned} \rho(v_{n+1}, q) &\le \rho(\Omega(Pv_n, Rz_n, \gamma_n), q) \\ &\le \gamma_n \rho(Pv_n, q) + (1-\gamma_n)\rho(Rz_n, q) \\ &\le \gamma_n \rho(Pv_n, q) + (1-\gamma_n)\rho(Pv_n, Rz_n) + (1-\gamma_n)\rho(Pv_n, q) \\ &\le \rho(Pv_n, q) + (1-\gamma_n)\rho(Pv_n, Rz_n) \end{aligned}$$

which implies

$$c \le \lim_{n\to\infty} \inf d(Tx_n, p)$$

so that (7) gives that

$$\lim_{n\to\infty} d(Tx_n, p) = c.$$

In turn,

$$\begin{aligned} d\left(Tx_n, p\right) &\leq d\left(Tx_n, Sy_n\right) + d(Sy_n, p) \\ &\leq d\left(Tx_n, Sy_n\right) + d(y_n, p) \end{aligned}$$

implies

$$\sigma \leq \liminf_{n\to\infty} \rho(z_n, q).$$

Thus $\sigma = \lim_{n\to\infty} \rho(z_n, q) = \lim_{n\to\infty} \rho(\Omega(v_n, Pv_n, \delta_n), q)$ gives by Lemma 1.2

$$\lim_{n\to\infty} \rho\left(v_n, Pv_n\right) = 0. \tag{9}$$

Now

$$\begin{aligned} \rho(z_n, v_n) &= \rho(\Omega(v_n, Pv_n, \delta_n), v_n) \\ &\leq (1-\delta_n)\rho\left(v_n, Pv_n\right) \end{aligned}$$

imply by (9) that

$$\lim_{n\to\infty} \rho(z_n, v_n) = 0. \tag{10}$$

Using (8), (9) and (10), we have

$$\begin{aligned} \rho\left(v_n, Rv_n\right) &\leq \rho\left(v_n, Pv_n\right) + \rho\left(Pv_n, Rz_n\right) + \rho\left(Rz_n, Rv_n\right) \\ &\leq \rho\left(v_n, Pv_n\right) + \rho\left(Pv_n, Rz_n\right) + \rho\left(z_n, v_n\right) \end{aligned}$$

and so

$$\lim_{n\to\infty} \rho\left(v_n, Rv_n\right) = 0.$$

Theorem 2.1 *Let (Z, ρ) be a convex metric space and let V be a nonempty, closed and convex subset of Z. Let $P : V \to V$ be an α-nonexpansive mapping and let $R : V \to V$ be a mapping satisfying condition (C). Moreover, let the sequence $\{v_n\}$ be defined as in (2). If $F \neq \emptyset$, then $\triangle - \lim_n v_n = v \in F$.*

Proof From Lemma 2.1, we know the sequence $\{v_n\}$ is bounded. Therefore the sequence $\{v_n\}$ has a unique asymptotic center, i.e., $C\left(\{v_n\}\right) = \{v\}$. For any subsequence $\{z_n\}$ of $\{v_n\}$, Lemma 1.1 gives that $C\left(\{z_n\}\right) = \{z\}$. Using Lemma 2.2, we have

$$\lim_{n\to\infty} \rho\left(z_n, Pz_n\right) = \lim_{n\to\infty} \rho\left(z_n, Rz_n\right) = 0.$$

From Lemma 1.4, we have $z \in F$. We will show that $v = z$. Let's accept the opposite. Using the uniqueness of asymptotic centers, we have

$$\begin{aligned}\limsup_{n\to\infty} \rho\left(z_n, z\right) &< \limsup_{n\to\infty} \rho\left(z_n, v\right)\\ &\le \limsup_{n\to\infty} \rho\left(v_n, v\right)\\ &< \limsup_{n\to\infty} \rho\left(v_n, z\right)\\ &= \limsup_{n\to\infty} \rho\left(z_n, z\right).\end{aligned}$$

This is a contradiciton. Therefore $C\left(\{z_n : \{z_n\} \text{ is any subsequences of } \{v_n\}\ \}\right) = \{v\}$ .This requires that $\Delta - \lim_n v_n = v \in F$.

Now, we will give the following concepts in order to obtain some strong convergence results.

A self-mapping $P : V \to V$ is semi-compact if for any bounded sequence $\{v_n\}$ in V with $\rho\left(v_n, Pv_n\right) \to 0$, we must have that $\{v_n\}$ has a convergent subsequence in V.

Let $P : V \to V$ and $R : V \to V$ be two mapping and let F be a nonempty subset F of V. They are said to satisfy condition (II) if there exists a nondecreasing function g on $[0, \infty)$ with $g(0) = 0$ and $g(s) > 0$ for $s \in (0, \infty)$ such that

$$\frac{1}{2}\left[\rho\left(v, Pv\right) + \rho\left(v, Rv\right)\right] \ge g(\rho(v, F))$$

for all $v \in V$.

Finally, we will give the following strong convergence results without proofs.

Corollary 2.1 *Let (Z, ρ) be a convex metric space and let V be a nonempty, closed and convex subset of Z. Let $P : V \to V$ be an α-nonexpansive mapping and let $R : V \to V$ be a mapping satisfying condition (C). Moreover, let the sequence $\{v_n\}$ be defined as in (2). If $F \neq \emptyset$ and either P or R is semi-compact, then $\{v_n\}$ converges strongly to a fixed point.*

Corollary 2.2 *Let (Z, ρ) be a convex metric space and let V be a nonempty, closed and convex subset of Z. Let $P : V \to V$ be an α-nonexpansive mapping and let $R : V \to V$ be a mapping satisfying condition (C). Moreover, let the sequence $\{v_n\}$ be defined as in (2). If $F \neq \emptyset$ and P and R satisfy condition (II), then $\{v_n\}$ converges strongly to a fixed point.*

Conclusion 2.1 In this paper, we showed that some fixed point theorems on strong convergence and Δ-convergence of the iteration process (2) to common fixed point of a mapping satisfying condition (C) and an α- nonexpansive mapping in a convex metric space. Since Hyperbolic spaces, $CAT(0)$ spaces and Hilbert spaces are convex metric spaces, therefore our results also hold in such spaces. Furthermore, our results generalize the corresponding ones in [2, 7, 9] etc.

References

1. Agarwal R.P, O'Regan D., Sahu D.R.: Iterative construction of fixed points of nearly asymptotically nonexpansive mappings, J. Nonlinear Convex Anal.,8(2007), no. 1, 61–79.
2. Akbulut S., Gunduz, B.: Strong and Δ-converges of a faster iteration process in hyperbolic space, Commun. Korean Math. Soc. 30 (2015), No. 3, pp. 209–219.
3. Alimohammady, M., Ramazannejad, M., Bagheric, Z., Shahkoohid, R.J.: Common Zero Points of Two Finite Families of Maximal Monotone Operators Via Iteration Methods, Iranian Journal of Mathematical Sciences and Informatics, 12(2), (2017), 73–99.
4. Dhompongsa, S., Inthakon, W., Takahashi, W.: A Weak Convergence Theorem for Com mon Fixed Points of Some Generalized Nonexpansive Mappings and non-Spreading Mappings in a Hilbert Space, Optimization: A Journal of Mathematical Programming and Operations Research, 60, (2011), 769–779.
5. Fukhar-ud-din, H.: Iterative Process for an α -Nonexpansive Mapping and a Mapping Satisfying Condition (C) in a Convex Metric Space, Iranian Journal of Mathematical Sciences and Informatics, Vol. 14, No. 1 (2019), pp 167–179.
6. Fukhar-ud-din, H.: Existence and Approximation of Fixed Points in Convex Metric Spaces, Carpathian J. Math., 30, (2014), 175–185.
7. Fukhar-ud-din, H.: One Step Iterative Scheme for a Pair of Nonexpansive Mappings in a Convex Metric Space, Hacet. J. Math.Stat., 44, (2015), 1023–1031.
8. Karahan, İ., Jolaoso, L.O.: A three steps iterative process for approximating the fixed points of multivalued generalized α -nonexpansive mappings in uniformly convex hyperbolic spaces, Sigma J Eng & Nat Sci 38 (2), 2020, 1031–1050.
9. Khamsi, M. A.: Approximate Fixed Point Sequences of Nonlinear Semigroups in Metric Spaces, Canad. Math. Bull., 58 (2), (2015), 297–305.
10. Khan, S.H., Yildirim, I., Ozdemir, M.: Convergence of an implicit algorithm for two families of nonexpansive mappings, Computers & Mathematics with Applications 59 (9), 3084–3091.
11. Sahebi, H.R., Razani, A.: An Explicit Viscosity Iterative Algorithm for Finding Fixed Points of Two Noncommutative Nonexpansive Mappings, Iranian Journal of Mathematical Sciences and Informatics, 11(1) (2016), 69–83.
12. Shimizu, T., Takahashi, W.: Fixed Points of Multivalued Mappings in Certain Convex Metric Spaces, Topol. Methods Nonlinear Anal., 8, (1996), 197–203.
13. Suzuki, T.: Fixed Point Theorems and Convergence Theorems for Some Generalized Nonexpansive Mappings, J. Math. Anal. Appl., 341, (2008), 1088–1095.
14. Takahashi, W., Tamura, T.: Convergence Theorems for a Pair of Nonexpansive Mappings, J. Convex Anal., 5, (1998), 45–56
15. Takahashi, W.: A convexity in metric space and nonexpansive mappings, Kodai Math. Sem. Rep, 22 (1970) 142–149.
16. Wattanawitoon, K., Khamlae, Y.: Weak and Strong Convergence Theorems for an α-Nonexpansive Mapping and a Generalized Nonexpansive Mapping in Hilbert Spaces, Thai J. Math., 11, (2013), 633–643.

Finite Element Approximation of Eigenvibration of a Coupled Vibro-Acoustic System Motivated by Phonation into Tubes

Jan Valášek, Petr Sváček, and Jaromir Horáček

Abstract The mathematical model of the vibro-acoustic problem representing human phonation into tube is formulated. It consists of acoustic and structural problem description as well as their mutual coupling. Here, the vocal folds vibrations are modelled using linear elasticity theory and the Helmholtz equation is used for frequency characterization of acoustic waves propagation in the vocal tract model. The both subproblems are numerically approximated by finite element method. The preliminary results compare the acoustic eigenfrequencies of vocal tract with tube and the eigenfrequencies of the coupled vibro-acoustic system. The eigenmodes with significant acoustic contribution are identified. The first eigenfrequency of coupled system with prevailing acoustic part is substantially increased indicating strong interaction with the elastic structure.

Keywords Finite element method · Vibroacoustics · Helmholtz equation · Eigenfrequencies of coupled system.

1 Introduction

Phonation into tubes of various dimensions is a popular technique used by voice professionals for voice training and therapy purposes, see [1]. The favourable effect of the exercise can be explained by lowering the needed power for vocal fold (VF) phonation process enabled by advantageous interaction of structural vibrations with the propagating acoustic waves, see [2]. This interaction is particularly strong for the case when the prolongation of the vocal tract (VT) decreases the first VT acoustic resonance to the vicinity of the first eigenfrequency of VF vibration. Then it was measured as well as computed— see [1, 2], that the first acoustic resonance frequency

J. Valášek (✉) · P. Sváček
Faculty of Mechanical Engineering, CTU in Prague,, Czech Republic Prague
e-mail: Jan.Valasek@fs.cvut.cz

J. Horáček
Institute of Thermomechanics, Czech Academy of Sciences,, Czech Republic Prague

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_19

of coupled system is essentially increased due to the aforementioned interaction. In that case a more complex model of vocal tract acoustics coupled to vocal folds vibrations needs to be considered what motivates our study of the vibro-acoustic problem.

Articles [1] and [2] employ a simple 1D model of transfer matrix approach coupled with a single degree-of-freedom model of elastic body. Such model can not fully describe spatial shape of the VF vibration eigenmode and its coupling contribution to the acoustic problem. Therefore we present here the 2D model based on the coupled partial differential equations and its numerical solution as one of the standard approaches to coupled vibro-acoustic problem, see e.g. books [3] or [4]. Such 2D coupled model can reveal us additional information about spatial distribution of coupled eigenmodes, etc.

In this article the vocal folds vibrations are modelled as linear isotropic elastic body and the Helmholtz equation is used for the frequency characterization of acoustic waves propagation in the vocal tract model. The standard coupling conditions on the common interface are considered, see [3]. The both subproblems are numerically approximated by finite element method (FEM). The final coupled system of equations in frequency domain represents a generalized eigenvalue problem and its solution reduces to finding a subset of the lowest eigenfrequencies and corresponding eigenmodes.

The preliminary results compare the acoustic resonance frequencies of vocal tract alone and the resonance frequencies of the coupled vibro-acoustic system. The first eigenfrequency of coupled system with the dominance of acoustic part is substantially increased compared to solely acoustic model. The associated eigenmode shape is shown.

The paper outline is following: In the next section the mathematical model of vibro-acoustic model is described. Then the finite element approximation is given. In the end the preliminary results of eigenfrequencies of coupled system are presented.

2 Mathematical Model

Let us consider a two-dimensional vibro-acoustic problem in domain Ω consisting of elastic structure domain Ω^s (vocal folds) and acoustic domain Ω^a. The acoustic domain models human vocal tract with length L_1 and thin tube inserted into mouth of length L_2 and diameter d_2, see Fig. 1. The boundary of domain Ω is composed of mutually disjoint parts: Γ^a_{Dir}, Γ^s_{Dir} and Γ^a_{Neu}. Moreover by Γ_{W} is denoted the common interface between domains Ω^s and Ω^a, see Fig. 1.

Let us consider the problem description in frequency domain, i.e. we assume that all involved quantities depend on the spatial coordinates x and angular frequency ω, e.g. $\hat{p}(x, \omega)$, if it is not mentioned differently.

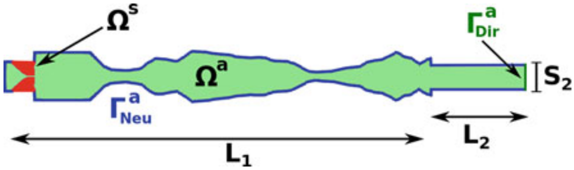


Fig. 1 Scheme of structure domain Ω^s and acoustic domain Ω^a together with marked boundaries of $\partial\Omega^a$

2.1 Elastic Structure

The elastic structure displacement $\hat{\mathbf{u}}(x, \omega) = (\hat{u}_1, \hat{u}_2)$ is modelled by

$$\omega^2 \rho^s \hat{u}_i + \frac{\partial \hat{\tau}^s_{ij}}{\partial x_j} = 0, \qquad \text{in }^s, \quad (\mathrm{i} = 1, 2), \tag{1}$$

where ρ^s is the structure density and $\hat{\tau}^s_{ij}$ denote the components of the Cauchy stress tensor. Under assumption of isotropic body the stress tensor components can be expressed with help of the Hooke's law as

$$\hat{\tau}^s_{ij} = \lambda^s \mathrm{div}\, \hat{\mathbf{u}}\, \delta_{ij} + 2\mu^s \hat{e}^s_{ij}(\hat{\mathbf{u}}), \tag{2}$$

where $\mathbb{I} = (\delta_{ij})$ is Kronecker's delta, $\hat{e}^s_{ij}(\mathbf{u}) = \frac{1}{2}\left(\frac{\partial \hat{u}_j}{\partial x_i} + \frac{\partial \hat{u}_i}{\partial x_j}\right)$ denotes the small strain tensor and parameters λ^s, μ^s are the Lamé coefficients, see e.g. [5]. The elastic body is firmly clapped at the bottom given by Dirichlet boundary condition

$$\hat{\mathbf{u}}(x, \omega) = \mathbf{0}, \qquad \text{for } x \in \Gamma^s_{\mathrm{Dir}}. \tag{3}$$

2.2 Acoustics

The sound propagation through homogeneous medium at rest is described in frequency domain by the Helmholtz equation for acoustic pressure $\hat{p}(x, \omega)$, see [6],

$$-\frac{\omega^2}{c_0^2}\hat{p} - \Delta \hat{p} = \hat{f}^a(x, \omega), \qquad \text{in } \Omega^a, \tag{4}$$

where c_0 is the speed of sound and function $\hat{f}^a$ describes possible generic sound sources. The following boundary conditions are considered

$$\begin{aligned} &(\mathrm{a}) \qquad \hat{p}(x, \omega) = 0, \qquad &&\text{for } x \in \Gamma^a_{\mathrm{Neu}}, \\ &(\mathrm{b}) \qquad \frac{\partial \hat{p}}{\partial \mathbf{n}}(x, \omega) = 0, \qquad &&\text{for } x \in \Gamma^a_{\mathrm{Dir}}, \end{aligned} \tag{5}$$

where vector $\mathbf{n} = (n_j)$ is unit outer normal to $\partial \Omega^a$. Condition (5a) is so called sound soft boundary condition which models the free end of tube. The second condition (5b) is the sound hard condition and it represents fully reflecting walls, see [6].

2.3 Vibro-Acoustic Coupling

The vibro-acoustic problem, represented by equations (1) and (4), is coupled by boundary conditions on the common interface Γ_W. The boundary condition for acoustic pressure $\hat{p}$ has the form, see e.g. [4],

$$\frac{\partial \hat{p}}{\partial \mathbf{n}}(x, \omega) = \rho^a \omega^2 \hat{\mathbf{u}}(x, t) \cdot \mathbf{n}, \qquad x \in \Gamma_W. \tag{6}$$

where ρ^a is the density of air. This boundary condition can be understood as the acoustic emission given by normal acceleration of vibrating surface Γ_W.

The boundary condition prescribed for elastic body is based on the stress continuity in normal direction and it reads

$$\hat{\tau}^s_{ij}(x, \omega)\, n_j = \hat{p}(x, \omega)\, n_i, \qquad x \in \Gamma_W, \tag{7}$$

where $\mathbf{n} = (n_j)$ is unit outer normal to Γ_W pointing from Ω^s to Ω^a.

3 Numerical Modelling

The FEM is used for spatial discretization of both subproblems (1) and (4).

3.1 Elastic Structure

The standard weak formulation of (1) together with considered boundary conditions (3) and (7) leads to problem – find $\hat{\mathbf{u}} \in \mathbf{V}$ such that

$$-\omega^2 \left(\rho^s \hat{\mathbf{u}}, \boldsymbol{\psi}\right)_{\Omega^s} + \left(\lambda^s (\operatorname{div} \hat{\mathbf{u}})\, \mathbb{I} + 2\mu^s \mathbf{e}^s(\hat{\mathbf{u}}), \mathbf{e}^s(\boldsymbol{\psi})\right)_{\Omega^s} = -\left(\hat{p}\mathbf{n}, \boldsymbol{\psi}\right)_{\Gamma_W}, \tag{8}$$

holds for all $\boldsymbol{\psi} \in \mathbf{V} = V \times V$, $V = \{f \in H^1(\Omega^s) | f = 0 \text{ on } \Gamma^s_{\mathrm{Dir}}\}$. The scalar product of functions from $L^2(\mathcal{D})$ is denoted by $(\cdot, \cdot)_{\mathcal{D}}$.

Finite element approximation can be expressed as a linear combination of basis functions $\boldsymbol{\psi}_i$ from FE space $\mathbf{V}_h$, i.e. $\mathbf{u}_h = \sum_{i=1}^{2N_h} \hat{\alpha}^s_i \boldsymbol{\psi}_i(x)$. It leads to linear algebraic system of equations for unknown vector $\hat{\boldsymbol{\alpha}}^S = (\hat{\alpha}^s_i)$ for given $\omega \in \mathbb{R}$

$$-\omega^2 \mathbb{M}^s \hat{\boldsymbol{\alpha}}^s + \mathbb{K}^s \hat{\boldsymbol{\alpha}}^s + \mathbb{C}^a \hat{\boldsymbol{\alpha}}^a = \mathbf{0}, \tag{9}$$

where vector $\hat{\boldsymbol{\alpha}}^a$ denotes unknowns of acoustic part of the problem. The elements of matrices $\mathbb{M}^s = (m^s_{ij})$, $\mathbb{K}^s = (k^s_{ij})$ and $\mathbb{C}^a = (c^a_{ij})$ are given by

$$m^s_{ij} = (\rho^s \boldsymbol{\psi}_j, \boldsymbol{\psi}_i)_{\Omega^s}, \quad k^s_{ij} = (\lambda^s (\operatorname{div} \boldsymbol{\psi}_j)\, \delta_{ij} + 2\mu^s \mathbf{e}^s(\boldsymbol{\psi}_j), \mathbf{e}^s(\boldsymbol{\psi}_i))_{\Omega^s}, \tag{10}$$
$$c^a_{ij} = \left(\eta_i \, \mathbf{n}, \; \boldsymbol{\psi}_j\right)_{\Gamma_{\mathrm{W}}},$$

where η_i denotes FE basis functions of the acoustic FE approximation space Y_h.

3.2 Acoustics

The weak formulation of problem (4) together with conditions (5) and (6) in functional space $Y = \{f \in H^1(\Omega^a) | f = 0 \text{ on } {}^{\mathrm{a}}_{\mathrm{Dir}}\}$ reads: find $\hat{p} \in Y$ that

$$-\omega^2 \left(\frac{1}{c^2}\hat{p}, \eta\right)_{\Omega^a} + \left(\nabla \hat{p}, \nabla \eta\right)_{\Omega^a} + \omega^2 \left(\rho^a \hat{\mathbf{u}} \cdot \mathbf{n}, \eta\right)_{\Gamma_{\mathrm{W}}} = \left(\hat{f}^a, \eta\right)_{\Gamma_{\mathrm{W}}} \tag{11}$$

is satisfied for any $\eta \in Y$. The same discretization procedure by the FEM yields

$$-\omega^2 \mathbb{M}^a \hat{\boldsymbol{\alpha}}^a + \mathbb{K}^a \hat{\boldsymbol{\alpha}}^a + \omega^2 \mathbb{C}^s \hat{\boldsymbol{\alpha}}^s = \mathbf{b}^a(\omega), \tag{12}$$

where matrices $\mathbb{M}^a$ and $\mathbb{K}^a$ are the mass and stiffness matrices, respectively, and the components of right hand side vector $\mathbf{b}^a(\omega) = (b^a_j)$ and the coupling matrix elements $\mathbb{C}^s = (c^s_{ij})$ are given by

$$b^a_j = \left(\hat{f}^a, \eta_j\right)_{\Omega^a}, \qquad c^s_{ij} = \left(\rho^a \boldsymbol{\psi}_i \cdot \mathbf{n}, \eta_j\right)_{\Gamma_{\mathrm{W}}}. \tag{13}$$

The acoustic and the structure meshes are chosen to be consistent across the interface Γ_{W}. Then no additional treatment during computation of the coupling matrices is needed.

3.3 Numerical Solution of Coupled Vibro-Acoustic Problem

Collecting all terms of (9) and (12) in one system yields

$$\left(-\omega^2 \begin{pmatrix} \mathbb{M}^s & 0 \\ -\mathbb{C}^s & \mathbb{M}^a \end{pmatrix} + \begin{pmatrix} \mathbb{K}^s & \mathbb{C}^a \\ 0 & \mathbb{K}^a \end{pmatrix}\right) \begin{pmatrix} \hat{\boldsymbol{\alpha}}^s \\ \hat{\boldsymbol{\alpha}}^a \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{b}^a \end{pmatrix}, \tag{14}$$

where further vector $\mathbf{b}^a := \mathbf{0}$ is assumed. Then this problem represents the generalized eigenvalue problem and it can be solved by e.g. mathematical library ARPACK.

4 Numerical Experiments

First the eigenfrequencies of separated acoustic and structural model are described. Then the resonance frequencies of the coupled system is computed.

4.1 Acoustic Resonance Frequencies of Vocal Tract Without and with Tube

Vocal tract (VT) model is motivated by MRI measurement of vowel [u:] of [7] where the glottal space is additionally included resulting in $L_1 = 0.1931$ m. Further as in [1] the VT is prolonged by the tube of dimensions $L_2 = 0.264$ m, $d_2 = 6.77$ mm. This geometry setting is in next referred as *VT+tube*. Boundary condition (6) on Γ_W is replaced for the case of sole acoustic problem by sound hard boundary condition (5b). The speed of sound is chosen as $c_0 = 343$ m/s and density is $\rho^a = 1.2\,\mathrm{kg/m}^3$.

The first four computed eigenfrequencies of VT model without and with the tube is listed in Table 1. In our problem setting the VT or VT+tube can be represented very approximately for low frequencies as an acoustic quarter-wave resonator, see [8]. Then inclusion of the tube leading to significant decrease of eigenfrequencies of acoustic system, see Table 1, can be satisfactory explained by the increase of acoustic resonator length ($L \approx L_1$ vs. $L \approx 0.4571$ m for VT+tube).

4.2 Structural Resonance Frequencies of Vocal Fold Model

The vocal fold (VF) model is based on the geometric and material settings published in [9] with the initial glottal gap equal to 2.0 mm. Structural density is chosen as $\rho^s =$

Table 1 Computed eigenfrequencies (in Hz) of the vocal tract models without and with tube prolongation. The eigenfrequencies of vocal fold model is in last column

	VT	VT+tube	Vocal folds
F_1	374.8	139.8	121.2
F_2	1522.4	559.8	216.7
F_3	1866.8	881.0	275.3
F_4	2671.6	1347.2	388.1

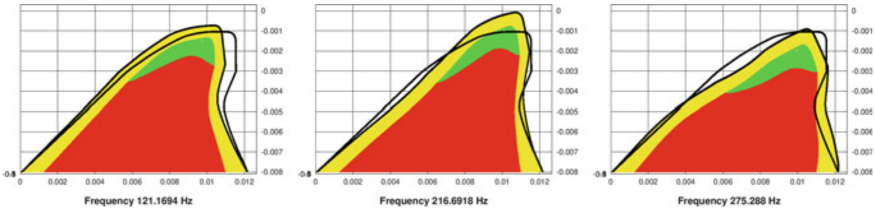


Fig. 2 The first three eigenmodes of chosen VF model with the corresponding eigenfrequencies 121.2 Hz, 216.7 Hz and 275.3 Hz. The reference VF shape is marked by black line. Different colors reproduce regions with different material settings

$1020\,\mathrm{kg/m^3}$. Boundary condition (7) on Γ_W in the case of sole structure problem is replaced by zero Neumann boundary condition $\hat{\tau}^s_{ij}(x,\omega)\,n_j = 0$.

The four lowest VF eigenfrequencies are shown in Table 1 and the corresponding (three) eigenmodes are plotted on Fig. 2. The lowest structural eigenfrequency is close to the first acoustic frequencies of VT+tube suggesting the need to consider their mutual coupling as it is performed in the next paragraph.

4.3 Resonance Frequencies of Coupled System

Let us solve now eigenvalue problem (14). It has many eigenfrequencies however we are interested only in such ones where the acoustic part of eigenmode has a significant contribution. In our case with high ratio of considered densities the eigenfrequencies of the structural part are negligibly influenced whereas the eigenfrequencies of acoustic part are significantly impacted indicating strong interaction with the elastic structure, see [3]. The eigenfrequencies with significant acoustic contribution can be found by skipping all solely structural eigenfrequencies (as obtained in previous paragraph) from the list of computed coupled eigenfrequencies of system (14).

The first two coupled eigenmodes with significant acoustic contribution are plotted in Figs. 3 and 4. The considering of coupling leads to significant increase of the first eigenfrequency ($F_1 = 139.8$ Hz of purely acoustic problem to $F_1^c = 365.3$ Hz of coupled problem, $F_2 = 559.8$ Hz vs. $F_2^c = 692.4$ Hz) similarly as analysed by [1]. Very interesting is the change of acoustic resonator character of the coupled system. Figures 3 and 4 suggest that vocal tract (with tube) behaves rather as half-wave resonator when coupled with the vocal folds. This observation is missing in reference [1].

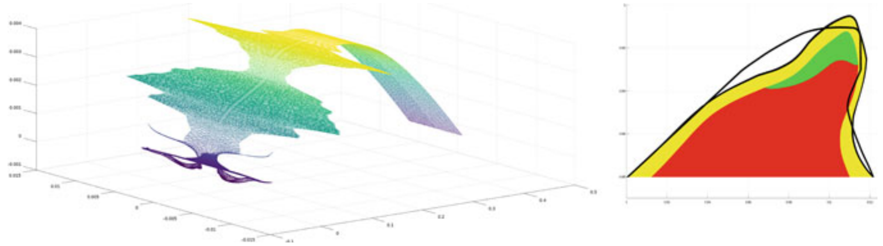


Fig. 3 The first eigenmode of coupled system with significant acoustic contribution at $F_1^c = 365.3$ Hz. *Left:* The amplitude of acoustic pressure and the magnitude of structural displacement is plotted on z-axis. *Right:* The structural part of the first coupled eigenmode is shown

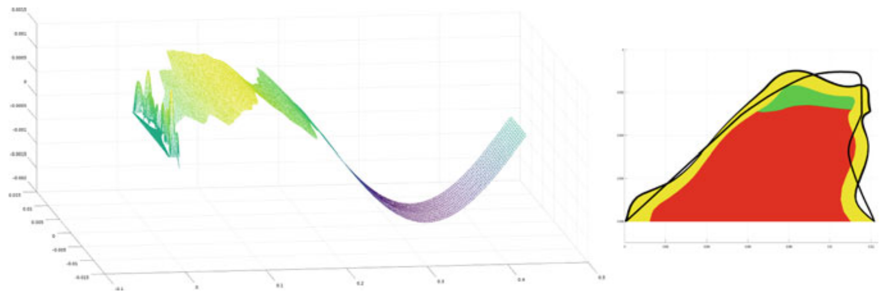


Fig. 4 The second (acoustic) eigenmode of coupled system at $F_2^c = 692.4$ Hz. *Left:* The amplitude of acoustic pressure and the magnitude of structural displacement is plotted on z-axis. *Right:* The structural part of the second coupled eigenmode is shown

5 Conclusion

The mathematical model of the vibro-acoustic problem in frequency domain representing phonation into tube was described and this problem was discretized in space by the FEM leading to the generalized eigenvalue problem. The eigenfrequencies of the coupled system provide interesting results highlighting the importance of structural coupling inclusion into model. The first eigenfrequency of coupled system with acoustic dominance was substantially increased compared to solely acoustic problem in correspondence with reference [1]. The change of acoustic resonator character is newly pointed out. An open question remains to determine the influence of vocal fold stiffness to the resulting eigenfrequencies of coupled system.

Acknowledgements The financial support of this study was provided by the Grant No. SGS19/154/OHK2/3T/12 of the Grant Agency of the CTU in Prague and by the Czech Science Foundation under the Grant No. GA 19-04477S.

References

1. J. Horáček, V. Radolf, A.M. Laukkanen, Biomedical Signal Processing and Control **37**, 39 (2017). https://doi.org/10.1016/j.bspc.2017.02.004
2. J. Horáček, V. Radolf, A.M. Laukkanen, in *Journal of Speech, Language, and Hearing Research* (American Speech-Language-Hearing Association, 2019), pp. 1–18
3. F. Ihlenburg, *Finite element analysis of acoustic scattering*, vol. 132 (Springer, 1998)
4. M. Kaltenbacher, *Numerical simulation of mechatronic sensors and actuators: finite elements for computational multiphysics* (Springer, 2015)
5. W.S. Slaughter, *Linearized Elasticity Problems* (Springer, 2002). https://doi.org/10.1007/978-1-4612-0093-2
6. S.W. Rienstra, A. Hirschberg, *An introduction to acoustics* (Eindhoven University of Technology, 2015).
7. B.H. Story, I.R. Titze, E.A. Hoffman, The Journal of the Acoustical Society of America **100**(1), 537 (1996)
8. I.R. Titze, *Principles of Voice Production* (Prentice Hall, 1994)
9. J. Valášek, M. Kaltenbacher, P. Sváček, Flow, Turbulence and Combustion **102**(1), 129 (2019). https://doi.org/10.1007/s10494-018-9900-z

On Mathematical Modelling of Flow Induced Vocal Folds Vibrations During Phonation

Petr Sváček

Abstract In this paper the problem of mathematical modelling of phonation is discussed. The main attention is paid to the treatment of vocal fold vibrations including the periodical appearance of their contacts. A simplified mathematical model is presented, numerically analyzed and discussed. The Hertz impact forces are used in the structural part. In order to treat the contact phenomena in the fluid model a strategy based on fictitious porous media is introduced. The numerical discretization is described and numerical results are presented.

Keywords Aeroelasticity · Navier-Stokes equations · Finite element method

1 Introduction

The fluid-structure-acoustic interaction problems are usually associated with technical applications as aeroelasticity, see [1]. However, couplings between fluid flow, elastic structure deformation and acoustics are involved also in biomechanics of voice, see [2]. Voice production is a complex process, which involves airflow induced vibrations of vocal folds generating a sound source. The fundamental sound is further modified by the acoustic resonances in the vocal tract cavities. The vocal folds start to oscillate at the so-called phonation onset (flutter instability) given by certain airflow rate and a certain prephonatory vocal folds position, see [3]. For higher flow rates, the glottis is closing during VFs vibration and the VFs collide loading the tissue periodically by the contact stress. Consequently, the mathematical modelling of phonation process is challenging task, it addresses flow field, structure deformation as well as acoustics, see e.g. [4].

The financial support was provided by the *Czech Science Foundation* under the *Grant No. 19-04477S*.

P. Sváček (✉)
Faculty of Mechanical Engineering, Department of Technical Mathematics, Czech Technical University in Prague, Karlovo nám. 13, 121 35 Praha 2, Czech Republic
e-mail: Petr.Svacek@fs.cvut.cz

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_20

In this paper an attention is paid the mathematical modelling of the voice production. As during voice creation the airflow velocity in the human glottal region is lower than 100 m/s, one can use separately the incompressible Navier-Stokes model for the fluid flow and the Lighthill's acoustic analogy for the acoustic wave propagation, see [5]. The considered problem is characterized as a problem of fluid-structure interaction and an attention is paid to the problem of glottis closure (glottis is the narrowest part between the vibrating vocal folds).

Computational modelling can help with analysis of the physical background of the phonation processes. These involve the interactions of the fluid flow with solid body deformation, the contact problem and acoustics. One of the possible approaches is using of a simplified model as the 2-mass model of the vocal folds of [6], where a simplified air flow model is used. Such aeroelastic models [3] has applications in simulation of vowels and in estimation of the vocal fold loading by impact stress and inertial forces.

Here, a simplified lumped VF model with the Hertz contact model is considered in order to more easily address the phenomena of fluid-structure interactions with the contact of the vibrating structure similarly as in [7] and [8]. Due to the same reason a suitable modification of the inlet boundary condition is used. The novelty of this paper lies in verification of the problem formulation with modified boundary conditions, where a simplified stationary model problem is analyzed. Further, more consistent formulation of the porous media term is used in the present paper compared to the approach proposed in [7]. The applied numerical method is described and numerical results are shown.

2 Mathematical Model

We consider two-dimensional model of incompressible fluid flow in an interaction with a simplified model of vocal fold, whose deformation is described as motion of an equivalent mechanical systems with two-degrees of freedom, see [3].

2.1 Flow Model

First, the flow model through the two dimensional model of the computational domain Ω_t during the phonation onset phase is introduced. In this case only small amplitudes of the vibrations of vocal folds appear and thus the flow in the domain Ω_t can be treated with the aid of Arbitrary Lagrangian Eulerian (ALE) method, see [9]. The computational domain Ω_t is shown at Fig. 1, where the additional assumption of a symmetric flow and symmetric vibrations of vocal folds are made. The boundary $\partial\Omega_t$ is assumed to consist of the inlet Γ_I, the outlet Γ_O, the axis of symmetry Γ_S and the time dependent part of boundary Γ_t consisting of its fixed Γ_F and deformable part Γ_{Wt}, which corresponds to the surface of the vibrating vocal fold.

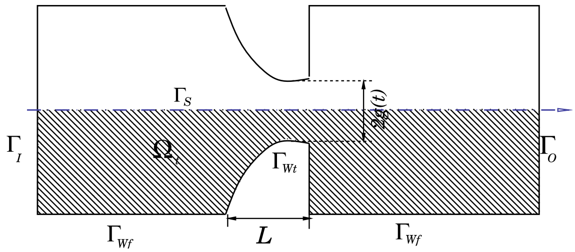


Fig. 1 The computational domain Ω_t with specification of the boundary parts

The flow in the computational domain Ω_t is modelled as incompressible fluid flow described by the system of the incompressible Navier-Stokes equations (cf. [10]) written in the ALE form.

$$\frac{D^{\mathcal{A}}\boldsymbol{u}}{Dt} + ((\boldsymbol{u} - \boldsymbol{w}_D)\cdot\nabla)\boldsymbol{u} = \mathrm{div}\boldsymbol{\tau}^f, \\ \nabla\cdot\boldsymbol{u} = 0, \tag{1}$$

where $\boldsymbol{u}$ denotes the fluid velocity vector $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{\tau}^f = (\tau_{ij}^f)$ is the fluid stress tensor given as $\boldsymbol{\tau}^f = -p\mathbb{I} + \nu\left(\nabla\boldsymbol{u} + \nabla^T\boldsymbol{u}\right)$, p is the kinematic pressure (means pressure divided by the constant fluid density ρ) and $\nu > 0$ denotes the constant kinematic fluid viscosity (the viscosity divided by the density). Further, $\boldsymbol{w}_D$ denotes the domain velocity (i.e. the velocity of the point with a fixed reference), and $\frac{D^{\mathcal{A}}\boldsymbol{u}}{Dt}$ is the ALE derivative, i.e. the derivative with respect to the reference configuration Ω_{ref}. Both the domain velocity $\boldsymbol{w}_D$ as well as the ALE derivative depends on the ALE mapping $\mathcal{A}_t$ describing the deformation of a reference domain Ω_{ref} onto the computational domain Ω_t.

The system (1) is then equipped with an initial condition and with the following boundary conditions are prescribed

$$\begin{aligned} &\text{(a) } \boldsymbol{u} = \boldsymbol{w}_D \text{ on } \Gamma_{Wt}, \\ &\text{(b) } u_2 = 0, -\tau_{12}^f = 0 \text{ on } \Gamma_S, \\ &\text{(c) } \frac{1}{2}(\boldsymbol{u}\cdot\boldsymbol{n})^-\boldsymbol{u} - \boldsymbol{n}\cdot\boldsymbol{\tau}^f = \frac{1}{\varepsilon}(\boldsymbol{u} - \boldsymbol{u}_I) \text{ on } \Gamma_I, \\ &\text{(d) } \frac{1}{2}(\boldsymbol{u}\cdot\boldsymbol{n})^-\boldsymbol{u} - \boldsymbol{n}\cdot\boldsymbol{\tau}^f = p_{ref}\boldsymbol{n} \text{ on } \Gamma_O, \end{aligned} \tag{2}$$

where $\boldsymbol{u}_I$ is a given reference inlet velocity, p_{ref} is a reference outlet pressure, $\boldsymbol{n}$ denotes the unit outward normal vector to $\partial\Omega_t^f$, α^- denotes the negative part of a real number α. Here, the boundary condition (2c) weakly imposes the inlet velocity $\boldsymbol{u}_I$ using the penalization parameter $\varepsilon > 0$.

