UJMA

Universal Journal of Mathematics and Applications

VOLUME IV
ISSUE I

ISSN 2619-9653

http://dergipark.gov.tr/ujma

VOLUME IV ISSUE I
ISSN 2619-9653

March 2021
http://dergipark.gov.tr/ujma

UNIVERSAL JOURNAL OF MATHEMATICS AND APPLICATIONS

Contents

Universal Journal of Mathematics and Applications, 4 (1) (2021) 1-8
Research paper

UJMA

Universal Journal of Mathematics and Applications

Journal Homepage: www.dergipark.gov.tr/ujma
ISSN 2619-9653
DOI: https://doi.org/10.32323/ujma.833414

Additive Refinements and Reverses of Young's Operator Inequality Via a Result of Cartwright and Field

Silvestru Sever Dragomir[1,2*]

[1]*Mathematics, College of Engineering & Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia*
[2]*School of Computer Science & Applied Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg 2050, South Africa*

Article Info

Keywords: *Arithmetic mean-Geometric mean inequality, Hölder-McCarthy operator inequality, Young's inequality.*
2010 AMS: *47A63, 47A30, 15A60, 26D15; 26D10.*
Received: *30 November 2020*
Accepted: *1 March 2021*
Available online: *22 March 2021*

Abstract

In this paper we obtain some new additive refinements and reverses of Young's operator inequality via a result of Cartwright and Field. Comparison with other additive Young's type inequalities are also provided.

1. Introduction

We have the following inequality that provides a refinement and a reverse for the celebrated Young's inequality

$$\frac{1}{2}\nu(1-\nu)\frac{(b-a)^2}{\max\{a,b\}} \leq (1-\nu)a+\nu b-a^{1-\nu}b^{\nu} \leq \frac{1}{2}\nu(1-\nu)\frac{(b-a)^2}{\min\{a,b\}} \tag{1.1}$$

for any $a,b>0$ and $\nu\in[0,1]$.
This result was obtained in 1978 by Cartwright and Field [1] who established a more general result for n variables and gave an application for a probability measure supported on a finite interval.
Throughout this paper A, B are positive invertible operators on a complex Hilbert space $(H,\langle\cdot,\cdot\rangle)$. We use the following notations for operators

$$A\nabla_{\nu}B := (1-\nu)A+\nu B,$$

the *weighted operator arithmetic mean* and

$$A\sharp_{\nu}B := A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{\nu}A^{1/2},$$

the *weighted operator geometric mean.* When $\nu=\frac{1}{2}$ we write $A\nabla B$ and $A\sharp B$ for brevity, respectively.
The famous *Young inequality* for positive invertible operators A, B says that if $\nu\in[0,1]$, then

$$A\sharp_{\nu}B \leq A\nabla_{\nu}B. \tag{1.2}$$

The inequality (1.2) is also called ν*-weighted arithmetic-geometric operator mean inequality*.
In the recent paper [12], by the use of Cartwright and Field inequality (1.1), Minculete and Furuichi showed amongst other that

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)\left(AB^{-1}A-2A+B\right) &\leq A\nabla_{\nu}B-A\sharp_{\nu}B \\ &\leq \frac{1}{2}\nu(1-\nu)\left(BA^{-1}B-2B+A\right),\end{aligned} \tag{1.3}$$

Email address and ORCID number: sever.dragomir@vu.edu.au, 0000-0003-2902-6805, (S.S. Dragomir)

provided that $A \leq B$, and

$$\frac{1}{2}\nu(1-\nu)\left(BA^{-1}B-2B+A\right) \leq A\nabla_\nu B - A\sharp_\nu B \leq \frac{1}{2}\nu(1-\nu)\left(AB^{-1}A-2A+B\right), \tag{1.4}$$

provided that $B \leq A$.
For other inequalities between the operator means $A\sharp_\nu B$ and $A\nabla_\nu B$ see [2]- [11], [13]- [14] and the references therein.
In this paper, several other lower and upper bounds for the *Young's difference* $A\nabla_\nu B - A\sharp_\nu B$ under various boundedness assumptions for the involved operators A and B are given. Comparison with other additive Young's type inequalities are also provided.

2. A Refinement and Reverse of Young's Inequality

We have:

Theorem 2.1. *Let A, B be positive invertible operators and $M > m > 0$ such that*

$$MA \geq B \geq mA. \tag{2.1}$$

Then for any $\nu \in [0,1]$ we have

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)c(m,M)A &\leq \frac{1}{2}\frac{\nu(1-\nu)}{\max\{M,1\}}(B-A)A^{-1}(B-A),\\ &\leq A\nabla_\nu B - A\sharp_\nu B,\\ &\leq \frac{1}{2}\frac{\nu(1-\nu)}{\min\{m,1\}},(B-A)A^{-1}(B-A),\\ &\leq \frac{1}{2}\nu(1-\nu)C(m,M)A,\end{aligned} \tag{2.2}$$

where

$$c(m,M) := \begin{cases} (M-1)^2 \text{ if } M<1,\\ 0 \text{ if } m \leq 1 \leq M,\\ \frac{(m-1)^2}{M} \text{ if } 1<m \end{cases}$$

and

$$C(m,M) := \begin{cases} \frac{(m-1)^2}{m} \text{ if } M<1,\\ \frac{1}{m}\max\left\{(m-1)^2,(M-1)^2\right\} \text{ if } m \leq 1 \leq M,\\ (M-1)^2 \text{ if } 1<m. \end{cases}$$

In particular,

$$\begin{aligned}\frac{1}{8}c(m,M)A &\leq \frac{1}{8\max\{M,1\}}(B-A)A^{-1}(B-A) \leq A\nabla B - A\sharp B\\ &\leq \frac{1}{8\min\{m,1\}}(B-A)A^{-1}(B-A) \leq \frac{1}{8}C(m,M)A.\end{aligned} \tag{2.3}$$

Proof. If we write the inequality (1.1) for $a=1$ and $b=x$ we get

$$\frac{1}{2}\nu(1-\nu)\frac{(x-1)^2}{\max\{x,1\}} \leq 1-\nu+\nu x - x^\nu \leq \frac{1}{2}\nu(1-\nu)\frac{(x-1)^2}{\min\{x,1\}} \tag{2.4}$$

for any $x>0$ and for any $\nu \in [0,1]$.
If $x \in [m,M] \subset (0,\infty)$, then $\max\{x,1\} \leq \max\{M,1\}$ and $\min\{m,1\} \leq \min\{x,1\}$ and by (2.4) we get

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)\frac{\min_{x\in[m,M]}(x-1)^2}{\max\{M,1\}} &\leq \frac{1}{2}\nu(1-\nu)\frac{(x-1)^2}{\max\{M,1\}}\\ &\leq 1-\nu+\nu x - x^\nu\\ &\leq \frac{1}{2}\nu(1-\nu)\frac{(x-1)^2}{\min\{m,1\}}\\ &\leq \frac{1}{2}\nu(1-\nu)\frac{\max_{x\in[m,M]}(x-1)^2}{\min\{m,1\}}\end{aligned} \tag{2.5}$$

for any $x \in [m,M]$ and for any $\nu \in [0,1]$.

Observe that

$$\min_{x\in[m,M]}(x-1)^2=\begin{cases}(M-1)^2 \text{ if } M<1,\\ 0 \text{ if } m\le 1\le M,\\ (m-1)^2 \text{ if } 1<m\end{cases}$$

and

$$\max_{x\in[m,M]}(x-1)^2=\begin{cases}(m-1)^2 \text{ if } M<1,\\ \max\left\{(m-1)^2,(M-1)^2\right\} \text{ if } m\le 1\le M,\\ (M-1)^2 \text{ if } 1<m.\end{cases}$$

Then

$$\frac{\min_{x\in[m,M]}(x-1)^2}{\max\{M,1\}}=\begin{cases}(M-1)^2 \text{ if } M<1,\\ 0 \text{ if } m\le 1\le M,\\ \frac{(m-1)^2}{M} \text{ if } 1<m\end{cases}=c(m,M)$$

and

$$\frac{\max_{x\in[m,M]}(x-1)^2}{\min\{m,1\}}=\begin{cases}\frac{(m-1)^2}{m} \text{ if } M<1,\\ \frac{1}{m}\max\left\{(m-1)^2,(M-1)^2\right\} \text{ if } m\le 1\le M,\\ (M-1)^2 \text{ if } 1<m.\end{cases}=C(m,M)$$

Using the inequality (2.5) we have

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)c(m,M)&\le\frac{1}{2}\nu(1-\nu)\frac{(x-1)^2}{\max\{M,1\}}\\&\le 1-\nu+\nu x-x^{\nu}\\&\le\frac{1}{2}\nu(1-\nu)\frac{(x-1)^2}{\min\{m,1\}}\\&\le\frac{1}{2}\nu(1-\nu)C(m,M)\end{aligned}\tag{2.6}$$

for any $x\in[m,M]$ and for any $\nu\in[0,1]$.
If we use the continuous functional calculus for the positive invertible operator X with $mI\le X\le MI$, then we have from (2.10) that

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)c(m,M)I&\le\frac{1}{2}\frac{\nu(1-\nu)}{\max\{M,1\}}(X-I)^2\\&\le(1-\nu)I+\nu X-X^{\nu}\\&\le\frac{1}{2}\frac{\nu(1-\nu)}{\min\{m,1\}}(X-I)^2\\&\le\frac{1}{2}\nu(1-\nu)C(m,M)I\end{aligned}\tag{2.7}$$

for any $\nu\in[0,1]$.
If we multiply (2.1) both sides by $A^{-1/2}$ we get $MI\ge A^{-1/2}BA^{-1/2}\ge mI$.
By writing the inequality (2.7) for $X=A^{-1/2}BA^{-1/2}$ we obtain

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)c(m,M)I&\le\frac{1}{2}\frac{\nu(1-\nu)}{\max\{M,1\}}\left(A^{-1/2}BA^{-1/2}-I\right)^2\\&\le(1-\nu)I+\nu A^{-1/2}BA^{-1/2}-\left(A^{-1/2}BA^{-1/2}\right)^{\nu}\\&\le\frac{1}{2}\frac{\nu(1-\nu)}{\min\{m,1\}}\left(A^{-1/2}BA^{-1/2}-I\right)^2\\&\le\frac{1}{2}\nu(1-\nu)C(m,M)I\end{aligned}\tag{2.8}$$

for any $\nu\in[0,1]$.

If we multiply the inequality (2.8) both sides with $A^{1/2}$, then we get

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)c(m,M)A &\leq \frac{1}{2}\frac{\nu(1-\nu)}{\max\{M,1\}}A^{1/2}\left(A^{-1/2}BA^{-1/2}-I\right)^2 A^{1/2} \\ &\leq (1-\nu)A+\nu B - A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{\nu}A^{1/2} \\ &\leq \frac{1}{2}\frac{\nu(1-\nu)}{\min\{m,1\}}A^{1/2}\left(A^{-1/2}BA^{-1/2}-I\right)^2 A^{1/2} \\ &\leq \frac{1}{2}\nu(1-\nu)C(m,M)A,\end{aligned} \tag{2.9}$$

and since

$$\begin{aligned}A^{1/2}\left(A^{-1/2}BA^{-1/2}-I\right)^2 A^{1/2} &= A^{1/2}\left(A^{-1/2}(B-A)A^{-1/2}\right)^2 A^{1/2} \\ &= A^{1/2}A^{-1/2}(B-A)A^{-1/2}A^{-1/2}(B-A)A^{-1/2}A^{1/2} \\ &= (B-A)A^{-1}(B-A),\end{aligned}$$

then by (2.9) we get the desired result (2.2). □

When the operators A and B are bounded above and below by constants we have the following result as well:

Corollary 2.2. *Let A, B be two positive operators and m, m', M, M' be positive real numbers. Put $h := \frac{M}{m}$ and $h' := \frac{M'}{m'}$.*
(i) If $0 < mI \leq A \leq m'I < M'I \leq B \leq MI$, then

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)\frac{(h'-1)^2}{h}A &\leq \frac{1}{2}\frac{\nu(1-\nu)}{h}(B-A)A^{-1}(B-A) \\ &\leq A\nabla_\nu B - A\sharp_\nu B \\ &\leq \frac{1}{2}\nu(1-\nu)(B-A)A^{-1}(B-A) \\ &\leq \frac{1}{2}\nu(1-\nu)(h-1)^2 A,\end{aligned} \tag{2.10}$$

and, in particular,

$$\begin{aligned}\frac{(h'-1)^2}{8h}A &\leq \frac{1}{8h}(B-A)A^{-1}(B-A) \leq A\nabla B - A\sharp B \\ &\leq \frac{1}{8}(B-A)A^{-1}(B-A) \leq \frac{1}{8}(h-1)^2 A.\end{aligned} \tag{2.11}$$

(ii) If $0 < mI \leq B \leq m'I < M'I \leq A \leq MI$, then

$$\begin{aligned}\frac{1}{2}\nu(1-\nu)\left(\frac{h'-1}{h'}\right)^2 A &\leq \frac{1}{2}\nu(1-\nu)(B-A)A^{-1}(B-A) \\ &\leq A\nabla_\nu B - A\sharp_\nu B \\ &\leq \frac{1}{2}\nu(1-\nu)h(B-A)A^{-1}(B-A) \\ &\leq \frac{1}{2}\nu(1-\nu)\frac{(h-1)^2}{h}A\end{aligned} \tag{2.12}$$

and, in particular,

$$\begin{aligned}\frac{1}{8}\left(\frac{h'-1}{h'}\right)^2 A &\leq \frac{1}{8}(B-A)A^{-1}(B-A) \leq A\nabla B - A\sharp B \\ &\leq \frac{1}{8}h(B-A)A^{-1}(B-A) \leq \frac{(h-1)^2}{8h}A.\end{aligned} \tag{2.13}$$

Proof. We observe that $h, h' > 1$ and if either of the condition (i) or (ii) holds, then $h \geq h'$.
If (i) is valid, then we have

$$A < h'A = \frac{M'}{m'}A \leq B \leq \frac{M}{m}A = hA, \tag{2.14}$$

while, if (ii) is valid, then we have

$$\frac{1}{h}A \leq B \leq \frac{1}{h'}A < A. \tag{2.15}$$

If we use the inequality (2.2) and the assumption (i), then we get (2.10).
If we use the inequality (2.2) and the assumption (ii), then we get (2.12). □

3. Bounds in Term of Kantorovich's Constant

We consider the *Kantorovich's constant* defined by

$$K(h) := \frac{(h+1)^2}{4h},\ h>0. \tag{3.1}$$

The function K is decreasing on $(0,1)$ and increasing on $[1,\infty)$, $K(h)\geq 1$ for any $h>0$ and $K(h)=K\left(\frac{1}{h}\right)$ for any $h>0$.
Observe that for any $h>0$

$$K(h)-1=\frac{(h-1)^2}{4h}=K\left(\frac{1}{h}\right)-1.$$

Also, if $a,b>0$ then

$$K\left(\frac{b}{a}\right)-1=\frac{(b-a)^2}{4ab}.$$

Since $\min\{a,b\}\max\{a,b\}=ab$ if $a,b>0$, then

$$\frac{(b-a)^2}{\max\{a,b\}}=\frac{\min\{a,b\}(b-a)^2}{ab}=4\min\{a,b\}\left[K\left(\frac{b}{a}\right)-1\right]$$

and

$$\frac{(b-a)^2}{\min\{a,b\}}=\frac{\max\{a,b\}(b-a)^2}{ab}=4\max\{a,b\}\left[K\left(\frac{b}{a}\right)-1\right]$$

and the inequality (1.1) can be written as

$$\begin{aligned}2\nu(1-\nu)\min\{a,b\}\left[K\left(\frac{b}{a}\right)-1\right]&\leq(1-\nu)a+\nu b-a^{1-\nu}b^{\nu}\\&\leq 2\nu(1-\nu)\max\{a,b\}\left[K\left(\frac{b}{a}\right)-1\right]\end{aligned} \tag{3.2}$$

for any $a,b>0$ and $\nu\in[0,1]$.
For positive invertible operators A,B we define

$$A\nabla_{\infty}B:=\frac{1}{2}(A+B)+\frac{1}{2}A^{1/2}\left|A^{-1/2}(B-A)A^{-1/2}\right|A^{1/2}$$

and

$$A\nabla_{-\infty}B:=\frac{1}{2}(A+B)-\frac{1}{2}A^{1/2}\left|A^{-1/2}(B-A)A^{-1/2}\right|A^{1/2}.$$