2.2 Vocal Fold Vibrations

The vocal fold vibrations is modelled using the mechanically equivalent two degrees of freedom model characterized by three masses m_1, m_2 and m_3. The two masses m_1 and m_2 are displaced by the length $l = L/2$ from the center of the vocal fold, where the mass m_3 is located. The three masses m_1, m_2, m_3 were determined by

$$m_{1,2} = \frac{1}{2l^2}(I + m\,e^2 \pm m\,e\,l), \qquad m = m_1 + m_2 + m_3, \tag{3}$$

with $l = L/2$ being the distance of the masses m_1 and m_2 from the center, m denotes the total mass $m = m_1 + m_2 + m_3$, e is the eccentricity and I is the inertia moment, see Fig. 2. The parameters e, m and I are determined using the density $\rho_{VF} = 1020$ kg/m^3, the length (depth of the channel in the third dimension) $h = 18$mm and the (parabolic) shape of the surface of the vocal fold

$$a_m(x) = 1.858\,x - 159.861\,x^2 \text{ [m]} \tag{4}$$

for $x \in \langle 0, L\rangle$ [m] with L being the thickness of the vocal fold $L = 6.8$ mm.

The vibration of the vocal fold is modelled by two degrees of freedom, see Fig. 2, which are the displacements $w_1(t)$ and $w_2(t)$ of the masses m_1 and m_2, respectively. The governing equation of motion reads

$$\mathbb{M}\ddot{\boldsymbol{w}} + \mathbb{B}\dot{\boldsymbol{w}} + \mathbb{K}\boldsymbol{w} = -\boldsymbol{F}, \tag{5}$$

where $\mathbb{M}$ is the mass matrix of the system, $\mathbb{K}$ is the stiffness matrix of the system characterized by the spring constants k_1, k_2, see [3] for details. The matrices are given by

$$\mathbb{M} = \begin{pmatrix} m_1 + \frac{m_3}{4} & \frac{m_3}{4} \\ \frac{m_3}{4} & m_2 + \frac{m_3}{4} \end{pmatrix}, \quad \mathbb{K} = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix}, \quad \mathbb{B} = \varepsilon_1\mathbb{M} + \varepsilon_2\mathbb{K} \tag{6}$$

The vector $\boldsymbol{F} = \boldsymbol{F}_{imp} + \boldsymbol{F}_{aero}$ consists of the impact forces $\boldsymbol{F}_{imp}$ and the aerodynamical forces $\boldsymbol{F}_{aero} = (F_1, F_2)^T$ (downward positive) acting at the masses m_1 and m_2

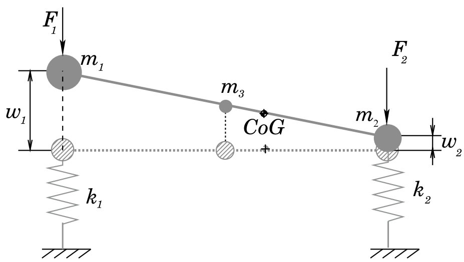


Fig. 2 Two degrees of freedom model (with masses m_1, m_2, m_3) in displaced position (displacements w_1 and w_2). The acting aerodynamic forces F_1 and F_2 are shown

evaluated from the aerodynamical forces as surface integrals using the (kinematic) pressure p and derivatives flow velocity $\boldsymbol{u} = (u_1, u_2)$, see [7].

Moreover, the displacement of the structure surface Γ_{Wt} determines the boundary condition for the construction of the ALE mapping and the domain velocity $\boldsymbol{w}_D$ at Γ_{Wt} is determined using by time derivatives of w_1, w_2.

2.3 Contact Problem

The treatment of the contact the vocal folds in the flow model requires to address not only the inlet boundary condition, but also the periodical topological changes of the flow domain. It can be realized more easily for the simplified situation of the symmetric domain, but this concept can be extended to a more complicated case. First, in this section the computational domain Ω_t is assumed to be formed of the subdomain Ω_t^f, which is really occupied by the fluid(air), and the subdomain Ω_t^p, which is still part of the computational flow domain but it should be occupied the elastic vocal fold Ω_t^{VF}. In practical implementation, the domain Ω_t^p is determined as the intersection of the domain Ω_t with the deformed vocal fold domain Ω_t^{VF}. The geometrical modification of the motion of the surface Γ_{Wt} is based on the deformation of the surface at the contact region, see Fig. 3, where the deformation is locally modified not to violate the minimal gap condition. At these points the surface of the vocal fold is shifted in order to guarantee the minimal gap (g_{min}) condition, see Fig. 3.

The part of the fluid domain Ω_t^p is assumed to be domain of porous media, and the flow is then assumed to be governed by the modified equations

$$\frac{D^{\mathcal{A}}\boldsymbol{u}}{Dt} + ((\boldsymbol{u} - \boldsymbol{w}_D) \cdot \nabla)\boldsymbol{u} + \boldsymbol{\sigma}_P \boldsymbol{u} = \mathrm{div}\boldsymbol{\tau}^f, \tag{7}$$

where the tensor coefficient $\boldsymbol{\sigma}_P$ corresponds to the artificial porosity of the fictitious porous media, see [11], or it can be understand as penalization, see [12]. Here the tensor is chosen to act only the x-direction, i.e. the choice $\boldsymbol{\sigma}_P = \frac{P}{\nu}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \chi_{\Omega_t^p}$ was used,

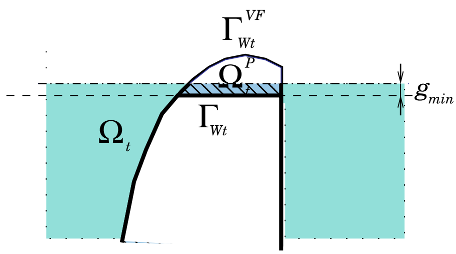


Fig. 3 The detail of the porous media flow domain Ω_t^p

where P denotes the artificial porosity coefficient and $\chi_{\Omega_t^p}$ denotes the characteristic function of the set Ω_t^p which is equal to one on Ω_t^p and zero otherwise.

Although this approach can be written generally, for the presented model problem it can be described more specifically: The reference (undeformed) shape of the vocal fold Ω_{ref}^{VF} is given by

$$\Omega_{ref}^{VF} = \{[x, y] \in \mathbb{R}^2 : x \in (0, L), -H < y < a_m(x) - H\}, \tag{8}$$

where $H = g_0 + H_{VF}$ denotes the half-height of the inlet channel given as sum of the initial halfgap g_0 and the height of the vocal fold $H_{VF} = \max_{x\in[0,L]} a_m(x)$. The deformation of the vocal fold Ω_t^{VF} is described using w_1, w_2 by the Lagrangian mapping $\mathcal{L}_t(x, y) = (x, y_{new})$ with

$$y_{new} = y + \frac{w_1 + w_2}{2} + \frac{w_2 - w_1}{L}x \tag{9}$$

for $[x, y] \in \Omega_{ref}^{VF}$. In particularly, the position of the vocal fold surface Γ_t^{VF} (interface between the fluid and structure domain) is given as

$$\Gamma_t^{VF} = \{[x, y] \in \mathbb{R}^2 : x \in [0, L], y = a_m(x) - H + \frac{w_1 + w_2}{2} + \frac{w_2 - w_1}{L}x\}. \tag{10}$$

The domain Ω_t^P can be characterized as all points $[x, y] \in \Omega_t^{VF}$ which would violate the condition $g(t) \geq g_{min}$ or $y > -g_{min}$. Consequently, the porous media domain can be specified as

$$\Omega_t^p = \{[x, y] \in \mathbb{R}^2 : x \in (0, L), -g_{min} < y < a_m(x) - H + \frac{w_1 + w_2}{2} + \frac{w_2 - w_1}{L}x\}, \tag{11}$$

see Fig. 3.

Let us mention, that for the half-gap $g(t)$ (i.e. the oriented distance of the vocal fold and the symmetry axis) satisfying $g(t) \geq g_{min} > 0$ (phonation onset) such an intersection is naturally empty and in this case the mathematical model is equivalent to the mathematical model presented in [15] and the presented numerical method then leads to the same results, which well determines the flutter velocity.

3 Existence and Uniqueness of a Stationary Solution

In order to discuss the penalization boundary condition we shall start with a simplified stationary problem on two-dimensional domain $\Omega \subset \mathbb{R}^2$ with the Lipschitz-continuous boundary $\partial\Omega$. The system of Navier-Stokes equations is written in the form

$$\begin{aligned} -\nu\triangle u_i + \boldsymbol{u}\cdot\nabla u_i + \frac{\partial p}{\partial x_i} = f_i, \qquad i=1,2, \quad \text{in } \Omega \\ \nabla\cdot\boldsymbol{u} = 0, \quad \text{in } \Omega \end{aligned} \tag{12}$$

with the boundary conditions prescribed on the mutually disjoint parts $\partial\Omega = \Gamma_0 \cup \Gamma_1 \cup \Gamma_2 \cup \Gamma_S$, as

$$\begin{aligned} \boldsymbol{u} = 0, \qquad & \text{on } \Gamma_0, \\ -\nu\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}} + (p - p_{ref})\boldsymbol{n} - \frac{1}{2}(\boldsymbol{u}\cdot\boldsymbol{n})^-\,\boldsymbol{u} = 0, \qquad & \text{on } \Gamma_1\cup\Gamma_2, \end{aligned} \tag{13}$$

with $p_{ref} = p_i$ on Γ_i, $i = 1, 2$.

The stationary problem of Navier-Stokes system of equations reads: Find $\boldsymbol{u} \in \mathcal{X}$ such that for all $\boldsymbol{z} \in \mathcal{X}$ and $q \in \mathcal{Q}$

$$\begin{aligned} \nu\Big(\nabla\boldsymbol{u}, \nabla\boldsymbol{z}\Big)_\Omega + c(\boldsymbol{u};\boldsymbol{u},\boldsymbol{z}) - \Big(p, \nabla\cdot\boldsymbol{z}\Big)_\Omega + \Big(q, \nabla\cdot\boldsymbol{u}\Big)_\Omega + \\ + \int_{\Gamma_1\cup\Gamma_2} \frac{1}{2}(\boldsymbol{u}\cdot\boldsymbol{n})^+\boldsymbol{u}\cdot\boldsymbol{z}\,dS = \Big(\boldsymbol{f}, \boldsymbol{z}\Big)_\Omega - \sum_{k=1}^{2}\int_{\Gamma_k} p_k(\boldsymbol{z}\cdot\boldsymbol{n})\,dS. \end{aligned} \tag{14}$$

In order to prove the existence and uniqueness of the solution let us consider the subspace $\mathcal{X}_{div} \subset \mathcal{X}$ defined as

$$\mathcal{X}_{div} = \{\boldsymbol{\varphi} \in \mathcal{X}, \nabla\cdot\boldsymbol{\varphi} = 0\}.$$

Any solution $\boldsymbol{u} \in \mathcal{X}$ of Equation (14) satisfies $\boldsymbol{u} \in \mathcal{X}_{div}$ and moreover the equation

$$\nu\Big(\nabla\boldsymbol{u}, \nabla\boldsymbol{z}\Big)_\Omega + c(\boldsymbol{u};\boldsymbol{u},\boldsymbol{z}) + \int_{\Gamma_1\cup\Gamma_2} \frac{1}{2}(\boldsymbol{u}\cdot\boldsymbol{n})^+\boldsymbol{u}\cdot\boldsymbol{z}\,dS = \Big(\boldsymbol{f}, \boldsymbol{z}\Big)_\Omega - (p_1 - p_2)\int_{\Gamma_1}(\boldsymbol{z}\cdot\boldsymbol{n})\,dS. \tag{15}$$

holds for all $\boldsymbol{z} \in \mathcal{X}_{div}$.

Theorem 1 (Existence and uniqueness of the solution) *Let $C_F\|\boldsymbol{f}\|_{0,2,\Omega} + C_1|p_2 - p_1| < \frac{\nu^2}{\widetilde{C}}$, where C_F is the constant from Friedrichs inequality, C_1 is the constant from the trace theorem and $\widetilde{C}$ is the constant from the continuity of the trilinear form c. Then there exists an unique solution $\boldsymbol{u} \in \mathcal{X}_{div}$, which satisfies Equation (15) for all $\boldsymbol{z} \in \mathcal{X}_{div}$.*

Proof 1. First, we consider for any $\boldsymbol{w} \in \mathcal{X}_{div}$ the problem: find $\boldsymbol{u} \in \mathcal{X}_{div}$

$$\mathcal{A}_{\boldsymbol{w}}(\boldsymbol{u}, \boldsymbol{z}) = \mathcal{L}(\boldsymbol{z}) \qquad \text{for all } \boldsymbol{z} \in \mathcal{X}_{div}.$$

The bilinear form $\mathcal{A}_{\boldsymbol{w}}(\cdot,\cdot)$ is defined by

$$\mathcal{A}_{\boldsymbol{w}}(\boldsymbol{u}, \boldsymbol{z}) = \nu\Big(\nabla\boldsymbol{u}, \nabla\boldsymbol{z}\Big)_{\Omega} + c(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{z}) + \int\limits_{\Gamma_1\cup\Gamma_2} \frac{1}{2}(\boldsymbol{u}\cdot\boldsymbol{n})^{+}|\boldsymbol{u}|^2 dS,$$

and the form $\mathcal{L}(\cdot)$ is defined by

$$\mathcal{L}(\boldsymbol{z}) = \Big(\boldsymbol{f}, \boldsymbol{z}\Big)_{\Omega} - (p_1 - p_2)\int_{\Gamma_1}(\boldsymbol{z}\cdot\boldsymbol{n})dS,$$

because for any $\boldsymbol{z} \in \mathcal{X}_{div}$ holds

$$\sum_{k=1}^{2}\int_{\Gamma_k} p_k(\boldsymbol{z}\cdot\boldsymbol{n})dS = (p_1 - p_2)\int_{\Gamma_1}(\boldsymbol{z}\cdot\boldsymbol{n})dS.$$

The bilinear form $\mathcal{A}_{\boldsymbol{w}}(\cdot,\cdot)$ is continuous and coercive on $\mathcal{X}_{div}$, the linear form $\mathcal{L}(\cdot)$ is continuous on $\mathcal{X}_{div}$. Thus for any $\boldsymbol{w}^* \in \mathcal{X}_{div}$ there exists solution $\boldsymbol{z}^* \in \mathcal{X}_{div}$ such that

$$\mathcal{A}_{\boldsymbol{w}^*}(\boldsymbol{z}^*, \boldsymbol{z}) = \mathcal{L}(\boldsymbol{z}) \qquad \text{for all } \boldsymbol{z} \in \mathcal{X}_{div}.$$

With the choice of $\boldsymbol{z} = \boldsymbol{z}^*$ we get the following apriori bound

$$\nu|\boldsymbol{z}^*|^2_{1,\Omega} \le \mathcal{A}_{\boldsymbol{w}}(\boldsymbol{z}^*, \boldsymbol{z}^*) = \mathcal{L}(\boldsymbol{z}^*) \le \Big(C_F\|\boldsymbol{f}\|_{0,2,\Omega} + C_1|p_2 - p_1|\Big)|\boldsymbol{z}^*|_{1,\Omega}$$

thus

$$|\boldsymbol{z}^*|_{1,\Omega} \le \frac{1}{\nu}\Big(C_F\|\boldsymbol{f}\|_{0,2,\Omega} + C_1|p_2 - p_1|\Big).$$

2. We define the mapping $\Psi : \boldsymbol{w} \to \boldsymbol{z}$ from $\mathcal{K}$ onto $\mathcal{K}$, where

$$\mathcal{K} = \left\{\boldsymbol{z} \in \mathcal{X}_{div}, |\boldsymbol{z}|_{1,\Omega} \le \frac{1}{\nu}\Big(C_F\|\boldsymbol{f}\|_{0,2,\Omega} + |p_2 - p_1|C_1\Big)\right\},$$

where C_1 is the constant from the trace theorem. Further, we will show that the mapping Ψ is the contractive mapping on $\mathcal{K}$. Let us take $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{K}$ and denote $\boldsymbol{z}_1 = \Psi(\boldsymbol{w}_1)$ and $\boldsymbol{z}_2 = \Psi(\boldsymbol{w}_2)$. Thus the following equations are satisfied

$$\begin{aligned}\mathcal{A}_{\boldsymbol{w}_1}(\boldsymbol{z}_1, \boldsymbol{z}_2 - \boldsymbol{z}_1) &= \mathcal{L}(\boldsymbol{z}_2 - \boldsymbol{z}_1),\\ \mathcal{A}_{\boldsymbol{w}_2}(\boldsymbol{z}_2, \boldsymbol{z}_2 - \boldsymbol{z}_1) &= \mathcal{L}(\boldsymbol{z}_2 - \boldsymbol{z}_1).\end{aligned}$$

Now by subtracting both equations we get from the continuity of the trilinear form c

$$\begin{aligned}\nu|\boldsymbol{z}_2 - \boldsymbol{z}_1|^2_{1,\Omega} &= c(\boldsymbol{w}_2 - \boldsymbol{w}_1; \boldsymbol{z}_2, \boldsymbol{z}_2 - \boldsymbol{z}_1)\\ &\le \tilde{C}|\boldsymbol{w}_2 - \boldsymbol{w}_1|_{1,\Omega}|\boldsymbol{z}_2|_{1,\Omega}|\boldsymbol{z}_2 - \boldsymbol{z}_1|_{1,\Omega}\end{aligned}$$

and with $z_2 \in \mathcal{K}$ we have

$$|z_2 - z_1|^2_{1,\Omega} \le \frac{\tilde{C}}{\nu^2}\Big(C_F \|\boldsymbol{f}\|_{0,2,\Omega} + |p_2 - p_1| C_1\Big)|\boldsymbol{w}_2 - \boldsymbol{w}_1|_{1,\Omega}.$$

Thus the mapping Ψ is a contractive mapping from $\mathcal{K}$ in $\mathcal{K}$, and there exists a fixed point of the mapping Ψ, which is the unique solution of the problem (15).

4 Numerical Approximation

In this section the numerical approximation of the flow model is introduced: an equidistant partition $t_j = j\Delta t$ of the time interval I with a constant time step $\Delta t > 0$ is considered. At time instants t_j, $j = 0, 1, \dots$ the approximations of velocity and pressure are sought $\boldsymbol{u}^j \approx \boldsymbol{u}(\cdot, t_j)$ and $p^j \approx p(\cdot, t_j)$, respectively. The domain velocity at time instant t_j is denoted by $\boldsymbol{w}^j_D$. For the time discretization the formally second order backward difference formula is used, i.e. the ALE derivative is approximated at $t = t_{n+1}$ as

$$\frac{D^{\mathcal{A}}\boldsymbol{u}}{Dt}\Big|_{t_{n+1}} \approx \frac{3\boldsymbol{u}^{n+1} - 4\tilde{\boldsymbol{u}}^n + \tilde{\boldsymbol{u}}^{n-1}}{2\Delta t} \tag{16}$$

where at a given time instant $t = t_{n+1}$ by $\tilde{\boldsymbol{u}}^k$ the transformation of the velocity $\boldsymbol{u}^k$ defined on Ω_{t_k} onto $\Omega_{t_{n+1}}$ is denoted.

In order to apply finite element method the weak form of Eqs. (1) is derived in a standard form, where the ALE derivative is approximated using Eq. (16). The stabilized weak at time instant t_{n+1} form then reads: find finite element approximations $U = (\boldsymbol{u}, p) := (\boldsymbol{u}^{n+1}, p^{n+1})$ such that $\boldsymbol{u}$ satisfy the boundary condition (2a) and

$$a(U; U, V) + a_S(U; U, V) + \mathcal{P}_S(U, V) = L(V) + L_S(V) \tag{17}$$

holds for any test functions $V = (z, q)$ from the finite element spaces, [7] for details. The Galerkin forms a and L are defined for any $U = (\boldsymbol{u}, p)$, $\overline{U} = (\overline{\boldsymbol{u}}, \overline{p})$ and $V = (z, q)$ by

$$\begin{aligned} a(\overline{U}; U, V) = &\int_\Omega \left((\frac{3}{2\Delta t} + \boldsymbol{\sigma}_P)\boldsymbol{u} + ((\overline{\boldsymbol{w}} \cdot \nabla)\boldsymbol{u} \right) \cdot z dx - \int_{\Gamma_{I,O}} \frac{1}{2}(\overline{\boldsymbol{u}} \cdot \boldsymbol{n})^- \boldsymbol{u} \cdot z dS \\ &+ \int_{\Gamma_I} \frac{1}{\varepsilon}\boldsymbol{u} \cdot z dS + \int_\Omega (2\nu(\nabla \boldsymbol{u} : \nabla z) - (\nabla \cdot z)p)\, dx \end{aligned} \tag{18}$$

and

$$L(V) = \int_\Omega \frac{4\tilde{\boldsymbol{u}}^n - \tilde{\boldsymbol{u}}^{n-1}}{2\Delta t} \cdot z dx + \int_{\Gamma_I} \frac{1}{\varepsilon}\boldsymbol{u}_I \cdot z dS - \int_{\Gamma_O} p_{ref}(\boldsymbol{n} \cdot z) dS, \tag{19}$$

where $\overline{\boldsymbol{w}} = \overline{\boldsymbol{u}} - \boldsymbol{w}^{n+1}_D$, $\Omega := \Omega_{t_{n+1}}$.

The terms $a_S(U; U, V)$ and $L_S(U; V)$ are the SUPG/PSPG stabilization terms and the term $\mathcal{P}_S$ denotes the div-div stabilization term. The stabilization terms are defined locally on each element K of the employed triangulation $\mathcal{T}_\Delta$ and summed together, i.e.

$$
\begin{aligned}
a_S(\overline{U}; U, V) &= \sum_{K \in \mathcal{T}_\Delta} \delta_K \left((\frac{3}{2\Delta t} + \boldsymbol{\sigma}_P)\boldsymbol{u} - \mu \Delta \boldsymbol{u} + (\overline{\boldsymbol{w}} \cdot \nabla)\, \boldsymbol{u} + \nabla p, (\overline{\boldsymbol{w}} \cdot \nabla) z + \nabla q \right)_K \\
L_S(\overline{U}; V) &= \sum_{K \in \mathcal{T}_\Delta} \delta_K \left(\frac{4\tilde{\boldsymbol{u}}^n - \tilde{\boldsymbol{u}}^{n-1}}{2\Delta t}, (\overline{\boldsymbol{w}} \cdot \nabla) z + \nabla q \right)_K \\
\mathcal{P}_S(U, V) &= \sum_{K \in \mathcal{T}_\Delta} \tau_K \left(\nabla \cdot \boldsymbol{u}, \nabla \cdot z \right)_K,
\end{aligned}
$$

where δ_K, τ_K are stabilization parameters chosen similarly as in [15].

The problem (17) is linearized, strongly coupled with the structural solver and the described treatment of the contact problem is used.

5 Numerical Results

This section presents the numerical results for the described aeroelastic model. The parabolic vocal fold shape $a_m(x)$ given by Eq. (4) was used and the computations with the initial half-gap chosen as $g_0 = 0.2$mm and the inflow velocity $U_\infty = 0.65 m/s$ were performed. These conditions the phonation onset occurs, see [15]. The following parameters were used: the mass $m = 4.812 \times 10^{-4}$ kg, the inertia moment $I = 2.351 \times 10^{-9}$ kg / m^2 and the eccentricity $e = 0.771 \times 10^{-3}$ m. The stiffness constants were chosen as $k_1 = 56$N/m and $k_2 = 174.3$ N/m. The proportional damping constants were set to $\varepsilon_1 = 120.35\,\mathrm{s}^{-1}$ and $\varepsilon_2 = 6.12 \times 10^{-5}$ s. The stiffness constants give the natural frequencies of the structural model $f_1 = 100$Hz, $f_2 = 160$Hz, see [13, 14]. The fluid density was $\rho = 1.2$ kg/m^3 and the kinematic viscosity $\nu = 1.58 \times 10^{-5}\,\mathrm{m}^2/\mathrm{s}$.

The numerical results are shown in terms of a typical aeroelastic reponse for the aeroelastically unstable system in Fig. 4. The vibration of the vocal fold is shown by the graph of displacements w_1 and w_2 in time domain. The graph Fig. 4a) corresponds to the phonation onset case, the gap between the vocal folds at the glottis is still wide opened and the modification of the mathematical model to treat the contact phenomena is not needed.

The vocal fold vibrations grows further and the increase of amplitudes finally leads to the almost periodical mutual contact of the vocal folds. The appearance of the impact forces leads to almost a limit cycle of oscillations, see Fig. 4b).

The computations were performed either for the case of porosity coefficient P equal zero (which corresponds to the open space flow) or a prescribed fixed value ($P = 10^6 \nu$) of porosity. Figure 6 shows the comparison of these two computations

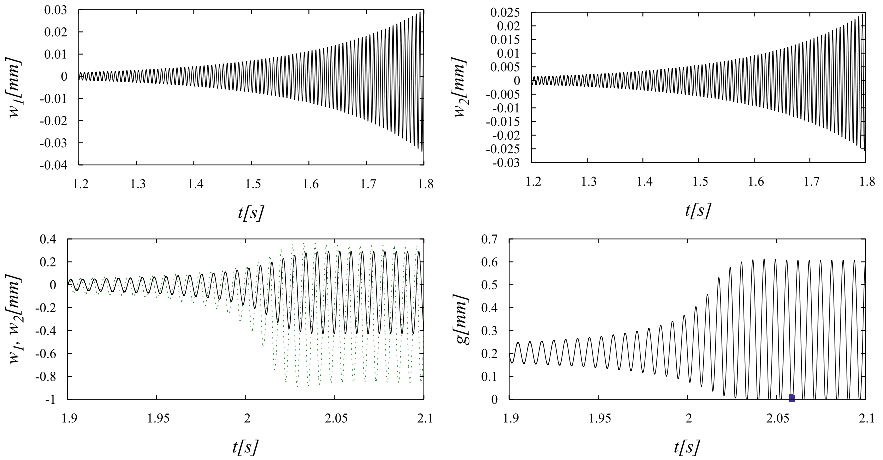


Fig. 4 The aeroelastic response of the structure for flow velocity $U_\infty = 0.65$ m/s: **a** phonation onset in terms of the displacements $w_1(t)$ and $w_2(t)$ (top), **b** phonation with the glottis closure in terms of the displacements $w_1(t)$ (solid/black line) and $w_2(t)$ (dashed/green line) on the left and the half-gap $g(t)$ on the right

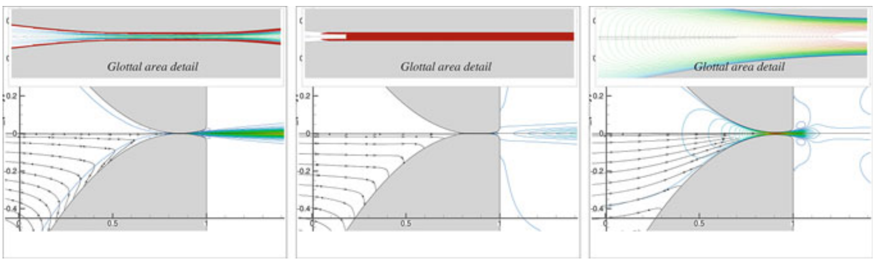


Fig. 5 The flow patterns in terms of flow velocity magnitude and instantenous streamlines during the closing and the reopening phase. The axis show the non-dimensional coordinates x/L, y/L with L being the width of the vocal fold (length in x-direction)

in terms of the inlet quantities. This graph confirms that use of this modified mathematical model also really well addresses the real gap closing similarly as in [7]. The use of the anisotropic porosity has almost no influence on the inlet values of velocity or pressure, see Fig. 6, still the x-component of the gap velocity shown in Fig. 7 becomes zero in the case of the fictitious porosity approach employed. This is also confirmed in Fig. 5, where the flow stops during the closure period (see the middle part of Fig. 5).

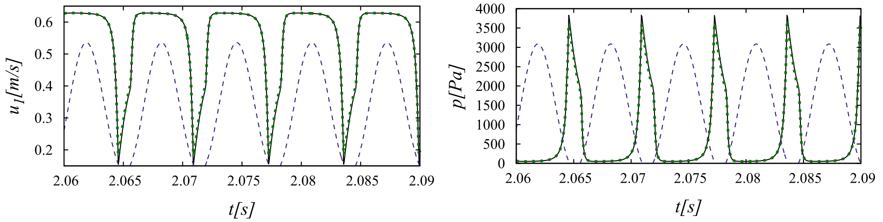


Fig. 6 Inlet velocity (left) and pressure (right) - comparison of the quantities for zero porosity (dashed/green line) and non-zero porosity (solid/black line). The dotted/blue line shows the (scaled) half-gap in dependence on time

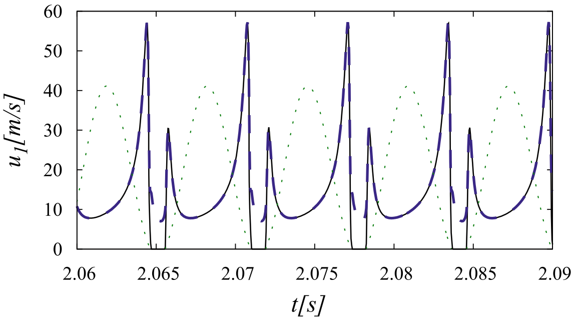


Fig. 7 The flow velocity x-component in the glottal area - comparison of the results for zero porosity (dashed/blue line) and non-zero porosity (solid/black line). The dotted/green line shows the (scaled) half-gap in dependence on time

6 Conclusion

This paper focuse on analysis and an improvement of the mathematical model of human phonation process previously suggested in [7]. The mathematical model is based on the incompressible flow model strongly coupled with a system of ordinary equations describing the motion of the vocal fold model. In order to treat the vocal folds contact the inlet boundary conditions are prescribed by the penalization approach, the geometrical modification of the computational domain is made and in the artificially created part of the computational domain the fictitious anisotropic porous media flow model is used. The analysis of a stationary problem with corresponding boundary conditions is presented. The proposed concept of anisotropic porous media flow is applied , the problem is numerically discretized by an in-house software by the stabilized finite element method. Numerical results are shown proving that the suggested approach is applicable and robust.

References

1. E.H. Dowell, R.N. Clark, *A modern course in aeroelasticity*. Solid mechanics and its applications (Kluwer Academic Publishers, Dordrecht, Boston, 2004)
2. I.R. Titze, *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, U.S.A., 2006)
3. J. Horáček, P. Šidlof, J.G. Švec, Journal of Fluids and Structures **20**(6), 853 (2005)
4. J. Valášek, M. Kaltenbacher, P. Sváček, Flow Turbulence Combust **102**, 129–143 (2019)
5. G. Link, M. Kaltenbacher, M. Breuer, M. Döllinger, Computation Methods in Applied Mechanical Engineering **198**, 3321 (2009)
6. K. Ishizaka, J.L. Flanagan, The Bell System Technical Journal **51**, 1233 (1972)
7. P. Sváček, J. Horáček, Journal of Computational and Applied Mathematics **393**, 113529 (2021). https://doi.org/10.1016/j.cam.2021.113529
8. P. Sváček, **61**(SI) (2021)
9. T. Nomura, T.J.R. Hughes, Computer Methods in Applied Mechanics and Engineering **95**, 115 (1992)
10. M. Feistauer, *Mathematical Methods in Fluid Dynamics* (Longman Scientific & Technical, Harlow, 1993)
11. Burman, Erik, Fernández, Miguel A., Frei, Stefan, ESAIM: M2AN **54**(2), 531 (2020). https://doi.org/10.1051/m2an/2019072
12. P. Angot, C.H. Bruneau, P. Fabrie, Numer Math **81** (1999)
13. J. Horáček, J.G. Švec, AMD, American Society of Mechanical Engineers, Applied Mechanics Division **253**(2), 1043 (2002)
14. J. Horáček, J.G. Švec, Journal of Fluids and Structures **16**(7), 931 (2002)
15. P. Sváček, J. Horáček, Communications in Computational Physics **12**(3), 789 (2012)

Team-Based Learning Collaborative, Is Possible Online?

Cristina M. R. Caridade

Abstract In the proposal presented in this paper, the use of the Team-Based Learning Collaborative methodology during online classes was explored, with a view to training in mathematics for the (future) engineer. The objective was to answer the rhetorical question of the paper's title. During the online classes of Calculus 1, fixed teams of 4–6 students were created that collaboratively interacted and solved 3 challenges. The proposed problems followed 3 phases: Preparation, Learning assurance process and Application. Classroom experiences are reported here using GeoGebra software and Miro collaborative working technology. Activities proposed covering concepts related to solving nonlinear equations and numerical integration, situations that address notions of numerical methods related to topics that are part of the curriculum. It is hoped that these experiences will serve as an inspiration for more active mathematics teaching practice.

Keywords Active learning · Team-Based Learning · Collaborative learning · Math for engineers · Online course

1 Introduction

In the formation of future engineers, who are increasingly innovative, autonomous and entrepreneurs, more significant pedagogical interventions are needed, which allow them to develop, in addition to technical knowledge, the ability to argue, teamwork, management, innovation and problem solving. The application and development of new teaching-learning methodologies that motivate and help students to build knowledge have been much studied and developed [1]. Within these methodologies, active learning stands out [2]. Active learning is a teaching approach in which the student takes an active role in their own learning, participating in activities, making and reflecting on their own learning, often working in collaboration with their peers

C. M. R. Caridade (✉)
Institute of Engineering, Polytechnic of Coimbra, Coimbra, Portugal
e-mail: caridade@isec.pt

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_21

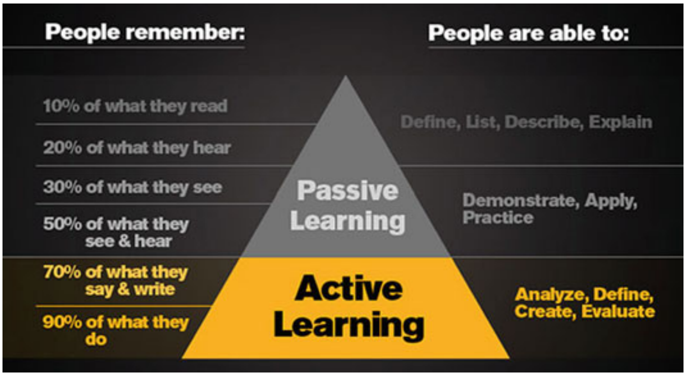


Fig. 1 Active versus Passive Learning. Adapted from https://teachonline.asu.edu/2013/03/how-does-active-learning-support-student-success/

[3]. Active learning can happen in many different ways, but it usually involves 4 main steps (analyze, define, create, evaluate) that can occur individually, in pairs or in groups (Fig. 1). One of these ways is Team-Based Learning. Team-Based Learning (TBL) is an active learning strategy developed at the University of Oklahoma Business School by Dr. Larry Michaelsen when his classes grew from 40 to 120 students. Team-Based Learning is one of the learning technologies that have been used in several areas of education, culture, arts, health, among others [4, 5]. TBL is now used around the world in various schools of medicine, dentistry, pharmacy and other health science disciplines. In mathematics teaching, the TBL methodology has also been applied in secondary courses [6–9] and in higher education, although less frequently, especially in the case of engineering courses [10, 11]. TBL not only encourages individual effort, but also team involvement to learn in an academic environment, also called Team-Based Learning Collaborative (TBL-C) characterized by a three-phase approach (Fig. 2). The TBL-C begins to emerge in collaborative learning teaching strategy [12]. As more courses/disciplines start using online platforms to teach content, universities are becoming more interested in learning how to integrate TBL-C into the online learning environment [5]. Learning is transformed into online learning and the TBL-C methodology will have to accompany this transformation, TBL-C learning is possible online.

2 Methodology

During the transition regime towards COVID-19, the teaching and learning environment changed from face-to-face in a traditional classroom to online mode in a virtual classroom, in order to provide continuity of learning, given the impossibility of them having in-person classes. Traditional higher education institutions also

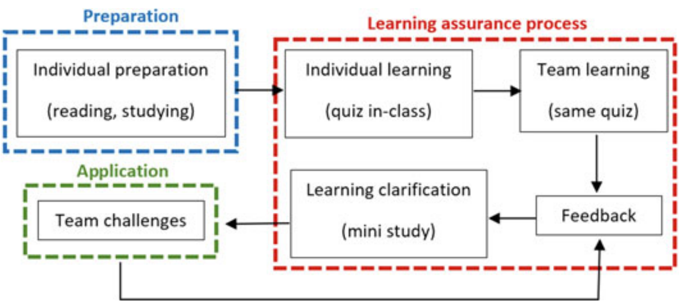


Fig. 2 Methodology process adapted from [16]

had to change overnight from a predominantly face-to-face teaching mode to a fully on-line mode. This required a fundamental change in their core teaching and assessment processes. Students' learning experiences and teachers' way of teaching were impacted and everyone had to adapt to this new reality [13–15]. This paper reports a case study developed in the classes of Calculus 1 of the Degree in Electrotechnical Engineering at Coimbra Institute of Engineering, in which the TBL-C methodology (face-to-face) had to be adapted to online TBL-C. About sixty students from the first-year who attended Calculus 1 mathematics classes in 2020/2021 participated in this study. Three online TBL-C projects (designated by challenges) were carried out that addressed different content and that follow a line of activities proposed to students and represented schematically in the Fig. 2. In the first phase (Preparation), students get the information they need to meet the learning objectives. Information can include book chapter readings, learning guides, online information, research articles, or even information given in classes or labs. In the second phase, called the Learning Assurance Process, students must review and be prepared to use the information during class. At this stage, to ensure that all students have mastered the pre-class material, individual and team readiness tests are conducted. The individual and team test is the same, consisting of 3 or 4 multiple choice questions. Afterwards, the individual and teams results were presented to the students, as well as the correct answers and their resolution. It is at this point that, if necessary, students will have to review some concepts. The final phase (Application), where students share information they 've acquired in teams to solve real-world problems, apply the learned information, and ultimately provide feedback to peers [5, 17]. In this phase, when students as a team solve the proposed challenge, they follow 5 to 6 essential steps: briefly present the studied content, give examples of application of that content, visualize and solve the proposed challenge, present the results to colleagues, and in the final give an opinion on the activity developed. Throughout this process, students need to be supported and guided to learn the necessary skills and achieve learning outcomes. Without that support and direction, learning will not succeed. The teacher's primary role is that of a facilitator: supporting and advising when needed, providing the necessary structure and competencies.