If we consider the continuous functions $f_{\infty},f_{-\infty}:[0,\infty)\to[0,\infty)$ defined by

$$f_{\infty}(x)=\max\{x,1\}=\frac{1}{2}(x+1)+\frac{1}{2}|x-1|$$

and

$$f_{-\infty}(x)=\max\{x,1\}=\frac{1}{2}(x+1)-\frac{1}{2}|x-1|,$$

then, obviously, we have

$$A\nabla_{\pm\infty}B=A^{1/2}f_{\pm\infty}\left(A^{-1/2}BA^{-1}\right)A^{1/2}. \tag{3.3}$$

If A and B are commutative, then

$$A\nabla_{\pm\infty}B=\frac{1}{2}(A+B)\pm\frac{1}{2}|B-A|=B\nabla_{\pm\infty}A.$$

Theorem 3.1. *Let A, B be positive invertible operators and $M>m>0$ such that the condition (2.1) holds. Then we have*

$$\begin{aligned}2\nu(1-\nu)g(m,M)A\nabla_{-\infty}B&\leq A\nabla_{\nu}B-A\sharp_{\nu}B\\&\leq 2\nu(1-\nu)G(m,M)A\nabla_{\infty}B,\end{aligned} \tag{3.4}$$

where

$$g(m,M):=\begin{cases}K(M)-1 \text{ if } M<1,\\ 0 \text{ if } m\leq 1\leq M,\\ K(m)-1 \text{ if } 1<m\end{cases}$$

and

$$G(m,M) := \begin{cases} K(m)-1 \text{ if } M<1, \\ \max\{K(m),K(M)\}-1 \text{ if } m\le 1\le M, \\ K(M)-1 \text{ if } 1<m. \end{cases}$$

In particular,

$$\frac{1}{2}g(m,M)A\nabla_{-\infty}B \le A\nabla B - A\sharp B \le \frac{1}{2}G(m,M)A\nabla_{\infty}B. \tag{3.5}$$

Proof. From (3.2) we have for $a=1$ and $b=x$ that

$$\begin{aligned} 2\nu(1-\nu)\min\{1,x\}[K(x)-1] &\le 1-\nu+\nu x - x^{\nu} \\ &\le 2\nu(1-\nu)\max\{1,x\}[K(x)-1] \end{aligned} \tag{3.6}$$

for any $x>0$.
From (3.6) we then have

$$\begin{aligned} 2\nu(1-\nu)f_{-\infty}(x)\min_{x\in[m,M]}[K(x)-1] &\le 1-\nu+\nu x - x^{\nu} \\ &\le 2\nu(1-\nu)f_{\infty}(x)\max_{x\in[m,M]}[K(x)-1] \end{aligned} \tag{3.7}$$

for any $x\in[m,M]$.
Observe that

$$\max_{x\in[m,M]}[K(x)-1] = \begin{cases} K(m)-1 \text{ if } M<1, \\ \max\{K(m),K(M)\}-1 \text{ if } m\le 1\le M, \\ K(M)-1 \text{ if } 1<m. \end{cases} = G(m,M)$$

and

$$\min_{x\in[m,M]}[K(x)-1] = \begin{cases} K(M)-1 \text{ if } M<1, \\ 0 \text{ if } m\le 1\le M, \\ K(m)-1 \text{ if } 1<m. \end{cases} = g(m,M).$$

Therefore by (3.7) we get

$$2\nu(1-\nu)f_{-\infty}(x)g(m,M) \le 1-\nu+\nu x - x^{\nu} \le 2\nu(1-\nu)f_{\infty}(x)G(m,M) \tag{3.8}$$

for any $x\in[m,M]$ and $\nu\in[0,1]$.
If we use the continuous functional calculus for the positive invertible operator X with $mI\le X\le MI$, then we have from (3.8) that

$$\begin{aligned} 2\nu(1-\nu)f_{-\infty}(X)g(m,M) &\le (1-\nu)I+\nu X - X^{\nu} \\ &\le 2\nu(1-\nu)f_{\infty}(X)G(m,M) \end{aligned} \tag{3.9}$$

for any $x\in[m,M]$ and $\nu\in[0,1]$.
By writing the inequality (2.7) for $X=A^{-1/2}BA^{-1/2}$ we obtain

$$\begin{aligned} 2\nu(1-\nu)f_{-\infty}\left(A^{-1/2}BA^{-1/2}\right)g(m,M) &\le (1-\nu)I+\nu A^{-1/2}BA^{-1/2} - \left(A^{-1/2}BA^{-1/2}\right)^{\nu} \\ &\le 2\nu(1-\nu)f_{\infty}\left(A^{-1/2}BA^{-1/2}\right)G(m,M) \end{aligned} \tag{3.10}$$

for any $\nu\in[0,1]$.
If we multiply (3.10) both sides by $A^{1/2}$ we get

$$\begin{aligned} 2\nu(1-\nu)A^{1/2}f_{-\infty}\left(A^{-1/2}BA^{-1/2}\right)A^{1/2}g(m,M) &\le (1-\nu)A+\nu BA - A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{\nu}A^{1/2} \\ &\le 2\nu(1-\nu)A^{1/2}f_{\infty}\left(A^{-1/2}BA^{-1/2}\right)A^{1/2}G(m,M) \end{aligned}$$

for any $\nu\in[0,1]$, which, by (3.3) produces the desired result (3.4). □

We have:

Corollary 3.2. *Let A, B be two positive operators and m, m', M, M' be positive real numbers. Put $h := \frac{M}{m}$ and $h' := \frac{M'}{m'}$. If either of the conditions (i) or (ii) from Corollary 2.2 holds, then*

$$2\nu(1-\nu)\left[K\left(h'\right)-1\right]A\nabla_{-\infty}B \le A\nabla_{\nu}B - A\sharp_{\nu}B \le 2\nu(1-\nu)\left[K(h)-1\right]A\nabla_{\infty}B. \tag{3.11}$$

In particular,

$$\frac{1}{2}\left[K\left(h'\right)-1\right]A\nabla_{-\infty}B \le A\nabla B - A\sharp B \le \frac{1}{2}\left[K(h)-1\right]A\nabla_{\infty}B. \tag{3.12}$$

Proof. If (i) is valid, then we have

$$A < h'A = \frac{M'}{m'}A \le B \le \frac{M}{m}A = hA.$$

By using the inequality (3.4) we get (3.11).
If (ii) is valid, then we have

$$\frac{1}{h}A \le B \le \frac{1}{h'}A < A.$$

By using the inequality (3.4) we get

$$2\nu(1-\nu)\left[K\left(\frac{1}{h'}\right)-1\right]A\nabla_{-\infty}B \le A\nabla_{\nu}B - A\sharp_{\nu}B \le 2\nu(1-\nu)\left[K\left(\frac{1}{h}\right)-1\right]A\nabla_{\infty}B,$$

and since $K\left(\frac{1}{h'}\right) = K(h')$ and $K\left(\frac{1}{h}\right) = K(h)$, the inequality (3.11) is also obtained. □

4. Comparison with Other Additive Inequalities

Kittaneh and Manasrah [9], [10] provided a refinement and a reverse for Young's scalar inequality as follows:

$$r\left(\sqrt{a}-\sqrt{b}\right)^2 \le (1-\nu)a + \nu b - a^{1-\nu}b^{\nu} \le R\left(\sqrt{a}-\sqrt{b}\right)^2, \tag{4.1}$$

where $a, b > 0$, $\nu \in [0,1]$, $r = \min\{1-\nu, \nu\}$ and $R = \max\{1-\nu, \nu\}$. The case $\nu = \frac{1}{2}$ reduces (4.1) to an identity and is of no interest.
In [2] we obtained the following logarithmic upper bound for the Young's difference

$$(0 \le)\,(1-\nu)a + \nu b - a^{1-\nu}b^{\nu} \le \nu(1-\nu)(a-b)(\ln a - \ln b) \tag{4.2}$$

for any $a, b > 0$ and $\nu \in [0,1]$, while in the subsequent paper [3] we obtained the following refinement and reverse of Young's inequality

$$\frac{1}{2}\nu(1-\nu)(\ln a - \ln b)^2 \min\{a,b\} \le (1-\nu)a + \nu b - a^{1-\nu}b^{\nu} \le \frac{1}{2}\nu(1-\nu)(\ln a - \ln b)^2 \max\{a,b\}, \tag{4.3}$$

for any $a, b > 0$ and $\nu \in [0,1]$.
Consider the following functions of two variables obtained from the upper bounds in inequalities (1.1), (4.1), (4.2) and (4.3) for $a = 1$, $b = x \in (0,\infty)$ and $\nu = y \in (0,1)$, namely

$$U_1(x,y) := \frac{1}{2}y(1-y)\frac{(x-1)^2}{\min\{x,1\}},$$

$$U_2(x,y) := \max\{y, 1-y\}\left(\sqrt{x}-1\right)^2,$$

$$U_3(x,y) := y(1-y)(x-1)\ln x \text{ and}$$

$$U_4(x,y) := \frac{1}{2}y(1-y)\max\{x,1\}\ln^2 x.$$

We observe that the 3D plots of the differences $U_1(x,y) - U_2(x,y)$ on $(0,10)\times(0,1)$, $U_1(x,y) - U_3(x,y)$ on $(2,4)\times(0,1)$, $U_2(x,y) - U_3(x,y)$ on $(2,4)\times(0,1)$, $U_2(x,y) - U_4(x,y)$ on $(2,4)\times(0,1)$ and $U_3(x,y) - U_4(x,y)$ on $(3,6)\times(0,1)$ show that they take both negative and positive values, meaning that neither of the corresponding upper bounds are better in general.
It appears that $U_1(x,y) > U_4(x,y)$ on the box $(0,10)\times(0,1)$ suggesting that the upper bound in (4.3) is better than the one from (1.1). However we do not have an analytic proof for it in general.
Similar conclusions may be derived for the lower bounds, however the details are left to the interested reader.

5. Conclusion

In this paper we obtained some new additive refinements and reverses of Young's operator inequality via a result of Cartwright and Field. Comparison with other additive Young's type inequalities were also provided.

References

[1] D.I. Cartwright, M.J. Field, *A refinement of the arithmetic mean-geometric mean inequality*, Proc. Amer. Math. Soc., **71** (1978), 36-38.
[2] S.S. Dragomir, *Some new reverses of Young's operator inequality*, Preprint RGMIA Res. Rep. Coll. **18** (2015), Art. 130. [http://rgmia.org/papers/v18/v18a130.pdf].
[3] S.S. Dragomir, *On new refinements and reverses of Young's operator inequality*, Preprint RGMIA Res. Rep. Coll. **18** (2015), Art. 135. [http://rgmia.org/papers/v18/v18a135.pdf].
[4] S.S. Dragomir, *Some inequalities for operator weighted geometric mean*, Preprint RGMIA Res. Rep. Coll. **18** (2015), Art. 139. [http://rgmia.org/papers/v18/v18a139.pdf].
[5] S.S. Dragomir, *Refinements and reverses of Hölder-McCarthy operator inequality*, Preprint RGMIA Res. Rep. Coll. **18** (2015), Art. 143. [http://rgmia.org/papers/v18/v18a143.pdf].
[6] S.S. Dragomir, *Some reverses and a refinement of Hölder operator inequality*, Preprint *RGMIA Res. Rep. Coll.* **18** (2015), Art. 147. [http://rgmia.org/papers/v18/v18a147.pdf].
[7] S. Furuichi, *Refined Young inequalities with Specht's ratio*, J. Egyptian Math. Soc. **20**(2012), 46-49.
[8] S. Furuichi, *On refined Young inequalities and reverse inequalities*, J. Math. Inequal. **5** (2011), 21-31.
[9] F. Kittaneh, Y. Manasrah, *Improved Young and Heinz inequalities for matrix*, J. Math. Anal. Appl., **361** (2010), 262-269
[10] F. Kittaneh, Y. Manasrah, *Reverse Young and Heinz inequalities for matrices*, Linear Multilinear Algebra., **59** (2011), 1031-1037.
[11] W. Liao, J. Wu, J. Zhao, *New versions of reverse Young and Heinz mean inequalities with the Kantorovich constant*, Taiwanese J. Math. **19**(2) (2015), 467-479.
[12] N. Minculete, S. Furuichi, *Several applications of Cartwright-Field's inequality*, Int. J. Pure Appl. Math., **71**(1) (21), 19-30.
[13] M. Tominaga, *Specht's ratio in the Young inequality*, *Sci. Math. Japon.*, **55** (2002), 583-588.
[14] G. Zuo, G. Shi, M. Fujii, *Refined Young inequality with Kantorovich constant*, *J. Math. Inequal.,* **5** (2011), 551-556.

Universal Journal of Mathematics and Applications, 4 (1) (2021) 9-25
Research paper

UJMA

Universal Journal of Mathematics and Applications

Journal Homepage: www.dergipark.gov.tr/ujma
ISSN 2619-9653
DOI: https://doi.org/10.32323/ujma.751868

Generalized Rayleigh-Quotient Formulas for the Real Parts, Imaginary Parts, and Moduli of the Eigenvalues of General Matrices

Ludwig Kohaupt

Department of Mathematics, Beuth University of Technology Berlin, Berlin, Germany

Article Info

Keywords: *Asymptotic stability of dynamical systems, Circular damped eigenfrequencies, Moduli of eigenvalues, Rayleigh quotient, Real and imaginary parts of eigenvalues, Weighted norm*

2010 AMS: *11E39, 15A18, 15B57, 15B99, 65F35, 65J05*

Received: *12 June 2020*

Accepted: *13 January 2021*

Available online: *22 March 2021*

Abstract

In the present paper, generalized Rayleigh-quotient formulas for the real parts, imaginary parts, and moduli of the eigenvalues of general (not necessarily diagonalizable) matrices are derived by using quotients of the form $(Au,v)/(u,v)$ instead of $(Au,u)/(u,u)$. These formulas are new and correspond to similar formulas for diagonalizable matrices obtained recently. Numerical examples underpin the theoretical findings. We point out that, in the case of general matrices, the principal vectors of largest stage of matrix A^* take over the role of the eigenvectors in the case of diagonalizable matrices. So, even though the formulas in both cases look very similar, the result is somehow unexpected and surprising.

1. Introduction

For self-adjoint matrices, there are formulas for the eigenvalues in the form of generalized Rayleigh quotients; more precisely, max-, min-, min-max-, and max-min-formulas have been derived by the author in [8].

Recently, corresponding formulas could be carried over to formulas for the real parts, imaginary parts, and moduli of diagonalizable matrices in [10].

The aim of the present paper is to extend these results to general matrices.

We mention also that the presentation of this paper parallels that of [8] and [10]. So, similarities in the formulation do not happen by accident, but are intended in order to underline the similarities. As a consequence, many verbatim passages in the formulations are taken from there.

As has already been said in [9], first, the obtained formulas are of interest on their own in Linear Algebra. Second, these are also of potential interest, for example, in the theory of linear dynamical systems. The reason for this is as follows. The real parts of the eigenvalues multiplied by the time are equal to the arguments of the exponential functions that describe the decay behavior of the solution (see, e.g., [7, Section 7.1, p.2011, Formulas (89), (90)]). Further, the system is asymptotically stable if the real parts of all eigenvalues are negative. Moreover, when the eigenvalues are pairwise conjugate-complex, then the moduli of the imaginary parts are the circular damped eigenfrequencies of the system (see, e.g., [7, Section 7.1, p. 2011, (89)]).

The paper is structured as follows.

In Sections 2 - 4, the new generalized Rayleigh-quotient formulas for the real parts, imaginary parts, and moduli for general matrices are stated, as the case may be. In Section 5, the special case of general matrices with real eigenvalues is treated. Section 6 contains an application and Section 7 the definitions of new generalized numerical ranges. In Sections 8 and 9, numerical examples are presented that underpin the obtained findings. In the first example, matrix A is taken as the non-diagonalizable system matrix of a linear dynamical problem. In the second example, we choose a non-diagonalizable matrix with real eigenvalues. Finally, Section 10 contains the conclusion. The References are restricted to those that are cited in this paper augmented by those used in [8] and [10], the latter being [2], [3], [12], [13], and [14].