3 Experience in Classes

The classroom is no longer a physical place within the university, but an online space anywhere, where students on that day and time connect and have classes over the internet in a virtual environment. Teamwork is not carried out at tables, with students around, writing in notebooks, computers and sharing ideas, but in simultaneous rooms using collaborative whiteboards and digital resources. In this experience, students formed permanent teams throughout the semester of 4 to 6 students, who for 1h30 every week were housed in simultaneous zoom rooms, where they shared their learning and ideas. Simultaneously, in the Miro portal [18], collaborative areas were created for the registration, exchange and sharing ideas of teamwork. On this platform the teams could develop creative, dynamic and innovative alternatives of the solutions proposed to the challenges presented. The teacher's role was to facilitate the entire learning process, by analysing, detecting, and verifying whether the learning objectives were being achieved. Students were presented with 3 challenges: Create a non-linear equation and solve it; Find the Batman symbol area and Find the Small Lagoon area.

3.1 First Challenge—Create a Non-linear Equation and Solve It

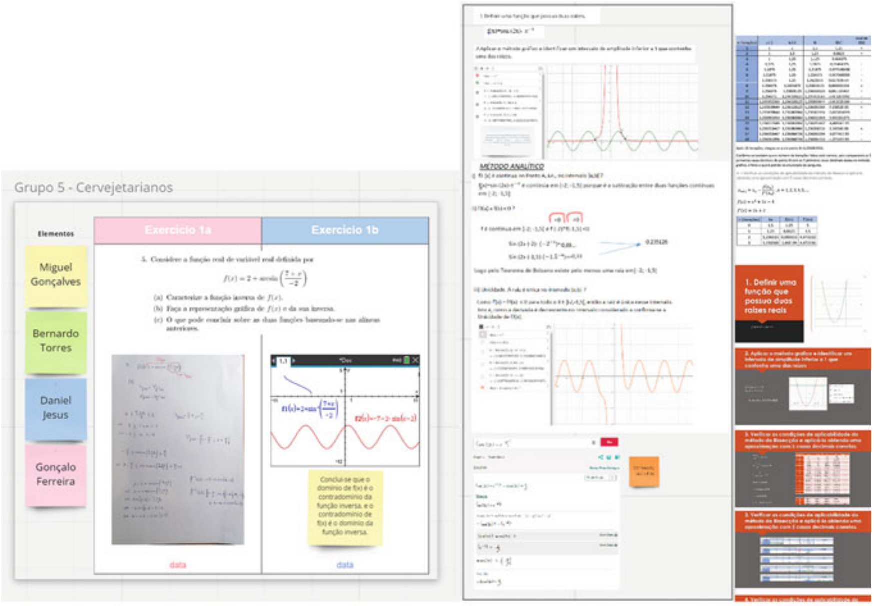
The first challenge presented to the students was the creation of a non-linear equation and the calculation of one of its solutions using the bisection method and the Newton method. During individual preparation, students studied alone based on the course notes. In class, they answered a quiz about each method individually and then in groups. In this first experience, the results (individual/groups) were not very good. Additional clarification were made about: how many iterations are necessary to obtain a given approximation using a method or which one should be the initial chosen approximation for method to converge. Figure 3 depicts Miro's mural with the work of the 12 teams (which correspond to the 12 columns of the mural). In this first phase, the students had to get to know each other (contacts were placed with sticks on the wall), and to become familiar with the collaborative use of Miro. It is necessary to learn how to use the platform with its dynamic whiteboards, where it is possible to register the team's structured ideas, visualize and interact with the tasks, providing a contextualization of the team work. After a first contact, the teams started to solve their challenge. The Fig. 4 represents the solutions obtained by three different teams. Team A with 4 elements (names indicated on the sticks on the left), created a problem and solved it using prints of the calculator machine and its resolutions on paper. In the case of team B and C, others mathematical applications were used (GeoGebra, wolfram alpha, excel) and presentations in word and powerpoint. It is worth noting the diversity of materials and tools that the teams used together with the fixed annotations

Fig. 3 Miro's mural for challenge 1

(through scratches, underlines, sticks, text balloons, etc.), which reinforces that it is a team effort.

3.2 *Second Challenge—Find the Batman Symbol Area*

The second challenge was to use numerical integration, more specifically the trapezoid rule, to calculate the area of the Batman symbol. In this challenge, the teams were already more interconnected, the students already knew each other and were already starting to become more familiar with the collaborative platform. The involvement, the time to complete the challenge and the results obtained were frankly more positive. The individual preparation focused on trapezoid rule for numerical integration, using the links and documents available on Miro. The quiz was carried out with 3 questions: (1) calculate the approximate value of the area defined by points, (2) find the smallest number of points for the area calculation with a given error (3) compare the results obtained by the 2 methods. Some difficulties still arose, especially regarding the number of points to be considered (even/odd points; many/few points) and the the error obtained in the approximation. These difficulties were overcome when the questions were asked in groups. In Fig. 5 it is possible to find the complete mural of this challenge. With the 12 teams working in real time (about 50 to 60 students),



(a) (b) (c)

Fig. 4 Examples of the challenge 1: **a** team A, **b** team B and **c** team C

using the same model (template available in Miro [18]) and more organized, the classes were very motivating and enriching for everyone involved in this learning process. The Fig. 6 shows two illustrative examples made by teams D and E. In the left column a summary of the trapezoid rule for the calculation of numerical integration was presented, in the center on top an example of the applicability of this rule; the resolution of the challenge and its application in GeoGebra were placed in the right on top and in the center on bottom, and in the right on bottom the opinions of the team about the challenge: what they learned, what they liked the most and least and the opinion.

3.3 *Third Challenge—Find the Little Lagoon Area*

In the third challenge it was applied Simpson's rule, to calculate the approximate area of the Little Lagoon, located in "Serra da Estrela", Portugal. It was necessary to use Google Maps to access a real image of the lagoon and obtain real points and distances to calculate the respective area. In the individual preparation there were no doubts. The results of the individual and group evaluation were much more satisfying, therefore, no further clarification was needed. This challenge was the one that went better, the teams were already working as a whole and the learning environment

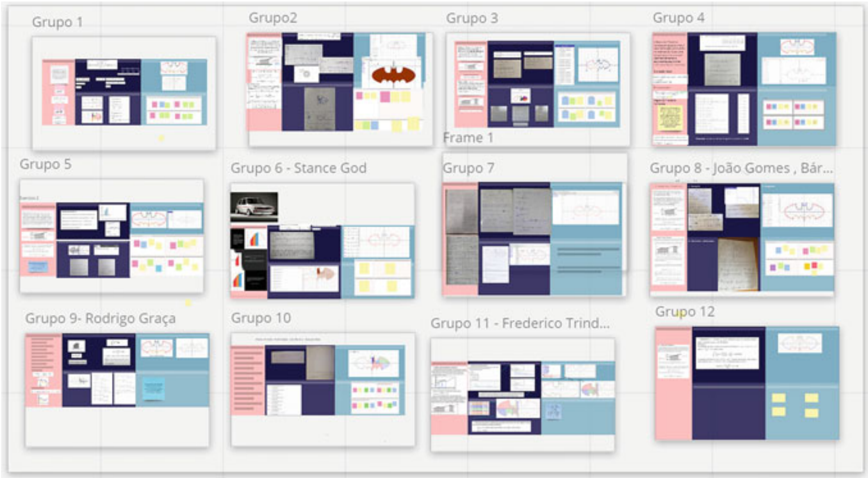


Fig. 5 Miro's mural for challenge 2

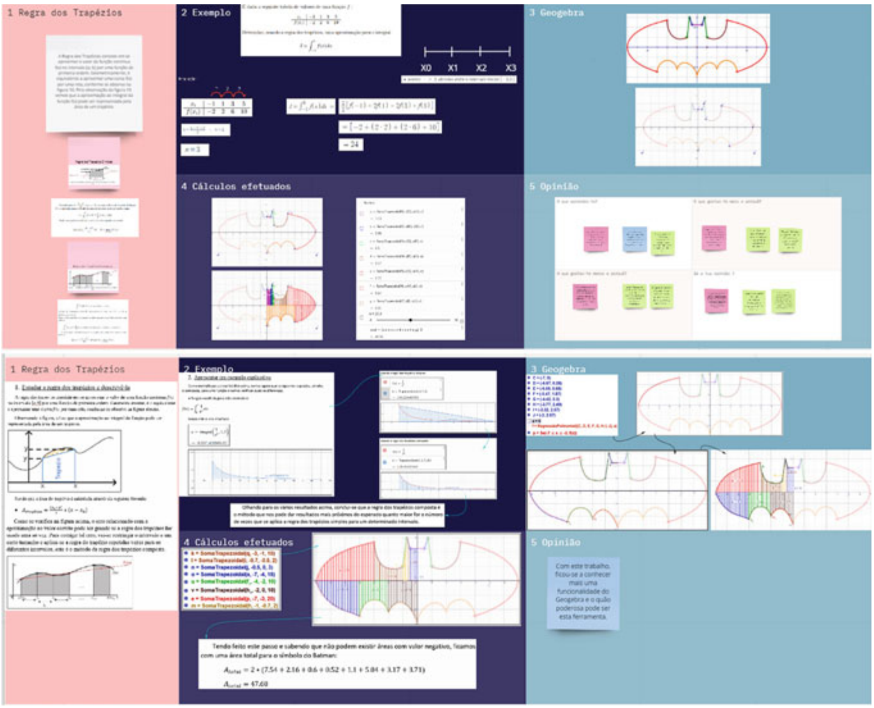


Fig. 6 Examples of challenge 2: Team D (top); Team E (bottom); theory summary (left); example (center-top); GeoGebra (right-top); calculations (center-bottom) and opinions (right-bottom)

Fig. 7 Miro's mural for challenge 3

Fig. 8 Example of challenge 3 performed by team F: a real image (left-top); GeoGebra (right-top); calculations (left-bottom) and opinions (right-bottom)

was already completely dominated by the students. The teacher cooperation with the teams was almost non-existent, only at the beginning to assess the acquired knowledge and at the end to analyze the presentations and opinions. Figure 7 shows the complete mural and Fig. 8 the example of team F. The template is divided into 4 parts (from left to right, top to bottom): aerial image of the Little Lagoon, with the scale; point identification and calculations in GeoGebra; challenge resolution and students' opinions. The teams were given the freedom to use the tools and technologies they liked best and that would be appropriate to solve the challenge.

3.4 Students' Opinion

After the first challenge, some oral opinions were obtained from the students, which allowed for the adjustment and improvement of the proposal for the second challenge. Hence, in the following challenges, the opinion of those involved was part of the challenge itself. Analyzing these opinions, it is verified that the students learned the contents and applied them autonomously using new tools (without the teacher) in a group. What they liked most was the use of new support technologies (GeoGebra and Miro), the existing dynamics of working in groups and interacting and getting to know their teammates and other teams. The lack of technology practice and autonomous studies were the aspects that the students liked the least. As for the general opinions about the activities, called challenges, students feel that they are innovative and interesting activities, and that they surprisingly learn math. They claim that this type of activity has a very important role in the times we live in because it promotes the interaction of students with each other and promotes the spirit of collaborative work. One of the students even comments: "I feel very fulfilled with the work done" and another says "This was the best work in terms of knowledge and application of new technologies that I have ever done in a mathematics curricular unit".

4 Conclusions

There are several teaching techniques that can be used in the classroom. However, the urgency of using new teaching/learning methodologies in higher education and more specifically in mathematics that is taught to engineering is noticeable. There is a need to intensify research on significant pedagogical strategies aimed at developing engineers better prepared for their integration into the labor market and society.

The results of this study demonstrate the interrelationships between the different elements of TBL-C. The results provide new insights into how to apply TBL-C in math classes and what students' perceptions are towards this new learning methodology. Collaborative Team Based Learning (TBL-C) helps instructors (facilitators) develop an active teaching approach to the classroom by collaboratively working in groups in the same environment. With TBL-C, students are involved in their learning and continuous assessment process through problem solving (challenges), through team discussions and peer feedback, ensuring greater responsibility for everyone. Team learning activities should be regular and assessments should be ongoing to document student progression and verify that desired end goals have been achieved. Using TBL-C, students are the agents of their teaching/learning process, highlighting the following advantages that were perceived in the experience reported here: motivation to work as a team, closer contact with students, innovation methodologies, improvement in the student/student and student/teacher relationship and learning based on technical knowledge and social skills. As a first experience of applying the TBL-C didactic innovation, in completely distance learning classes, it was perceived that it is necessary to improve some aspects and adjust others, such as: increasing the workload to carry out this type of activities, at least initially when the students need it to become familiar with the technologies used; sending video lessons and supplementary materials at the beginning of the semester to facilitate access to these resources earlier and apply a quiz assessment in addition to the assessment of oral presentations.

A popular misconception is that active learning can only take place in a face-to-face learning environment. As it was possible to verify with the experience described here, active learning can also occur in courses with many students, hybrid courses and in online courses. The key to the success of this type of learning in the online classroom is to assess students before and throughout the process in order to verify that knowledge has been effectively achieved and to create learning tasks that involve students in teams and that allow for a whole process of thinking, discussing, reflecting and concluding. In the specific case of what was previously presented in this paper, it appears that the practice of TBL-C is possible, and it is possible in online courses, thus answering the rhetorical question presented in the title. The experiences reported here and lived are intended to serve as inspiration in the teaching practice of other trainers (facilitators), with similar proposals for learning other contents. And that in the near future these experiences can be shared by all those who teach mathematics to engineers.

References

1. Caridade, C.M.R.: STEM (Science, Technology, Engineering, and Mathematics) education to teach solids of revolution, ICERI2020 Proceedings, 374–381 (2020).
2. Doucet, A., Netolicky, D., Timmers, K., Tuscano, F. J.: Thinking about pedagogy in an unfolding pandemic (An Independent Report on Approaches to Distance Learning during COVID-19 School Closure). Work of Education International and UNESCO.(2020). https://issuu.com/educationinternational/docs/2020_research_covid-19_eng
3. Arizona State University – Teach online. Available: https://teachonline.asu.edu/2013/03/how-does-active-learning-support-student-success/ [Accessed July 2021].
4. Morales, D.P., Castillo, E., Parrilla, L., García A., Otín, A.: Towards Project-Based Learning applied to the Electronic Engineering studies. 2015 Conference on Design of Circuits and Integrated Systems (DCIS), 1–5, (2015). https://doi.org/10.1109/DCIS.2015.7388615.
5. Levine, R.E., Hudes, P.D.: How-to guide for Team-Base Learning. IAMSE Manuals. Springer. (2021). ISBN 978-3-030-62923-6
6. Naughton, L., Butler, R., Parkes, A., Wilson, P., Gascoyne A.: Raising aspirations using elements of team-based learning in mathematics: a pilot study. International Journal of Mathematical Education in Science and Technology, (2020). https://doi.org/10.1080/0020739X.2020.1770880
7. Abdullah, N.I., Tarmizi, R.A., Abub, R.: The Effects of Problem Based Learning on Mathematics Performance and Affective Attributes in Learning Statistics at Form Four Secondary Level. Procedia Social and Behavioral Sciences 8, 370–376 (2010).
8. Lai, T.L., Lin, F.T., Yueh, H.P.: The effectiveness of team-based flipped learning on a vocational high school economics classroom. Interactive Learning Environments, 28:1, 130–141, (2020). https://doi.org/10.1080/10494820.2018.1528284

9. Sparks D.M., Jackson, D.: Incorporating project-Based Learning into the secondary mathematics and science classroom. Preparing STEAM teachers: The UTeach Replication Model. 387–400, (2020). e-book:978-1-64802-168-8
10. Peters, T., Johnston, E., Bolles, H., Ogilvie, C., Knaub, A., Holme, T.: Benefits to Students of Team-Based Learning in Large Enrollment Calculus. PRIMUS, 30:2, 211–229, (2020). https://doi.org/10.1080/10511970.2018.15424170
11. Campbell, C.G., Taylor, T.Z.: Assessing student learning and knowledge transfer using principles of team-based learning and bridge assignments in large introductory statistics classes. Scholarship of Teaching and Learning in Psychology, 6(1), 15–23. (2020). https://doi.org/10.1037/stl0000167
12. Team-Based Learning™ Collaborative. Available: http://www.teambasedlearning.org/ [Accessed July 2021].
13. Pokhrel S., Chhetri R.: A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. Higher Education for the Future.;8(1):133–141, (2021). https://doi.org/10.1177/2347631120983481.
14. Caridade, C.M.R.: From face-to-face to remote learning: the engineering student's view, ICERI2020 Proceedings, 382–389 (2020).
15. Caridade, C.M.R.: A visão do aluno de engenharia sobre a aprendizagem da matemática em transição. Higher Education in Transition: Students online. CINEP/IPC. 10, 101–120 (2021). ISBN 978-989-54520-8-8
16. Najdanovic-Visak, V.: Team-based Learning for First Year Engineering Students. Education for Chemical Engineers, 18, 26–34. (2017). http://dx.doi.org/10.1016/j.ece.2016.09.001
17. Whitley, H.P., Bell, E., Eng, M., Fuentes, D.G., Helms, K.L., Maki, E.D., Vyas, D.: Practical Team-Based Learning from Planning to Implementation. Am J Pharm Educ. 79(10): 149, (2015). https://doi.org/10.5688/ajpe7910149.
18. Miro: An Online Whiteboard & Visual Collaboration Platform. https://miro.com

GeoGebra Augmented Reality: Ideas for Teaching and Learning Math

Cristina M. R. Caridade

Abstract The methodologies of teaching mathematics in engineering must be active, and directed to the area of courses, so that freshman students are motivated in their learning and thus be able to overcome their difficulties. Augmented reality (AR) is an immersing experience that can be introduced into learning content such as integral calculus and solids of revolution. This paper presents ideas for teaching and learning math using GeoGebra AR. The results presented and visualized show that students engage in this virtual environment integrated into their real world and that the contents explored in a creative way are learned by students. Learning math is done within augmented reality!

Keywords Active methodologies · Mathematics · Engineering · Augment reality · GeoGebra

1 Introduction

Poor performance in mathematics is not a specific case and, although secondary education is a more advanced stage of teaching, a large number of students do not master basic mathematical concepts. In this way, students, upon entering higher education, have serious difficulties in understanding the contents that depend on previous concepts [1]. It is common knowledge that today's students feel a lack of motivation and poor preparation in mathematics, which makes their integration into an engineering course difficult [2]. In higher education, mathematics plays a prominent role in technology-based courses. In the particular case of engineering, mathematics plays an important role, since many aspects of engineering activity include formulating problems and choosing suitable methods to solve them [3]. For students, mathematics is often seen as difficult and many students do not enroll in science, technology, engineering and mathematics subjects, closing the doors to

C. M. R. Caridade (✉)
Institute of Engineering, Polytechnic of Coimbra, Coimbra, Portugal
e-mail: caridade@isec.pt

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_22

scientific, engineering and technology careers [4]. Fortunately, much of the math content can be presented using practical applications to justify its study and spark interest in students. The manipulation of mathematical concepts and their connection can be explored using methodological strategies in mathematical disciplines [5, 6]. Learning is, therefore, an active process that requires participation, commitment and involvement of students throughout the learning process [7].

Virtual reality (VR) and Augmented reality (AR) are two things in common, both aim to expand an individual's sensory environment by mediating reality through technology [8].

GeoGebra AR [9] displays 3D graphics and objects in real-world environments and makes the connection between the real world and the world of abstract mathematics [10, 11]. It can also place math objects on any surface, allow the user to walk around them and take screenshots from different angles. With the GeoGebra feature, students can now model 3D virtual objects into real world objects directly in AR applications [11, 12]. Using the AR feature, students visualize mathematical objects such as geometric solids or solids of revolution that in real life are part of their quotidian. It is also possible to enter the virtual environment and take screenshots from different perspectives, which is a unique experience for students [13]. Many works have been developed with GeoGebra AR in a learning context [14, 15], essentially in the learning of mathematics. Projects aimed at exploring and (re)discovering the mathematical world in different age groups, different levels of education and different contents have been reported in recent years. The main objective of these projects is to stimulate the engagement and enthusiasm for mathematics through its connection with the contents of other areas of knowledge, such as history, art or architecture, these projects facilitate the integration and adaptation of multidisciplinary content to curricular and educational.

In the line of research that has been carried out [16] in order to innovate in the teaching methodologies of mathematics for engineering, this paper presents some ideas for teaching and learning mathematics using GeoGebra AR in engineering education.

2 Motivation

The main motivation of this paper is the need felt as an educator to find "something" that really matters to students, something that keeps them attentive and interested in the classroom. It is to experience the teaching and learning of mathematics to promote the discovery of subject content and connect them with other subjects or areas, improving the learning of Science, Technology, Engineering, Arts and Mathematics (STEAM) through the use of digital manipulation tools [17].

3 Methodology

During classes in the first semester and as a first step upon the arrival of freshman engineering students, the curricular unit of Calculus 1 is presented. With 5 h per week, students must learn basic mathematics essential to their progression and integration into an engineering course. But the lack of prior mathematical knowledge and the lack of interest in this area of knowledge lead students to de-motivation and disinterest.

This paper presents ideas for teaching and learning mathematics to engineers using active methodologies that aim to involve students in their learning and show them that mathematics is essential for their training as an engineer.

4 Experience in Classes

The students of the first year of Electrotechnical Engineering at the Coimbra Institute of Engineering in the academic year 2020/2021, during the first semester's math classes in Calculus 1, collaborated in the experience reported in this paper. Students were organized into groups of 2 and each group interacts collaboratively. They must download the GeoGebra 3D application (app) and the AR tool on their cell phones or tablets to handle and adjust the geometric models. Concrete instructions for the activities are provided to students based in five separate stages: Real object as a solid of revolution; Object dimensions and scale; 3D object (Planar regions and limit curves; 3D solid by rotating); Calculations and Augment Reality represented in the flowchart of Fig. 1.

The first step is to find an everyday object that can be considered a solid of revolution, such as a ball, a glass, a lamp or a jar. In the second step, the object is

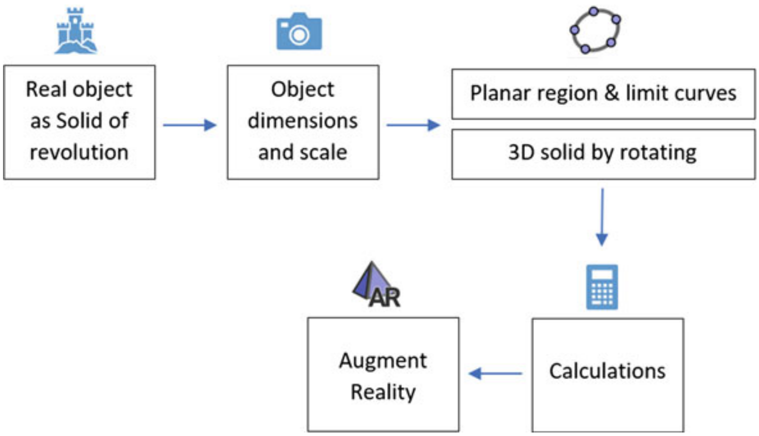


Fig. 1 Flowchart of the steps followed in the applied methodology

Fig. 2 Examples of everyday objects that are considered to be a solid of revolution. Courtesy by Author

photographed and its dimensions calculated. In the next step, the three-dimensional solid corresponding to the real object chosen, is created in GeoGebra. Then all calculations are performed to determine areas and lengths of planar regions and the volume of the 3D solid. In the last step and using the GeoGebra AR application the created 3D object is inserted into reality.

4.1 Real Object as a Solid of Revolution

Given a region R in the xy-plane, it is possible to generate a solid, by revolve the R region about a vertical or horizontal axis of revolution. A solid generated this way is often called a solid of revolution. In our daily life, there are many objects with these characteristics, for example a ball that is geometrically defined by a sphere can be obtained by rotating a semi-circle around an axis of revolution. Figure 2 shows some examples of real objects that can be considered as solids of revolution.

4.2 Object Dimensions and Measurement Scale

After acquiring the photograph of the chosen object, it is important to define the scale between the real object and the object's image. For this, using a measuring instrument, measurements are taken of the object and the image of the object to find the scale. In Fig. 3 some of these measurements are represented.

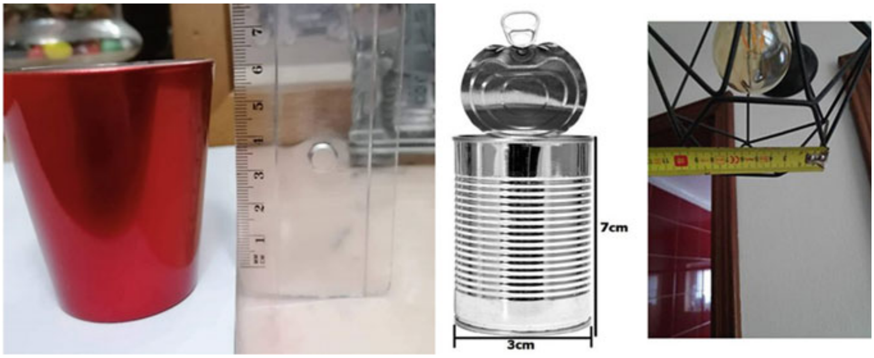


Fig. 3 Objects dimensions and measurement scale. Courtesy by Author

4.3 3D Solid

The photograph is then uploaded to GeoGebra, some geometric transformations are carried out so that the object is aligned with the coordinate axes and so that the entire planar region is located in one of the quadrants of the xoy plane. Then it is necessary to define the boundary of the planar region. This task can be performed in two different ways: by defining a composition of curves (polynomials—Fig. 4 (right)) or by defining a set of points that allow defining one or more interpolating polynomials (Fig. 4 (left)).

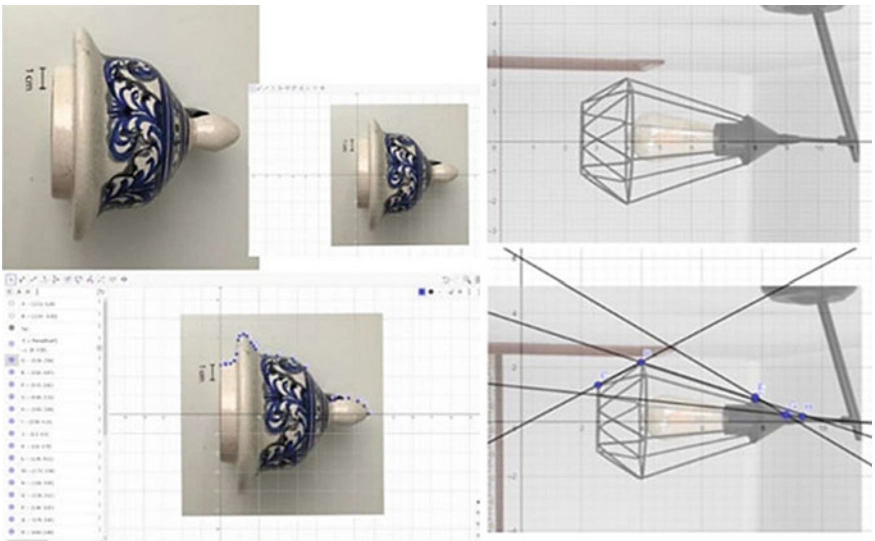


Fig. 4 Boundary of planar regions: by interpolating polynomials (left) or by curves (right)

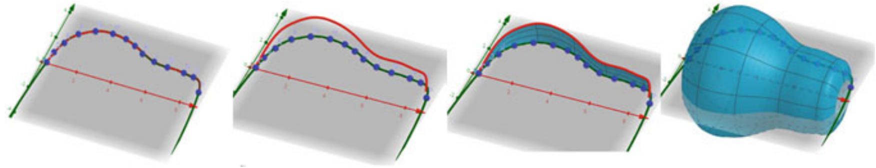


Fig. 5 Image sequence for building a 3D object

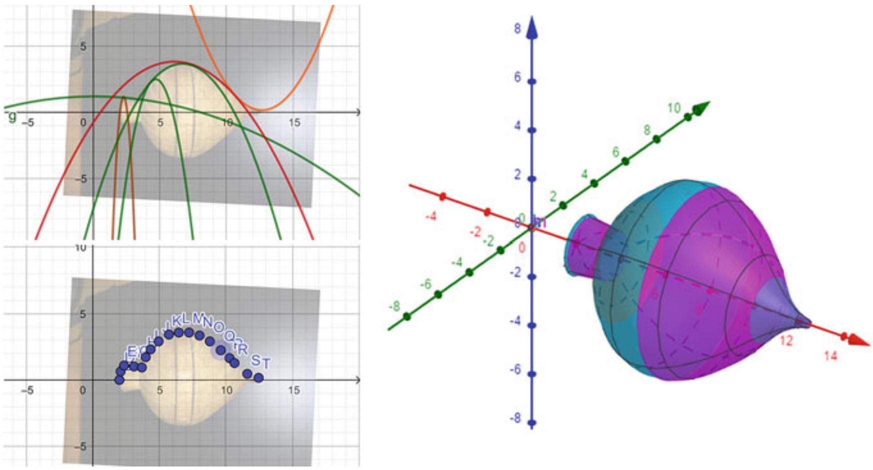


Fig. 6 3D object created: planar region (left) and 3D object (right)

Once the planar region is defined, using GeoGebra a 3D solid can be constructed by rotating that region around an axis. In the case presented in Fig. 5, some steps in the construction of a 3D object can be observed by rotating the planar region around the x-axis.

Another example can be seen in Fig. 6, where on the left side the definition of the planar region by the two processes (composition of curves—top and interpolating polynomial—bottom) is shown, and on the right side the visualization of the three-dimensional object constructed.

4.4 Calculations

The introduction of these active learning techniques into application-oriented freshman engineering courses allows students to discover, practice, master and integrate their knowledge of integral calculus. Thus, students must also present the mathematical calculations necessary to determine the area and perimeter of the planar region, and the volume of the solid of revolution created. In Fig. 7 on the left are presented the calculations made by hand and on the right using the calculator. Calculations

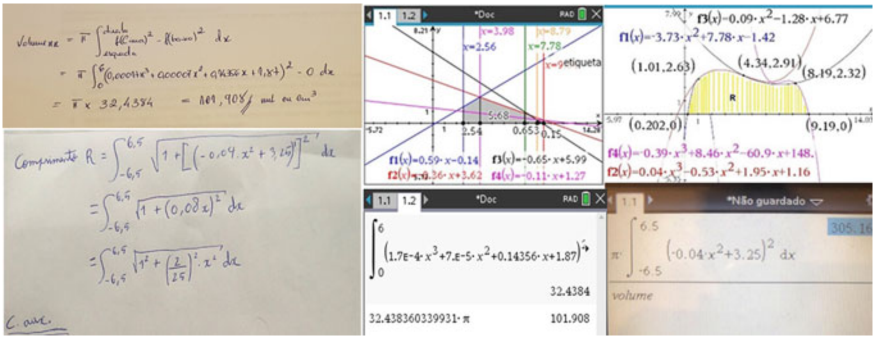


Fig. 7 Calculations performed: by hand (left) and using a calculator (right)

must be done by hand and confirmed using calculators or other mathematical applications. In some cases, students built the planar region on the calculating machine before performing the calculations.

4.5 Augment Reality

The final and most motivating phase consists of introducing the object created by the student into reality. An interactive experience of a real world environment where some of the objects placed there were created by the students on the computer.

The student uses the GeoGebra AR tool that he downloaded to his smart-phone or tablet and as he films the environment that surrounds him, he inserts the 3D object in that environment, literally joining the real world with the abstract world of mathematics. In this way, the application allows the student to explore 3D mathematical objects placed virtually in the students' environments, while they can walk around them and observe them from different perspectives (Fig. 9) or pass through the objects (Fig. 8) and observe their geometric characteristics. Furthermore, it is possible to look for symmetries, rotations, translations or other transformations and take screenshots of different perspectives, as if it were a mathematical walk.

Finally, in Fig. 10, screenshots of the video made by a student are presented with the addition of music and special effects on the solid of revolution. Creativity and motivation are strongly present in this very simple example.

5 Conclusions

First-year math study is often seen as irrelevant and distracting by engineering students who are more interested in applied engineering disciplines [18]. Many students are disconnected for this reason and present a performance in mathematics at levels

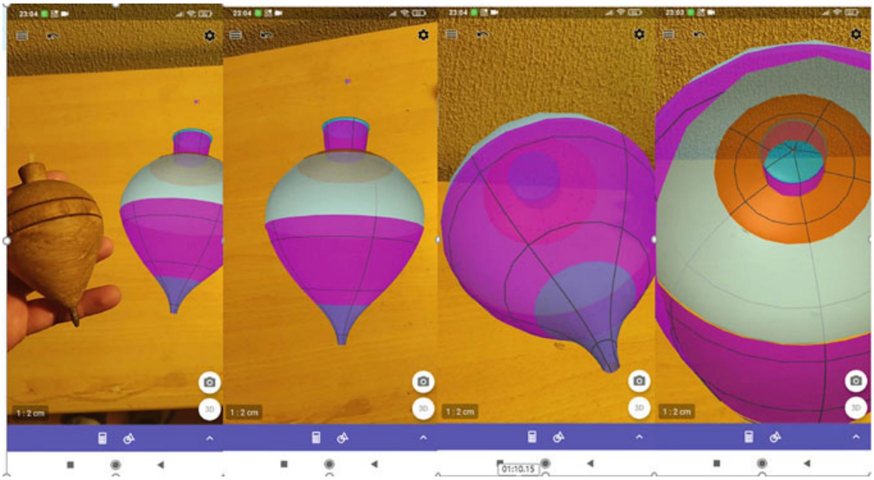


Fig. 8 GeoGebra AR: Screenshots from video

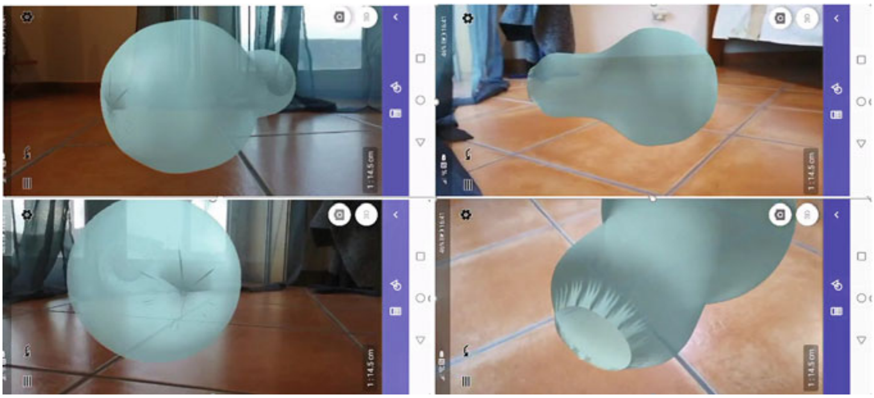


Fig. 9 GeoGebra AR: Screenshots from video

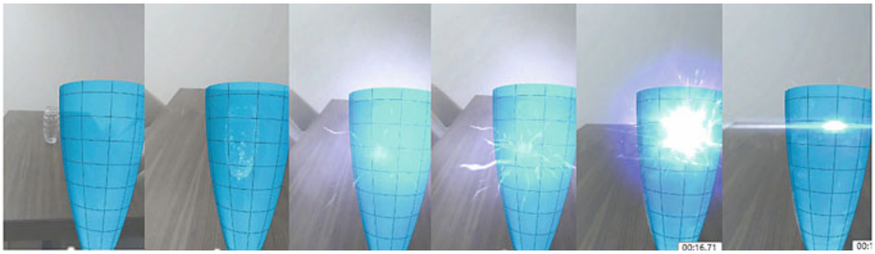


Fig. 10 GeoGebra AR: Screenshots from video

well below their abilities, which will be reflected in the subjects of subsequent years where the contents depend on the basic contents of the 1st year mathematics subjects. In this sense, it is increasingly important that mathematics teachers use active teaching methodologies that can reveal to students the importance of mathematics in their engineering course. Captivating students through the implementation of projets or activities in their area of study where the mathematics content taught to students is present there and can be explore in a more enriching and motivating way. It is in this line of teaching/learning that the study described in this paper is located. It was intended to present concrete examples of activities carried out in the classes of Calculus 1 of the first year of the degree in engineering.

The results of the application of these technologies such as GeoGebra, GeoGebra AR and the calculating machine, is the creation of a richer learning environment, which makes the student more interested, applied and motivated. Working in different scenarios allows for more concise and rich learning with more creative solutions and, therefore, greater student motivation. By using mathematical modeling, students are involved in designing and solving problems that typically originate outside of mathematics.

The results visualized and presented above confirm that this type of ideas applied to teaching and learning mathematic are possible and that the teaching of engineering students is a teaching that must be multi-methodological, by the gradual and hierarchical construction of capacities over time that create bases for engineering students have developed skills and competencies in various areas of knowledge (STEAM) as an Engineer presupposes.

As future projects to be applied in the next school year and using augmented reality as a background, it is planned to develop the possibility of inserting in Augmented reality objects that are composed of more than a solid of revolution, such as a vase with flowers, a cup of coffee with saucer or a graph and a knife. It will also be possible to create a virtual environment with solids of revolution created by different students in groups, for example placing a set of objects on a desk or table.