***Email address and ORCID number*: kohaupt@bht-berlin.de, 0000-0003-1781-0600, (L. Kohaupt)**

2. Generalized Rayleigh-Quotient Formulas for the Real Parts of the Eigenvalues of a General Matrix

In this section, we want to derive formulas for the representation of the real parts of the eigenvalues of a general matrix $A \in \mathbb{C}^{n\times n}$ by Rayleigh quotients that generalize existing ones. More precisely, max-, min-, min-max-, and max-min-representations are obtained in the form of more general Rayleigh quotients corresponding to associated formulas for the eigenvalues of diagonalizable matrices assembled in [10]. The difference to the results obtained in [9] is that here we use the scalar product $(\cdot,\cdot)$ in $\mathbb{C}^n$ instead of a weighted scalar product $(\cdot,\cdot)_R$.

First, we formulate the following conditions $(C1')$ - $(C4')$:

$(C1')$ $A \in \mathbb{C}^{n\times n}$
$(C2')$ $\lambda_i,\ i=1,\cdots,r$ are the eigenvalues of A corresponding to the Jordan blocks $J_i(\lambda_i) \in \mathbb{C}^{m_i\times m_i},\ i=1,\cdots,r$ with the chains of principal vectors $p_1^{(i)},\cdots,p_{m_i}^{(i)},\ i=1,\cdots,r$
$(C3')$ ${u_1^{(i)}}^*,\cdots,{u_{m_i}^{(i)}}^*,\ i=1,\cdots,r$ are the principal vectors of A^* corresponding to the eigenvalues $\overline{\lambda_i},\ i=1,\cdots,r$ of the Jordan blocks $J_i(\overline{\lambda_i}) \in \mathbb{C}^{m_i\times m_i},\ i=1,\cdots,r$
$(C4')$ $\lambda_i \neq \lambda_j,\ i\neq j,\ i,j=1,\cdots,r$

We mention that, even though condition $(C4')$ may be omitted (see [6, Theorem 4]), it is nevertheless useful here since it will turn out to be fulfilled in the numerical examples in Sections 8 and 9 and since the biorthogonal system in Theorem 2.1 can be constructed more easily than without this condition. One has the following theorem.

Theorem 2.1. (Biorthogonality relations for principal vectors)
Let the conditions $(C1')$-$(C4')$ be fulfilled. Then, the systems $\{p_1^{(1)},\cdots,p_{m_1}^{(1)};\cdots;p_1^{(r)},\cdots,p_{m_r}^{(r)}\}$ and $\{{u_1^{(1)}}^*,\cdots,{u_{m_1}^{(1)}}^*;\cdots;{u_1^{(r)}}^*,\cdots,{u_{m_r}^{(r)}}^*\}$ can be constructed such that the following biorthogonality relations hold:

$$(p_k^{(i)},{u_l^{(i)}}^*) = \begin{cases} 1, & l=m_i-k+1 \\ 0, & l\neq m_i-k+1 \end{cases} \tag{2.1}$$

$k=1,\cdots,m_i,\ i=1,\cdots,r$ *and*

$$(p_k^{(i)},{u_l^{(j)}}^*) = 0,\ i\neq j, \tag{2.2}$$

$k=1,\cdots,m_i,\ l=1,\cdots,m_j,\ i,j=1,\cdots,r.$
So, with

$$ {v_l^{(i)}}^* := {u_{m_i-l+1}^{(i)}}^*, \tag{2.3}$$

$l=1,\cdots,m_i,\ i=1,\cdots,r$ *one has the biorthogonality relations*

$$(p_k^{(i)},{v_l^{(i)}}^*) = \delta_{kl}, \tag{2.4}$$

$k,l=1,\cdots,m_i,\ i=1,\cdots,r,$ *and*

$$(p_k^{(i)},{v_l^{(j)}}^*) = 0,\ i\neq j, \tag{2.5}$$

$k=1,\cdots,m_i,\ l=1,\cdots,m_j,\ i,j=1,\cdots,r.$

Proof. See proof of [5, Theorem 2] or [6, Theorem 4]. □

Next, we want to derive a relation corresponding to that of [10, Formula (12)]. This is done in the following Formula (2.19).
First, with the identity matrix E, we introduce the abbreviation

$$N_{\lambda_j(A)} := \{u\in\mathbb{C}^n \,|\, (A-\lambda_j(A)E)u=0\},\ j=1,\ldots,r \tag{2.6}$$

for the geometric eigenspaces so that

$$N_{\lambda_j(A)} := [p_1^{(j)}] = [p_j],\ j=1,\ldots,r.$$

Herewith, we define

$$N_{\sigma(A)} := \bigoplus_{j=1}^{r} N_{\lambda_j(A)}. \tag{2.7}$$

Further, we define the following subspaces of $\mathbb{C}^n$. For every $k=1,\cdots,r$, let

$$N_{p,k} := \left\{u\in\mathbb{C}^n \,|\, u=\sum_{j=1}^{k}\alpha_j p_j \text{ with } \alpha_j\in\mathbb{C},\ j=1,\cdots,k\right\} =: [p_1,\cdots,p_k] \tag{2.8}$$

and

$$N_{p,k,\mathbb{R}} := \left\{u\in\mathbb{C}^n \,|\, u=\sum_{j=1}^{k}\beta_j p_j \text{ with } \beta_j\in\mathbb{R},\ j=1,\cdots,k\right\} =: [p_1,\cdots,p_k]_{\mathbb{R}} \tag{2.9}$$

as well as

$$N_p := N_{p,r} := \left\{ u \in \mathbb{C}^n \,|\, u = \sum_{j=1}^{r} \alpha_j p_j \; with \; \alpha_j \in \mathbb{C}, \; j = 1, \cdots, r \right\} =: [p_1, \cdots, p_r] \tag{2.10}$$

and

$$N_{p,\mathbb{R}} := N_{p,r,\mathbb{R}} := \left\{ u \in \mathbb{C}^n \,|\, u = \sum_{j=1}^{r} \beta_j p_j \; with \; \beta_j \in \mathbb{R}, \; j = 1, \cdots, r \right\} =: [p_1, \cdots, p_r]_{\mathbb{R}} \tag{2.11}$$

where $N_{p,\mathbb{R}}$ is apparently isomorphic to $\mathbb{R}^r$ and N_p is isomorphic to $\mathbb{C}^r$.
We mention that all these spaces (2.8) - (2.11) are subspaces of the geometric eigenspace $N_{\sigma(A)}$.
In [10], we have defined the further spaces $N_{u^*,k}$, $N_{u^*,k,\mathbb{R}}$, N_{u^*}, and $N_{u^*,\mathbb{R}}$ which are subspaces of the geometric eigenspace $N_{\sigma(A^*)}$. Here, however, we need different spaces. For this, we begin with the abbreviations

$$v_j^* := v_1^{(j)^*} = u_{m_j-1+1}^{(j)}{}^* = u_{m_j}^{(j)^*}, \tag{2.12}$$

$j = 1, \cdots, r$ that are principal vectors of stage m_j pertinent to the eigenvalue $\lambda_j(A^*) = \overline{\lambda_j(A)}, \; j = 1, \cdots, r$. Herewith, for every $k = 1, \cdots, r$, we define

$$N_{v^*,k} := \left\{ u \in \mathbb{C}^n \,|\, u = \sum_{j=1}^{k} \alpha_j v_j^* \; with \; \alpha_j \in \mathbb{C}, \; j = 1, \cdots, k \right\} =: [v_1^*, \cdots, v_k^*] \tag{2.13}$$

and

$$N_{v^*,k,\mathbb{R}} := \left\{ u \in \mathbb{C}^n \,|\, u = \sum_{j=1}^{k} \beta_j v_j^* \; with \; \beta_j \in \mathbb{R}, \; j = 1, \cdots, k \right\} =: [v_1^*, \cdots, v_k^*]_{\mathbb{R}} \tag{2.14}$$

as well as

$$N_{v^*} := N_{v^*,r} := \left\{ u \in \mathbb{C}^n \,|\, u = \sum_{j=1}^{r} \alpha_j v_j^* \; with \; \alpha_j \in \mathbb{C}, \; j = 1, \cdots, n \right\} =: [v_1^*, \cdots, v_r^*] \tag{2.15}$$

and

$$N_{v^*,\mathbb{R}} := N_{v^*,r,\mathbb{R}} := \left\{ u \in \mathbb{C}^n \,|\, u = \sum_{j=1}^{r} \beta_j v_j^* \; with \; \beta_j \in \mathbb{R}, \; j = 1, \cdots, r \right\} =: [v_1^*, \cdots, v_r^*]_{\mathbb{R}} \tag{2.16}$$

where $N_{v^*,\mathbb{R}}$ is apparently isomorphic to $\mathbb{R}^r$ and N_{v^*} is isomorphic to $\mathbb{C}^r$.
After these preparations, we are able to prove the following lemma.

Lemma 2.2. *Let the conditions (C1′)-(C4′) be fulfilled.*
Then, with the denotations of Theorem 2.1 and (2.12),

$$(A\,u, v) = \sum_{j=1}^{r} \lambda_j(A) \sum_{k=1}^{m_j} (u, v_k^{(j)^*})(p_k^{(j)}, v) + \sum_{j=1}^{r} \sum_{k=2}^{m_j} (u, v_k^{(j)^*})(p_{k-1}^{(j)}, v), \; u, v \in \mathbb{C}^n \tag{2.17}$$

leading to

$$(A\,u, v) = \sum_{j=1}^{r} \lambda_j(A)\,(u, v_j^*)(p_j, v), \; u \in N_{\sigma(A)}, \; v \in \mathbb{C}^n \tag{2.18}$$

and thus to

$$Re(A\,u, v) = \sum_{j=1}^{r} Re\lambda_j(A)\,(u, v_j^*)(p_j, v), \; u \in N_{p,\mathbb{R}}, \; v \in N_{v^*,\mathbb{R}}. \tag{2.19}$$

Proof. First, we prove (2.17). For this, let $u \in \mathbb{C}^n$. Then, with the denotations of Theorem 2.1,

$$u = \sum_{j=1}^{r} \sum_{k=1}^{m_j} (u, v_k^{(j)^*})\, p_k^{(j)} \tag{2.20}$$

leading to

$$\begin{aligned} Au &= \sum_{j=1}^{r} \sum_{k=1}^{m_j} (u, v_k^{(j)^*})\, A p_k^{(j)} \\ &= \sum_{j=1}^{r} \sum_{k=1}^{m_j} (u, v_k^{(j)^*})\, [\lambda_j p_k^{(j)} + p_{k-1}^{(j)}] \\ &= \sum_{j=1}^{r} \lambda_j \sum_{k=1}^{m_j} (u, v_k^{(j)^*})\, p_k^{(j)} + \sum_{j=1}^{r} \sum_{k=2}^{m_j} (u, v_k^{(j)^*})\, p_{k-1}^{(j)} \end{aligned} \tag{2.21}$$

since $p_0^{(j)} = 0,\ j = 1,\ldots,r$.
Further, for every $v \in \mathbb{C}^n$,

$$v = \sum_{l=1}^{r}\sum_{s=1}^{m_l}(v, p_s^{(l)})\, v_s^{(l)^*}. \tag{2.22}$$

This leads to

$$\begin{aligned}(Au,v) &= \sum_{j=1}^{r}\lambda_j\sum_{k=1}^{m_j}(u,v_k^{(j)^*})\,(p_k^{(j)},v) + \sum_{j=1}^{r}\sum_{k=2}^{m_j}(u,v_k^{(j)^*})\,(p_{k-1}^{(j)},v)\\ &= \sum_{j=1}^{r}\lambda_j\sum_{k=1}^{m_j}(u,v_k^{(j)^*})\sum_{l=1}^{r}\sum_{s=1}^{m_l}\overline{(v,p_s^{(l)})}\,\underbrace{(p_k^{(j)},v_s^{(l)^*})}_{\delta_{lj}\delta_{sk}} + \sum_{j=1}^{r}\sum_{k=2}^{m_j}(u,v_k^{(j)^*})\sum_{l=1}^{r}\sum_{s=1}^{m_l}\overline{(v,p_s^{(l)})}\,\underbrace{(p_{k-1}^{(j)},v_s^{(l)^*})}_{\delta_{s,k-1}\delta_{lj}}\end{aligned}$$

implying

$$(Au,v) = \sum_{j=1}^{r}\lambda_j\sum_{k=1}^{m_j}(u,v_k^{(j)^*})\,(p_k^{(j)},v) + \sum_{j=1}^{r}\sum_{k=2}^{m_j}(u,v_k^{(j)^*})\,(p_{k-1}^{(j)},v)$$

so that (2.17) follows.
Now, for $u \in N_{\sigma(A)}$, we have

$$(u, v_k^{(j)^*}) = 0,\ k = 2,\cdots,m_j$$

since $(p_s, v_k^{(j)^*}) = (p_1^{(s)}, v_k^{(j)^*}) = 0$ if $k \neq 1, ; s, j = 1,\cdots,r$. Thus, from (17), we deduce that

$$(Au,v) = \sum_{j=1}^{r}\lambda_j(u,v_1^{(j)^*})\,(p_1^{(j)},v). \tag{2.23}$$

□

With the abbreviation (2.12), we obtain (2.18).
Relation (2.19) is a direct consequence of equation (2.18) since $(u,v_j^*) \in \mathbb{R}$, $u \in N_{p,\mathbb{R}}$ and $(p_j,v) \in \mathbb{R}$, $v \in N_{v^*,\mathbb{R}}$.
Next, as in [10], we define the vector spaces $M_{p,k,\mathbb{R}}$, namely:

$$\begin{aligned}M_{p,1,\mathbb{R}} &:= N_{p,\mathbb{R}} = [p_1,\cdots,p_r]_{\mathbb{R}},\\ M_{p,k,\mathbb{R}} &:= \{u \in N_{p,\mathbb{R}} \,|\, (u,u_j^*) = 0,\ j = 1,2,\cdots,k-1\},\ k = 2,\cdots,r.\end{aligned} \tag{2.24}$$

Instead of the spaces $M_{u^*,k,\mathbb{R}}$ in [10], we need the spaces $M_{v^*,k,\mathbb{R}}$, i.e.,

$$\begin{aligned}M_{v^*,1,\mathbb{R}} &:= N_{v^*,\mathbb{R}} = [v_1^*,\cdots,v_r^*]_{\mathbb{R}},\\ M_{v^*,k,\mathbb{R}} &:= \{u \in N_{v^*,\mathbb{R}} \,|\, (u,p_j) = 0,\ j = 1,2,\cdots,k-1\},\ k = 2,\cdots,r.\end{aligned} \tag{2.25}$$

The next lemma characterizes these spaces.

Lemma 2.3. *Let the conditions (C1′)-(C4′) be fulfilled as well as $\{p_1,\cdots,p_r\}$ be the eigenvectors of A and $\{v_1^*,\cdots,v_r^*\}$ be principal vectors of A^* defined by (2.12) with the property*

$$(p_i, v_j^*) = \delta_{ij},\ i,j = 1,\cdots,r.$$

Then,

$$M_{p,k,\mathbb{R}} = [p_k, p_{k+1},\cdots,p_n]_{\mathbb{R}},\ k = 1,\cdots,r \tag{2.26}$$

and

$$M_{v^*,k,\mathbb{R}} = [v_k^*, v_{k+1}^*,\cdots,v_r^*]_{\mathbb{R}},\ k = 1,\cdots,r. \tag{2.27}$$

Proof. The proof is similar to that of [10, Lemma 3]. □

Similarly to [10, (21)], we suppose that the eigenvalues $\lambda_1(A),\cdots,\lambda_r(A)$ of matrix A are arranged such that

$$Re\lambda_1(A) \geq Re\lambda_2(A) \geq \cdots \geq Re\lambda_r(A). \tag{2.28}$$

Further, let $u \in N_{p,\mathbb{R}}$ with $u = \sum_{k=1}^{r}\alpha_k p_k$ and $v \in N_{v^*,\mathbb{R}}$ with $v = \sum_{k=1}^{r}\beta_k v_k^*$. Then, due to Theorem 2.1, as in [10],

$$(u,v) = \sum_{k=1}^{r}\alpha_k\beta_k. \tag{2.29}$$

In order to facilitate the manner of speaking, we say that *the scalar product (u,v) of u and v is strongly positive* if $\alpha_k\beta_k \geq 0$, $k = 1,\cdots,r$ and $\sum_{k=1}^{r}\alpha_k\beta_k > 0$. For short, we write $(u,v) \gg 0$.

Remark 2.4. *One has* $\alpha_k = (u, v_k^*)$, $u \in N_{p,\mathbb{R}}$ *and* $\beta_k = (p_k, v)$, $v \in N_{v^*,\mathbb{R}}$ *for* $k = 1, \cdots, r$. *Therefore,* $(u, v) \gg 0$ *means* $(u, v_k^*)(p_k, v) \geq 0$, $k = 1, \cdots, r$ *and* $(u, v) = \sum_{k=1}^{r} (u, v_k^*)(p_k, v) > 0$.