References

1. Oliveira, A.J.: Aprender Matemática no Ensino Superior: Desafios e Superação. Revista Científica Multidisciplinar Núcleo do Conhecimento. 5(3), 94–103, (2018). ISSN:2448-0959
2. Caridade, C.M.R., Faulhaber, M.C..: e-MAIO: Módulos de Aprendizagem Interativa Online, Ambientes Virtuais no Ensino Superior, pp. 91–113 CINEP/IPC, (2016). ISBN: 978-989-98679-7-0 (ebook).
3. Alves, M., Coutinho, C., Rocha, A.M., Rodrigues, C.: Fatores que influenciam a aprendizagem de conceitos matemáticos em cursos de engenharia: Um estudo exploratório com estudantes da Universidade do Minho, Revista Portuguesa de Educação 29(1), (2016). http://dx.doi.org/10.21814/rpe.5998
4. Li, Y., Schoenfeld, A.H.: Problematizing teaching and learning mathematics as "given" in STEM education. IJ STEM 6(44), (2019). https://doi.org/10.1186/s40594-019-0197-9

5. Caridade, C.M.R.: Applications (Ideas) in Linear Algebra with Digital Image Processing. Can we Do, Teach, Motivate and Evaluate? Journal of Information Systems Engineering & Management, 4(4), em0103, (2019). https://doi.org/10.29333/jisem/6347
6. Kounlaxay, K. et. al: Learning Media on Mathematical Education based on Augmented Reality. KSII Transactions on Internet and Information Systems 15(3), (2021). https://www.koreascience.or.kr/article/JAKO202116739518290.pdf
7. Krishnan, S.: Students' Perceptions of Learning Mode in Mathematics, The Malaysian Online Journal of Educational Sciences, 4(2), (2016).
8. Widada W., et al: Augmented Reality assisted by GeoGebra 3-D for geometry learning. J. Phys.: Conf. Ser. 1731 012034 (2021). https://doi.org/10.1088/1742-6596/1731/1/012034
9. GeoGebra. https://www.geogebra.org/
10. Tomaschko M., Hohenwarter, M.: Augmented Reality in Mathematics Education: The Case of GeoGebra AR, Augmented Reality in Educational Settings, 325–346, (2019).
11. Trappmair A., Hohenwarter, M.: Driving augmented reality: GeoGebra's new AR features in teaching mathematics, in Proc. of Conference on Technology in Mathematics Teaching–ICTMT 14, (2019).
12. Ancochea, B., Cárdenas, M.I.: Exploring Real World Environments using Potential of GeoGebra AR. In M. Ludwig, S. Jablonski, A. Caldeira, A. Moura (Eds.), Research on Outdoor STEM Education in the digiTal Age. Proceedings of the ROSETA Online Conference in June 2020, Münster: WTM, 41–46, (2020). https://doi.org/10.37626/GA9783959871440.0.05
13. Budinski N., Lavicza, Z,: Teaching Advanced Mathematical Concepts with Origami and GeoGebra Augmented Reality, in Proc. of Bridges 2019 Conference Proceedings, 387–390, (2019).
14. Brzezinski, T.: GeoGebra 3D with AR (Google): Explorations & Lesson Ideas. Online page, visited in july 2021. https://www.geogebra.org/m/qbxbcmqw
15. Brzezinski, T.: Augmented Reality: Ideas for Student Explorations. Online page, visited in july 2021. https://www.geogebra.org/m/RKYFdQJy
16. Caridade, C.M.R.: STEM (Science, Technology, Engineering, and Mathematics) education to teach solids of revolution, Proceedings of 13th annual International Conference of Education - ICERI2020, 374–381, (2020).
17. Agrawal, V.J., Patel, J.: A Review: Augmented Reality and Its Working, International Journal of Engineering and Technology, 4(5), (2017).
18. Loch, B., Lamborn, J.: How to make mathematics relevant to first-year engineering students: perceptions of students on student-produced resources, International Journal of Mathematical Education in Science and Technology 47(1): 29–44, (2016). https://doi.org/10.1080/0020739X.2015.1044043

Understanding the Level of Mathematics Knowledge of Students Who Joined ISEC

M. E. Bigotte de Almeida, J. R. Branco, L. Margalho, A. Queiruga-Dios, and M. J. Cáceres

Abstract The low pass rate in the curricular units of Differential and Integral Calculus, taught throughout the first semester of the first year of engineering degrees at the Coimbra Institute of Engineering, led to the development of a diagnostic test with the intention of identifying the degree of knowledge considered essential for the full integration [1]. This diagnostic test was constructed considering the guidelines of European Society for Engineering Education. Regarding the minimum knowledge recommended for entering higher education for an engineering course and according to the basic and secondary education program of Portugal, the 20 questions were grouped taking into account the topics algebra, analysis and calculus, geometry and trigonometry. The analysis of the data, from academic years 2013/2014 to 2019/2020, allows us to analyze the level of knowledge in terms of mathematics content of the students as well as to realize in which topics they show the greatest difficulties.

Keywords Mathematics knowledge · Diagnostic test · Engineering · Differential · Integral calculus

M. E. Bigotte de Almeida (✉) · J. R. Branco · L. Margalho
Coimbra Institute of Engineering/Polytechnic Institute of Coimbra, Coimbra, Portugal
e-mail: ebigotte@isec.pt

J. R. Branco
e-mail: jrbranco@isec.pt

L. Margalho
e-mail: lmelo@isec.pt

A. Queiruga-Dios · M. J. Cáceres
University of Salamanca, Salamanca, Spain
e-mail: queirugadios@usal.es

M. J. Cáceres
e-mail: majocac@usal.es

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_23

1 Introduction

It's not new the debate about the failure of mathematics in the teaching of engineering and its relationship to the knowledge obtained in high school. In 1995, the London Mathematical Society, in collaboration with the Institute of Mathematics and its Applications and the Royal Statistical Society, produced the report *Tackling the Mathematics Problem* [2], which investigated the existing concerns among mathematicians, scientists and engineers in the mathematical preparation of students entering higher education. The final report, produced by the UK Engineering Council, showed strong evidence of a "steady decline" in basic mathematical competencies and "increasing inhomogeneity in mathematical attainment and knowledge" [3]. One of the main recommendations of this report is to apply a diagnostic test, aiming at:

– identifying students at risk of failing because of their mathematical deficiencies,
– targeting remedial help,
– designing programmes and modules that take account of general levels of mathematical attainments, and
– removing unrealistic staff expectations.

It also states that "students embarking on mathematics-based degrees should have a diagnostic test on entry", emphasizing that this test must be understood as a means to an end: "Diagnostic test should be seen as part of a two-stage process. Prompt and effective follow-up is essential to deal with both individual weaknesses and those of the whole cohort".

This gap has been, at an international level, one of the main reasons given for failure in higher education, a situation regularly referred to as the "Maths Problem" [4]. Commonly cited characteristics of this problem include lack of basic math skills, fragmented understanding, inadequate knowledge of concepts, and inability to successfully solve math problems. The "Maths Problem" has been highlighted as an issue of concern in Ireland, UK, Australia and US [4–6].

This discussion led to the definition of multiple strategies aimed at overcoming the difficulties detected and the consequent analysis of the impact of implementing the measures [7–12].

In order to discuss the European systems of education and training in engineering, the European Society for Engineering Education (SEFI) created, in 1982, the Mathematics Working Group (MWG), which aims to promote a forum for discussion and guidance aimed at all those interested in the mathematics education of engineering students in Europe.

In this context, given the decline in mathematical skills and knowledge of students upon entry into higher education and the heterogeneity of candidates for engineering degrees, it was created, in 1992, the first curriculum guidance document to formulate a detailed and structured list of topics, organized by levels, which correspond to specific content essential for learning mathematics in engineering degrees. The main concern of this document was to define the appropriate contents of mathematics in engineering education. After 5 meetings held in several countries of the European

Community, MWG produced a set of recommendations on *A core curriculum in Mathematics for the European engineer* [13].

Based on the 1965 OECD report [14], this document summarizes the reasons for the importance of mathematics in an engineering curriculum:

1. mathematics provides a training in rational thinking and justifies confidence in the value of such thinking;
2. it is the principal tool for the derivation of quantitative information about natural systems;
3. it is the second language of human discourse and parallels natural language, providing a means of communication for ideas, as evidenced by contents of technical papers;
4. it facilitates the analyses of natural phenomena;
5. it is important in assisting the engineer to generalise from experience;
6. it trains the imagination and inquisitiveness of the student if properly taught;
7. it is an education for adaptation to the future.

To these aspects, there is also the evolution of technology that has substantially reinforced the role of mathematics on engineering.

The document also states that the prior knowledge of students at the beginning of the degree is essential, so it is necessary to define the set of prerequisites in pure mathematics required when accessing higher education. Core Zero was thus established, which integrates contents of arithmetic, algebra, geometry and trigonometry, and differential and elementary integral calculus. In 2002, MWG reformulated [13], aiming to learn outcomes rather than a simple list of topics. Regarding the minimum knowledge recommended for entering higher education for an engineering course, these are detailed by areas and identified by topics in the Core Zero section from *Mathematics for the European Engineer—A Curriculum for the Twenty-First Century* [15].

The insufficient preparation that students have when they arrive to higher education is also a problem of the Portuguese education system, and is compounded by the heterogeneity of the undergraduate students in engineering.

According to Programme for International Student Assessment (PISA) 2018 report, Portugal scored around the OECD average in mathematics. Mean performance improved since 2003, although in 2018 was close to the level observed over the period 2009-15. The report also remarks that career expectations of the highest-achieving 15-year-old students reflect strong gender stereotypes. Amongst high-performing students in mathematics or science, about one in two boys expects to work as an engineer or science professional at the age of 30, while only about one in seven girls expects to do so.

Higher education Portuguese institutions used several alternative processes to attract new students, which increase the heterogeneity of personal and motivational features of students, that access through very different means (scientific-humanistic graduates, professional undergraduates courses, technological undergraduates courses, technological specialization undergraduate courses, over 23 years old, etc). This multiplicity of students have originated heterogeneous mathematical skills, asymmetries in the essential mathematics knowledge and difficulties on integration into higher education, and motivate the definition of alternative paths that allowed those students to follow a positive learning process.

The Department of Physics and Mathematics of Coimbra Institute of Engineering has promoted and developed a set of strategies to reverse this situation, reorganizing the operation and evaluation of courses, building tools that facilitate the learning process and implementing strategies for student engagement. Despite all this effort, students have not met the expectations and continue to exhibit a high failure rate and high abstention rates (to classes and also to evaluations). Assuming that students are not learning what they should learn, because of the enormous gap in the basic knowledge, we should construct pedagogical tools that can contribute to the diagnosis, acquisition and consolidation of mathematical knowledge and skills needed in engineering, as well as develop resources that will give engineering students the best possible learning experience.

Differential and Integral Calculus courses have been referred in many studies, and the difficulties experienced by students in elementary and basic contents, essentials for full integration in higher education, are major concerns expressed by many teachers, leading to adaptations of curricular reorganization and definition of actions that allow modify this situation.

The aim of this paper it to describe the experience carried out with the mathematics diagnostic test, the partnership established with the Dublin Institute of Technology and to analyze the results obtained by students on the last 6 academic years.

2 The Diagnostic Test at ISEC: The Evolution

2.1 ISEC—Coimbra Institute of Engineering

Coimbra Institute of Engineering (ISEC) is an organic unit of the Polytechnic Institute of Coimbra that offers engineering undergraduate degrees: Biomedical, Bioengineering, Civil, Electrical (normal and post-work regime), Electromechanical, Engineering and Industrial Management, Informatics (normal, post-work regime and European course) and Mechanical. In the 2018/2019 academic year, ISEC also created a non engineering degree in Sustainable Cities Management, which was presented with a curricular plan focused on economic, environmental and social stainability of learning methodologies that allow students to develop professional skills.

All these 9 degrees include a Differential and Integral Calculus curricular unit, with a common syllabus: trigonometric and inverse trigonometric functions, integration, definite and improper integrals. This syllabus constitute the core of mathematics knowledge that teachers understand as essential. However, poor results from students on access to higher education national exams are frequently observed, specially on self-proposed students [16], which means that students who access higher education generally have difficulties on basic and elementary contents of mathematics, that prevent a full integration on Differential and Integral Calculus curricular unit and on the degree itself.

ISEC also provides a “Year Zero”, a start programme available for students who have not joined in higher education but attend to optional undergraduate courses at ISEC, to deepen knowledge in basic engineering subjects (mathematics and physics), to prepare them to the qualifying exams needed to access to higher education in the following year as well as preparation for attending classes.

Research group on didactics of mathematics in engineering (GIDiMatE) is a group formed within the mathematic scientific area of ISEC, established in 2011, that implements competency assessment/improvement actions and develops teaching tools and instruments to contribute to the acquisition and consolidation of basic and complementary mathematical knowledge, essentials for the Differential and Integral Calculus curricular units.

2.2 The Diagnostic Test

Given the previous context and the real situation found at ISEC, it became evident for maths teachers from ISEC to build a diagnostic tool, applied at the entry of higher education, which helps to identify flaws in the set of essential knowledge in mathematics that students need to master, so that, based on the results obtained, strategies can be defined to overcome the gaps detected, thus ensuring the understanding of the syllabus contents in the Differential and Integral Calculus curricular units.

To study this question, since the school year 2011/2012 a diagnostic test (DT) has been held annually, in the first week of classes of the first semester, from students who attended to classes of Differential and Integral Calculus curricular units from ISEC’s first year degrees.

DT had suffered successive alterations. Nowadays we consider a stabilized version, that allows to compare results from different years.

Of multiple reflections that mathematic teachers from ISEC have maintained throughout their academic career, it has been found that students bring large gaps in elementary knowledge necessary for successful integration in the curricular units of mathematics of engineering degrees.

So, in 2011/2012 school year they carried out the first edition of the DT. Topics were addressed that, based on the experience accumulated by those teachers, should be the most relevant in the acquisition of knowledge and skills in mathematics in higher education, particularly for degrees in engineering. Five topics were considered: equations, functions, rationals, geometry and trigonometry and derivatives. The questions were of multiple choice, with six options each, in order to assess the most common shortcomings of students. The 36 questions of the test were distributed as showed at Table 1. The results were obtained by arithmetic mean of correct answers in each of the 5 topics.

Table 1 Distribution by topic of the 36 questions from 1st version of the DT (2011/2012)

Topic	Number of questions
Equations	8
Functions	5
Rationals	7
Geometry/Trigonometry	10
Derivatives	6

Faced with the existence of SEFI document [15], GIDiMatE group decided to compare the structure of the DT with the structure suggested by MWG. Among the suggested areas and according to the program of the elementary and secondary education of Portugal, GIDiMatE gave special attention to algebra, analysis and calculus, geometry and trigonometry. Those areas were considered as the most significant, because they are essential for most of the mathematic curricular units. According to the guidelines of SEFI the 36 initial questions included in the diagnostic test were regrouped in the topics listed in Table 2.

Table 2 Distribution by topic, according to SEFI area, of the 36 questions from 1st version of the DT (2011/2012)

Topic	Number of questions
Arithmetic of real numbers	
Algebraic expressions and formulae	3
Linear laws	3
Quadratics, cubics, polynomials	2
Arithmetic of real numbers	2
Analysis and Calculus	
Functions and their inverses	8
Logarithmic and exponential functions	3
Rates of change and differentiation	6
Complex Numbers	2
Geometry and Trigonometry	
Geometry	2
Trigonometric functions and applications	4
Trigonometric identities	1

On 2012/2013 school year, GIDiMatE decided to reduce the DT both in size and in time, and changed some issues in order to contribute to a better understanding of the difficulties faced by students. Questions were reduced from 36 to 25, such that there was an uniform distribution on the topics proposed by GIDiMatE and answers were reduced from 6 to 4, including the elimination of the option “none”. It was also decided to remove all questions about complex numbers, since it was found this content was not uniformly addressed in secondary education, putting students in unequal situation in the diagnostic evaluation. From the 36 original questions of 2011/2012 diagnostic test only 12 were included in the revised version.

From contacts with Dublin Institute of Technology (DIT) and the joint reflection that was done, it was found to be important to introduce some issues in common with the Irish diagnostic test in order to be able to make comparisons of results and proposed actions in partnership [17]. As a result of this cooperative work, on 2013/2014 school year 9 common issues were constructed or modified, as listed in Table 3, and the number of questions was reduced to 20.

Table 3 Distribution by topic of the 9 question common with DIT DT (2013/2014)

Topic	Number of questions
Algebra	
Algebraic expressions and formulae	1
Analysis and Calculus	
Functions and their inverses	3
Rates of change and differentiation	3
Geometry and Trigonometry	
Geometry	2

Additionally, Danish KOM project led by Niss [18], organized a detailed and systematic description of what we should expected to obtain with the teaching of mathematics, using the concept of competence which influenced the description of the learning objectives reflected in studies of the OECD-PISA, [19]: "Possessing mathematical competence means having knowledge of, understanding, doing and using mathematics and having a well-founded opinion about it, in a variety of situations and contexts where mathematics plays or can play a role".

KOM project identified a list of mathematical competencies such as "the ability to ask and answer questions in and with mathematics, focus on mathematical thinking, problem handling, modelling and reasoning" and "the ability to deal with mathematical language and tools, focus on representation, symbols and formalism, communication competency".

In this context we decided to integrate a question to evaluate the competence in mathematical modeling. For that purpose we selected a statement of a problem which reflects a linear system of two equations and two unknowns.

According to SEFI [15], the final 20 questions were regrouped by the areas listed in Table 4. We remark that 3 of the questions are evaluating more than 1 subject.

This is the final version of the DT, so it was not modified since that.

Table 4 Distribution by topic of the 20 questions from 3rd version of the DT (2013/2014)

Topic	Number of questions
Algebra	
Arithmetic of real numbers	3
Algebraic expressions and formulae	4
Linear laws	2
Quadratics, cubics, polynomials	1
Analysis and Calculus	
Functions and their inverses	3
Logarithmic and exponential functions	1
Rates of change and differentiation	3
Geometry and Trigonometry	
Geometry	2
Trigonometric functions and applications	2
Trigonometric identities	1
Mathematical modeling	1

2.3 The Methodology

This study follows a quantitative research methodology, considering the observation of data collection instruments. Taking into account that the analysis may allow us to conclude the level of mathematical knowledge that students have when they access the ISEC, as well as in which topics they show greater difficulties, the case study approach will be made according to an interpretive paradigm. Thus, it is intended, without exercising any type of control over the situation, to obtain conclusions that can lead to the implementation of strategies at the level of teaching, learning and assessment that contribute to the promotion of success in those Differential and Integral Calculus curricular units.

To carry out the DT, certain requirements were required, such as:

– failure to submit calculations or justifications;
– the existence of a grid of answers where the respective letter considered correct by the student was placed;
– the unequivocal presentation of the answer, under penalty of being annulled;
– not using any calculator.

The duration of the DT was 60’, being given in the 1st week of the semester to students attending Differential and Integral Calculus curricular units.

3 Sample

DT was given to all ISEC's engineering degrees, from 2013/14 to 2019/20, with the exception of the school years 2018/2019 and 2019/2020 which, due to various constraints, was only given to informatics and biomedical students, which the authors teach at Differential and Integral Calculus curricular units.

Sample is formed by 2231 results of students from ISEC degrees (Biomedical, Bioengineering, Civil, Electrical, Electromechanical, Engineering and Industrial Management, Informatics, Mechanical and Sustainable City Management) and also students from "Year Zero".

4 Results

4.1 Global Results

Results of the DT are present at Table 5. We present global results and also results, mean value and standard deviation, from each school year, from 2013/14 to 2019/20. According to Portuguese system, results are presented to 20 values.

As DT evaluates basic knowledge on mathematics, a mean value of 10.26 points shows that mathematical knowledge is far below what one would expect.

There is no stated trend, neither of growth nor of degrowth, but it is to notice a significant improvement on results on 2019/2020 school year, as presented on Fig. 1. The standard deviation is high, being in the order of 20% in all years.

Table 5 DT results, from 2013/2014 to 2019/2020 school years

School year	Dimension	Mean value	St. deviation
2013/2014	432	9.52	(4.14)
2014/2015	398	9.92	(3.82)
2015/2016	495	10.33	(4.49)
2016/2017	397	10.57	(3.79)
2017/2018	225	10.21	(4.25)
2018/2019	93	9.55	(4.42)
2019/2020	191	12.20	(4.48)
All	2231	10.26	(4.22)

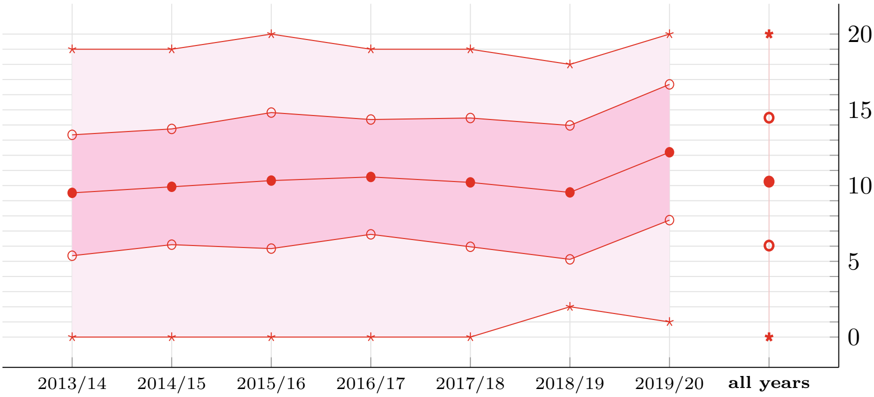


Fig. 1 DT results, from 2013/2014 to 2019/2020 school years

4.2 Results Depending on the Number of Enrollments

On Fig. 2 we present the results obtained on the DT depending on the number of enrollments. The goal is to analyze if students that put apart Differential and Integral curricular units lost any kind of basic knowledge of mathematics, that is essential to these curricular units and, of course, for the degree. As the majority of the students from our sample are from informatics (1324 students), for this analysis we only consider students from this degree, to avoid bias.

Results show a decreasing trend as number of enrollments increases, which reveals that students that put aside the mathematics curricular units lose basic mathematical knowledge that is essential for that course. The only exception are students on the 2nd enrollment, that show a very irregular behaviour: usually worse than students with

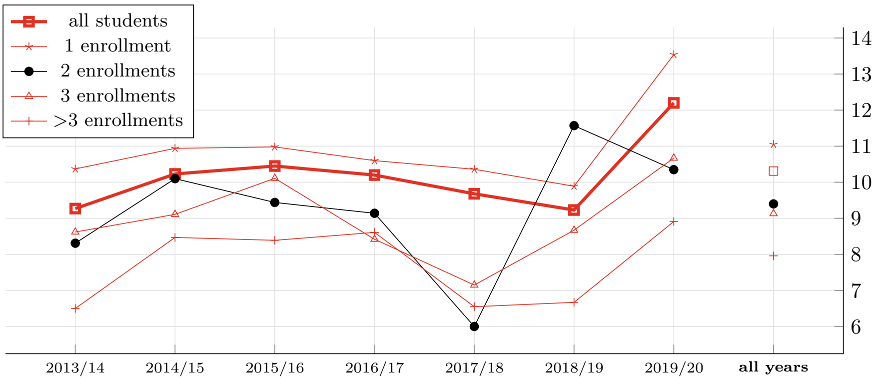


Fig. 2 Informatics DT results depending on the number of enrolments, from 2013/2014 to 2019/2020 school years

1 enrollment, but sometimes worse than students with 3 or even more enrollments and sometimes even better than students with only 1 enrollment!

In terms of time dependency, we observe similar evolution than global results, even the improvement on results on the 2019/2020 school year.

4.3 Results According to Topic

On Fig. 3 we present the results according to the year and topic. Since the number of questions in each topic is different, to facilitate the comparative analysis of results, we present them converted to a scale from 0 to 20.

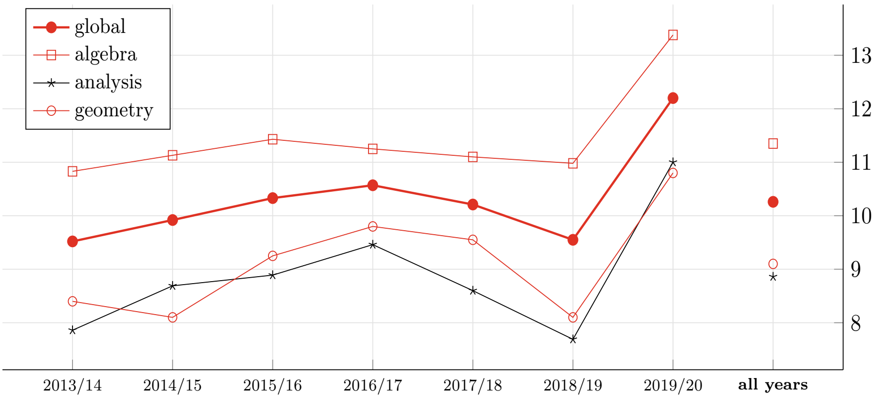


Fig. 3 DT results depending on the topic, from 2013/2014 to 2019/2020 school years

Naturally, the pattern shown by year is the same as the global results.

Results on algebra are always higher than the average result, and the results on analysis and geometry are lower than the average. The poor results on analysis are particularly worrisome, since this is one of the fundamental pillars of the mathematics curricular units.

5 Conclusions and Future Work

Results show that, in average terms, basic knowledge on mathematics is insufficient for engineering degree students.

Even with poor basic knowledge on mathematics, they tend to get worse if students do not monetize them in the first years of the degrees. That shows the importance to students monetize their mathematical basic knowledge in the first enrollment.

Geometry is where students perform worst, but it's also the least important. Behavior on analysis is particularly worrisome, because this is the most important topic for differential and Integral calculus curricular units.

DT result provides information on the mathematical content that should be worked with the student, which allows to define an individual working plan. This plan is prepared according to [15] and will describe the evolution of the student's learning, through the monitoring and reformulation. The goal is to encourage autonomous work and to adapt to personal learning style, study method and cognitive development.

In the future, we intend to study regression models, to analyze the determinants of students' mathematical knowledge.

References

1. Bigotte de Almeida, M.E., Queiruga-Dios, A., Cáceres, M.J.: Differential and Integral Calculus in First-Year Engineering Students: A Diagnosis to Understand the Failure. Mathematics, 9(1), 61 (2021). https://doi.org/10.3390/math9010061.
2. Tackling the Mathematics Problem. London Mathematical Society, Institute of Mathematics and its Applications, Royal Statistical Society (1995).
3. Savage, M., Kitchen, A., Sutherland, R., Porkess, R.: Measuring the Mathematics Problem. Engineering Council (2000).
4. Lawson, D., Croft, T., Waller, D.: Mathematics support past, present and future. EE 2012 - International Conference on Innovation, Practice and Research in Engineering Education, Conference Proceedings (2012).
5. Gill, O., O'Donoghue, J., Faulkner, F., Hannigan., A.: Trends in Performance of Science and Technology Students (1997-2008) in Ireland. International Journal of Mathematical Education in Science and Technology 41(3), 323–339 (2010).
6. Treacy, P., Faulkner, F.: Trends in Basic Mathematical Competencies of Beginning Undergraduates in Ireland, 2003-2013. International Journal of Mathematical Education in Science and Technology, 46(8), 1182–1196 (2015). https://doi.org/10.1080/0020739X.2015.1050707.
7. Koch, D., Herrin, G.D.: Intervention strategy for improving success rates in calculus. Proceedings of the 2006 ASEE Annual Conference and Exposition, 775 (2006).
8. Gill, O., O'Donoghue, J.: The mathematical deficiencies of students entering third level: An item by item analysis of student diagnostic tests. Proceedings of Second National Conference on Research in Mathematics Education, St. Patrick's College, Dublin, 228–239 (2007).
9. Hieb, J., Lyle, K., Ralston, P., Chariker, J.: Predicting performance in a first engineering calculus course: implications for interventions. International Journal of Mathematical Education in Science and Technology, 46(1), 40–55 (2015).
10. Nite, S.B., Capraro, R.M., Capraro, M.M., Allen, G.D., Pilant, M., Morgan, J.: A bridge to engineering: A personalized precalculus (bridge) program. 2015 IEEE Frontiers in Education Conference (FIE), 1–6 (2015). https://doi.org/10.1109/FIE.2015.7344358.
11. Bigotte, E., Gomes, A., Branco, J.R., Pessoa, T.: The influence of educational learning paths in academic success of mathematics in engineering undergraduate. 2016 IEEE Frontiers in Education Conference, 1–6 (2016).
12. O'Sullivan, C., Bhaird, C., Fitzmaurice, O., Fhloinn, E.: An Irish Mathematics Learning Support Network (IMLSN) Report on Student Evaluation of Mathematics Learning Support: Insights from a large scale multi-institutional study. National Centre for Excellence in Mathematics and Science Teaching and Learning (NCEMSTL):Ireland (2014).
13. SEFI: A core curriculum in Mathematics for the European engineer. Mathematics Working Group (1992). http://sefi.htw-aalen.de/Curriculum/CoreCurriculum1992.pdf.

14. OEC: Mathematical Education of Enginners. OECD Report (1965).
15. SEFI: Mathematics for the European Engineer-A Curriculum for the Twenty-First Century. Mathematics Working Group (2002). http://sefi.htw-aalen.de/Curriculum/sefimarch2002.pdf.
16. Direção-Geral da Educação. Relatórios/Estatísticas. https://www.dge.mec.pt/relatoriosestatisticas-0.
17. Carr, M., Fidalgo, C., Bigotte, E., Branco, J.R., Santos, V., Murphy, E., Fhloinn, E.N.: Mathematics diagnostic testing in Engineering: an international comparison between Ireland and Portugal. European Journal of Engineering Education 40(5), 546-556 (2015).
18. Niss, M.: Quantitative literacy and mathematical competencies. Quantitative literacy: why numeracy matters for school and colleges. Princeton: National Council on Education and the Disciplines, 215-220 (2003).
19. OECD: PISA 2009 Results: Executive Summary (2010). https://www.oecd.org/pisa/pisaproducts/46619703.pdf.

From Fibonacci Sequence to More Recent Generalisations

Paula Catarino and Helena Campos

Abstract Number sequences have been the subject of several research studies. From the algebraic properties to the generating matrices and generating functions of these sequences, all these topics and many others have been studied by several researchers and a vast bibliography covers this type of sequences. When in 2013 the first author had an initial contact with this topic, she soon turned to other sequences of numbers and began a careful investigation with great enthusiasm. A group of researchers interested in these themes was formed, based on the most recent works on this subject of number sequences, they developed new generalisations of some of them, introducing the respective definitions, properties and some results concerning those sequences. In this paper we propose a tour of our work involving the number sequences we have been studying over the past eight years, either jointly or individually.

Keywords k-Fibonacci numbers · k-Jacobsthal numbers · k-Pell numbers · Balancing numbers · k-Telephone numbers · Hyper k-Pell numbers · Incomplete numbers

1 Introduction

About the number sequences there is a vast literature studying several properties, ones involving the well-known Binet's formula, Catalan's, Cassini's and d'Ocagne's identities and there is also a vast literature dedicated to the study of other properties involving each sequence. The Fibonacci succession has been one of the geneses

P. Catarino (✉) · H. Campos (✉)
Department of Mathematics, School of Science and Technology, University of Trás-os-Montes e Alto Douro, Vila Real, Portugal
e-mail: pcatarin@utad.pt

H. Campos
e-mail: hcampos@utad.pt

P. Catarino
Research Centre of Mathematics, University of Minho-Polo CMAT-UTAD, Vila Real, Portugal

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_24

of multiple researches that resulted in the creation of other number sequences and respective research. Many papers are dedicated to Fibonacci and Lucas sequences, such as the works of Dil and Mezo [1], Hoggatt [2], Vorobiov [3], among so many others. The modern science is interested in the application of the Golden Section and Fibonacci numbers [4]. It is well known that the ratio of two consecutive Fibonacci numbers converges to the Golden Mean, or Golden Section, $\frac{1+\sqrt{5}}{2}$, which appears in problems related with physics of the high energy particles or theoretical physics.

Based on the most recent works on the subject related to number sequences, we have developed new generalisations of these recurrence sequences, introducing the definition, properties and some theorems concerning the k-Fibonacci and k-Lucas numbers, as well as k-Jacobsthal and k-Jacobsthal-Lucas [5], not also forgetting the k-Pell, k-Pell-Lucas, and Modified k-Pell numbers [6], as well as balancing and cobalancing numbers [7] and some of their generalisations. More recently, the definition and properties of the Incomplete sequence of numbers [8], the k-Telephone numbers [9] and the Leonardo numbers [10–12] have resulted in new papers already published or already submitted. Also, the Hyper k-Pell, Hyper k-Pell-Lucas, and Hyper Modified k-Pell [13] sequences are introduced, as well as their generating functions and some properties of each one. Furthermore, properties about the relationship between them and some of their generating functions are presented in our works.

As we have mentioned before, in 2013 the first author had a first contact with this type of topic. Since that time a group of researchers interested in these themes was formed and one of the first sequences studied was one of the generalizations of the Fibonacci sequence. In this context, based on the most recent works on this subject of number sequences, we develop new generalisations of some of them and the aim of this paper is to visit some of this work that we have done in the last eight years in individual and in group research. Therefore, the next section is divided into subsections according to different number sequences that are presented. The final section is dedicated to give some final remarks about this tour.

2 A Tour Around Some of the Sequences We Worked On

In 1965, Horadam studied some properties of sequences of the type, $\{w_n\} \equiv \{w_n(a, b; p, q)\}$, where a, b are nonnegative integers and p, q are arbitrary integers, see [14, 15]. Such sequences are defined by the recurrence relations of second order: $w_n = pw_{n-1} - qw_{n-2}$, $(n \geq 2)$, rule that will allow the next terms to be calculated according to their predecessors. The initial conditions $w_0 = a$ and $w_1 = b$. For example, the Fibonacci, the Lucas, the Pell, the Pell-Lucas and the Modified Pell sequences can be considered as special cases of sequences of this type, having the following for these cases:

The **Fibonacci sequence**: $\{w_n(0, 1; 1, -1)\} \equiv \{F_n\}$, with initial conditions $F_0 = 0$, $F_1 = 1$. The **Lucas sequence**: $\{w_n(2, 1; 1, -1)\} \equiv \{L_n\}$, with initial conditions

$L_0 = 2$, $L_1 = 1$. The **Pell sequence**: $\{w_n(1, 2; 2, -1)\} \equiv \{P_n\}$, with initial conditions $P_0 = 1$, $P_1 = 2$.

The **Modified Pell sequence**: $\{w_n(1, 1; 2, -1)\} \equiv \{q_n\}$, with initial conditions $q_0 = 1$, $q_1 = 1$. The **Pell-Lucas sequence**: $\{w_n(2, 2; 2, -1)\} \equiv \{Q_n\}$, with initial conditions $Q_0 = 2$, $Q_1 = 2$.

Also, the **Jacobsthal** and the **Jacobsthal-Lucas sequences** can be considered as $\{w_n(0, 1; 1, -2)\} \equiv \{J_n\}$, with initial conditions $J_0 = 0$, $J_1 = 1$ and $\{w_n(2, 1; 1, -2)\} \equiv \{j_n\}$, with initial conditions $j_0 = 0$, $j_1 = 1$, respectively.

Next, we find the explicit formula for the term of order n of each number sequences using the well-known results involving recursive recurrences. Consider the following characteristic equation, associated to Fibonacci and Lucas recurrence relation given by $r^2 - r - 1 = 0$ with two distinct roots, $r_{1,2} = \frac{1 \pm \sqrt{5}}{2}$. Using these roots, we can have the Binet formulas for each sequence as follows:

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right] = \frac{1}{\sqrt{5}}\left(r_1^2 - r_2^2\right);$$

$$L_n = \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n = r_1^2 + r_2^2.$$

About the Pell, Pell-Lucas and Modified Pell sequences, the respective characteristic equation is given by $r^2 - 2r - 1 = 0$ with also two distinct roots, $r_{1,2} = 1 \pm \sqrt{2}$. Hence the Binet formulas for these sequences are

$$P_n = \frac{1}{2\sqrt{2}}\left(\left(1+\sqrt{2}\right)^n - \left(1-\sqrt{2}\right)^n\right);$$

$$Q_n = \left(1+\sqrt{2}\right)^n + \left(1-\sqrt{2}\right)^n; \quad q_n = \frac{\left(1+\sqrt{2}\right)^n + \left(1-\sqrt{2}\right)^n}{2},$$

respectively.

Finally, for the Jacobsthal and Jacobsthal-Lucas sequences, the characteristic equation is $r^2 - r - 2 = 0$ with two distinct roots, $r_1 = -1$ and $r_2 = 2$, being

$$J_n = \frac{1}{3}\left(2^n - (-1)^n\right); \quad j_n = 2^n + (-1)^n,$$

the respective Binet's formulas.

It should be noted that many other results related to these sequences can be found in the literature.

Some of these sequences were generalised for any positive real number k. The studies of k-Fibonacci sequence, k-Lucas sequence, k-Pell sequence, k-Pell-Lucas sequence, Modified k-Pell sequence, k-Jacobhstal and k-Jacobhstal-Lucas sequence are examples of these generalizations and a look about them is done in the next subsections.

Also quaternions and octonions sequences as well as polynomials sequences were and continue to be studied by the authors of this work. However, in this article it is our purpose to make a retrospective of only some numerical sequences introduced and studied by this research group.

2.1 The Sequences of k-Pell, k-Pell-Lucas and Modified k-Pell Numbers

In 2013, the first author published the first paper in this topic related with the sequence of k-Pell numbers (see [16], for more detail). In this research work, the Binet's formula for k-Pell numbers is provided and as a consequence some properties for k-Pell numbers are established. Also the generating function for k-Pell sequences and another expression for the general term of the sequence, using the ordinary generating function, are presented. Following the Horadam's idea, the k-Pell sequence is the special case given by $\{w_n(0, 1; 2, -k)\} \equiv \{P_{k,n}\}$, with initial conditions $P_{k,0} = 0$, $P_{k,1} = 1$.

Also in 2013, the author and some colleagues from the research group studied other generalizations of these sequences, namely the k-Pell-Lucas and the Modified k-Pell sequences (see [17–19], for more detail). In the case of Modified k-Pell sequence, it is defined by $\{w_n(1, 1; 2, -k)\} \equiv \{q_{k,n}\}$, with initial conditions $q_{k,0} = 1$, $q_{k,1} = 1$ and for the k-Pell-Lucas, we have $\{w_n(2, 2; 2, -k)\} \equiv \{Q_{k,n}\}$, with initial conditions $Q_{k,0} = 2$, $Q_{k,1} = 2$. About the characteristic equation, we have $r^2 - 2r - k = 0$ with two distinct roots, $r_{1,2} = 1 \pm \sqrt{1+k}$, giving for Binet's formulas,

$$P_{k,n} = \frac{{r_1}^n - {r_2}^n}{r_1 - r_2}; \quad Q_{k,n} = {r_1}^n + {r_2}^n; \quad q_{k,n} = \frac{{r_1}^n + {r_2}^n}{2}.$$

As a curiosity, for $k = 1$, we obtain that r_1 is the silver ratio which is related with the Pell number sequence. Silver ratio is the limiting ratio of consecutive Pell numbers. Sometimes some basic properties of these sequences come from the Binet's formula. In other paper [20], these sequences are considered to achieve some sums and certain products involving their terms. For instance, the next results (see Propositions 1 and 2 in [20]) can give an idea of some of the properties studied.

Proposition 1 *For any positive real number k, if $P_{k,j}$, $Q_{k,j}$ and $q_{k,j}$ are the jth k-Pell, k-Pell-Lucas and Modified k-Pell numbers, respectively then we have:*

(1) $\sum_{j=0}^{n} \left(P_{k,j} + Q_{k,j}\right) = \frac{1}{k+1}\left[-1 + (1+2k)P_{k,n+1} + P_{k,n+2}\right];$
(2) $\sum_{j=0}^{n} \left(P_{k,j} + q_{k,j}\right) = \frac{1}{k+1}\left[-1 + kP_{k,n} + (k+2)\,P_{k,n+1}\right];$
(3) $\sum_{j=0}^{n} \left(Q_{k,j} + q_{k,j}\right) = \frac{3}{2(k+1)}\left(Q_{k,n+1} + kQ_{k,n}\right) = \frac{3}{k+1}\left(q_{k,n+1} + kq_{k,n}\right).$

Proposition 2 *For any positive real number k, if $P_{k,j}$, $Q_{k,j}$ and $q_{k,j}$ are the jth k-Pell, k-Pell-Lucas and Modified k-Pell numbers, respectively then we have:*

(1) $\sum_{j=0}^{n} P_{k,j} Q_{k,j} = \frac{1}{3-k}\left[\frac{2}{k+1}(-1 + P_{k,2n+1} + kP_{k,2n}) - kP_{k,2n}\right]$;

(2) $\sum_{j=0}^{n} P_{k,j} q_{k,j} = \frac{1}{2(3-k)}\left[\frac{2}{k+1}(-1 + P_{k,2n+1} + kP_{k,2n}) - kP_{k,2n}\right]$; ;

(3) $\sum_{j=0}^{n} Q_{k,j} q_{k,j} = (1-k)\left(q_{k,2} + q_{k,2}q_{k,3} + \cdots + q_{k,n-1}q_{k,n}) + q_{k,n}q_{k,n+1} - k\right)$.