Remark 2.5. *More generally, in the sequel, one could admit linear combinations* $u = \sum_{k=1}^{r} \alpha_k p_k$ *and* $v = \sum_{k=1}^{r} \beta_k v_k^*$ *with* $\alpha_k, \beta_k \in \mathbb{C}$ *such that* $\alpha_k \overline{\beta_k} = |\alpha_k \beta_k|$ *and* $\sum_{k=1}^{r} |\alpha_k \beta_k| > 0$. *For example, all elements* $\alpha_k, \beta_k \in \mathbb{C}$ *with* $\alpha_k = |\alpha_k| e^{i\varphi_k}$ *and* $\beta_k = |\beta_k| e^{i\varphi_k}$ *where* φ_k *is in* $0 \leq \varphi_k < 2\pi$ *for* $k = 1, \cdots, r$ *would be acceptable. But, we do not want to pursue this aspect in more detail.*

Comparing relation (2.19) with [10, (12)], it is clear that one can obtain similar generalized max-, min-, min-max-, and max-min-representations for the real parts, imaginary parts, and moduli as in the case of diagonalizable matrices. Therefore, we state them without proofs.

Theorem 2.6. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (2.28). Moreover, let the vector spaces* $M_{p,k,\mathbb{R}}$ *and* $M_{v^*,k,\mathbb{R}}$ *for* $k = 1, \cdots, n$ *be defined by (2.24), (2.25) or (2.26),(2.27).*
Then,

$$Re\lambda_k(A) = \max_{\substack{(u,v)\gg 0 \\ u\in M_{p,k,\mathbb{R}}, v\in M_{v^*,k,\mathbb{R}}}} \frac{Re(Au, v)}{(u, v)}, \; k = 1, 2, \cdots, r. \tag{2.30}$$

The maximum is attained for $u = p_k$, $v = v_k^*$.

Theorem 2.7. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (2.28).*
Then, for every $j = 1, \cdots, r$ *and every subspace* $M_p \subset N_{p,\mathbb{R}}$ *and* $M_{v^*} \subset N_{v^*,\mathbb{R}}$ *with* $dim M_p = dim M_{v^*} = m = r + 1 - j$, *the following inequalities are valid:*

$$Re\lambda_j(A) \leq \max_{\substack{(u,v)\gg 0 \\ u\in M_p, v\in M_{v^*}}} \frac{Re(Au, v)}{(u, v)} \leq Re\lambda_1(A), \tag{2.31}$$

and the following representation formulas hold:

$$Re\lambda_j(A) = \min_{\substack{dim M_p = m \\ dim M_{v^*} = m}} \max_{\substack{(u,v)\gg 0 \\ u\in M_p, v\in M_{v^*}}} \frac{Re(Au, v)}{(u, v)}. \tag{2.32}$$

Remark 2.8. *From (2.31), it follows*

$$\frac{Re(Au, v)}{(u, v)} \leq \nu[A] = \max_{j=1,\cdots,r} Re\lambda_j(A), \; (u, v) \gg 0, \; u \in N_{p,\mathbb{R}}, \; v \in N_{v^*,\mathbb{R}}.$$

Theorem 2.9. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (2.28). Moreover, let the vector spaces* $N_{p,k,\mathbb{R}}$ *and* $N_{v^*,k,\mathbb{R}}$ *for* $k = 1, \cdots, r$ *be defined by (2.9) and (2.14).*
Then,

$$Re\lambda_k(A) = \min_{\substack{(u,v)\gg 0 \\ u\in N_{p,k,\mathbb{R}}, v\in N_{v^*,k,\mathbb{R}}}} \frac{Re(Au, v)}{(u, v)}, \; k = 1, 2, \cdots, r. \tag{2.33}$$

The minimum is attained for $u = p_k$, $v = v_k^*$.

Theorem 2.10. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (2.28).*
Then, for every $j = 1, \cdots, r$ *and all subspaces* $N_p \subset N_{p,\mathbb{R}}$ *and* $N_{v^*} \subset N_{v^*,\mathbb{R}}$ *with* $dim N_p = dim N_{v^*} = j$, *the following inequalities are valid:*

$$Re\lambda_r(A) \leq \min_{\substack{(u,v)\gg 0 \\ u\in N_p, v\in N_{v^*}}} \frac{Re(Au, v)}{(u, v)} \leq Re\lambda_j(A), \tag{2.34}$$

and the following representation formulas hold:

$$Re\lambda_j(A) = \max_{\substack{dim N_p = j \\ dim N_{v^*} = j}} \min_{\substack{(u,v)\gg 0 \\ u\in N_p, v\in N_{v^*}}} \frac{Re(Au, v)}{(u, v)}. \tag{2.35}$$

Remark 2.11. *From (2.34), it follows*

$$\frac{Re(Au, v)}{(u, v)} \geq -\nu[-A] = \min_{j=1,\cdots,r} Re\lambda_j(A), \; (u, v) \gg 0, \; u \in N_{p,\mathbb{R}}, \; v \in N_{v^*,\mathbb{R}}.$$

3. Generalized Rayleigh-Quotient Formulas for the Imaginary Parts of the Eigenvalues of a General Matrix

In this section, we want to state formulas for the representation of the imaginary parts of the eigenvalues of a general matrix $A \in \mathbb{C}^{n\times n}$ by Rayleigh quotients corresponding to those for the real parts. More precisely, max-, min-, min-max-, and max-min-representations are obtained corresponding to those in Section 2.
First, we want to state a relation corresponding to that of (2.19).

Lemma 3.1. *Let the conditions (C1′)-(C4′) be fulfilled. Then, with the denotations of Theorem 2.1 and (2.12),*

$$Im(A\,u,v) = \sum_{j=1}^{r} Im\lambda_j(A)\,(u,v_j^*)\,(p_j,v),\; u \in N_{p,\mathbb{R}},\; v \in N_{v^*,\mathbb{R}}. \tag{3.1}$$

Proof. Equation (3.1) follows directly from Lemma 2.2, Formula (2.18). □

Similarly to (2.28), we suppose that the eigenvalues $\lambda_1(A),\cdots,\lambda_r(A)$ of matrix A are arranged such that

$$Im\lambda_1(A) \geq Im\lambda_2(A) \geq \cdots \geq Im\lambda_r(A). \tag{3.2}$$

Then, we have the following series of theorems.

Theorem 3.2. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (3.2). Moreover, let the vector spaces $M_{p,k,\mathbb{R}}$ and $M_{v^*,k,\mathbb{R}}$ for $k = 1,\cdots,r$ be defined by (2.24), (2.25) or (2.26),(2.27).*
Then,

$$Im\lambda_k(A) = \max_{\substack{(u,v)\gg 0\\ u\in M_{p,k,\mathbb{R}},v\in M_{v^*,k,\mathbb{R}}}} \frac{Im(Au,v)}{(u,v)},\; k = 1,2,\cdots,r. \tag{3.3}$$

The maximum is attained for $u = p_k,\; v = v_k^*$.

Theorem 3.3. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (3.2).*
Then, for every $j = 1,\cdots,r$ and every subspace $M_p \subset N_{p,\mathbb{R}}$ and $M_{v^*} \subset N_{v^*,\mathbb{R}}$ with $dim M_p = dim M_{v^*} = m = r+1-j$, the following inequalities are valid:

$$Im\lambda_j(A) \leq \max_{\substack{(u,v)\gg 0\\ u\in M_p,v\in M_{v^*}}} \frac{Im(Au,v)}{(u,v)} \leq Im\lambda_1(A), \tag{3.4}$$

and the following representation formulas hold:

$$Im\lambda_j(A) = \min_{\substack{dim M_p=m\\ dim M_{v^*}=m}} \max_{\substack{(u,v)\gg 0\\ u\in M_p,v\in M_{v^*}}} \frac{Im(Au,v)}{(u,v)}. \tag{3.5}$$

Remark 3.4. *From (3.4), it follows*

$$\frac{Im(Au,v)}{(u,v)} \leq \max_{j=1,\cdots,r} Im\lambda_j(A),\; (u,v) \gg 0,\; u \in N_{p,\mathbb{R}},\; v \in N_{v^*,\mathbb{R}}.$$

Theorem 3.5. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (3.2). Moreover, let the vector spaces $N_{p,k,\mathbb{R}}$ and $N_{v^*,k,\mathbb{R}}$ for $k = 1,\cdots,r$ be defined by (2.9) and (2.14).*
Then,

$$Im\lambda_k(A) = \min_{\substack{(u,v)\gg 0\\ u\in N_{p,k,\mathbb{R}},v\in N_{v^*,k,\mathbb{R}}}} \frac{Im(Au,v)}{(u,v)},\; k = 1,2,\cdots,r. \tag{3.6}$$

The minimum is attained for $u = p_k,\; v = v_k^*$.