In 2015, in [21], the authors present some identities involving terms of k-Pell, k-Pell-Lucas and Modified k-Pell sequences. Also some results on the column and row norms of Hankel matrices which entries are numbers of these sequences are given.

In 2019, in [22], the authors define the generalised Fibonacci sequence and k-Pell sequence. After that, by using these sequences, the authors delineated the generalised Fibonacci matrix sequence and k-Pell matrix sequence. Also, results by some matrix technique for both general sequences as well as for matrix sequences are provided. Also in 2019, in [23] was presented the (s,t)-type sequences and the (s,t)-type matrices sequences and then the authors study the case of (s,t)-Generalised Pell Sequence and its Matrix Sequence. Gaussian Modified Pell sequence is defined by the authors in the study [24]. Some properties involving this sequence, including the Binet-style formula and the generating function are also presented.

In 2020, in the paper [6], the authors defined the k-Pell, k-Pell-Lucas and Modified k-Pell numbers with arithmetic indexes and they study some properties of these sequences of numbers and give their generating functions. These are another type of generalisation of Fibonacci numbers.

2.2 The Sequence of k-Fibonacci and k-Lucas Numbers

In 2014, once more the first author published a research article related with the sequence of k-Fibonacci numbers (see [25], for more detail). We obtain some identities for k-Fibonacci numbers by using their Binet's formulae. Also, another expression for the general term of the sequence, using the ordinary generating function, is provided in this paper.

This sequence is defined as follows $\{w_n(0, 1; k, -1)\} \equiv \{F_{k,n}\}$, with initial conditions $F_{k,0} = 0$, $F_{k,1} = 1$, characteristic equation $r^2 - kr - 1 = 0$ with two distinct roots, $r_{1,2} = \frac{k \pm \sqrt{k^2+4}}{2}$ and the Binet formulas

$$F_{k,n} = \frac{{r_1}^n - {r_2}^n}{r_1 - r_2}; \quad L_{k,n} = {r_1}^n + {r_2}^n.$$

These sequences generalise, between others, both the classical Fibonacci sequence and the Pell sequence. Some identities are stated and studied as we can see in the next results (see Propositions 3 and 4, that correspond, respectively to Propositions 2 and 4 of [25]):

Proposition 3 (Catalan's identity)

$$F_{k,n-r}F_{k,n+r} - F_{k,n}^2 = (-1)^{n+1-r}F_{k,r}^2.$$

Proposition 4 (d'Ocagne's identity) *If* $m > n$ *then*

$$F_{k,m}F_{k,n+1} - F_{k,m+1}F_{k,n} = (-1)^n F_{k,m-n}.$$

Using the generating function, other expression of the general term of k-Fibonacci sequence was also analysed as we can see in what follows (see Proposition 5, that is the Proposition 7 of [25])).

Proposition 5 *Let us consider* $f(x) = \sum_{n=0}^{\infty} F_{k,n}\, x^n$, *for* $x \in \left]-\frac{1}{\alpha_1}, \frac{1}{\alpha_1}\right[$. *Then we have that*

$$F_{k,n} = \frac{f^{(n)}(0)}{n!}$$

where $f^{(n)}(x)$ *denotes the* nth *order derivative of the function* f.

Once more in 2014, in the paper [26], some basic properties of the k-Fibonacci and k-Lucas sequences are provided. Their relationship allows us to obtain some new properties involving sums and products of terms of these sequences. Furthermore, as a consequence of the Binet's formula for each one, we also proved some identities involving these sequences. In the paper [27], the authors derive the explicit formula for the term of order n of the k-Fibonacci and k-Lucas sequences and also get the well-known Cassini's identity using some tools of linear algebra. The Binet's formulas of both sequences are also deduced from the diagonalization of the respective generating matrices.

In 2018, in the paper [28] the first author with other researchers introduced a new generalisation of Fibonacci sequence and that was called as k-Fibonacci-Like sequence. They studied some fundamental properties for k-Fibonacci-Like sequence and also they present some relations among k-Fibonacci-Like sequence, k-Fibonacci sequence and k-Lucas sequence by some algebraic methods.

2.3 The Sequence of k-Jacobsthal and k-Jacobsthal-Lucas Numbers

In 2014, in the paper [5], it was presented the sequence of the k-Jacobsthal-Lucas numbers that generalizes the Jacobsthal-Lucas sequence introduced by Horadam in 1988. For this new sequence we established an explicit formula for the term of order n, the well-known Binet's formula, Catalan's and d'Ocagne's Identities and a generating function. Using the Horadam's idea, these sequences are defined by $\{w_n(0, 1; k, -2)\} \equiv \{J_{k,n}\}$ with initial conditions $J_{k,0} = 0$, $J_{k,1} = 1$, for the case of k-Jacobsthal sequence, and $\{w_n(2, k; k, -2)\} \equiv \{j_{k,n}\}$ with initial conditions $j_{k,0} = 2$, $j_{k,1} = k$, for the case of k-Jacobsthal-Lucas. The characteristic equation, for both cases, is $r^2 - kr - 2 = 0$, that has two distinct roots, $r = \frac{k\pm\sqrt{k^2+8}}{2}$ and the respective Binet's formulas are the following

$$J_{k,n} = \frac{{r_1}^n - {r_2}^n}{r_1 - r_2}; \quad j_{k,n} = {r_1}^n + {r_2}^n.$$

The use of the Binet's formula of each sequence and the fact that $r_1 r_2 = -2$ allows us to obtain Catalan's Identity for the k-Jacobsthal-Lucas as we have in the next result, Proposition 6 (see Proposition 2.2 of [5]).

Proposition 6 (Catalan's identity)

$$j_{k,n-r}\, j_{k,n+r} - j_{k,n}^2 = (-2)^{n-r}\left(j_{k,r}^2 - (-2)^{r+2}\right).$$

Finally, we have one more identity involving the k-Jacobsthal-Lucas sequence, Proposition 7 (see Proposition 2.4 of [5]).

Proposition 7 (d'Ocagne's identity) *For $m > n$,*

$$j_{k,m}\, j_{k,n+1} - j_{k,m+1}\, j_{k,n} = (-2)^n\sqrt{k^2+8}\left(j_{k,m-n} - 2^{n-m+1}\left(k + \sqrt{k^2+8}\right)^{m-n}\right).$$

The generating function for the k-Jacobsthal-Lucas sequence is given by Proposition 8 (see Proposition 3.1 of [5]).

Proposition 8

$$j_k(x) = \frac{2 - kx}{1 - kx - 2x^2}.$$

In the paper [29], new families of sequences that generalize the Jacobsthal and the Jacobsthal-Lucas numbers are presented and some identities are established. The authors also give a generating function for a particular case of the sequences presented. The new sequence is given as follows in Definition 1 of [29].

Definition 1 Let n be a nonnegative integer and k be a natural number. By the division algorithm there exist unique numbers m and r such that $n = mk + r$ $(0 \leq r < k)$. Using these parameters we define the new generalized Jacobsthal and generalized Jacobsthal-Lucas numbers

$$J_n^{(k)} = \frac{1}{(r_1 - r_2)}\ \left(r_1^{m+1} - r_2^{m+1}\right)^r\ \left(r_1^m - r_2^m\right)^{k-r}$$

and

$$j_n^{(k)} = (r_1^{m+1} + r_2^{m+1})^r \ \ (r_1^m + r_2^m)^{k-r}$$

where $r_1 = 2, r_2 = -1$, respectively.

In 2018, the aim of the work [30] was to introduce new difference sequences by the application of the concept of difference relation to the sequences of k-Jacobsthal and k-Jacobsthal-Lucas numbers. Some algebraic properties of these sequences were studied and in addition, the author have given the Binet formula and generating functions satisfied by each of these sequences.

2.4 The Balancing, Cobalancing, Lucas-Balancing and Lucas-Cobalancing Numbers

Many authors have dedicated their research to the study of these sequences and also to the generalisations of the theory of the sequences of balancing, cobalancing, Lucas-balancing and Lucas-cobalancing numbers. Behera and Panda [31] observed that n is a balancing number if and only if n^2 is a triangular number, that is $8n^2 + 1$ is a perfect square and the square of a balancing number is a square triangular number, that is, ${B_n}^2 = ST_n$, where ST_n denotes the nth square triangular number. The recurrence relations and the initial conditions of these sequences are the following: $B_{n+1} = 6B_n - B_{n-1} (n \geq 1)$, with $B_0 = 0$, $B_1 = 1$; $b_{n+1} = 6b_n - b_{n-1} + 2 (n \geq 1)$, with $b_1 = 0, b_2 = 1$; $C_{n+1} = 6C_n - C_{n-1} (n \geq 1)$, with $C_0 = 1, C_1 = 3$; $c_{n+1} = 6c_n - c_{n-1} (n \geq 1)$, with $c_0 = 1, c_1 = 7$, where B_n, b_n, C_n, c_n denotes the n^{th} term of balancing, cobalancing, Lucas-balancing and Lucas-cobalancing, respectively.

In all cases the characteristic equation associated to the recurrence relations is given by $r^2 - 6r + 1 = 0$ with two distinct roots, $r_{1,2} = 3 \pm 2\sqrt{2}$. The Binet formula of two sequences mentioned before are

$$B_n = \frac{{r_1}^n - {r_2}^n}{r_1 - r_2}; C_n = \frac{{r_1}^n + {r_2}^n}{2}.$$

Because of $r_{1,2} = {\sigma_{1,2}}^2$ with $\sigma_{1,2}$ the two distinct roots of $r^2 - 2r - 1 = 0$, then the Binet formulas of the last two sequences are

$$b_n = \frac{{\sigma_1}^{2n-1} + {\sigma_2}^{2n-1}}{4\sqrt{2}} - \frac{1}{2}; c_n = \frac{{\sigma_1}^{2n-1} + {\sigma_2}^{2n-1}}{2}.$$

As a consequence of the Binet formula for balancing, cobalancing, Lucas-balancing and Lucas-cobalancing numbers, in [7], the authors provide some formulas for these sequences explicitly, which can have certain importance or applications in most recent investigations in this area. Also the authors give another expression for the general term of each sequence, using the ordinary generating function.

2.5 The Incomplete k-Pell, Incomplete k-Pell-Lucas and Incomplete Modified k-Pell Numbers

In the paper [8], it was defined the incomplete k-Pell, k-Pell-Lucas and Modified k-Pell numbers, it was studied the recurrence relations, some properties of these sequences of integers and their generating functions. In the definition of incomplete k-Pell number we use a combinatorial tool as follows: with k any integer, the Incomplete k-Pell numbers are defined by

$$P_{k,n}^{l} := \sum_{j=0}^{l} \binom{n-1-j}{j} 2^{n-1-2j} k^{j}, \quad 0 \leq l \leq \left\lfloor \frac{n-1}{2} \right\rfloor; \; n \in \mathbb{N}.$$

From this definition, we present a few incomplete k-Pell numbers as we can see in the following Table 1 (Table of Incomplete numbers of [8]).

In the next result, Proposition 9 (Proposition 3.2 of [8]), we present the recurrence relation verified by these numbers:

Proposition 9 *With k any integer, the recurrence relation of the incomplete k-Pell numbers $P_{k,n}^{l}$ is*

$$P_{k,n+2}^{l+1} = 2P_{k,n+1}^{l+1} + kP_{k,n}^{l}, \quad \left(0 \leq l \leq \frac{n-2}{2}; \; n \in \mathbb{N}\right).$$

Table 1 The incomplete $P_{k,n}^{l}$ for $1 \leq n \leq 8$

n	j			
	0	1	2	3
1	1			
2	2			
3	4	$4+k$		
4	8	$8+4k$		
5	16	$16+12k$	$k^2+12k+16$	
6	32	$32+32k$	$6k^2+32k+32$	
7	64	$64+80k$	$24k^2+80k+64$	$k^3+24k^2+80k+64$
8	128	$128+192k$	$80k^2+192k+128$	$8k^3+80k^2+192k+128$

3 Conclusions

It is our intention to continue to work in this topic, if possible to introduce new sequences of numbers, polynomials, quaternions, octonions, etc., studing several algebraic properties, matrices whose entries are elements of these sequences and also some combinatorial aspect involving them. We do not forget their applications in several fields, in particular in Cryptography.

Acknowledgements The first author is member of the Research Centre CMAT-UTAD and also a collaborating member of the Research Centre CIDTFF. The second author is member of the Research Centre CIDTFF. This research was partially financed by Portuguese Funds through FCT -Fundação para a Ciência e a Tecnologia, within the Projects UIDB/00013/2020, UIDP/00013/2020 and UIDB/00194/2020.

References

1. Dil, A., Mezo,I., A Symmetric algorithm hyperharmonic and Fibonacci Numbers. Appl. Math. Comp., (206): 942–951, 2006.
2. Hoggatt, V. E., Fibonacci and Lucas Numbers, A publication of the Fibonacci, Association. University of Santa Clara, Santa Clara. Houghton Mifflin Company, 1969.
3. Vorobiov, N. N., Números de Fibonacci, Editora MIR, URSS, 1974.
4. Vajda, S., Fibonacci and Lucas numbers, and the Golden section, theory and applications. Ellis Horwood limited, 1989.
5. Campos, H., Catarino, P., Aires, A. P., Vasco, P. and Borges, A., On Some Identities of k-Jacobsthal-Lucas Numbers, Int. Journal of Math. Analysis, 8(10): 489–494, 2014.
6. Catarino, P., Vasco, P. and Campos, H., A note on k-Pell, k-Pell-Lucas and modified k-Pell Numbers with arithmetic indexes, Acta Mathematica Universitatis Comenianae, 89(1): 97–107, 2020.
7. Catarino, P., Vasco, P., and Campos, H., On some identities for balancing and cobalancing numbers, Annales Mathematicae et Informaticae, 45: 11–24, 2015.
8. Catarino, P. and Campos, H., Incomplete k-Pell, k-Pell-Lucas and modified k-Pell numbers, Hacettepe journal of mathematics and statistics, 46(3): 361–372, 2017.
9. Catarino, P., Morais, E. and Campos, H., A note on k-Telephone and Incomplete k-Telephone numbers, 2021. (to appear)
10. Catarino, P. and Borges, A., On Leonardo numbers, Acta Mathematica Universitatis Comenianae, 89(1): 75–86, 2019.
11. Catarino, P. and Borges, A., A Note on Incomplete Leonardo numbers, Integers, 20, 2020.
12. Vieira, R. P. M., Alves, F. R. V. and Catarino, P., Relações Bidimensionais e Identidades da sequência de Leonardo, Revista Sergipana de Matemática e Educação Matemática, 2: 156–173, 2019.
13. Catarino, P., Alves, F. R. V. and Campos, H., On Hyper k-Pell, Hyper k-Pell-Lucas and Hyper Modified k-Pell sequences, 2021. (to appear)
14. Horadam, A. F., Basic properties of a certain generalized sequence of numbers. Fibonacci Quart. 3(3): 161–176, 1965.
15. Horadam, A. F., Generating functions for powers of a certain generalized sequences of numbers. Duke Math. J. 32(3): 437–446, 1965.
16. Catarino, P., On Some Identities and Generating Functions for k-Pell Numbers. Int. J. Math. Anal. (Ruse), 7(38): 1877–1884, 2013.

17. Catarino, P. and Vasco, P. On some Identities and Generating Functions for k-Pell-Lucas sequence. Appl. Math. Sci, 7(98): 4867–4873, 2013.
18. Catarino, P. and Vasco, P. Modified k-Pell Sequence: Some Identities and Ordinary Generating Function. Appl. Math. Sci, 7(121): 6031–6037, 2013.
19. Catarino, P. and Vasco, P. Some basic properties and a two-by-two matrix involving the k-Pell Numbers. Int. J. Math. Anal. (Ruse), 7(45): 2209–2215, 2013.
20. Vasco, P. and Catarino, P. Sums and products involving terms of k-Pell, k-Pell-Lucas and Modified k-Pell sequences, JP Journal of Algebra, Number Theory and Applications, 32(2): 87–98, 2014.
21. Vasco, P., Catarino, P., Campos,H., Aires, A. P. and Borges, A., k-Pell, k-Pell-Lucas and modified k-Pell numbers: Some identities and norms of Hankel matrices, International Journal of Mathematical Analysis, 9(1): 31–37, 2015.
22. Wani, A. A., Badshah, V.H., Rathore,G. B. S. and Catarino, P., Generalized Fibonacci and k-Pell matrix sequences, Punjab University Journal of Mathematics, 50(1): 17–28, 2019.
23. Wani, A. A., Catarino,P. and Halici, S. On a Study of (s, t)-Generalized Pell Sequence and its Matrix Sequence, Punjab University Journal of Mathematics, 51(9): 17–32, 2019.
24. Catarino, P. and Campos, H., A note on Gaussian Modified Pell numbers, Journal of Information and Optimization Sciences, 39(6): 1363–1371, 2018.
25. Catarino, P., On some identities for k-Fibonacci sequence, Int. J. Contemp. Math. Sciences, 9(1): 37–42, 2014.
26. Catarino, P., Vasco, P., Borges, A., Campos,H. and Aires, A. P., Sums, Products and identities involving k-Fibonacci and k-Lucas sequences, JP Journal of Algebra, Number Theory and Applications, 32(2): 63–77, 2014.
27. Borges, A., Catarino, P., Aires, A. P., Vasco, P. and Campos, H., Two-by-Two Matrices Involving k-Fibonacci and k-Lucas Sequences, Applied Mathematical Sciences, 34(8): 1659–1666, 2014.
28. Wani, A. A., Catarino, P. and Rafiq, R., On the Properties of k-Fibonacci-Like Sequence, Int. J. Math. And Appl., 6(1-A): 187–198, 2018.
29. Catarino, P., Vasco P., Campos, H., Aires, A. P. and Borges, A., New families of Jacobsthal and Jacobsthal-Lucas numbers. Algebra and Discrete Mathematics, 20(1): 40–54, 2015.
30. Catarino, P., On Jacobsthal difference sequences, Acta Math. Univ. Comenianae, LXXXVII(2): 267–276, 2018.
31. Behera, A. and Panda, G. K., On the square roots of triangular numbers, Fibonacci Quart., 37(2): 98–105, 1999.

Extension of Leap Condition in Approximate Stochastic Simulation Algorithms of Biological Networks with 2nd and 3rd order Taylor Expansion

Saliha Demirbüken and Vilda Purutçuoğlu

Abstract The approximate stochastic simulation algorithms are the alternative methods to simulate the complex biological systems with a loss in accuracy by acquiring from computational demand. These methods depend on the leap condition. Here, the study aims to construct an actual and close confidence interval for the parameter denoting the number of simultaneously reaction in the system, by expanding the leap condition and the hazard function by second and third order Taylor expansion in the same time. To reach the goal, we use the poisson τ-leap and approximate Gillespie algorithm. Moreover, we derive the maximum likelihood estimators (MLE) and the method of moment estimators (MME) of the simulation parameters and construct confidence interval estimators at a given significance level α for these extended version of algorithms. Finally, we theoretically present that the obtained k can generate more narrower results [1–5, 7, 10].

Keywords Approximate stochastic simulation algorithms · Leap condition · Confidence interval · Taylor expansion

1 Introduction

There are many occuring chemical reactions in the biological systems. Stochastic Simulation Algorithms (SSAs) enable to simulate these reactions in the time evolution. Three main SSAs are commonly used in the literature. These are Gillespie

S. Demirbüken
Scientific Computing, Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey
e-mail: salihademirbuken@gmail.com

Department of Mathematics, Faculty of Science, Gebze Technical University, Kocaeli, Turkey

V. Purutçuoğlu (✉)
Department of Statistics, Faculty of Arts and Sciences, Middle East Technical University, Ankara, Turkey
e-mail: vpurutcu@metu.edu.tr

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_25

algorithm, the first reaction method and the next reaction method. Although these algorithms give exact generation of the system, they are computationally costly. So approximate SSAs can be used since they are faster than SSA. Nevertheless, approximate SSAs generate less accurate results. Primarily, these methods depend on the leap condition. In other words, there should be no significant change in the values of the propensity function during the time change from t to $t+\tau$, i.e, $[t, t+\tau)$ [4]. That is [11],

$$|h_j(Y+\bar{\lambda})(Y,\tau) - h_j(Y)| \leq \epsilon h_0(Y),$$

where $h_0(Y) = \sum_{j=1}^{r} h_j(Y)$ is the sum of all hazard functions $h_j(Y)$, ϵ denotes the error control parameter and $\bar{\lambda}$ presents

$$\bar{\lambda}(Y,\tau) = \sum_{j=1}^{r} [h_j(Y)\tau]\nu_j = \tau\xi(Y).$$

The above expression shows the expected net change in the state for the given time interval when the system has r members of reactions. In this equation, ν_j represents the stoichiometric coefficients of the reaction j which corresponds to the jth row of the net effect matrix V.

Basically, estimators are used to find the plausible value of $\hat{Y}$ by using unknown population parameter. Hereby, there are different kinds of estimators. In this study, especially, maximum likelihood estimators (MLE) and method of moment estimators (MME) are handled to deduce the parameters such as k, which is the number of simulatenous reactions in the system and τ which is the time interval to specify k. Shortly, the likelihood function can be written as follows

$$L_n(\theta) = L_n(\theta, y) = f_n(y, \theta),$$

where the observed data set is denoted by $y = (y_1, y_2, \ldots, y_n)$, associated with a vector $\theta = [\theta_1,\ \theta_2,\ \ldots,\ \theta_k]^{\mathsf{T}}$ of parameters that indices the probability distribution within a parametric family$\{f(\cdot\,;\theta) \mid \theta \in \Theta\}$. Thus, this is called the parameter space and $f_n(y;\theta)$ is the product of univariate density functions. Therefore, the aim of MLE is to find the values of the model parameters that maximize the likelihood function over the parameter space, i.e,

$$\hat{\theta} = \underset{\theta\in\Theta}{\arg\max}\ \widehat{L}_n(\theta\,; y).$$

On the other hand, in addition to MLE, MME can be described in the following way. Suppose that the problem is to estimate k unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$ characterizing the distribution $f_W(w;\theta)$ of the random variable W. Also, assume that the first k moments of the true distribution, i.e, the "population moments", can be expressed as the functions of θ's via

$$
\begin{aligned}
\mu_1 &\equiv \mathrm{E}[W] = g_1(\theta_1, \theta_2, \ldots, \theta_k),\\
\mu_2 &\equiv \mathrm{E}[W^2] = g_2(\theta_1, \theta_2, \ldots, \theta_k),\\
&\vdots\\
\mu_k &\equiv \mathrm{E}[W^k] = g_k(\theta_1, \theta_2, \ldots, \theta_k).
\end{aligned}
$$

Accordingly, if a sample of size n is drawn, resulting in the values $w_1, \ldots, w_n$, $j = 1, \ldots, k$, the estimated mean of these values μ_j can be found by

$\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} w_i^j$ as the jth sample moment. As a result, the method of moments estimator for $\theta_1, \theta_2, \ldots, \theta_k$ denoted by $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k$ can be described as the solution of the equations:

$$
\begin{aligned}
\widehat{\mu}_1 &= g_1(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k),\\
\widehat{\mu}_2 &= g_2(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k),\\
&\vdots\\
\widehat{\mu}_k &= g_k(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k).
\end{aligned}
$$

Furthermore, a confidence interval can be defined as an estimated range of values which is likely to include an unknown population parameter. This estimated range is evaluated from a given set of sample data.

Hence, in this study, by expanding net change of hazard function via 2nd and 3rd order Taylor expansion, the aim is to construct the confidence intervals for the population parameters k and τ in the two approximate SSAs, namely, the poisson tau-leap and the approximate Gillespie method, by using MLE and MME approaches. All these approaches are based on the leap condition. In the current literature, the k and τ in these three simulation approaches have been used via a conservative one-sided confidence interval without controlling the significance level α. Hence, this study suggests realistic and accurate confidence intervals for both parameters by controlling α and by using the MLE and MME of the modal parameters so that narrower and more accurate confidence intervals can be obtained theoretically. Thereby, in the organization of the paper, we demonstrate the selected three major approximate SSAs in Sect. 2. We show our confidence intervals inserted to the leap condition of the underlying approximate SSAs in Sect. 3. Finally, we conclude our results in Sect. 4 [7, 8].

2 Approximate Stochastic Simulation Algortihms

Despite the fact that advancement of the SSA described previously gives helpful results, however, focused on being efficient, it is slow and computational costly. To speed the time of the simulation, there could be some sacrifice in the exactness of the SSA. Using approximate SSAs are one way to do this [5]. These fast algorithms are depend on a time discretisation of the Markov process [12]. Here, the main concept is that, firstly, the time is seperated into small different pieces, leap. Then, to be possible to continue with improvement of the state from the start of one piece to another, the basis kinetics are approximated [12]. Mostly, these algorithms act with the assumption of leap condition. In other words, the selected time interval τ should be satisfied that there is no notable change in the propensity function during the time change from t to $t + \tau$. In addition, since SSAs are computational costly, approximate SSAs make possible to obtain less computational demand.

2.1 Poisson τ-Leap Method

Under the leap condition, the aim of this method is improve intervals between selecting times with selected the time interval τ as large as possible [4, 11]. Here, for each reaction channel R_j, a random value k_j is generated from a Poisson distribution by $Poi(h_j(Y)\tau)$ in the time interval $[t, t + \tau]$, where $Y(t) = Y$ is a state vector. Then, an acceptable τ is obtained by substituting it boundary for the difference between

$$h_j(Y + \lambda(Y)) - h_j(Y),$$

where $\lambda(Y) = \sum_{j=1}^{r} k_j v_j$ denotes the net change in the state of the system in $[t, t + \tau]$. As $k_j \sim Poi(h_j\tau)$, mean of k_j equals $h_j\tau$, i.e., $E(k_j) = h_j(Y)\tau$ and

$$\bar{\lambda}(Y, \tau) = \sum_{j=1}^{r} [h_j(Y).\tau]v_j = \tau\xi(Y) \tag{1}$$

that gives the expected net change in the state for the given time interval. In this equality, j represents the stoichiometric coefficients of the reaction j corresponding to the jth row of the net effect matrix V as defined beforehand and $h_j(Y)$ corresponds the hazard function of the jth reaction which is found by the product of the rate constant c_j and distinct molecular reactant combination of underlying reaction. Subsequently, $\xi(Y) = \sum_{j=1}^{r} h_j v_j$ can be represented as the mean or expected state change in a unit of time by an n-dimensional vector where each ith component, $\xi_j(Y)$, subtends to the expected change of the ith species in an unit of time. Afterwards, the following inequality is obtained

$$|h_j(Y + \bar{\lambda})(Y, \tau) - h_j(Y)| \le \epsilon h_0(Y) \tag{2}$$

by using $\lambda(Y, \tau)$ in Eq. (1). It can be inferred from Eq. (2) in the following. The expected changes in the hazard functions in time τ is restricted by a fraction ϵ, error control parameter lying $0 < \epsilon < 1$, and the sum of all hazard functions $h_0 = \sum_{j=1}^{r} h_j(Y)$. In fact, this inequality presents a leap condition. Consequently, using the first order Taylor expansion helps to predict the difference on the left hand side of Eq. (2). After the application of the Taylor expansion, the following equality is obtained.

$$|h_j(Y + \bar{\lambda}(Y, \tau)) - h_j(Y)| \approx \bar{\lambda}(Y, \tau)h_j(Y) = \sum_{i=1}^{n} \tau\xi_j(Y)\frac{\partial h_j(Y)}{\partial Y_i}. \tag{3}$$

Then, setting $b_{ji}(Y) = \frac{\partial h_j(Y)}{\partial Y_i}$ $(i = 1, \ldots, n,\ j = 1, \ldots, r)$, the below inequality can be shown

$$\tau|\sum_{i=1}^{n} \xi_j(Y)b_{ji}(Y) \leq \epsilon h_0(Y)|.$$

Consequently, the largest value of τ satisfying the leap condition for the given Y and the preselected ϵ is calculated by

$$\tau = min\left\{\frac{\epsilon h_0(Y)}{|\sum_{i=1}^{n} \xi_j(Y)b_{ji}(Y)|}\right\}. \tag{4}$$

After all, using the exact SSA is more preferable rather than Eq. (4) since the obtained value of τ is favorable for the leap size. The obtained value of τ in Eq. (4) would not be selected if $\tau \leq \frac{1}{h_0(Y)}$ as $\tau = \frac{1}{h_0(Y)}$ is given from SSA. Not considering of the computational cost, the time interval in the Poisson τ-leap methos is more suitable than the time of SSA. Essentially, there is an incremental difference between them.

In final step, there is an update of the current state in the Poisson τ-leap methods by replacing t by $t := t + \tau$. In addition, for Y, there is a necessity to determine the largest value of τ and to be adaptable with the leap condition.

Beside this, in the long-run simulations, Poisson τ-leap algorithm can create problem of negative molecular populations from the application of this method in various systems. To overcome this problem, there are some suggested solutions. One of the well-known alternative solutions is the Binomial τ-leap approach. Although it does not give accurate result to have smooth approximation of exact SSAs, it overcomes the negativity problem [6, 7, 9, 11].

2.2 Approximate Gillespie Algorithm

As an alternative approaches of the poisson τ-leap method, the approximate Gillespie algorithm can be considered. Basically, the approximate Gillespie algorithm [11], depending on the extension of the exact Gillespie method, states that k numbers of reactions, generated from the Gamma distribution with a parameter $\sum_{j=1}^{r} h_j(Y)$, where each of them occurs in an exponential time step t, which is performed rather than a single reaction at a time. Then, it can be shown that

$\tau \sim \Gamma(k, h_0(Y))$, where τ presents the time interval of k reactions in the total hazard, $h_0(Y)$, $h_0(Y) := \sum_{j=1}^{r} h_j(Y)$. In this case, the system is updated by replacing t by $t := t + \tau$ and by changing the current state Y by $Y := Y + \lambda(Y)$, where the net change in the state is found via $\lambda(Y) = \sum_{j=1}^{r} k_j v_j$. In this expression v_j is the net effect of the jth reaction by showing the jth row of the net effect matrix V as used previously. By this way, it is assumed that the essential time for every reaction correspons to that of Gillespie. Under this assumption, the total number of reactions during the interval τ is determined by controlling k in each time interval. For this purpose, it is initially described a k satisfying the leap condition in each time step. Then, the change in hazard function $\Delta h_j(Y)$, $(j = 1, ..., r)$ is approximated by the first order Taylor expansion in the time interval $[t, t + \tau]$, in a such way that the following equality can be obtained as performed in the poisson τ-leap approach.

$$\Delta h_j(Y) = h_j(Y + \lambda(\bar{Y}, \tau)) - h_j(Y) \approx \lambda(\bar{Y}, \tau) h_j(Y) = \sum_{i=1}^{n} \lambda(\bar{Y}, \tau) \frac{\partial h_{ij}(Y)}{\partial Y_i} \quad (5)$$

in which the expected change in the state by regarding k simultaneous reaction computed by

$$\lambda(\bar{Y}, \tau) = Y(t + \tau) - Y(t) = \sum_{j=1}^{r} k_j \nu_j.$$

In the above expression, k_j shows the number of times of the jth reaction and ν_j is the net effect of the jth reaction by denoting the jth row of the net effect matrix V as before. Hence, by using a gamma distribution, we can show $\tau \sim \Gamma(k, h_0(Y))$ where $k = E(\tau).h_0(Y)$. In this expression, $E(\tau)$ illustrates the expected τ on average.

Then, by inserting this k into Eq. (5), the approximation of

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} f_{jj'}(Y) \tau h_0(Y),$$

is acquired where the total change in hazard of reaction j' is described in terms of $f_{jj'}(Y)$ via

$$f_{jj'}(Y) = \sum_{i=1} \frac{\partial h_j(Y)}{\partial Y_i} \nu_{ij}$$

for the execution of the reaction j'. Finally, in order to obtain the confidence interval, the following expression is written as

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm \sqrt{Var(\Delta h_j(Y))}, \quad (6)$$

where $Var(.)$ denotes the variance of the given random variable. Then, the statistics for $\Delta h_j(Y)$ can be shown by

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r} f_{jj'}(Y)E(\tau)h_0(Y) = \sum_{j=1}^{r} f_{jj'}(Y)\frac{k}{h_0(Y)}h_0(Y) = k\sum_{j=1}^{r} f_{jj'}(Y) \tag{7}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} f_{jj'}^2(Y)Var(\tau)h_0(Y) = \sum_{j=1}^{r} f_{jj'}^2(Y)\frac{k}{h_0(Y)}h_0(Y) = k\sum_{j=1}^{r} f_{jj'}^2(Y). \tag{8}$$

By substituting Eqs. (7) and (8) into the required leap condition, the below expression can be found

$$|k|\sum_{j'=1}^{r} f_{jj'}(Y) \leq \epsilon h_0(Y)$$

and

$$\sqrt{\left(k\sum_{j'=1}^{r} f_{jj'}^2(Y)\right)} \leq \epsilon h_0(Y).$$

Accordingly, the optimal k is computed from

$$k = \min_{j\in[1,r]}\left\lceil \left|\frac{\epsilon h_0(Y)}{\sum_{j'=1}^{r} f_{jj'}(Y)}\right|, \left|\frac{\epsilon^2 h_0^2(Y)}{\sum_{j'=1}^{r} f_{jj'}^2(Y)}\right| \right\rceil. \tag{9}$$

Indeed, mostly, inserting the distributions feature of k and τ into the leap condition and finding a conservative confidence interval has been derived for the poisson distribution too [6]. But in these studies, the confidence intervals are constructed one-sided and by taking a fixed significance level α which roughly sets the tabulated value to 1. Furthermore, they generate large intervals which decreases the accuracy of the approximate algorithms. Hereby, the following part introduces an extension where the confidence intervals can produce more accurate results regarding previous studies [8, 9].

3 Leap Condition

In this study, $h_0(Y)$ is expanded by 2nd and 3rd order Taylor formula to improve the accuracy of the leap condition [9]. Thus, it aims to obtain an acceptable k by forming confidence interval with parameters of estimation deriving MLE and MME under Poisson and Gamma distribution, seperately.

3.1 Expansion of 2nd Order Taylor Expansion

The change hazard function $\Delta h_j(Y)$ is approximated by the second order Taylor expansion in the time interval $[t, t+\tau]$ in a such way that

$$h_j(Y+\bar{\lambda}(Y,\tau)) = h_j(Y) + \bar{\lambda}(Y,\tau)h'_j(Y) + \frac{\bar{\lambda}^2}{2}h''_j(Y) + \mathcal{O}^3(h_j(Y)),$$

$$h_j(Y+\bar{\lambda}(Y,\tau)) - h_j(Y) \approx \bar{\lambda}(Y,\tau)h'_j(Y) + \frac{\bar{\lambda}^2}{2}h''_j(Y)$$

$$\Delta h_j(Y) \approx \bar{\lambda}(Y,\tau)h'_j(Y) + \frac{\bar{\lambda}^2}{2}h''_j(Y), \tag{10}$$

where $h'_j(Y) = \sum_{i=1}^{n} \frac{\partial h_{ij}(Y)}{\partial Y_i}$ and $h''_j(Y) = \sum_{i=1}^{n} \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}$

$\bar{\lambda}(Y,t) = Y(t+\tau) - Y(t) = \sum_{j=1}^{r} k_j \nu_j$

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} k_j \nu_j h'_j(Y) + \frac{1}{2}\Big(\sum_{j=1}^{r} k_j \nu_j\Big)^2 h''_j(Y)$$

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} k_j \nu_j h'_j(Y) + \frac{1}{2}\sum_{j=1}^{r} k_j \nu_j \sum_{j=1}^{r} k_j \nu_j h''_j(Y) \tag{11}$$

Then, considering τ is generated from gamma distribution with parameters k and $h_0(Y)$, i.e., $\tau \sim \Gamma(k, h_0(Y))$, k can be written as $k = \tau.h_0(Y)$ by using its expectation. After substituting into Eq. (11), the following expressions are obtained.

$$\begin{aligned}\Delta h_j(Y) &\approx \sum_{j=1}^{r} \tau h_0(Y) h'_j(Y)\nu_j + \frac{1}{2}\sum_{j=1}^{r} \tau h_0(Y)\nu_j \sum_{j=1}^{r} \tau h_0(Y)\nu_j h''_j(Y) \\ &= \tau h_0(Y)\sum_{j=1}^{r} h'_j(Y)\nu_j + \frac{1}{2}\tau^2 h_0^2(Y)\sum_{j=1}^{r}\nu_j \sum_{j=1}^{r}\nu_j h''_j(Y).\end{aligned} \tag{12}$$

Setting $f_{jj'}(Y) = \sum_{i=1}^{r} \frac{\partial h_{ij}(Y)}{\partial Y_i}\nu_{ij} = h'_j(Y)\nu_j$ and $g_{jj'}(Y) = \nu_j \sum_{j=1}^{r} \nu_j \sum_{i=1}^{r} \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}\nu_{ij} = \nu_j \sum_{j=1}^{r} h''_j(Y)\nu_j$.

Accordingly, $\Delta h_j(Y)$ can be obtained in the following way.

$$\begin{aligned}\Delta h_j(Y) &\approx \tau h_0(Y)\sum_{j=1}^{r} f_{jj'}(Y) + \frac{1}{2}\tau^2 h_0^2(Y)\sum_{j=1}^{r} g_{jj'}(Y)\\ &= \sum_{j=1}^{r}\big(\tau h_0(Y) f_{jj'}(Y) + \frac{h_0^2(Y)}{2}\tau^2 g_{jj'}(Y)\big).\end{aligned} \tag{13}$$

We have that

$$\Delta h_j(Y) \approx E(\Delta h_j(Y)) \pm z_{\alpha/2}\sqrt{Var(\Delta h_j(Y))}. \tag{14}$$

Then, the mean of Eq. (13), $E(\Delta h_j(Y))$, is calculated by

$$\begin{aligned}E(\Delta h_j(Y)) &\approx E\Bigg(\sum_{j=1}^{r}\big(\tau h_0(Y) f_{jj'}(Y) + \frac{h_0^2(Y)}{2}\tau^2 g_{jj'}(Y)\big)\Bigg)\\ &= \sum_{j=1}^{r}\Bigg(E(\tau) h_0(Y) f_{jj'}(Y) + E(\tau^2)\frac{h_0^2(Y) g_{jj'}(Y)}{2}\Bigg).\end{aligned} \tag{15}$$

Since $E(\tau) = \frac{k}{h_0(Y)}$ and $E(\tau^2) = \frac{k(k+1)}{h_0^2(Y)}$,

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r}\Bigg(\Bigg(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\Bigg)^2 - \Bigg(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\Bigg)^2\Bigg). \tag{16}$$

Hence, by substituting this final expression into the leap condition, the underlying inequality can be acquired as below.

$$\Bigg|\sum_{j=1}^{r}\Bigg(\Bigg(k\sqrt{\frac{g_{jj'}(Y)}{2}} + \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\Bigg)^2 - \Bigg(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\Bigg)^2\Bigg)\Bigg| \le \epsilon h_0(Y) \tag{17}$$

Then, we get the following expressions.

$$k \le \frac{\sqrt{\epsilon h_0(Y) + \Bigg(\frac{(f_{jj'}(Y)+\frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}\Bigg)^2} - \sum_{j=1}^{r}\frac{(f_{jj'}(Y)+\frac{g_{jj'}(Y)}{2})}{\sqrt{2g_{jj'}(Y)}}}{\sum_{j=1}^{r}\sqrt{\frac{g_{jj'}(Y)}{2}}}. \tag{18}$$

In a similar way of calculation of $E(\Delta h_j(Y))$, $Var(\Delta h_j(Y))$ is derived in the following approach.

$$Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \left(h_0^2(Y) f_{jj'}(Y) Var(\tau) + \frac{h_0^4 g_{jj'}(Y)}{4} Var(\tau^2) + h_0^3(Y) f_{jj'}(Y) g_{jj'}(Y) Cov(\tau, \tau^2) \right). \tag{19}$$

As $\tau \sim \Gamma(k, h_0(Y)$, it is known that $Var(\tau) = \frac{k}{h_0^2(Y)}$, $Var(\tau^2) = \frac{4k^3+10k^2+6k}{h_0^4(Y)} = \frac{2k(k+1)(2k+3)}{h_0^4(Y)}$ and $Cov(\tau, \tau^2) = \frac{2k(k+1)}{h_0^3(Y)}$. By substituting these values into the equation of $Var(\Delta h_j(Y))$,i.e., (Eq. (19)), the following form is found.

$$Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \left(f_{jj'}^2(Y) k + \frac{g_{jj'}^2(Y) 2k(k+1)(2k+3)}{4} + f_{jj'}(Y) g_{jj'}(Y) 2k(k+1) \right). \tag{20}$$

After plugging Eq. (20) into the leap condition, we have that

$$k = \min_{j \in [1,r]} \left\lceil \frac{\sqrt{\epsilon h_0(Y) + \left(\frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2 g_{jj'}(Y)}} \right)^2} - \sum_{j=1}^{r} \frac{(f_{jj'}(Y) + \frac{g_{jj'}(Y)}{2})}{\sqrt{2 g_{jj'}(Y)}}}{\sum_{j=1}^{r} \sqrt{\frac{g_{jj'}(Y)}{2}}}, \right.$$
$$\left. \frac{\sqrt{\epsilon^2 h_0^2(Y) - A(Y) + \sum_{j=1}^{r} B^2(Y)} - \sum_{j=1}^{r} B(Y)}{\sum_{j=1}^{r} g_{jj'}(Y)} \right\rceil, \tag{21}$$

where $A(Y) = \sum_{j=1}^{r} \left(2 f_{jj'}(Y) g_{jj'}(Y) + f_{jj'}^2(Y) + \frac{3}{2} g_{jj'}^2(Y) \right)$ and $B(Y) = \left(\frac{\frac{5}{2} g_{jj'}(Y) + 2 f_{jj'}(Y) g_{jj'}(Y)}{2 g_{jj'}(Y)} \right)$. Hereby, although it is hard to compare directly Eq. (9) with Eqs. (21) and 21) is expected to give more accurate result due to the higher order expansion in the derivation. Whereas, as a future work, these values will be analysed via simulations.

Confidence Intervals with Poisson by Distribution Using MLE From MLE under the Poisson distribution, $\tau \sim Poi(k)$, it is known that the value of k equals to $k = \tau$. Then, this value is substituted into $\Delta h_j(Y)$ in Equation (11). Thus, $\Delta h_j(Y)$ can be found as

$$\begin{aligned} \Delta h_j(Y) &\approx \sum_{j=1}^{r} \tau \nu_j h_j'(Y) + \frac{1}{2} \sum_{j=1}^{r} \tau \nu_j \sum_{j=1}^{r} \frac{\tau}{n h_0(Y)} \nu_j h_j''(Y) \\ &= \sum_{j=1}^{r} \left(\tau f_{jj'}(Y) + \frac{\tau^2}{2} g_{jj'}(Y) \right), \end{aligned} \tag{22}$$

where $f_{jj'}(Y) = \sum_{i=1}^{r} \frac{\partial h_{ij}(Y)}{\partial Y_i}\nu_{ij} = h'_j(Y)\nu_j$ and $g_{jj'}(Y) = \nu_j \sum_{j=1}^{r} \nu_j \sum_{i=1}^{r} \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}\nu_{ij} = \nu_j \sum_{j=1}^{r} h''_j(Y)\nu_j$.

By evaluating $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ as in the following equation,

$$E(\Delta h_j(Y)) \approx= \sum_{j=1}^{r} \left(E(\tau) f_{jj'}(Y) + \frac{E(\tau^2)}{2} g_{jj'}(Y) \right). \tag{23}$$

Since $\tau \sim Poi(k)$, the equalities of $E(\tau) = k$ and $E(\tau^2) = k + k^2$ are known. Therefore,

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \left(\left(\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}} k + \frac{2 f_{jj'}(Y) + g_{jj'}(Y)}{\sqrt{2} g_{jj'}(Y)} \right)^2 - \left(\frac{2 f_{jj'}(Y) + g_{jj'}(Y)}{\sqrt{2} g_{jj'}(Y)} \right)^2 \right). \tag{24}$$

Similarly, $Var(\tau) = k$, $Var(\tau^2) = k + 6k^2 + 4k^3$ and $Cov(\tau, \tau^2) = 2k^2 + k$ can be computed too. Then, the following equality can be obtained

$$\begin{aligned} Var(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r} \Big(& g_{jj'}^2(Y) k^3 + \big(\frac{3}{2} g_{jj'}^2(Y) + 2 f_{jj'}(Y) g_{jj'}(Y)\big) k^2 \\ & + \big(f_{jj'}^2(Y) \frac{g_{jj'}(Y)}{4} + f_{jj'}(Y) g_{jj'}(Y) \big) k \Big) \end{aligned} \tag{25}$$

After substituting Eqs. (24) and (25) into the leap condition, an admissable value of k can be chosen by

$$\begin{aligned} k = \min_{j \in [1,r]} \Bigg\lceil & \frac{\sqrt{\epsilon h_0^2(Y) + \sum_{j=1}^{r} \left(\frac{2 f_{jj'}(Y) + g_{jj'}(Y)}{\sqrt{2} g_{jj'}(Y)} \right)^2} - \sum_{j,j'=1}^{r} \left(\frac{2 f_{jj'}(Y) + g_{jj'}(Y)}{\sqrt{2} g_{jj'}(Y)} \right)}{\sum_{j,j'=1}^{r} \sqrt{\frac{g_{jj'}(Y)}{2}}}, \\ & \frac{\sqrt{\epsilon^2 h_0^2(Y) \sum_{j=1}^{r} \left(\frac{9}{14} g_{jj'}^2(Y) + f_{jj'}^2(Y) + \frac{f_{jj'}(Y) g_{jj'}(Y)}{2} - \frac{f_{jj'}^2(Y) g_{jj'}(Y)}{4} \right)}}{\sum_{j=1}^{r} g_{jj'}(Y)} \\ & - \frac{\sum_{j=1}^{r} \left(\frac{3}{4} g_{jj'}(Y) + f_{jj'}(Y) \right)}{\sum_{j=1}^{r} g_{jj'}(Y)} \Bigg\rceil. \end{aligned} \tag{26}$$

Indeed it can be difficult to make a direct comparison between Eq. (26) and its previous value. However, as in our derivation, we apply the MLE expression in the derivation of τ and k and since MLE is a sufficient statistic, as a consequence of

Rao-Blackwellized theorem, we expect toget lower variance in the expression of k. Therefore, thereotically, we presume that thi final expression of k's Eq. (26) is narrowe than k in Eq. (9).

In addition to these , the confidence interval under the Poisson distribution for the value of k estimated from MLE is obtained by using the approximation of $k \approx \tau \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\tau}{n}}$. Then, plugging this interval into Eq. (11) gives the expression below.

$$\begin{aligned}\Delta h_j(Y) &\approx \sum_{j,j'=1}^{r}\left(\left(\tau \pm z_{\alpha/2}\sqrt{\frac{\tau}{n}}\right)f_{jj'}(Y) + \frac{1}{2}\left(\tau \pm z_{\alpha/2}\sqrt{\frac{\tau}{n}}\right)^2 g_{jj'}(Y)\right)\\ &= \sum_{j,j'=1}^{r}\left(\frac{g_{jj'}(Y)}{2}\tau^2 \pm z_{\alpha/2}\frac{g_{jj'}(Y)}{\sqrt{n}}\tau\sqrt{\tau} + \left(f_{jj'}(Y) + z_{\alpha/2}^2\frac{g_{jj'}(Y)}{2n}\right)\tau \pm z_{\alpha/2}\frac{f_{jj'}(Y)}{\sqrt{n}}\sqrt{\tau}\right),\end{aligned} \tag{27}$$

where $f_{jj'}(Y) = \sum_{i=1}^{r}\frac{\partial h_{ij}(Y)}{\partial Y_i}\nu_{ij} = h'_j(Y)\nu_j$ and $g_{jj'}(Y) = \nu_j\sum_{j=1}^{r}\nu_j\sum_{i=1}^{r}\frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}\nu_{ij} = \nu_j\sum_{j=1}^{r}h''_j(Y)\nu_j$.

Then, the values of $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by using the following equations.

$$\begin{aligned}E(\tau\sqrt{\tau}) &\approx k\sqrt{k},\ \text{having a similar process of obtaining } E(\sqrt{\tau}) \text{ (See the Appendix)},\\ Var(\sqrt{\tau}) &\approx 0,\\ Var(\tau\sqrt{\tau}) &\approx k + 3k^2,\\ Cov(\tau^2, \tau\sqrt{\tau}) &\approx -k^2\sqrt{k},\\ Cov(\tau^2, \tau) &= 2k^2 + k,\\ Cov(\tau^2, \sqrt{\tau}) &= -k\sqrt{k},\\ Cov(\tau\sqrt{\tau}) &\approx 0,\\ Cov(\tau\sqrt{\tau}, \sqrt{\tau}) &\approx k,\\ Cov(\tau, \sqrt{\tau}) &\approx 0.\end{aligned}$$

Therefore,

$$\begin{aligned}E(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r}&\left(\frac{g_{jj'}(Y)}{2}(k + k^2) \pm z_{\alpha/2}\frac{g_{jj'}(Y)}{\sqrt{n}}(k\sqrt{k})\right.\\ &\left.+\left(f_{jj'}(Y) + z_{\alpha/2}^2\frac{g_{jj'}(Y)}{2n}\right)k \pm z_{\alpha/2}\frac{f_{jj'}(Y)}{\sqrt{n}}\sqrt{k}\right)\end{aligned} \tag{28}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r} \Bigg(\frac{g_{jj'}^2(Y)}{4}(k+6k^2+4k^3) + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{n}(k+3k^2)$$
$$+ \left(f_{jj'}(Y) + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n} \right)^2 k + z_{\alpha/2}^2 \frac{f_{jj'}^2(Y)}{n} Var(\tau) - z_{\alpha/2} \frac{g_{jj'}^2(Y)}{\sqrt{n}} k^2 \sqrt{k}$$
$$+ g_{jj'}(Y)\big(f_{jj'}(Y) + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n}\big)(2k^2+k) - \frac{f_{jj'}(Y) g_{jj'}(Y)}{\sqrt{n}} k\sqrt{k}$$
$$+ z_{\alpha/2}^2 \frac{2 f_{jj'}(Y) g_{jj'}(Y)}{n} k \Bigg). \tag{29}$$

From Eqs. (28) and (29), by getting k alone is effortful since the equations have not integer powers. In other words, the equations have roots powers. Hence, the process of constructing confidence interval for k at the beginning and inserting Eq. (11) is likely to be inefficient as a workload.

Confidence Intervals with Poisson Distribution by Using MME From MME under the Poisson distribution, $\tau \sim Poi(k)$, it is known that the value of k equals to $k = \frac{\tau}{n}$. Then, this value is inserted $\Delta h_j(Y)$ in (Eq. 11) via

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} \frac{\tau}{n} \nu_j h_j'(Y) + \frac{1}{2} \sum_{j=1}^{r} \frac{\tau}{n} \nu_j \sum_{j=1}^{r} \frac{\tau}{n} \nu_j h_j''(Y)$$
$$= \sum_{j=1}^{r} \left(\frac{\tau}{n} f_{jj'}(Y) + \frac{\tau^2}{2n^2} g_{jj'}(Y) \right), \tag{30}$$

where $f_{jj'}(Y) = \sum_{i=1}^{r} \frac{\partial h_{ij}(Y)}{\partial Y_i} \nu_{ij} = h_j'(Y)\nu_j$ and $g_{jj'}(Y) = \nu_j \sum_{j=1}^{r} \nu_j \sum_{i=1}^{r} \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2} \nu_{ij} = \nu_j \sum_{j=1}^{r} h_j''(Y)\nu_j$.

Then, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by the following way.

$$E(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r} \left(\left(\frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}} k + \frac{2 f_{jj'}(Y) n + g_{jj'}(Y)}{2n\sqrt{2 g_{jj'}(Y)}} \right)^2 - \left(\frac{2 f_{jj'}(Y) n + g_{jj'}(Y)}{2n\sqrt{2 g_{jj'}(Y)}} \right)^2 \right) \tag{31}$$

and

$$Var(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r} \Bigg(\frac{g_{jj'}^2(Y)}{n^4} k^3 + \left(\frac{3 g_{jj'}(Y)}{2n^4} + \frac{2 f_{jj'}(Y) g_{jj'}(Y)}{n^3} \right) k^2$$
$$+ \left(\frac{f_{jj'}^2(Y)}{n^2} + \frac{g_{jj'}^2(Y)}{4n^4} + \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^3} + \right) k \Bigg). \tag{32}$$

Putting Eqs. (31) and (32) into the leap condition gives the plausible k via

$$k = \min_{j \in [1,r]} \left\lceil \frac{\sqrt{\epsilon h_0(Y) + \sum_{j,j'=1}^{r} \left(\frac{2 f_{jj'}(Y) n + g_{jj'}(Y)}{2n\sqrt{2 g_{jj'}(Y)}} \right)^2} - \sum_{j,j'=1}^{r} \left(\frac{2 f_{jj'}(Y) n + g_{jj'}(Y)}{2n\sqrt{2 g_{jj'}(Y)}} \right)}{\sum_{j,j'=1}^{r} \frac{\sqrt{g_{jj'}(Y)}}{n\sqrt{2}}} \right.$$
$$\left. \frac{\sqrt{\epsilon^2 h_0^2(Y) - \sum_{j=1}^{r} \left(\frac{g_{jj'}^2(Y) - g_{jj'}(Y)}{4n^4 - \frac{f_{jj'}(Y) g_{jj'}(Y) - 6 f_{jj'}(Y)}{n^3}} - \frac{3 f_{jj'}^2(Y)}{n^2} \right)} - \sum_{j=1}^{r} \left(\frac{3}{2n^2} - \frac{2 f_{jj'}(Y)}{n} \right)}{\sum_{j=1}^{r} \frac{g_{jj'}(Y)}{n^2}} \right\rceil . \quad (33)$$

Indeed here, although Eq. (33) can be simulated to determine which k value gives more accurate result. We cannot directly say that the expression in Eq. (33) is narrower than the expression of k in Eq. (26). Because, theoretically it is not guarantee that all produce sufficient statistics as the maximum likelihood estimators. Hence, this final expression is one of the future work by evaluating its value in simulation studies. Theoretically, it is not straightforward to do this.

Confidence Intervals with Gamma Distribution by Using MLE From MLE, it is known that the value of k equals to $k = \frac{\tau}{n h_0(Y)}$. Then, this value is inserted $\Delta h_j(Y)$ in (Eq. 11) via

$$\begin{aligned} \Delta h_j(Y) &\approx \sum_{j=1}^{r} \frac{\tau}{n h_0(Y)} \nu_j h_j'(Y) + \frac{1}{2} \sum_{j=1}^{r} \frac{\tau}{n h_0(Y)} \nu_j \sum_{j=1}^{r} \frac{\tau}{n h_0(Y)} \nu_j h_j''(Y) \\ &= \sum_{j=1}^{r} \left(\frac{\tau}{n h_0(Y)} f_{jj'}(Y) + \frac{\tau^2}{2n^2 h_0^2(Y)} g_{jj'}(Y) \right), \end{aligned} \quad (34)$$

where $f_{jj'}(Y) = \sum_{i=1}^{r} \frac{\partial h_{ij}(Y)}{\partial Y_i} \nu_{ij} = h_j'(Y) \nu_j$ and $g_{jj'}(Y) = \nu_j \sum_{j=1}^{r} \nu_j \sum_{i=1}^{r} \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2} \nu_{ij} = \nu_j \sum_{j=1}^{r} h_j''(Y) \nu_j$.

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \left(\left(\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2} n h_0^2(Y)} k + \frac{2n h_0^2(Y) f_{jj'}(Y) + g_{jj'}(Y)}{2n h_0^2 \sqrt{2 g_{jj'}(Y)}} \right)^2 - \left(\frac{2n h_0^2(Y) f_{jj'}(Y) + g_{jj'}(Y)}{2n h_0^2 \sqrt{2 g_{jj'}(Y)}} \right)^2 \right) \quad (35)$$

and

$$\begin{aligned} Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \Bigg(&\left(\frac{f_{jj'}(Y)}{n h_0(Y)} \right)^2 \frac{k}{h_0^2(Y)} + \left(\frac{g_{jj'}(Y)}{2n^2 h_0^2(Y)} \right)^2 \frac{2k(k+1)(2k+3)}{h_0^4(Y)} \\ &+ \frac{2 f_{jj'}(Y) g_{jj'}(Y)}{2n^3 h_0^3(Y)} \frac{2k(k+1)}{h_0^3(Y)} \Bigg). \end{aligned} \quad (36)$$

Similar to previous derivations, inserting Eqs. (35) and (36) into the required leap condition gives an acceptable value of k via

$$k = \min_{j\in[1,r]} \left\lceil \frac{\sqrt{\sum_{j=1}^{r}\left(\frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}\right)^2 - \epsilon h_0(Y)} - \frac{2nh_0^2(Y)f_{jj'}(Y)+g_{jj'}(Y)}{2nh_0^2\sqrt{2g_{jj'}(Y)}}}{\sum_{j=1}^{r}\frac{\sqrt{g_{jj'}(Y)}}{\sqrt{2}nh_0^2(Y)}} \right.$$
$$\left. \frac{\sqrt{\epsilon^2 h_0^2(Y) + \sum_{j=1}^{r}\left(\frac{19g_{jj'}^2(Y)}{4n^4h_0^8(Y)} - \frac{8f_{jj'}(Y)g_{jj'}(Y)}{n^3h_0^6(Y)} - \frac{3f_{jj'}^2(Y)}{n^2h_0^4(Y)}\right)} - \left(\frac{5g_{jj'}(Y)}{2n^2h_0^4(Y)} + \frac{2f_{jj'}(Y)}{nh_0^2(Y)}\right)}{\sum_{j=1}^{r}\frac{g_{jj'}(Y)}{n^2h_0^4(Y)}} \right\rceil. \quad (37)$$

Thus, as MLE gives the sufficient statistics and as a consequence of the Rao-Blackwellized theorem, we can theoretically say that Eq. (37) obtained from (37) can provides more accurate results than the estimators derived by the method of moments.

Moreover, the confidence interval for the value of k estimated from MLE is found as, previously, $k \approx \frac{k}{nh_0^2(Y)} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{k}{h_0^4(Y)n^3}}$. Then, insterting this interval into Eq. (11) gives the following approximation.

$$\Delta h_j(Y) \approx \sum_{j,j'=1}^{r}\left(\frac{g_{jj'}(Y)}{2n^2h_0^4(Y)}k^2 \pm z_{\frac{\alpha}{2}}\frac{2}{n^2\sqrt{n}h_0^4(Y)}k\sqrt{k} + \left(\frac{f_{jj'}(Y)}{nh_0^2(Y)} + z_{\frac{\alpha}{2}}^2\frac{g_{jj'}(Y)}{2n^3h_0^4(Y)}\right)k \pm z_{\frac{\alpha}{2}}\frac{f_{jj'}(Y)}{n\sqrt{n}h_0^2(Y)}\sqrt{k}\right), \quad (38)$$

where $f_{jj'}(Y) = \sum_{i=1}^{r}\frac{\partial h_{ij}(Y)}{\partial Y_i}\nu_{ij} = h_j'(Y)\nu_j$ and $g_{jj'}(Y) = \nu_j\sum_{j=1}^{r}\nu_j\sum_{i=1}^{r}\frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}\nu_{ij} = \nu_j\sum_{j=1}^{r}h_j''(Y)\nu_j$. As $k = \frac{\tau}{nh_0(Y)}$ is estimated from MLE, it is substituted into Eq. (38). Thereby,

$$\Delta h_j(Y) \approx \sum_{j,j'=1}^{r}\left(\frac{g_{jj'}(Y)}{2n^2h_0^4(Y)}\frac{\tau^2}{n^2h_0^2(Y)} \pm z_{\frac{\alpha}{2}}\frac{2}{n^2\sqrt{n}h_0^4(Y)}\frac{\tau}{nh_0(Y)}\sqrt{\frac{\tau}{nh_0(Y)}}\right.$$
$$\left. + \left(\frac{f_{jj'}(Y)}{nh_0^2(Y)} + z_{\frac{\alpha}{2}}^2\frac{g_{jj'}(Y)}{2n^3h_0^4(Y)}\right)\frac{\tau}{nh_0(Y)} \pm z_{\frac{\alpha}{2}}\frac{f_{jj'}(Y)}{n\sqrt{n}h_0^2(Y)}\sqrt{\frac{\tau}{nh_0(Y)}}\right) \quad (39)$$

Furthermore, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are calculated by using $E(\tau\sqrt{\tau}) \approx \left(\frac{k}{h_0(Y)}\right)^{3/2}$, which is evaluated by a similar process of finding $E(\sqrt{\tau})$, $Var(\sqrt{\tau}) \approx 0$, $Var(\tau\sqrt{\tau}) \approx \frac{3k^2+2k}{h_0^3(Y)}$, $Cov(\tau^2, \tau\sqrt{\tau}) \approx \left(\frac{k}{h_0(Y)}\right)^{7/2} - \frac{k^2+k}{h_0^2(Y)}\left(\frac{k}{h_0(Y)}\right)^{3/2}$, $Cov(\tau^2, \sqrt{\tau}) \approx \left(\frac{k}{h_0(Y)}\right)^{5/2} - \frac{k^2+k}{h_0^2(Y)}\left(\frac{k}{h_0(Y)}\right)^{1/2}$, $Cov(\tau\sqrt{\tau}, \tau) \approx 0$, $Cov(\tau\sqrt{\tau}, \sqrt{\tau}) \approx \frac{k}{h_0^2(Y)}$, $Cov(\tau^2, \tau) = \frac{2k(k+1)}{h_0^3(Y)}$ and $Cov(\tau, \sqrt{\tau}) \approx 0$. Thus,

$$
\begin{aligned}
E(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r} \Bigg(& \frac{g_{jj'}(Y)}{2n^4 h_0^6(Y)} \frac{k^2+k}{h_0(Y)} \pm z_{\alpha/2} \frac{2}{n^4 h_0^5(Y)\sqrt{h_0(Y)}} \frac{k^{3/2}}{h_0^{3/2}(Y)} \\
& + \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right) \frac{k}{h_0(Y)} \pm z_{\frac{\alpha}{2}} \frac{f_{jj'}(Y)}{n^2 h_0^2(Y)\sqrt{h_0(Y)}} \sqrt{\frac{k}{h_0(Y)}} \Bigg)
\end{aligned}
\tag{40}
$$

and

$$
\begin{aligned}
Var(\Delta h_j(Y)) \approx \sum_{j,j'=1}^{r} \Bigg(& \frac{g_{jj'}^2(Y)}{4n^8 h_0^{12}(Y)} \frac{6k^3+9k^2+6k-1}{h_0^4(Y)} + z_{\alpha/2} \frac{4(3k^2)+2k}{n^8 h_0^{14}(Y)} \\
& + \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right)^2 \frac{k}{h_0(Y)} \\
& \pm z_{\alpha/2} \frac{2g_{jj'}(Y)}{n^8 h_0^{11}(Y)\sqrt{h_0(Y)}} \left(\Big(\frac{k}{h_0(Y)}\Big)^{7/2} - \frac{k^2+k}{h_0^2(Y)} \Big(\frac{k}{h_0(Y)}\Big)^{3/2} \right) \\
& + \frac{g_{jj'}(Y)}{n^4 h_0^6(Y)} \left(\frac{f_{jj'}(Y)}{n^2 h_0^3(Y)} + z_{\alpha/2}^2 \frac{g_{jj'}(Y)}{2n^4 h_0^5(Y)} \right) \frac{2k(k+1)}{h_0^3(Y)} \\
& \pm z_{\alpha/2} \frac{f_{jj'}(Y) g_{jj'}(Y)}{n^6 h_0^8(Y)\sqrt{h_0(Y)}} \left(\Big(\frac{k}{h_0(Y)}\Big)^{5/2} - \frac{k^2+k}{h_0^2(Y)} \Big(\frac{k}{h_0(Y)}\Big)^{1/2} \right) \\
& \pm z_{\frac{\alpha}{2}}^2 \frac{4 f_{jj'}(Y)}{n^6 h_0^8(Y)} \frac{k}{h_0^2(Y)} \Bigg).
\end{aligned}
\tag{41}
$$

Like Eqs. (28) and (29), here, in Eqs. (40) and (41), it is not tractable to get alone k so finding analytic result can be obtained via numeric analysis.

3.2 Expansion of 3rd Order Taylor Expansion

Similar to previous part (3.1), here, the change in the hazard function $\Delta h_j(Y)$ is approximated by the third order Taylor expansion in the time interval $[t, t+\tau]$ in a such way that

$$
\begin{aligned}
h_j(Y+\bar{\lambda}(Y,\tau)) &= h_j(Y) + \bar{\lambda}(Y,\tau)h'_j(Y) + \frac{\bar{\lambda}^2}{2}h''_j(Y) + \frac{\bar{\lambda}^3}{6}h'''_j(Y) + \mathcal{O}^4(h_j(Y)) \\
h_j(Y+\bar{\lambda}(Y,\tau)) - h_j(Y) &\approx \bar{\lambda}(Y,\tau)h'_j(Y) + \frac{\bar{\lambda}^2}{2}h''_j(Y) + \frac{\bar{\lambda}^3}{6}h'''_j(Y),
\end{aligned}
\tag{42}
$$

where $h'_j(Y) = \sum_{i=1}^n \frac{\partial h_{ij}(Y)}{\partial Y_i}$,$h''_j(Y) = \sum_{i=1}^n \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}$ and $h'''_j(Y) = \sum_{i=1}^n \frac{\partial^3 h_{ij}(Y)}{\partial Y_i^3}$.

$\bar{\lambda}(Y,t) = Y(t+\tau) - Y(t) = \sum_{j=1}^{r} k_j \nu_j$. So,

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} k_j \nu_j h'_j(Y) + \frac{1}{2}\sum_{j=1}^{r} k_j \nu_j \sum_{j=1}^{r} k_j \nu_j h''_j(Y) + \frac{1}{6}\sum_{j=1}^{r} k_j \nu_j \sum_{j=1}^{r} k_j \nu_j \sum_{j=1}^{r} k_j \nu_j h'''_j(Y). \tag{43}$$

Similar to previous derivations, under the Gamma distribution, it is known that $k = \tau h_0(Y)$. Subsequently,

$$\Delta h_j(Y) \approx \sum_{j=1}^{r} \left(\tau h_0(Y) f_{jj'}(Y) + \frac{1}{2}\tau^2 h_0^2(Y) g_{jj'}(Y) + \frac{1}{6}\tau^3 h_0^3(Y) l_{jj'}(Y) \right), \tag{44}$$

setting $f_{jj'}(Y) = \sum_{i=1}^{r} \frac{\partial h_{ij}(Y)}{\partial Y_i}\nu_{ij} = h'_j(Y)\nu_j$, $g_{jj'}(Y) = \nu_j \sum_{j=1}^{r} \nu_j \sum_{i=1}^{r} \frac{\partial^2 h_{ij}(Y)}{\partial Y_i^2}\nu_{ij} = \nu_j \sum_{j=1}^{r} h''_j(Y)\nu_j$ and $l_{jj'}(Y) = \nu_j \sum_{j=1}^{r} \nu_j \sum_{j=1}^{r} \nu_j \sum_{i=1}^{r} \frac{\partial^3 h_{ij}(Y)}{\partial Y_i^3}\nu_{ij} = \nu_j \sum_{j=1}^{r} \nu_j \sum_{j=1}^{r} h'''_j(Y)\nu_j$.

Then, $E(\Delta h_j(Y))$ and $Var(\Delta h_j(Y))$ are derived by

$$E(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \left(k f_{jj'}(Y) + \frac{1}{2}k(k+1) g_{jj'}(Y) + \frac{1}{6}k(k+1)(k+2) l_{jj'}(Y) \right) \tag{45}$$

and

$$\begin{aligned} Var(\Delta h_j(Y)) \approx \sum_{j=1}^{r} \Bigg(& f_{jj'}^2(Y)k + \frac{g_{jj'}^2(Y)}{4}(6k^3 + 9k^2 + 6k - 1) \\ & + \frac{l_{jj'}^2(Y)}{36}(9k^5 + 72k^4 + 213k^3 + 180k^2 + 210k) \\ & + f_{jj'}(Y) g_{jj'}(Y)(2k(k+1)) \\ & + \frac{g_{jj'}(Y) l_{jj'}(Y)}{6}(k(k+1)(k+2)(6k+12)) \\ & + \frac{f_{jj'}(Y) l_{jj'}(Y)}{3}(3k(k+1)(k+2)) \Bigg). \end{aligned} \tag{46}$$

As $k \sim \Gamma(k, h_(0)(Y))$, it is known that

$$Var(\tau^2) = \frac{6k^3 + 9k^2 + 6k - 1}{h_0^4(Y)}$$
$$Var(\tau^3) = \frac{9k^5 + 72k^4 + 213k^3 + 180k^2 + 210k}{h_0^6(Y)}$$
$$Cov(\tau, \tau^2) = \frac{2k(k+1)}{h_0^3(Y)}$$
$$Cov(\tau^2, \tau^3) = \frac{k(k+1)(k+2)(6k+12)}{h_0^5(Y)}$$
$$Cov(\tau, \tau^3) = \frac{3k(k+1)(k+2)}{h_0^4(Y)}.$$

Similar to previous outputs, Eqs. (29) and (41), procuring an admissable k is tough in Eq. (46).

Finally, inserting the statements in the Eq. (45) into the required leap condition results that

$$k \leq \frac{\sqrt{\epsilon h_0(Y) + \sum_{j=1}^{r}\left(\frac{3(g_{jj'}(Y)+l_{jj'}(Y))^2}{2l_{jj'}(Y)} + f_{jj'}(Y) + \frac{l_{jj'}(Y)}{3}\right)} - \left(\frac{g_{jj'}(Y)+l_{jj'}(Y)}{2}\right)\sqrt{\left(\frac{6}{l_{jj'}(Y)}\right)}}{\sum_{j=1}^{r}\sqrt{\frac{l_{jj'}(Y)}{6}}}. \tag{47}$$

4 Conclusion

With the help of SSA, the reactions occuring in the biological systems can be simulated. Although the exact results can be obtained by SSA, the speed of acquiring the result is slow. When considering the speed, approximate SSA can be preferable as they are faster than SSA. Also, approximate SSA depends on leap condition which implies that the time step τ should be selected so that the propensity function do not change during the time interval $[t, t+\tau]$. In this study, to construct the confidence interval, we have used mainly Poisson τ-leap method and approximate Gillespie algorithm. We have expanded the function of net change in the hazard function with the 2nd and 3rd order Taylor expansions. Then, we have derived MLE and MME for the parameter k. After all, we have extended into leap condition to obtain an acceptable value of k which can produce theoretically more accurate result as a consequence of the Rao-Blackwell theorem. Because, we have used sufficient statistics in the derivation of k. As a result of these confidence interval, we have see that particularly the result found from MME are not tractable due to their high powers, whereas MLE can be tractable. On the other hand, as we have obtained theoretical outputs, we consider to evaluate all the expressions via simulation studies [9].

Acknowledgements The authors thank to the Scientific Research Project of Middle East Technical University (Project no: BAP- 10282) for their support.

Appendix

Let $g(\tau) = \sqrt{\tau}$ be a smooth function for $\tau \geq 0$ with $\tau \sim Poi(k)$. Then, by the Taylor expansion around the mean $\mu = E(\tau)$, the following expression can be obtained.

$$g(\tau) = g(\mu) + g'(\mu)(\tau - \mu) + \frac{g''(\mu)(\tau-\mu)^2}{2!} + \frac{g'''(\mu)(\tau-\mu)^3}{3!} + \cdots + \frac{g^t(\mu)(\tau-\mu)^t}{t!} + \ldots$$

Then, the mean can be derived as

$$\begin{aligned} E(g(\tau)) &= g(\mu) + g'(\mu)E(\tau - \mu) + \frac{g''(\mu)E(\tau-\mu)^2}{2!} + \frac{g'''(\mu)E(\tau-\mu)^3}{3!} + \cdots + \frac{g^t(\mu)E(\tau-\mu)^t}{t!} + \ldots \\ &= g(\mu) + g'(\mu)m_1 + \frac{g''(\mu)m_2}{2!} + \frac{g'''(\mu)m_3}{3!} + \cdots + \frac{g^t(\mu)m_t}{t!} + \ldots, \end{aligned}$$

where m_t is t-th central moment. In this case, considering just up to third order Taylor expansion, $m_1 = 0$ and $m_2 = m_3 = \mu$. So, we have

$$E[g(\tau)] = \sqrt{\mu} + 0 + \frac{1}{8}\mu^{-\frac{1}{2}} - \frac{1}{16}\mu^{-\frac{3}{2}}.$$

Then, $E[g(\tau)] = E[\mu] \approx \sqrt{\mu} = \sqrt{E(\tau)} = \sqrt{k}$ for $\mu >> 1$. Thus, $\sqrt{E(\tau)} \approx \sqrt{k}$. Similarly, we apply these processes for the gamma distribution with $\tau \sim \Gamma(k, h_0(Y))$ and by this way, tth moment for the gamma distribution is defined as $E(\tau^t) = \frac{(k+t-1)\ldots(1)}{h_0^t(Y)}$. Accordingly, we can obtain $E(\tau)$ as below.

$$E(\tau) = \sqrt{\frac{k}{h_0(Y)}} + \frac{1}{8}\left(\frac{1}{\sqrt{k.h_0(Y)}} - \frac{1}{h_0^4(Y).k\sqrt{k}}\right),$$

with $E[\tau - \mu] = 0$, $E[(\tau-\mu)^2] = V(\tau) = \frac{k}{h_0^2(Y)}$ and $E[(\tau-\mu)^3] = \frac{2k}{h_0^3(Y)}$. If $k \times h_0(Y) << 1$, then it is possible to reach that $E[\sqrt{\tau}] = \sqrt{\frac{k}{h_0(Y)}}$. Similarly, the equality $Var(\sqrt{\tau}) = \sqrt{\frac{k}{h_0^2(Y)}}$ can be obtained.

References

1. Gibson MA, Bruck J 2000 Efficient exact stochastic simulation of chemical systems with many species and many channels. Journal of Physical Chemistry, A (104):1876–1889
2. Gillespie DT 1977 Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry, 81(25):2340–23610
3. Gillespie DT 1992 A rigorous derivation of the chemical master equation. Physica A, 188:404–425
4. Gillespie DT 2001 Approximate accelerated stochastic simulation of chemically reacting systems. Journal of Chemical Physics, 115:1716–1733
5. Gillespie DT 2007 Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry, 58:35–55
6. Gillespie T, Petzold LR 2003 Improved Leap-Size Selection for Accelerated Stochastic Simulation. Journal of Chemical Physics, 119, 8229–8234
7. Demirbüken S, Purutçuoğlu V 2020 Extension of leap condition in approximate stochastic simulation algorithms of biological networks. 4th International Conference on Mathematics "An Istanbul Meeting for World Mathematicians", 288–298
8. Demirbüken S, Purutçuoğlu V Extension of leap condition in approximate stochastic simulation algorithms of biological networks. Turkish Journal of Mathematics and Computer Science (Unpublished)
9. Demirbüken S 2021 Extension of leap condition in approximate stochastic simulation algorithms of biological networks. Unpublished Master's Thesis, Middle East Technical University
10. Purutçuoğlu V, Wit E 2006 Exact and Approximate Stochastic Simulations of theMAPK Pathway and Comparisons of Simulations Results. Journal of Integrative Bioinformatics, 3, 1–13
11. Purutçuoğlu V, Wit E 2008 An approximation algorithm based on leap condition for stochastical simulation of biomedical systems. Proceeding of the 4th International Conference "Inverse problems: Modelling and Simulation", 151–152
12. Wilkinson DJ 2006 Stochastic modelling for systems biology, Chapman and Hall/CRC

Discrete Biorthogonal Systems and Equilibrium Condition in the Hardy Space of Unit Disc and Upper Half-Plane

Zsuzsanna Nagy-Csiha and Margit Pap

Abstract Starting from recent result of Fridli and Schipp, we introduce the dual system of the Malmquist–Takenaka complex orthonormal system on the upper half-plane, and prove the biorthogonality of the systems. We show that the nodal points of the biorthogonal systems on the unite disc and on the upper half-plane satisfy the similar equilibrium condition as it was proved previously for the discrete orthogonal system on the unite circle and on the real line. In this way we generalize the results obtained in [1, 2] by Pap and Schipp.

Keywords Hardy spaces · Malmquist–Takenaka systems · Discrete biorthogonality · Blaschke products · Equilibrium condition

1 Introduction

The Malmquist–Takenaka (M-T) system [3, 4] is an orthonormal system of rational functions-products of Blaschke factors–in the Hardy space of unit disc, which contains as special case the classical "trigonometric" system. It is frequently applied in system identification in order to approximate the transfer functions. Discretization results connected to M-T systems for unit disc and the upper half plane were published in [1, 2, 5, 6]. Based on these an analogue of discrete Fourier transform (DFT) was developed and the discrete versions was applied successfully for compression and representation of human ECG signals [7, 8].

Discretization nodes on unit circle and on the real line have similar properties: they satisfy some equilibrium conditions and are stationary points of some logarithmic potentials. The problem whether they are discrete energy minimizer configurations was formulated in [1, 2] and was answered positively recently in [6].

Z. Nagy-Csiha (✉) · M. Pap
University of Pécs, Ifjúság út 6, Pécs 7634, Hungary
e-mail: ncszsu@gamma.ttk.pte.hu

M. Pap
e-mail: papm@gamma.ttk.pte.hu

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_26

In paper [9] Fridli and Schipp introduced the dual of the Malmquist–Takenaka system on the unit disc and proved discrete biorthogonal property on a set of points of the unit disc. Based on the presented results connected to Malmquist–Takenaka systems and the discrete orthogonality on unit circle and real line in this paper we introduce the dual system of the Malmquist–Takenaka system on the upper half-plane and we prove discrete biorthogonality result on a set of discretization points on upper half plane. We study the properties of discretization points on disc and upper half-plane, and we prove that they satisfy analogue equilibrium conditions like on unit circle and real line.

Let $\mathbb{D}$ denote the open, and $\overline{\mathbb{D}}$ denote the closed unit disc, $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, $\overline{\mathbb{D}} := \{z \in \mathbb{C} : |z| \leq 1\}$, and let us denote the upper half-plane with $\mathbb{C}_+$, $\mathbb{C}_+ = \{z \in \mathbb{C} : \Im z > 0\}$. Let us denote the set of analytic functions over $\mathbb{D}$ and $\mathbb{C}_+$ with $A(\mathbb{D})$ and $A(\mathbb{C}_+)$, respectively, and consider the Hardy space of the unit disc and the Hardy space of the upper half-plane with

$$H^2(\mathbb{D}) = \left\{ f \in A(\mathbb{D}) : \|f\|_{H^2(\mathbb{D})} = \sup_{r<1} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(re^{it})|^2 dr \right)^{1/2} < \infty \right\},$$

$$H^2(\mathbb{C}_+) = \left\{ f \in A(\mathbb{C}_+) : \|f\|_{H^2(\mathbb{C}_+)} = \sup_{0<y} \left(\int |f(x+iy)|^2 dx \right)^{1/2} < \infty \right\}.$$

For every function $f \in H^2(\mathbb{D})$, for a.e. $t \in [-\pi, \pi)$ there exists the finite limit $f(e^{it}) := \lim_{r\to 1} f(re^{it})$. Moreover, the limit function f is in $L^2(\mathbb{T})$, and $\|f\|_{H^2(\mathbb{D})} = \|f\|_{L^2(\mathbb{T})}$. In the similar way for the Hardy space of the upper half plane if $f \in H_2(\mathbb{C}_+)$, for a.e. $x \in \mathbb{R}$ there exist the finite limit $f(x) := \lim_{y\to 0_+} f(x+iy)$, the limit function of f satisfies the following conditions $f \in L^2(\mathbb{R})$ and $\|f\|_{L^2(\mathbb{R})} = \|f\|_{H^2(\mathbb{C}_+)}$.

The Hardy space of the upper half-plane and the Hardy space of the unit disc may be connected through the Cayley transform, which is a conformal mapping from $\overline{\mathbb{C}_+}$ to $\overline{\mathbb{D}}$: $K_\theta(z) = e^{i\theta}\frac{z-i}{z+i}$. The maps $K_\theta : \mathbb{C}_+ \to \mathbb{D}$ and $K_\theta : \mathbb{R} \to \mathbb{T} \setminus \{e^{i\theta}\}$ are bijections, and the inverse of K_θ is $K_\theta^{-1}(z) = i\frac{1+ze^{-i\theta}}{1-ze^{-i\theta}}$. In the following, we choose $e^{i\theta} = -1$, and use the notation

$$K(z) := K_\pi(z) = \frac{i-z}{i+z} = \frac{1+iz}{1-iz}, \quad K^{-1}(z) := K_\pi^{-1}(z) = i\frac{1-z}{1+z}.$$

The following linear transformation is an isometry between $H^2(\mathbb{D})$ and $H^2(\mathbb{C}_+)$:

$$(Tf)(z) := \frac{1}{\sqrt{\pi}}\frac{1}{i+z} f(K(z)). \tag{1}$$

Let us define the b_a Blaschke-function and the r_a rational function:

$$b_a(z) := \frac{z-a}{1-\overline{a}z}, \quad r_a(z) := \frac{\sqrt{1-|a|^2}}{1-\overline{a}z} \quad (a \in \mathbb{D}, z \in \mathbb{D}).$$

$B_N^{\mathbf{a}}(z)$ is the finite product of Blaschke-functions,

$$B_N^{\mathbf{a}}(z) := B_N(z) := c \prod_{k=0}^{N-1} b_{a_k}(z)\,, (z \in \mathbb{D},\ \mathbf{a} = (a_0, \dots, a_{N-1}) \in \mathbb{D}^N, c \in \mathbb{T}).$$

The Malmquist–Takenaka (MT) system of the unit disc is given by the following formulas, for $z \in \mathbb{D},\ n \in \mathbb{N}^*$:

$$\phi_0^{\mathbf{a}}(z) = \phi_0(z) = r_{a_0}(z),\ \phi_n^{\mathbf{a}}(z) = \phi_n(z) = r_{a_n}(z) \prod_{k=0}^{n-1} b_{a_k}(z).$$

If the so called non Blaschke condition $\sum_{n=0}^{\infty}(1-|a_n|) = \infty$ holds, then the Malmquist–Takenaka system is complete orthonormal system in the Hardy space of the unite circle. The Christoffel–Darboux formula permits us to write the kernel function associated to the system in closed form:

$$K_N(z,\xi) = \sum_{k=0}^{N-1} \phi_k(z)\overline{\phi_k(\xi)} = \frac{1 - B_N(z)\overline{B_N(\xi)}}{1 - z\overline{\xi}} \quad (1 - z\overline{\xi} \neq 0).$$

Let us define the transform of the Malmquist–Takenaka system under the transformation defined in (1). We get that the system

$$\Psi_n(z) := (T\phi_n)(z) = \frac{1}{\sqrt{\pi}} \frac{1}{i+z} \phi_n(K(z))\ (\Im z \geq 0, n \in \mathbb{N}),$$

which is the analogue of the Malmquist–Takenaka system for the upper half-plane, is orthonormal in $L^2(\mathbb{R})$.

In more detailed let us introduce $\lambda_a := K^{-1}(a) = i\frac{1-a}{1+a},\ a \in \mathbb{D}$. Then

$$\frac{\sqrt{1-|a|^2}}{|1+\overline{a}|} = \sqrt{\Im \lambda_a}, \tag{2}$$

$$b_a(K(z)) = b_a(-1)\frac{z+i}{z-\overline{\lambda}_a}, \quad r_a(K(z)) = r_a(-1)\frac{z+i}{z-\overline{\lambda}_a}, \tag{3}$$

$$\Psi_0(z) = \frac{1}{\sqrt{\pi}} \frac{\phi_0(-1)}{z - \overline{\lambda}_{a_0}}, \quad \Psi_n(z) = \frac{1}{\sqrt{\pi}} \frac{\phi_n(-1)}{z - \overline{\lambda}_{a_n}} \prod_{k=0}^{n-1} \frac{z - \lambda_{a_k}}{z - \overline{\lambda}_{a_k}} \quad (n \in \mathbb{N}^*).$$

We will use the following normalization of the MT system on the upper half-plane (see in [5]):

$$\Psi_0^{\breve{}}(z) := \frac{\frac{\sqrt{\Im\lambda_0}}{\sqrt{\pi}}}{z-\overline{\lambda}_0}, \quad \Psi_n^{\breve{}}(z) := \frac{\frac{\sqrt{\Im\lambda_n}}{\sqrt{\pi}}}{z-\overline{\lambda}_n}\prod_{k=0}^{n-1}\frac{z-\lambda_k}{z-\overline{\lambda_k}} \quad (n \in \mathbb{N}^*). \tag{4}$$

We remark, that according to (2), and $|b_a(-1)| = 1$, we have $\left|\frac{1}{\sqrt{\pi}}\Phi_n(-1)\right| = \frac{\sqrt{\Im\lambda_n}}{\sqrt{\pi}}$. The secondly defined MT system differs from the firstly defined system with a constant, so we get that the second system is also an orthogonal system. Moreover, if the non-Blaschke condition $\sum_{k=0}^{\infty}\frac{\Im\lambda_k}{1+|\lambda_k|^2} = \infty$ for the upper half-plane is satisfied, then $(\Psi_n, n \in \mathbb{N})$ and $(\Psi_n^{\lambda}, n \in \mathbb{N})$ are complete orthonormal systems for $H^2(\mathbb{C}_+)$.

We will need for the proof Džrbašjan result, the analogue of the Cristophel–Darboux formula for the upper half plane, see in [10]. For $n \geq 0$ let us consider the function

$$\widetilde{B}_n(\omega) = \prod_{k=0}^{n-1}\frac{\omega-\lambda_k}{\omega-\overline{\lambda_k}}\tau_k, \quad \tau_k = \frac{|1+\lambda_k^2|}{1+\lambda_k^2}.$$

When $\lambda_k = i$, then by definition $\tau_k := 1$. For arbitrary values of the variables ω and w, and for any n, $0 \leq n < \infty$,

$$\sum_{k=0}^{n-1}\Psi_k^{\lambda}(\omega)\overline{\Psi_k^{\lambda}(w)} = \frac{1-\overline{\widetilde{B}_n(w)}\widetilde{B}_n(\omega)}{2i\pi(\overline{w}-\omega)}, \quad \omega \neq \overline{w}.$$

2 Previous Result

The discretization of the Malmquist–Takenaka system on the unite circle and on the real line was investigated in [1, 2, 5]. Recently, Fridli and Schipp generalized the discrete orthogonality proved by Pap and Schipp to the unite disc, where they proved discrete biorthogonality with respect to a discrete scalar product defined on a discrete subset of the unit disc [9]. At first, we recall Theorem 2.1. of Fridli and Schipp in [9]. Using this result we prove similar discrete biorthogonality for the Malmquist–Takenaka system on the upper half plane. We prove that the discretization nodes in both cases, in the unit disc and upper half plane, satisfy an equilibrium condition, proved previously on the circle and on the real line in [1, 2].

We consider the solutions of the equations

$$\begin{aligned} B_N^{\mathbf{a}}(z) &= u \quad (0 < |u| \leq 1) \textit{ and} \\ (B_N^{\mathbf{a}})'(z) &= 0. \end{aligned} \tag{5}$$

Let us introduce the set $\mathcal{Z}_{N,u}^{\mathbf{a}} := \{z \in \mathbb{C} : B_N^{\mathbf{a}}(z) = u\}$ $(0 < |u| \leq 1)$. Let $z^* = 1/\overline{z}$ and let $\mathcal{Q}$ denote the set of rational functions. For any $f \in \mathcal{Q}$ the domain will be extended to $\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ by $f(a) = \infty$, if a is a pole of f, and $f(\infty) :=$

$\lim_{z\to\infty} f(z)$. We introduce the following two types of inversions: $f^*(z) := (f(z))^*$, $f^\star(z) := f(z^*)$ $(z \in \overline{\mathbb{C}},\ f \in \mathcal{Q})$. It is obvious, that for any $z \in \mathbb{T}$, we have $z = z^*$. $f^*(z) = f^\star(z) = f(z)$ $(f \in \mathcal{Q})$. Moreover, in case of Blaschke-products, the operations coincide: $B_N^*(z) = B_N^\star(z) = B_N(z^*)$ $(z \in \overline{\mathbb{C}})$. Let us consider the following functions:

$$\phi_0^\star = \overline{z}\frac{\sqrt{1-|a_0|^2}}{\overline{z}-\overline{a}_0} = r_{a_0}^\star(z), \quad \phi_n^\star = \phi_n(z^*) = \overline{z}\frac{\sqrt{1-|a_n|^2}}{\overline{z}-\overline{a}_n}\prod_{k=0}^{n-1}\frac{1-a_k\overline{z}}{\overline{z}-\overline{a}_k} =$$
$$r_{a_n}^\star(z)\prod_{k=0}^{n-1} b_{a_k}^\star(z) \quad (n \in \mathbb{N}^*),\ z \in \mathbb{C}\setminus\mathbb{D}.$$

The system $\Phi^\star := ((\phi_n)^\star, n \in \mathbb{N})$ is called the dual of the MT-system $\Phi = (\phi_n^{\mathbf{a}}, n \in \mathbb{N})$. If $z \in \mathbb{T}$, then $\phi_n^\star = \phi_n$, $n \in \mathbb{N}$.

If $|u| \le 1$, then it is easy to see, that the equation $B_N(z) = u$ has exactly N solutions counting with multiplicities. In particular, if $u \in \mathbb{T}$, then all of the roots are of multiplicity one.

If $|u| \ge 1$, then $|u^*| \le 1$. In that case $B_N(z) = u$ if and only if $B_N^*(z) = u^*$. But $B_N^*(z) = B_N(z^*)$, so the equation $B_N(z) = u$ has N solutions. In the following we will consider that $u \in \mathbb{D}$ is a number for which the equation has N distinct roots, so the set $\mathcal{Z}_{N,u}^{\mathbf{a}}$ has N elements: $\mathcal{Z}_{N,u}^{\mathbf{a}} = \{z_k, k = 0,\dots,N-1\}$. It is easy to see, that similar lines of arguments hold in the case $|u| > 1$.

It is easy to verify (see the proof in [9]) that the following theorem holds not just for $0 < |u| \le 1$, as it is mentioned in [9], but for $u \in \mathbb{C}\setminus\{0\}$.

Theorem 1 *Let $u \in \mathbb{C}\setminus\{0\}$ be a parameter for which the set $\mathcal{Z}_{N,u}^{\mathbf{a}}$ has N different elements. Then the ϕ_n, $\phi_n^\star$ $(0 \le n < N)$ systems are biorthogonal:*

$$[\phi_n, \phi_m^\star]_{\mathbf{a},u} := \sum_{z\in\mathcal{Z}_{N,u}^{\mathbf{a}}} \phi_n(z)\overline{\phi_m^\star(z)}/K_N^{\mathbf{a}}(z,z^*) = \delta_{mn} \quad (0 \le m, n < N),$$

where $K_N^{\mathbf{a}}(z,z^*)$ is the Dirichlet kernel, $K_N^{\mathbf{a}}(z,z^*) = \sum_{k=0}^{N-1}\phi_k(z)\overline{\phi_k(z^*)}$.

3 New Results

3.1 Discrete Biorthogonality on the Upper Half-Plane

Now we give the isometric transform of the dual system $\Phi^\star := ((\phi_n)^\star, n \in \mathbb{N})$ to the upper half-plane. With straightforward computation it is easy to see, that

$$b_a^\star(K(z)) = b_a(-1)\frac{\overline{z}-\lambda_a}{\overline{z}-\overline{\lambda}_a} = b_a(K(\overline{z})), \quad r_a^\star(K(z)) = r_a(-1)\frac{\overline{z}+i}{\overline{z}-\overline{\lambda}_a} = r_a(K(\overline{z})).$$

Then

$$\widetilde{\Psi}_0(z) := (T\phi_0^\star)(z) = \frac{1}{\sqrt{\pi}}\frac{i+\overline{z}}{i+z}\frac{\phi_0(-1)}{\overline{z}-\overline{\lambda_{a_0}}} = \frac{i+\overline{z}}{i+z}\Psi_0(\overline{z}), \quad \widetilde{\Psi}_n(z) := (T\phi_n^\star)(z) =$$

$$\frac{1}{\sqrt{\pi}}\frac{1}{i+z}\phi_n^\star(K(z)) = \frac{1}{\sqrt{\pi}}\frac{i+\overline{z}}{i+z}\frac{\phi_n(-1)}{\overline{z}-\overline{\lambda_{a_n}}}\prod_{k=0}^{n-1}\frac{\overline{z}-\lambda_{a_k}}{\overline{z}-\overline{\lambda_{a_k}}} = \frac{i+\overline{z}}{i+z}\Psi_n(\overline{z}).$$

By definition, the dual system of (4) will be equal to

$$\widetilde{\Psi}_0^\lambda(z) := \frac{i+\overline{z}}{i+z}\frac{\frac{\sqrt{\Im\lambda}}{\sqrt{\pi}}}{\overline{z}-\overline{\lambda_0}} = \frac{i+\overline{z}}{i+z}\Psi_0^\lambda(\overline{z}),$$

$$\widetilde{\Psi}_n^\lambda(z) = \frac{i+\overline{z}}{i+z}\frac{\frac{\sqrt{\Im\lambda}}{\sqrt{\pi}}}{\overline{z}-\overline{\lambda_n}}\prod_{k=0}^{n-1}\frac{\overline{z}-\lambda_k}{\overline{z}-\overline{\lambda_k}} = \frac{i+\overline{z}}{i+z}\Psi_n^\lambda(\overline{z}).$$

For arbitrary values of the variables ω and w from $\mathbb{C}_+$ and for any n, $0 \le n < \infty$, the kernel function corresponding to the system (4) and its dual can be written also in closed form as follows:

$$\widetilde{K}_N(\omega, w) = \sum_{k=0}^{N-1}\Psi_k^\lambda(\omega)\overline{\widetilde{\Psi}_k^\lambda(w)} = \overline{\left(\frac{i+\overline{w}}{i+w}\right)}\sum_{k=0}^{N-1}\Psi_k^\lambda(\omega)\overline{\Psi_k^\lambda(\overline{w})} =$$

$$\frac{w-i}{\overline{w}-i}\frac{1-\overline{\widetilde{B}_N(\overline{w})}\widetilde{B}_N(\omega)}{2i\pi(w-\omega)}, \quad \omega \neq w.$$

$$\widetilde{K}_N(\omega,\omega) = \sum_{k=0}^{N-1}\Psi_k^\lambda(\omega)\overline{\widetilde{\Psi}_k^\lambda(\omega)} =: \frac{1}{\widetilde{\rho}_N(\omega)} = \frac{w-i}{\overline{w}-i}\sum_{k=0}^{N-1}\frac{\Im\lambda_k}{\pi(\omega-\lambda_k)(\omega-\overline{\lambda}_k)}.$$

Let us consider $a_k = K(\lambda_k) = \frac{i-\lambda_k}{i+\lambda_k}$. We assume that the Eq. (5) has N different solutions, denoted by z_k.

In what follows, we want to prove the analogue of Theorem 1 for the upper half-plane. Let us consider t_k, where $z_k = K(t_k) = \frac{i-t_k}{i+t_k}$ is the solution of the equation (5), and the following set of nodes on the upper half-plane:

$$\mathcal{C}_{N,u} = \{t_k : k = 0, \ldots, N-1\}. \tag{6}$$

Theorem 2 *The finite collection of* Ψ_n^λ, $(0 \le n < N)$ *and* $\widetilde{\Psi}_n^\lambda$, $(0 \le n < N)$ *are discrete biorthogonal systems with respect to the scalar product*

$$\langle F, G\rangle_N = \sum_{t\in\mathcal{C}_{N,u}} F(t)\overline{G(t)}\widetilde{\rho}_N(t),$$

namely $\langle\Psi_m^\lambda, \widetilde{\Psi}_n^\lambda\rangle_N = \delta_{mn} \quad (0 \le m, n < N)$.

Proof Let us denote by $\omega = K^{-1}(z) = i\frac{1-z}{1+z}$, $w = K^{-1}(\xi) = i\frac{1-\xi}{1+\xi}$, $a_k = K(\lambda_k) = \frac{i-\lambda_k}{i+\lambda_k}$, $z_k = K(t_k) = \frac{i-t_k}{i+t_k}$. Then

$$\overline{\left(\frac{i\frac{1-\xi}{1+\xi} - \lambda_k}{i\frac{1-\xi}{1+\xi} - \overline{\lambda_k}} \frac{|1+\lambda_k^2|}{1+\lambda_k^2}\right)} \frac{i\frac{1-z}{1+z} - \lambda_k}{i\frac{1-z}{1+z} - \overline{\lambda_k}} \frac{|1+\lambda_k^2|}{1+\lambda_k^2} = \overline{\left(\frac{\xi - a_k}{1-\overline{a_k}\xi}\right)} \frac{z-a_k}{1-\overline{a_k}z}. \tag{7}$$

According to (7) and the property $\overline{w} = K^{-1}(\xi^*)$ we get $\overline{\widetilde{B}_N(\overline{w})}\widetilde{B}_N(\omega) = \overline{B_N(\xi^*)}B_N(z)$. From this and the definition of the z_k it follows that

$$\overline{\widetilde{B}_N(\overline{t}_j)}\widetilde{B}_N(t_i) = \overline{B_N(z_j^*)}B_N(z_i) = \frac{u}{u} = 1.$$

From now the proof follows the same lines as the proof of Theorem 1. For $t_j \neq t_i$ we have

$$\sum_{k=0}^{N-1} \frac{\Psi_k^\lambda(t_i)\overline{\widetilde{\Psi}_k^\lambda(t_j)}}{\widetilde{K}_N(t_i,t_i)} = \frac{1}{\widetilde{K}_N(t_i,t_i)} \frac{t_j - i}{\overline{t}_j - i} \frac{1 - \overline{\widetilde{B}_N(\overline{t}_j)}\widetilde{B}_N(t_i)}{2i\pi(t_j - t_i)} = 0.$$

For $t_j = t_i$ we have $\sum_{k=0}^{N-1} \frac{\Psi_k^\lambda(t_i)\overline{\widetilde{\Psi}_k^\lambda(t_i)}}{\widetilde{K}_N(t_i,t_i)} = 1$. These relations imply that the matrices

$$A = [a_{ik}]_{i,k=0}^{N-1}, \ a_{ik} = \Psi_k^\lambda(t_i)/\widetilde{K}_N(t_i,t_i), \quad B = [b_{jk}]_{j,k=0}^{N-1}, \ b_{jk} = \widetilde{\Psi}_k^\lambda(t_j)$$

satisfy $AB^* = E$, where B^* means the adjoint matrix and E the identity matrix. From this it follows that $A = (B^*)^{-1}$, and then $B^*A = E$, which is equivalent to

$$\delta_{ij} = \sum_{k=0}^{N-1} b_{kj}a_{ki} = \sum_{k=0}^{N-1} \overline{\widetilde{\Psi}_j^\lambda(t_k)}\Psi_i^\lambda(t_k)/\widetilde{K}_N(t_k,t_k) = \langle \Psi_i^\lambda, \widetilde{\Psi}_j^\lambda \rangle_N.$$

For $\omega \in \mathbb{R}$, $\Psi_n^\lambda(\omega) = \widetilde{\Psi}_n^\lambda(\omega)$. If we choose in the proof of the theorem $u \in \mathbb{T}$, then the discretization points are all real numbers, $t_k \in \mathbb{R}$, $k = 0, \ldots, N-1$, and from Theorem 2 we reobtain the Theorem 2.2 of Eisner and Pap [5].

3.2 *Equilibrium Condition in* $\mathbb{D}$ *and on* $\mathbb{C}_+$

For $u \in \mathbb{T}$, in Pap and Schipp [1] it was proved that the elements of the set $\mathcal{Z}_{N,u}^{\mathbf{a}}$ satisfy an equilibrium condition. In what follows it is shown that for $u \in \mathbb{C} \setminus \{0\}$ the elements of $\mathcal{Z}_{N,u}^{\mathbf{a}}$ have similar property. i.e. the points are the solutions of similar equations.

For any complex number $z \in \mathbb{C}$ we introduce the polynomials

$$\omega_1(z) := \prod_{k=0}^{N-1}(z-a_k), \quad \omega_2(z) := \prod_{k=0}^{N-1}(1-\overline{a}_k z), \tag{8}$$

$$\omega(z) := \omega_1'(z)\omega_2(z) - \omega_2'(z)\omega_1(z) \ (z \in \mathbb{C}). \tag{9}$$

It is clear, that ω is a polynomial of degree $2N-2$. It is true (see in [1]), that if $\zeta \in \mathbb{C}$ is a root of ω, then ζ^* is also a root of ω with the same multiplicity. Denote by $\zeta_1, \zeta_1^*, \ldots, \zeta_s, \zeta_s^*$ the pairwise distinct roots of ω with the multiplicity $\nu_1, \nu_1, \ldots, \nu_s, \nu_s$. We mention, that the points $\zeta_1, \zeta_1^*, \ldots, \zeta_s, \zeta_s^*$ are in the same time critical points of the Blaschke product, i.e. the solution of equation $(B_N^{\mathbf{a}})'(z) = 0$.

Theorem 3 *In $B_N(z)$ let $c=1$, and let us set $u \in \mathbb{C} \setminus \{0\}$ such that the Eq. $B_N(z) = u$ has N different solution, and denote the set of the solutions by $\mathcal{Z}_{N,u}^{\mathbf{a}} = \{z_k,\ k = 0,\ldots,N-1\}$. Then the numbers $z_k \in \mathcal{Z}_{N,u}^{\mathbf{a}}$ $(k=0,\ldots,N-1)$ satisfy the following equations:*

$$\sum_{j=0,\ j\neq k}^{N-1} \frac{1}{z_k - z_j} = \sum_{\ell=1}^{s}\left(\frac{\nu_\ell}{2}\frac{1}{z_k-\zeta_\ell} + \frac{\nu_\ell}{2}\frac{1}{z_k-\zeta_\ell^*}\right) \quad (k=0,\ldots,N-1).$$

Proof Let $c=1$, and let u be a number, for which the equation $B_N(z) = u$ has N distinct solution, $\mathcal{Z}_{N,u}^{\mathbf{a}} = \{z_k, k=0,\ldots,N-1\}$. Denote $\varphi(z) := \prod_{j=0}^{N-1}\frac{z-a_j}{1-\overline{a}_j z} - u$. It is clear, that $\varphi(z) = 0$ if and only if $z = z_k$ $(k=0,\ldots,N-1)$. Let us denote by

$$f(z) := \prod_{k=0}^{N-1}(z-z_k), \quad g(z) = \omega_1(z) - u\cdot\omega_2(z). \tag{10}$$

Then $\varphi(z) = 0$ if and only if $g(z) = 0$. So the polynomials f and g have the same degree and roots, therefore $f = \lambda g$ with a constant $\lambda \in \mathbb{C}$. It is easy to see, that

$$\frac{1}{2}\frac{g''(z_k)}{g'(z_k)} = \frac{1}{2}\frac{f''(z_k)}{f'(z_k)} = \sum_{j=0,\ j\neq k}^{N-1}\frac{1}{z_k-z_j} \quad (k=0,\ldots,N-1). \tag{11}$$

By the definition of z_k

$$\prod_{j=0}^{N-1}\frac{z_k-a_j}{1-\overline{a}_j z_k} = \frac{\omega_1(z_k)}{\omega_2(z_k)} = u \quad (k=0,\ldots,N-1). \tag{12}$$

On the other hand according to (8)–(10) and (12) we get

$$\frac{g''(z_k)}{g'(z_k)} = \frac{\omega_1''(z_k) - u\cdot\omega_2''(z_k)}{\omega_1'(z_k) - u\cdot\omega_2'(z_k)} = \frac{u\cdot\omega_2(z_k)\omega_1''(z_k) - u^2\cdot\omega_2(z_k)\omega_2''(z_k}{u\cdot\omega_2(z_k)\omega_1'(z_k) - u^2\cdot\omega_2(z_k)\omega_2'(z_k)} = \frac{u\cdot\omega_2(z_k)\omega_1''(z_k) - u\cdot\omega_1(z_k)\omega_2''(z_k)}{u\cdot\omega_2(z_k)\omega_1'(z_k) - u\cdot\omega_1(z_k)\omega_2'(z_k)} = \frac{\omega'(z_k)}{\omega(z_k)}. \tag{13}$$

According to Lemma 1 from [1] the roots of ω are of the form $\zeta_1, \zeta_1^*, \dots, \zeta_s, \zeta_s^*$ with the multiplicity $\nu_1, \nu_1, \dots, \nu_s, \nu_s$. Consequently in ω'/ω every root appears with multiplicity 1 and thus we have the partial fraction decomposition

$$\frac{\omega'(z)}{\omega(z)} = \sum_{j=1}^{s}\left(\frac{A_j}{z-\zeta_j} + \frac{\tilde{A}_j}{z-\zeta_j^*}\right).$$

Consider $\omega(z) = (z-\zeta_j)^{\nu_j}P_j(z)$, where $P(\zeta_j)\neq 0$. Then $\frac{\omega'(z)}{\omega(z)} = \frac{\nu_j P_j(z) + (z-\zeta_j)P_j'(z)}{(z-\zeta_j)P_j(z)}$, Consequently $A_j = \lim_{z\to\zeta_j}(z-\zeta_j)\frac{\omega'(z)}{\omega(z)} = \nu_j$. Similarly, $\tilde{A}_j = \nu_j$, and by (11) and (13) we get

$$\frac{1}{2}\frac{g''(z_k}{g'(z_k)} = \frac{1}{2}\frac{\omega'(z_k)}{\omega(z_k)} = \sum_{j=1}^{s}\left(\frac{\nu_j/2}{z_k-\zeta_j} + \frac{\nu_j/2}{z_k-\zeta_j^*}\right),$$

and Theorem 3 is proved.

What kind of interpretation can we give for the points of $\mathcal{C}_{N,u}$, defined in (6)? For $u=1$, and then $t_k\in\mathbb{R}$ similar property was shown in [2], that we show in the previous theorem. For $\lambda_0,\dots,\lambda_{N-1}\in\mathbb{C}_+$ let us consider the polynomials

$$\varphi_1(z) := \prod_{k=0}^{N-1}(z-\lambda_k), \quad \varphi_2(z) := \prod_{k=0}^{N-1}(z-\overline{\lambda}_k)$$
$$\varphi(z) := \varphi_1'(z)\varphi_2(z) - \varphi_2'(z)\varphi_1(z) \quad (z\in\mathbb{C}).$$

It is clear, that φ is a polynomial of degree $2N-2$. It is easy to prove, that if d is a root of φ with multiplicity m, then $\overline{d}$ is also a root with the same multiplicity. Let us denote by $d_1, \overline{d}_1, \dots, d_{N-1}, \overline{d}_{N-1}$ the roots of φ, i.e.

$$\varphi(z) = \prod_{j=1}^{N-1}(z-d_j)(z-\overline{d}_j) \quad (z\in\mathbb{C}). \tag{14}$$

The numbers $a_k := K(\lambda_k)$ $(k=0,\dots,N-1)$ are in $\mathbb{D}$, and by (3)

$$B_N(K(z)) = B_N(-1)\prod_{k=0}^{N-1}\frac{z-\lambda_k}{z-\overline{\lambda}_k} = B_N(-1)\frac{\varphi_1(z)}{\varphi_2(z)}.$$

We note, that $\varphi(z) = 0$ if and only if $(B_N(K(z)))' = 0$. The functions $\omega_1, \varphi_1, \omega_2, \varphi_2$ and ω, φ can be expressed by each others:

$$(i+z)^N \omega_1(K(z)) = \omega_1(-1)\varphi_1(z), \quad (i+z)^N \omega_2(K(z)) = \omega_2(-1)\varphi_2(z),$$
$$(i+z)^{2N-2}\omega(K(z)) = -\omega_1(-1)\omega_2(-1)\varphi(z)$$

and consequently $\omega(K(d_j)) = 0$, if $d_j \neq -i$. Let $u \in \mathbb{D} \setminus \{0\}$ be such that the Eq. (5) has N different solutions, $\mathcal{Z}^{\mathbf{a}}_{N,u} = \{z_k, k = 0, \ldots, N-1\}$. Then the numbers $t_k := K^{-1}(z_k) \in \mathbb{C}_+$ are the solution of the equation

$$\frac{\varphi_1(z)}{\varphi_2(z)} = q := \frac{u}{B_N(-1)} \in \mathbb{D} \setminus \{0\}, \tag{15}$$

and we have

Theorem 4 *Let $q \in \mathbb{D} \setminus \{0\}$ and denote by $t_n \in \mathbb{C}_+$ $(n = 0, \ldots, N-1)$ the solutions of (15). Then the following equilibrium conditions are satisfied:*

$$\sum_{k=0,\ k\neq n}^{N-1} \frac{1}{t_n - t_k} = \frac{1}{2}\sum_{j=1}^{N-1}\left(\frac{1}{t_n - d_j} + \frac{1}{t_n - \overline{d}_j}\right) \quad (n = 0, \ldots, N-1).$$

The proof is the same as the proof of Theorem 5 in [2], with the condition $q \in \mathbb{D} \setminus \{0\}$ instead of $q \in \mathbb{T}$.

Proof By definition of t_k it follows that $g(z) := \varphi_1(z) - q\varphi_2(z) = 0$ if and only if $z = t_k$ $(k = 0, \ldots, N-1)$. Set $f(z) = \prod_{k=0}^{N-1}(z - t_k)$. The polynomials f and g have the same degree and roots, therefore $f = \lambda g$ with $\lambda \in \mathbb{C}$. It is easy to see, that

$$\frac{g''(t_n)}{2g'(t_n)} = \frac{f''(t_n)}{2f'(t_n)} = \sum_{k=0,\ k\neq n}^{N-1} \frac{1}{t_n - t_k} \quad (n = 0, \ldots, N-1).$$

By the definition of t_n $\frac{\varphi_1(t_n)}{\varphi_2(t_n)} = q \quad (n = 0, \ldots, N-1)$. On the other hand we get that

$$\frac{g''(t_n)}{g'(t_n)} = \frac{\varphi_1''(t_n) - q\varphi_2''(t_n)}{\varphi_1'(t_n) - q\varphi_2'(t_n)} = \frac{\varphi_2(t_n)\varphi_1''(t_n) - \varphi_1(t_n)\varphi_2''(t_n)}{\varphi_2(t_n)\varphi_1'(t_n) - \varphi_1(t_n)\varphi_2'(t_n)} = \frac{\varphi'(t_n)}{\varphi(t_n)}.$$

From (14) we have

$$\frac{\varphi'(t)}{\varphi(t)} = \sum_{k=1}^{N-1}\left(\frac{1}{t - d_k} + \frac{1}{t - \overline{d}_k}\right),$$

and the theorem is proved.

4 Conclusion

We introduced the dual system of the MT system on the upper half-plane, and similar as in [9], we obtained discrete biorthogonality result. We proved that the discretization nodes in both cases in the unite disc and upper half-plane satisfy an equilibrium condition. These properties can be considered as generalizations of the results obtained in [1, 2].

Notes and Comments. The research was financed by NKFIH in Hungary, within the framework of the 2020-4.1.1-TKP2020 3rd thematic programme of the University of Pécs.

Acknowledgements Zsuzsanna Nagy Csiha was supported by the ÚNKP-21-3 New National Excelence Program of the Ministry for Innovation and Technology.

References

1. Pap, M, and Schipp, F.: Malmquist-Takenaka systems and equilibrium conditions. Math. Pannon., 12, 2, pp 185–194, (2001)
2. Pap, M. and Schipp, F.: Equilibrium conditions for the Malmquist-Takenaka systems. Acta Sci. Math. (Szeged), 81, 3–4. pp 469–482, (2015)
3. Malmquist, F.: Sur la détermination d'une classe functions analytiques par leurs dans un esemble donné de points. Compute Rendus Six. Cong. math. scand. Kopenhagen Denmark, pp 253–259, (1925)
4. Takenaka, S.: On the orthogonal functions and a new formula of interpolation. Japanese J. Math. II. pp 129–145, (1925)
5. Eisner, T. and Pap, M.: Discrete orthogonality of the Malmquist-Takenaka system of the upper half plane and rational interpolation. J. Fourier Anal. Appl., 20, 4, pp 1–16, (2014)
6. Gaál, M. and Nagy, B. and Nagy-Csiha, Zs. and Révész, Sz. Gy.: Minimal energy point system on the unit circle and the real line. SIAM J. Math. Anal., 52, 6, pp 6281–6296, (2020)
7. Fridli, S., Lócsi, L. and Schipp, F.: Rational Function Systems in ECG Processing. In: R. Moreno-Diaz et al. (Eds.) EUROCAST 2011. pp 88–95, (2011)
8. Fridli, S., Kovács, P., Lócsi, L. and Schipp, F.: Rational modeling of multi-lead QRS complexes in ECG signals. Annales Univ. Sci, Budapest., Sect. Comp, 36, pp 145–155, (2012)
9. Fridli, S. and Schipp, F.: Discrete rational biorthogonal systems on the disc. Annales Univ. Sci. Budapest., Sect. Comp., 50, pp 127–134 (2020)
10. Džrbašjan, M. M.: Biorthogonal Systems of Rational Functions and best Approximation of the Cauchy Kernel on the Real Axis. Sbornik: Mathematics, 24, 3, pp 409–433, (1974)

A Modified Leslie–Gower Type Predation Model Considering Allee Effect on Prey and Competence Among Predators

Alejandro Rojas-Palma, Eduardo González-Olivares, and Paulo Tintinago-Ruiz

Abstract In this work, the analysis will be based on a Leslie–Gower type predation model, described by a two-dimensional system of ordinary differential equations, assuming that the prey population is influenced by an Allee effect, which modifies classical logistic equation. The functional response will be assumed linear, prey-dependent, and monotonously increasing. In turn, the equation of growth of predators will also be considered of like-logistic type, where the environmental carrying capacity for predators is assumed proportional to the prey population size. Among the most important results obtained is that for the same set of parameters, there are different behaviors of the system solutions, since two attractor singularities can appear simultaneously. Then, populations can coexist around fixed population sizes, or the prey population can become extinct.

Keywords Bifurcation · Stability · Limit cycle · Predator-prey model · Functional response

---

A. Rojas-Palma (✉)
Departamento de Matemática, Física y Estadística, Facultad de Ciencias Básicas, Universidad Católica del Maule, Talca, Chile
e-mail: amrojas@ucm.cl

E. González-Olivares
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
e-mail: eduardo.gonzalez@pucv.cl

P. Tintinago-Ruiz
Departamento de Matemática, Universidad del Quindío, Armenia, Colombia
e-mail: pctintinago@uniquindio.edu.co

F. Yilmaz et al. (eds.), *Mathematical Methods for Engineering Applications*,
Springer Proceedings in Mathematics & Statistics 384,
https://doi.org/10.1007/978-3-030-96401-6_27

1 Introduction

In Mathematical (or Theoretical) Ecology and in particular in Population Dynamics [1] the interaction between predators and their prey has been and will continue to be one of their fundamental subjects of study, as a consequence of their universal existence and high importance in nature [2].

In turn, the dynamics inherent to this relationship constitute an important field of study in Applied Mathematics, being approached by different perspectives and using different mathematical tools.

1.1 Brief Historical Review

The first predator-prey model was proposed by the Italian mathematician Vito Volterra (1860–1940) [1, 3], in a well-known monograph [4], which is described by a two-dimensional system of Ordinary Differential Equations. This model coincided with a model for biochemical interactions previously proposed by the American mathematician Alfred J. Lotka; so the formulated system is named as Lotka–Volterra model [5, 6].

The main feature of this first model is that the only positive equilibrium point is a center, that is, all paths are concentric closed orbits [1]. This implies that the population sizes of predators and their prey oscillate around a fixed point for any initial condition. These behaviors of the system's solutions were quite questioned, beginning so the debate concerning the realism of simplified models in Ecology, the question that is still a subject of scientific dispute [1]; moreover, in nature, there are no predation interactions with that characteristics.

From the work of Volterra, different proposals were followed to face and resolve the various objections raised. One of the first proposals to solve some of the objections to the Lotka–Volterra model was formulated by the Russian biologist Georgii F. Gause (1910–1986), who proposed in 1934 [7] a model that takes into account the intraspecific competition of dams, replacing the Malthusian growth incorporated in the Lotka–Volterra model, by the logistic growth function (also called Verhulst–Pearl equation [6])

An important result is a theorem proposed by the Russian mathematician Andrei N. Kolmogorov in 1936 [8], which ensures the existence of an ecologically stable periodic solution (mathematically, a solution called limit cycle), a phenomenon for which there is sufficient evidence in nature.

A different alternative is a model proposed by Patrick Holt Leslie (1900–1972) in 1948 [9], which does not fit the scheme of the Lotka–Volterra model [6]. Unlike the compartmentalized models of the Gause type [10], based on a principle of mass or energy transfer [2], Leslie's model is characterized in that the equation of growth of predators is of the logistic type, in which the environmental carrying capacity is proportional to the abundance of prey.

Clearly, the model of Leslie or Leslie–Gower [11], is not defined for $x = 0$, producing some anomalies in its predictions; it vaticinates that if the death rate of an individual predator is almost nil, the population of predators can grow, if the predator/prey ratio is very small, that is, if the predator population is even lower than the prey population [6].

1.2 Allee Effect

It is a social behavior [12–16], named in honor of the American ecologist Warder Clyde Allee (1885–1955), who was one of the first to study this phenomenon. It has other names in Population Dynamics such as: negative competition effect, inverse density dependence, positive density dependence or undercrowding [17]. In particular, in Fisheries Sciences, it is known as depensation [17, 18].

Undercrowding represents a negative effect of increased density on per capita growth [17] and it is defined as a positive relationship between a component of individual fitness and the number or density of conspecifics [13, 15, 16].

It is an ecological phenomenon that occurs at low population densities, where the per capita growth rate is a growing function of population abundance. At large population sizes, this rate is negative as it happens in the logistic equation for all population sizes.

Some populations may exhibit the Allee effect due to a wide range of biological phenomena, such as reduced vigilance antipredation, social thermoregulation, mating difficulty, and poor feeding at low densities. However, several other causes can lead to this phenomenon (see Table 1 on [12] or Table 2.1 on [19]).

The most common mathematical way to describe the Allee effect is the nonlinear differential equation.

$$\frac{dx}{dt} = r\,\left(1 - \frac{x}{K}\right)(x - m)\,x, \tag{1}$$

where $x = x\,(t)$ indicates the size of a population for $t \geq 0$. The parameter r is the intrinsic growth rate of and K indicates the environmental carrying capacity (having the same meanings as in the logistic equation); m is the parameter associated with the Allee effect.

If $m > 0$, the parameter is named the minimum of viable population or extinction threshold and it has a strong Allee effect. Clearly, if $x < m$ in Eq. (1), it has $\frac{dx}{dt} < 0$, implying that the population tends to extinction.

If $m \leq 0$, it has a weak Allee effect.

We note that equation (0) describes an Allee effect when the per capita growth rate is negative for values of the variable x close to zero, which happens, if and only if, $-K < m << K$.

1.3 The Functional Response

The functional response of predators or consumption function refers to the change in the density of prey attacked per unit of time per predator, when the density of prey changes [10]. They are classified into several types, depending only on the prey population size, or the size of both populations.

The first classification is due to the Canadian Ecologist Crawford S. Holling (1930–2019) in 1959 [20], who described three types of saturated functions, based on laboratory experiments; he considered them only dependent only on the prey population size (functional response prey-dependent). They are named as Holling, I, II, or III [6].

Laterly Robert J. Taylor in 1984 [21] proposed the functional response Holling type IV or non-monotonic [21] also prey-dependent, usually used to describe the ecological phenomenon called group defense formation.

We should note that there are many mathematical expressions to represent the saturated functions of Holling, except for Holling type I, which is described by

$$h(x) = \begin{cases} qx, & \text{if } x < a \\ qa, & \text{if } x \geq a \end{cases}$$

with $q > 0$.

Functional responses dependent on both populations have also been proposed, such as Beddington–DeAngelis, the rate-dependent, Crowley–Martin, Watt, or Hassel–Varley.

We note that the linear functional response $h(x) = qx$, with $q > 0$, does not correspond to any of the classifications proposed by Holling, because it is not saturated. However, it is considered in the Lotka–Volterra model [1], as well as in the Volterra [6] and Leslie–Gower [11] models.

The systems that describe these last two models are characterized by having a wide subset of parameters for which there is a unique positive equilibrium point that is globally asymptotically stable [22].

1.4 Competition Among Predators (CAP)

The oldest proposal for competition (or interference) among predators was formulated independently in 1975 by J. A. Beddington [23] and by Donald DeAngelis and collaborators [24], proposing a new functional response dependent on both interacting populations, which is described by the function

$$h(x, y) = \frac{qx}{a + bx + cy},$$

with $q, a, b, c > 0$.

A second form for the CAP was formulated by Herbert I. Freedman in 1979 [25], modifying the assumption of the usual models (the total rate of prey death prey, is a functional response-dependent only on prey population sizes, by the number or density of predators). Freedman proposed the function $h(x, y) = h(x)\, y^{\beta}$, where β is the mutual interference constant, so that $0 < \beta < 1$, and $h(x)$ is the functional response of the predator, which is prey-dependent.

Following Colin W. Clark [18], in bioeconomic models, the function $B(y) = y^{\beta}$ expresses the congestion among fishing vessels (men as predators) that harvest a bank of fish, resulting in a decrease in the catch rates [26].

The third way to describe the CAP is given by adding a negative quadratic term to the predator growth equation [3]. Therefore, the function that describes the mortality of predators takes the form $\varphi(y) = -cy - ey^2$, with c and $e > 0$. It is mainly used in predation models of the Gause type.

In this case, it is assumed that the population of predators can be reduced by other causes. One of them is the size of the appropriate habitat for the predator, to live and reproduce there [3].

2 Model Proposal

The model to be studied is a modification of the Leslie model proposed in 1948 [9] being described by an autonomous bi-dimensional nonlinear ordinary differential equation systems of the Kolmogorov type [10, 22]. The key factors of this interaction that we modify are:

(i) The prey population growth function, assuming that it is affected by the phenomenon called the Allee effect [13, 15, 16, 19].
(ii) The functional response [6, 10], including competition (or interference) among predators [3, 25, 27].
(iii) The predator population growth function, assuming that predators have an alternative food [28].

Considering the above mentioned, the model proposed is

$$G_{\mu}(x, y) : \begin{cases} \frac{dx}{dt} = \left(r\left(1 - \frac{x}{K}\right)(x - m) - q y^{\beta}\right) x \\ \frac{dy}{dt} = s\left(1 - \frac{y}{nx + c}\right) y, \end{cases} \tag{2}$$

where $x = x(t)$ e $y = y(t)$ are the prey and predator populations sizes and $\mu = (r, K, q, s, n, c, m, \beta) \in \mathbb{R}_{+}^{6} \times \,]-K, K[\, \times \,]0, 1[$. The parameters have the ecological meanings described in Table 1:

System (2) is defined throughout the first quadrant, that is, it is defined in

$$\Omega = \left\{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\right\}.$$

Table 1 Parameters and their ecological meanings in system (2)

Parameter	Meaning
r	Intrinsic prey growth rate or biotic potential
K	Prey environmental carrying capacity
q	Predators consumption rate
s	Intrinsic predator growth rate
n	Energy quality measure provided by the prey as food for predators
c	Amount of alternative food available for predators
m	Parameter describing the Allee effect
β	Mutual interference constant [25]

The equilibrium points of the system (1) or singularities of the vector field $G_\lambda(x, y)$ are: $(0, 0)$, $(m, 0)$, when $m > 0$, $(K, 0)$, $(0, c)$, and those who are at the intersection of the isoclines $y = nx + c$ and

$$y = \left(\frac{r}{q}\left(1 - \frac{x}{K}\right)(x - m)\right)^{\frac{1}{\beta}}.$$

2.1 Topologically Equivalent System

In order to simplify the calculations, a change of variables and a rescaling of the time is made, which is given in the following:

Lemma 1 *System (2) is topologically equivalent to the system*

$$G_\nu(u, v) : \begin{cases} \frac{du}{d\tau} = \left((1-u)\ (u-M)\ - Qv^\beta\right) u\,(u+C) \\ \frac{dv}{d\tau} = \ S\,(u + C - v\,)\,v \end{cases} \tag{3}$$

where $\nu = (Q, S, C, M, \beta) \in \mathbb{R}_+^3 \times\,]-1, 1[\, \times\,]0, 1[$ *with* $C = \frac{c}{nK}$, $M = \frac{m}{K}$, $Q = \frac{q(nK)^\beta}{rK}$ *and* $S = \frac{s}{rK}$.

Proof Let $x = Ku$ and $y = nKv$. Factoring, simplifying and it has effecting the rescaling of the time given by $\tau = \frac{rK}{u+\frac{c}{nK}}t$, $\frac{du}{dt} = \frac{du}{d\tau}\frac{d\tau}{dt}$ it has

$$V_\eta(u, v) : \begin{cases} \frac{rK}{u+\frac{c}{nK}}\frac{du}{d\tau} = rK\left((1-u)\left(u - \frac{m}{K}\right) - \frac{q(nK)^\beta}{rK}v^\beta\right)u \\ \frac{rK}{u+\frac{c}{nK}}\frac{dv}{d\tau} = s\left(1 - \frac{v}{u+\frac{c}{nK}}\right)v, \end{cases}$$

Finally,

$$V_\eta(u,v): \begin{cases} \frac{du}{d\tau} = \left((1-u)\left(u-\frac{m}{K}\right) - \frac{q(nK)^\beta}{rK}v^\beta\right)\left(u+\frac{c}{nK}\right)u \\ \frac{dv}{d\tau} = \frac{s}{rK}\left(u+\frac{c}{nK}-v\right)v, \end{cases}$$

Defining $M = \frac{m}{K}$, $S = \frac{s}{rK}$, $C = \frac{c}{nK}$ and $Q = \frac{q(nK)^\beta}{rK}$, system (3) is obtained. □

System (3) or vector field $G_\nu(u,v)$ is defined in $\bar{\Omega} = \left\{(u,v) \in \mathbb{R}^2 : u \geq 0, v \geq 0\right\}$.

Remark 1 We have built a diffeomorphism $\varphi : \bar{\Omega} \times \mathbb{R} \to \Omega \times \mathbb{R}$, such that

$$\varphi(u,v,\tau) = \left(Ku, nKv, \frac{u+\frac{c}{nK}}{rK}\tau\right) = (x,y,t).$$

Their differential is

$$D\varphi(u,v,\tau) = \begin{pmatrix} K & 0 & 0 \\ 0 & nK & 0 \\ \frac{1}{rK} & 0 & \frac{u+\frac{c}{nK}}{rK} \end{pmatrix},$$

and $\det D\varphi(u,v,\tau) = \frac{nK}{r}\left(u+\frac{c}{nK}\right)$, the diffeomorphism preserves the orientation of time. The topological equivalence between the systems ensures that the qualitative behavior between them is the same. For this reason, in what follows, we will consider the analytical results in the reparametrized system (3).

2.2 *Positively Invariant Region and Boundedness of Solutions*

For the system (3) or vector field $G_\nu(u,v)$ we have the following properties that we show below.

Lemma 2 *The set* $\bar{\Gamma} = \left\{(u,v) \in \bar{\Omega} : 0 \leq u \leq 1,\ v \geq 0\right\}$ *is a positively invariant region.*

Proof Where the system is of Kolmogorov type, the coordinate axes are invariant sets. If $u = 1$, we have that the first equation of the system (3) is $\frac{du}{d\tau} = -Qv^\beta(1+C) < 0$, which indicates that the solutions of the vector field cross the line $u = 1$, whatever the direction of the second component of the vector field $G_\nu(u,v)$. □

Remark 2 (i) The set $\bar{\Gamma}_C = \left\{(u,v) \in \bar{\Omega} : 0 \leq u \leq 1,\ 0 \leq v \leq u + C\right\} \subset \bar{\Gamma}$ is also a positively invariant region.

(ii) In the original system (2) the set $\Gamma = \{(x,y) \in \Omega : 0 \leq x \leq K,\ y \geq 0\}$, and also the subset $\Gamma_c = \{(x,y) \in \Omega : 0 \leq x \leq K,\ 0 \leq y \leq nx + c\} \subset \Gamma$ are positively invariant regions.

Lemma 3 *All solutions of the system (3) are uniformly bounded.*

Proof From the first equation of system (3) we have

$$\frac{du}{d\tau} \leq (1-u)(u-M)u,$$

for all $v \geq 0$, when $0 < u < 1$.

We also know that

$$\begin{cases} u \to 1, & \text{when } \tau \to \infty \text{ and } u < 1. \\ u \to 1, & \text{when } \tau \to \infty \text{ and } u > 1. \end{cases}$$

Let $L = max\,\{u\,(0)\,,1\}$ therefore $u\,(t) \leq L$, for all $t \geq 0$.

Defining the function $w\,(u,v) = \ u + \frac{1}{S}v$, it has, $w\,(u,v) > 0$, for all $t \geq 0$, and

$$\frac{dw}{d\tau} = \frac{du}{d\tau} + \frac{1}{S}\frac{dV}{d\tau} = (1-u)(u-M)u - Quv + (u + C - v\ )v.$$

Considering $\frac{dw}{d\tau} + \sigma w$, with $\sigma > 0$, we obtain

$$\frac{dw}{d\tau} + \sigma w = Mu^2 - Mu + u^2 - u^3 - Quv + (u+C)v - v^2 + \sigma u + \frac{\sigma}{S}v.$$

Thus,

$$\frac{dw}{d\tau} + \sigma w \leq (M+1)u^2 + \frac{\left(u+C+\frac{\sigma}{S}\right)^2}{4} + \sigma u \leq (M+1)L^2 + + \frac{\left(L+C+\frac{\sigma}{S}\right)^2}{4} + \sigma L$$

If we denote $N = (M+1)L^2 + \frac{\left(L+C+\frac{\sigma}{S}\right)^2}{4} + \sigma L$ then $0 < \frac{dw}{d\tau} + \sigma w \leq N$, which is a first-order differential inequality. Applying the Comparison Theorem for Differential Inequalities (page 30 in [29]), we obtain

$$0 < w\,(\tau)\,e^{\tau} \leq Ne^{\tau} + N_2.$$

when $\tau = 0$, it obtains $w\,(0) \leq N + N_2$; i.e., $N_2 \geq w\,(0) - N$.

Then, there is $p \in \mathbb{N}$ such that $N_2 \leq p\,(w\,(0) - N)$;

Thus, $w\,(\tau)\,e^{\tau} \leq Ne^{\tau} + p\,(w\,(0) - N)$, i.e., $w\,(t) \leq N + p\,(w\,(0) - N)\,e^{-t}$.

Accordingly, when $\tau \to \infty$ then $w\,(\tau) \leq N$. Therefore, all solutions are uniformly bounded. □

2.3 Equilibria Existence

The equilibrium points of the system (3) or vector field $G_\nu(u, v)$ are $(0, 0)$, $(1, 0)$, $(0, C)$, $(M, 0)$, when $M > 0$, and those that are at the intersection of the isoclines $v = u + C$ and

$$v = \left(\frac{1}{Q}(1-u)(u-M)\right)^{\frac{1}{\beta}}.$$

In what follows we will assume that $M > 0$.

We note that the second isocline is defined for $M \leq u \leq 1$, except for some particular cases of the parameter β.

The abscissa of the positive equilibrium points or at interior of the first quadrant satisfy the transcendent equation:

$$p(u) = Q(u+C)^\beta - (1-u)(u-M) = 0 \tag{4}$$

Lemma 4 *Equation (4) has at most two positive real roots, or one of multiplicity two, or none.*

Proof Considering the functions $g_1(u) = Q(u+C)^\beta$ and $g_2(u) = (1-u)(u-M)$. The function g_2 represents a parabola, and the intercepts on the $u - axis$ are $u = 1$ and $u = M$. Besides, it has a maximum value in $u_{\max} = \frac{1+M}{2}$; then, $g_2(u_{\max}) = \frac{1}{4}(1-M)^2$.

The functions g_1 and g_2 have intersection, if and only if, $g_1(0) < g_2(u_{max})$, i.e., $QC^\beta < \frac{1}{4}(1-M)^2$.

Now, the intersection of the curves will be determined considering the tangency of the graphics of these functions. When the curves are tangent, their derivatives must be equals.

Thus, $\frac{d}{du}(g_1(u)) = \beta Q(u+C)^{\beta-1}$ and $\frac{d}{du}(g_2(u)) = M - 2u + 1$.

Let u^* be the abscissa of the point of tangency of both curves, when it exists. Then

$$\beta Q\left(u^* + C\right)^{\beta-1} = M - 2u^* + 1.$$

Furthermore, u^* is solution of the equation (4), i.e.,

$$Q\left(u^* + C\right)^\beta = \left(1-u^*\right)\left(u^* - M\right).$$

Therefore, u^* is solution of the equation polynomial.

$$(2-\beta)(u^*)^2 + (2C - (1-\beta)(M+1))u^* - (M\beta + C(M+1)) = 0. \tag{5}$$

Whatever the sign of the coefficient $b_0 = 2C - (1-\beta)(M+1)$, the Eq. (5) has two real roots:

$$u_1^* = \frac{1}{2\,(2-\beta)}\left(b_0 - \sqrt{\Delta_T}\right) \text{ and } u_2^* = \frac{1}{2\,(2-\beta)}\left(b_0 + \sqrt{\Delta_T}\right),$$

with $u_1^* < 0 < u_2^*$, and

$$\Delta_T = (1-\beta)^2 M^2 + (2\,(2C+1+\beta\,(2-\beta)))\,M + \left((2C+1)^2 - \beta\,(2-\beta)\right),$$

Therefore, Eq. (4) can have up to two positive real roots. If they exist, they are $u_1 = u_2^* - \epsilon$ and $u_2 = u_2^* + \epsilon$, with $\epsilon > 0$. □

From the result above, the following statement is immediate

Corollary 1 *The system (3) has two positive equilibrium points, one or none.*

3 Main Results

3.1 Boundary Equilibria Stability

The Jacobian or community matrix of the system (3) is:

$$DG_\nu\,(u,v) = \begin{pmatrix} DG_\nu\,(u,v)_{11} & -Q\beta u\,(u+C)\,v^{\beta-1} \\ Sv & S\,(u+C-2v) \end{pmatrix}$$

where

$$\begin{aligned} DG_\nu\,(u,v)_{11} = {} & \left((1-u)\ (u-M)\ - Qv^\beta\right)u + \left((1-u)\ (u-M)\ - Qv^\beta\right)(u+C) \\ & + u\,(u+C)\,(M-2u+1)\,. \end{aligned}$$

The nature of the equilibrium points of the system (3) cannot be determined using the standard methodology. The Jacobian matrix is not defined for $v = 0$, although the system is defined for that value.

In order to determine that nature, we carry out the change of variable given by $w = v^\beta$. Thus, $\frac{dv}{d\tau} = \frac{1}{\beta} w^{\frac{1}{\beta}-1} \frac{dw}{d\tau}$.

Replacing in the second equation of the system (3) we obtain the new system

$$W_\eta\,(u,w) : \begin{cases} \frac{du}{d\tau} = ((1-u)\,(u-M) - Qw)\,(u+C)\,u \\ \frac{dw}{d\tau} = \beta S\left(u + C - w^{\frac{1}{\beta}}\right)w, \end{cases} \tag{6}$$

where $\eta = (Q, M, S, C, \beta) \in \mathbb{R}_+^3 \times]0,1[$. The equilibrium points are $(0,0)$, $(1,0)$, $(M,0)$, $\left(0, C^\beta\right)$ and $\left(u_e^*, \left(u_e^* + C\right)^\beta\right)$, with u_e determined by the intersection of the isoclines $h_1(u) = (u + C)^{\beta}$ and $h_2(u) = \frac{1}{Q}(1 - u)(u - M)$, or u_e satisfies equation (4)

$$p(u) = Q(u + C)^{\beta} - (1 - u)(u - M) = 0.$$

The Jacobian matrix of system (6) is:

$$DW_{\eta}(u, w) = \begin{pmatrix} DW_{\sigma}(u, w)_{11} & -Qu(u + C) \\ \beta S w & \beta S\left(\beta(u + C) - (1 + \beta) w^{\frac{1}{\beta}}\right) \end{pmatrix},$$

with $DW_{\sigma}(u, w)_{11} = u(u + C)(-1) + (1 - u - Qw)(u + C) + (1 - u - Qw)u$.

Clearly, the Jacobian matrix is defined for $v = 0$.

The nature of the equilibrium points of the system (6) on the $u - axis$ is determined by the following proposition:

Proposition 1 *(a) The equilibrium point* $(0, 0)$ *is a hyperbolic repeller.*
(b) The equilibrium point $(1, 0)$ *is a hyperbolic saddle point.*
(c) The equilibrium point $(M, 0)$ *is a hyperbolic repeller.*

Proof We prove only (a) because the other proofs are similar. The Jacobian matrix evaluated at (0, 0) is:

$$DW_{\eta}(0, 0) = \begin{pmatrix} C & 0 \\ 0 & \beta^2 C S \end{pmatrix}.$$

So, $\det(DW_{\eta})(0, 0) = \beta^2 C^2 S > 0$ and $\operatorname{tr}(W_{\eta})(0, 0) = C\left(1 + \beta^2 S\right) > 0$.

Therefore, by the trace and determinant theorem [30], the equilibrium point (0, 0) is a hyperbolic repeller. □

It can be concluded that for system (3), the equilibrium points on the $u - axis$ have the following nature.

Corollary 2 *The equilibrium point*

(a) $(0, 0)$ *is a non-hyperbolic repeller,*
(b) $(1, 0)$ *is a non-hyperbolic saddle point,*
(c) $(M, 0)$ *is a non-hyperbolic repeller.*

We observe that when $\beta \to 1$, the system (3) has a behavior similar to the case in which $\beta = 1$, model that was analyzed in [14].

To determine the nature of the other equilibrium points, we consider again the Jacobian matrix of the system (3).

Lemma 5 *The equilibrium point* $(0, C)$ *is a hyperbolic attractor for all parameter values.*

Proof The Jacobian matrix evaluated at the point $(0, C)$ is:
$$DY_{\lambda}(0, C) = \begin{pmatrix} -(M+\beta QC)C & 0 \\ SC & -SC \end{pmatrix}.$$
We have: $\det(DY_{\lambda})(0, C) = S\left(M + QC^{\beta}\right)C^2 > 0$,
and $\operatorname{tr}(DY_{\lambda})(0, C) = -\left(M + QC^{\beta}\right)C - SC < 0$. From the above, the point $(0, C)$ is a hyperbolic attractor. $\square$

3.2 *Nature of Positive Equilibria*

The Jacobian matrix evaluated at the points $(u_e, u_e + C)$ is

$$DG_{\nu}(u_e, u_e + C) = \begin{pmatrix} u_e(u_e + C)(M - 2u_e + 1) & -Q\alpha u_e(u_e + C)^{\beta} \\ S(u_e + C) & -S(u_e + C) \end{pmatrix},$$

Then,

$$\begin{aligned} \det DG_{\nu}(u_e, u_e + C) &= Su_e(u_e + C)\left(Q\alpha(u_e + C)^{\beta} - (M - 2u_e + 1)(u_e + C)\right) \\ &= Su_e(u_e + C)(u_e + C)^{\beta}\left(Q\beta - (M - 2u_e + 1)(u_e + C)^{1-\beta}\right), \end{aligned}$$

As $Q(u + C)^{\beta} = (1 - u)(u - M)$, then the sign of $\det(DG_{\nu})(u_e, u_e + C)$ depends on the factor $L_1 = Q\beta(u_e + C)^{\beta} - (M - 2u_e + 1)(u_e + C)$.

Further, $\operatorname{tr}(DG_{\nu})(u_e, u_e + C) = (u_e + C)((M - 2u_e + 1)u_e - S)$, whose sign depends on the factor $L_2 = (M - 2u_e + 1)u_e - S$.

Supposing there are two positive equilibrium points $(u_{e1}, u_{e1} + C)$ and $(u_{e2}, u_{e2} + C)$, with $M < u_{e1} < u_{e2} < 1$. Where
$u_{e1} = u_2^* - \epsilon$ and $u_{e1} = u_2^* + \epsilon$, con $\epsilon > 0$ and $u_2^* = \frac{1}{2(2-\beta)}\left(b_0 + \sqrt{\Delta_T}\right)$.

Considering $u_2^* = \frac{1}{2(2-\beta)}\left(b_0 + \sqrt{\Delta_T}\right)$, the abscissa of the tangency point, when exist; the abscissa of the positive positive equilibrium points can be expressed as $u_{e1} = u_2^* - \epsilon$ and $u_{e1} = u_2^* + \epsilon$, with $\epsilon > 0$.

Recalling equation (4) we have $Q(u_e + C)^{\beta} = (1 - u)(u - M)$.

Then,

$$\begin{aligned} L_1 &= \beta(1 - u)(u - M) - (M - 2u + 1)(u + C) \\ &= (2 - \beta)u^2 + (2C - (1 - \beta)(M + 1))u - (M\beta + C(M + 1)), \end{aligned}$$

which matches the equation for T, the factor that determines the tangency of the curves g_1 and g_2.

Lemma 6 *The equilibrium $(u_{e1}, u_{e1} + C)$ when it exists, is a hyperbolic saddle*

Proof Since $u_{e1} < u_2^* = \frac{1}{2(2-\beta)}\left(b_0 + \sqrt{\Delta_T}\right)$, with $b_0 = 2C - (1 - \beta)(M + 1)$, and

$$\Delta_T = (1 - \beta)^2 M^2 + (2(2C + 1 + \beta(2 - \beta)))M + \left((2C + 1)^2 - \beta(2 - \beta)\right).$$

Hence, the factor

$$L_1 = \beta (1 - u_{e1}) (u_{e1} - M) - (M - 2u_{e1} + 1) (u_{e1} + C) < 0.$$

Thus, it is immediate that $\det(DG_\nu)(u_{e1}, u_{e1} + C) < 0$, and the equilibrium point $(u_1, u_1 + C)$ is a hyperbolic saddle. □

Theorem 1 *Existence of a homoclinic curve and a non-infinitesimal limit cycle*
There are conditions on the parameter values for which:

(a) It exists a homoclinic curve determined by the stable and unstable manifold of point $(u_1, u_1 + C)$.
(b) It exists a non-infinitesimal limit cycle that bifurcates from the homoclinic [31] surrounding the point $(u_2, u_2 + C)$.

Proof (a) The path determined by the right unstable manifold W_+^u cannot intersect the line $u = 1$, since $\bar{\Gamma}$ is a positively invariant region. Then, its $\omega - limit$ should be:

(i) the point $(u_{e2}, u_{e2} + C)$, when this is an attractor;
(ii) a stable limit cycle, if $(u_{e2}, u_{e2} + C)$ is a repeller;
(iii) the point $(0, C)$.

On the other hand, the $\alpha - limit$ of W_+^s may be the point $(u_{e2}, u_{e2} + C)$, or an unstable limit cycle surrounding that point, or it may be to the right, in the $x - axis$ direction, outside of $\bar{\Gamma}$.

Therefore, by the theorem of existence and uniqueness of solutions and the geometry of the stable and unstable manifolds of the saddle point $(u_{e1}, u_{e1} + C)$, there is a subset in the parameter space for which W_+^u intersects with W_+^s, i.e., $W_+^s(u_{e1}, u_{e1} + C) \cap W_+^u(u_{e1}, u_{e1} + C) \neq \phi$, and a homoclinic curve is obtained.

(b) When the point $(u_2, u_2 + C)$ is an attractor and the $\omega - limit$ of the right unstable manifold $W_+^u(u_{e1}, u_{e1} + C)$ is the point $(0, C)$, there exists an unstable limit cycle dividing the behavior of trajectories in the neighborhood of $(u_2, u_2 + C)$, which is the frontier of the basin of attraction of that point. □

Theorem 2 *Denoting by $W_+^u(1, 0)$ the upper unstable manifold of saddle point $(1, 0)$ and $W_+^s(u_{e1}, u_{e1} + C)$ the upper unstable manifold of saddle point $(u_{e1}, u_{e1} + C)$. Let's consider the points $(u^*, v^u) \in W_+^u(1, 0)$ and $(u^*, v^s) \in W_+^u(u_{e1}, u_{e1} + C)$, with $M < u_{e1} < u^* < 1$.*

The relationship between v^u and v^s determines the stability of the point $(u_{e2}, u_{e2} + C)$ as follows:

(a) Assuming that $v^s > v^u$, then the equilibrium point $(u_{e2}, u_{e2} + C)$ is:
(a.1) a hyperbolic attractor, if and only if, $(M - 2u_{e2} + 1) u_{e2} < S$.
(a.2) a hyperbolic repeller, if and only if, $(M - 2u_{e2} + 1) u_{e2} > S$.
(a.3) a weak focus, if and only if, $(M - 2u_{e2} + 1) u_{e2} = S$.
(b) Assuming that $v^s < v^u$, then, it is a hyperbolic repeller node or focus and the trajectories of the system (2) have the point $(0, C)$ as their $\omega - limit$.
The point $(0, C)$ is an almost globally stable equilibrium [32].

Proof (a) It is also immediate, evaluating the Jacobian matrix, and considering the relative positions of the stable manifold of the point $(0, C)$ and the unstable manifold of the point $(1, 0)$.

(b) When $(M - 2u_{e2} + 1)\, u_{e2} > S$, there exists a limit cycle, whose diameter increases until to coincide with the heteroclinic curve joining the equilibrium points $(0, C)$ and $(1, 0)$.

Laterly, this heteroclinic breaks up and the singularity $(0, C)$ is an attractor for almost all trajectories [32]. □

Theorem 3 *The equilibrium point* $\left(u_2^*, u_2^* + C\right)$ *with* $u_2^* = \frac{1}{2(2-\beta)}\left(b_0 + \sqrt{\Delta_T}\right)$, *and*

$$\Delta_T = (1-\beta)^2 M^2 + (2\,(2C + 1 + \beta\,(2-\beta)))\, M + \left((2C+1)^2 - \beta\,(2-\beta)\right),$$

is (a) a saddle-node attractor, if and only if, $S > \frac{u_2^*\left(1-2u_2^*-A\right)}{A+u_2^*}$,

(b) a saddle-node repeller, if and only if, $S < \frac{u_2^*\left(1-2u_2^*-A\right)}{A+u_2^*}$,

(c) a cusp point, if and only if, $S = \frac{u_2^*\left(1-2u_2^*-A\right)}{A+u_2^*}$.

Proof Applying the conditions for the existence of the saddle-node equilibrium point when the positive equilibrium points coincide. □

4 Conclusions

In this work, we studied a Leslie–Gower type predation model, described by a two-dimensional system of ordinary differential equations, assuming that the prey population is affected by an Allee effect and competence among predators generalists. The inclusion of the Allee effect, affecting prey at low population densities has not only ecological importance but also mathematical importance since it originates in the models a much richer dynamic than that of the model that does not consider this effect.

In summary, the following aspects of the modified Leslie–Gower model [11] can be highlighted.

(i) There can be up to two positive equilibrium points.
(ii) There is a separatrix curve $\bar{\Sigma}$ in the phase plane (u, v), dividing the behavior of the solutions of the system. Two trajectories with very close initial conditions, but on a different side of that separatrix can have very far $\omega - limits$, such as the point $(0, C)$ or a positive equilibrium point, or else a stable limit cycle.

This implies that, for the same subset of parameters, there is a high sensitivity of the solutions to the initial conditions. So, for population sizes of prey and predators close to the separatrix curve can occur:

(a) the prey population can disappear and the predator population attains their maximum size C, or
(b) coexist with its predators in the long term, either at a fixed equilibrium point or oscillating around that point.

(iii) There are parameter values for which there is a homoclinic orbit generated by a hyperbolic saddle point. The breaking of this curve when the parameters vary originates a non-infinitesimal unstable limit cycle obtaining a Homoclinic Bifurcation [3].

Nonetheless, the inclusion of competition among predators has not had a significant impact on the dynamics of the model. Comparing our results with those obtained in [14], we see that the dynamics are similar and therefore it seems, that the parameter β does not have a strong impact on the behavior of the model. However, with the model studied in [33] there is a great difference since in this model the global stability of the unique positive equilibrium point is obtained.

It is well known that any model represents an abstraction of reality. The problem is not whether they are true, but whether the results obtained are valid and credible, under the stated assumptions underlying the model.

A credible predator-prey model must possess reasonable biological and ecological properties [2]. In addition, there are minimal attributes that determine various conditions for the growth equation of prey and predators [2].

We consider that the analyzed model satisfies these requirements and that the ecological interpretations of the analytical results obtained are consistent and congruent with the stated assumptions.

On the other hand, these results obtained can serve as a basis or be compared with models in which other types of mathematical tools are used.

References

1. Bacaër, N.: A Short History of Mathematical Population Dynamics. Springer London (2011). https://doi.org/10.1007/978-0-85729-115-8
2. Berryman, A.A., Gutierrez, A.P., Arditi, R.: Credible, parsimonious and useful predator-prey models: A reply to abrams, gleeson, and sarnelle. Ecology **76** (1995). https://doi.org/10.2307/1940728
3. Bazykin, A.D., Khibnik, A.I., Krauskopf, B.: Nonlinear Dynamics of Interacting Populations. WORLD SCIENTIFIC (1998). https://doi.org/10.1142/2284
4. Volterra, V.: Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. Memorie della Classe di Scienze Fisiche, Matematiche e Naturali. Societá anonima tipografica "Leonardo da Vinci" (1927)
5. May, R.: Stability and Complexity in Model Ecosystems. Princeton Landmarks in Biology. Princeton University Press (2019)
6. Turchin, P.: Complex Population Dynamics: A Theoretical/Empirical Synthesis (MPB-35). Monographs in Population Biology. Princeton University Press (2013)

7. Gause, G.: The Struggle for Existence. Dover Books on Biology. Dover Publications (2019)
8. Kolmogoroff, A.: Sulla teoria di volterra della lotta per l'esistenza. Giornale dell'Istituto Italiano degli Attuari **7** (1936)
9. Leslie, P.H.: Some further notes on the use of matrices in population mathematics. Biometrika **35** (1948). https://doi.org/10.2307/2332342
10. Freedman, H.: Deterministic Mathematical Models in Population Ecology. Monographs and textbooks in pure and applied mathematics. M. Dekker (1980)
11. Leslie, P.H., Gower, J.C.: The properties of a stochastic model for the predator-prey type of interaction between two species. Biometrika **47** (1960). https://doi.org/10.2307/2333294
12. Courchamp, L.B.E.A.F.: Multiple Allee effects and population management. Trends in Ecology & Evolution **22** (2007). https://doi.org/10.1016/j.tree.2006.12.002
13. Grenfell, F.C.T.C.B.B.: Inverse density dependence and the Allee effect. Trends in Ecology & Evolution **14** (1999). https://doi.org/10.1016/s0169-5347(99)01683-3
14. González-Olivares, E., Mena-Lorca, J., Rojas-Palma, A., Flores, J.D.: Dynamical complexities in the Leslie-Gower predator-prey model as consequences of the Allee effect on prey. Applied Mathematical Modelling **35** (2011). https://doi.org/10.1016/j.apm.2010.07.001
15. Sutherland, P.A.S.W.J.: Consequences of the Allee effect for behaviour, ecology and conservation. Trends in Ecology & Evolution **14** (1999). https://doi.org/10.1016/s0169-5347(99)01684-5
16. P. A. Stephens, W.J.S., Freckleton, R.P.: What is the Allee effect? Oikos **87** (1999). https://doi.org/10.2307/3547011
17. Hilborn, L.: Depensation: evidence, models and implications. Fish and Fisheries **2** (2001). https://doi.org/10.1046/j.1467-2979.2001.00029.x
18. Clark, C., Clark, P.: Mathematical Bioeconomics: The Optimal Management of Renewable Resources. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley (1990)
19. Courchamp, F., Berec, L., Gascoigne, J., Press, O.U.: Allee Effects in Ecology and Conservation. Oxford biology readers. OUP Oxford (2008)
20. Holling, C.S.: The components of predation as revealed by a study of small-mammal predation of the european pine sawfly. The Canadian Entomologist **91** (1959). https://doi.org/10.4039/Ent91293-5
21. Taylor, R.: Predation. Population and Community Biology. Springer Netherlands (2013)
22. Goh, B.: Management and Analysis of Biological Populations. ISSN. Elsevier Science (2012)
23. Beddington, J.: Mutual interference between parasites or predators and its effect on searching efficiency. Journal of Animal Ecology **44**, 331–340 (1975)
24. DeAngelis, D.L., Goldstein, R.A., O'Neill, R.V.: A model for tropic interaction. Ecology **56** (1975). https://doi.org/10.2307/1936298
25. Freedman, H.: Stability analysis of a predator-prey system with mutual interference and density-dependent death rates. Bulletin of Mathematical Biology **41** (1979). https://doi.org/10.1016/s0092-8240(79)80054-3
26. González-Olivares, E., Sáez, E., Stange, E., Szántó, I.: Topological description of a non-differentiable bioeconomics model. Rocky Mountain Journal of Mathematics **35**, 1133–1155 (2005)
27. González-Olivares, E., Cabrera-Villegas, J., Córdova-Lepe, F., Rojas-Palma, A.: Competition among predators and Allee effect on prey, their influence on a gause-type predation model. Mathematical Problems in Engineering **2019** (2019). https://doi.org/10.1155/2019/3967408
28. González-Olivares, E., Arancibia-Ibarra, C., Rojas-Palma, A., González-Yañez, B.: Dynamics of a modified Leslie-Gower predation model considering a generalist predator and the hyperbolic functional response. Mathematical Biosciences and Engineering **16**, 7995 (2019). https://doi.org/10.3934/mbe.2019403
29. Birkhoff, G., Rota, G.: Ordinary Differential Equations. Introductions to higher mathematics. Ginn (1962)
30. Murray, J.: Mathematical Biology: I. An Introduction. Interdisciplinary Applied Mathematics. Springer New York (2013)

31. Gaiko, V.: Global Bifurcation Theory and Hilbert's Sixteenth Problem. Mathematics and Its Applications. Springer US (2013)
32. Rantzer, A.: A dual to lyapunov's stability theorem. Systems & Control Letters **42** (2001). https://doi.org/10.1016/s0167-6911(00)00087-6
33. González-Olivares E.; Rojas-Palma, A.: Global stability in a modified Leslie-Gower type predation model assuming mutual interference among generalist predators. Mathematical Biosciences and Engineering **17**, 7708 (2020). https://doi.org/10.3934/mbe.2020392