Theorem 3.6. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (3.2).*
Then, for every $j = 1,\cdots,r$ and all subspaces $N_p \subset N_{p,\mathbb{R}}$ and $N_{v^*} \subset N_{v^*,\mathbb{R}}$ with $dim N_p = dim N_{v^*} = j$, the following inequalities are valid:

$$Im\lambda_r(A) \leq \min_{\substack{(u,v)\gg 0\\ u\in N_p,v\in N_{v^*}}} \frac{Im(Au,v)}{(u,v)} \leq Im\lambda_j(A), \tag{3.7}$$

and the following representation formulas hold:

$$Im\lambda_j(A) = \max_{\substack{dim N_p=j\\ dim N_{v^*}=j}} \min_{\substack{(u,v)\gg 0\\ u\in N_p,v\in N_{v^*}}} \frac{Im(Au,v)}{(u,v)}. \tag{3.8}$$

Remark 3.7. *From (3.7), it follows*

$$\frac{Im(Au,v)}{(u,v)} \geq \min_{j=1,\cdots,r} Im\lambda_j(A),\; (u,v) \gg 0,\; u \in N_{p,\mathbb{R}},\; v \in N_{v^*,\mathbb{R}}.$$

4. Generalized Rayleigh-Quotient Formula for the Moduli of the Eigenvalues of a General Matrix

Whereas in Sections 2 and 3 max-, min-, min-max-, and max-min-representations with generalized Rayleigh quotients for general matrices could be obtained, it seems that, for the moduli of eigenvalues, only a max-representation is possible. Some arguments why this is probably the case were already given in [10, Section 4].
Now, we want to state the max-representation. For this, we suppose that the eigenvalues $\lambda_1(A), \cdots, \lambda_r(A)$ of $A \in \mathbb{C}^{n\times n}$ are arranged such that

$$|\lambda_1(A)| \geq |\lambda_2(A)| \geq \cdots \geq |\lambda_r(A)|. \tag{4.1}$$

Herewith, one has the following theorem.

Theorem 4.1. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let the eigenvalues of A be arranged according to (4.1). Moreover, let the vector spaces $M_{p,k,\mathbb{R}}$ and $M_{v^*,k,\mathbb{R}}$ for $k = 1, \cdots, r$ be defined by (2.24), (2.25) or (2.26),(2.27).*
Then,

$$|\lambda_k(A)| = \max_{\substack{(u,v)\gg 0 \\ u\in M_{p,k,\mathbb{R}}, v\in M_{v^*,k,\mathbb{R}}}} \frac{|(Au,v)|}{(u,v)},\ k = 1,2,\cdots,r. \tag{4.2}$$

The maximum is attained for $u = p_k,\ v = v_k^*$.

5. Generalized Rayleigh-Quotient Formulas for a General Matrix with Real Eigenvalues

In Section 4, we have observed that, for the moduli of the eigenvalues of a general matrix, one obtains only a max-representation with generalized Rayleigh quotients. However, for $A \in \mathbb{C}^{n\times n}$ with

$$\sigma(A) \subset \mathbb{R},$$

one gets generalized Rayleigh-quotient formulas for the eigenvalues themselves. And it goes without saying that these imply Rayleigh-quotient representations for the moduli if all eigenvalues are nonnegative such as $\lambda_1(A^*A), \cdots, \lambda_r(A^*A)$ (where $r = n$).
So, let $A \in \mathbb{C}^{n\times n}$ with spectrum $\sigma(A) \subset \mathbb{R}$. Further, let the eigenvalues be arranged according to

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_r(A). \tag{5.1}$$

Then, we obtain the following series of corollaries following from Theorems 2.7 - 2.10, as the case may be.

Corollary 5.1. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let $\sigma(A) \subset \mathbb{R}$, and let the eigenvalues of A be arranged according to (5.1). Moreover, let the vector spaces $M_{p,k,\mathbb{R}}$ and $M_{v^*,k,\mathbb{R}}$ for $k = 1, \cdots, r$ be defined by (2.24), (2.25) or (2.26),(2.27).*
Then,

$$\lambda_k(A) = \max_{\substack{(u,v)\gg 0 \\ u\in M_{p,k,\mathbb{R}}, v\in M_{v^*,k,\mathbb{R}}}} \frac{(Au,v)}{(u,v)},\ k = 1,2,\cdots,r. \tag{5.2}$$

The maximum is attained for $u = p_k,\ v = v_k^*$.

Corollary 5.2. *Let the conditions (C1′)-(C4′) be fulfilled. Further, let $\sigma(A) \subset \mathbb{R}$, and let the eigenvalues of A be arranged according to (5.1).*
Then, for every $j = 1, \cdots, r$ and every subspace $M_p \subset N_{p,\mathbb{R}}$ and $M_{v^*} \subset N_{v^*,\mathbb{R}}$ with $dim M_p = dim M_{v^*} = m = r+1-j$, the following inequalities are valid:

$$\lambda_j(A) \leq \max_{\substack{(u,v)\gg 0 \\ u\in M_p, v\in M_{v^*}}} \frac{(Au,v)}{(u,v)} \leq \lambda_1(A), \tag{5.3}$$

and the following representation formulas hold:

$$\lambda_j(A) = \min_{\substack{dim M_p = m \\ dim M_{v^*} = m}} \max_{\substack{(u,v)\gg 0 \\ u\in M_p, v\in M_{v^*}}} \frac{(Au,v)}{(u,v)}. \tag{5.4}$$

Remark 5.3. *From (5.3), it follows*

$$\frac{(Au,v)}{(u,v)} \leq \max_{j=1,\cdots,r} \lambda_j(A),\ (u,v) \gg 0,\ u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}}.$$

Corollary 5.4. *Let the conditions $(C1')$-$(C4')$ be fulfilled. Further, let $\sigma(A) \subset \mathbb{R}$, and let the eigenvalues of A be arranged according to (5.1). Moreover, let the vector spaces $N_{p,k,\mathbb{R}}$ and $N_{v^*,k,\mathbb{R}}$ for $k = 1, \cdots, r$ be defined by (2.9) and (2.14).*
Then,

$$\lambda_k(A) = \min_{\substack{(u,v) \gg 0 \\ u \in N_{p,k,\mathbb{R}}, v \in N_{v^*,k,\mathbb{R}}}} \frac{(Au,v)}{(u,v)}, \ k = 1, 2, \cdots, r. \tag{5.5}$$

The minimum is attained for $u = p_k$, $v = u_k^*$.

Corollary 5.5. *Let the conditions $(C1')$-$(C4')$ be fulfilled. Further, let $\sigma(A) \subset \mathbb{R}$, and let the eigenvalues of A be arranged according to (5.1).*
Then, for every $j = 1, \cdots, r$ and all subspaces $N_p \subset N_{p,\mathbb{R}}$ and $N_{v^*} \subset N_{v^*,\mathbb{R}}$ with $dim N_p = dim N_{v^*} = j$, the following inequalities are valid:

$$\lambda_r(A) \leq \min_{\substack{(u,v) \gg 0 \\ u \in N_p, v \in N_{v^*}}} \frac{(Au,v)}{(u,v)} \leq \lambda_j(A), \tag{5.6}$$

and the following representation formulas hold:

$$\lambda_j(A) = \max_{\substack{dim N_p = j \\ dim N_{v^*} = j}} \min_{\substack{(u,v) \gg 0 \\ u \in N_p, v \in N_{v^*}}} \frac{(Au,v)}{(u,v)}. \tag{5.7}$$

Remark 5.6. *From (5.6), it follows*

$$\frac{(Au,v)}{(u,v)} \geq \min_{j=1,\cdots,r} \lambda_j(A), \ (u,v) \gg 0, \ u \in N_{p,\mathbb{R}}, \ v \in N_{v^*,\mathbb{R}}.$$

6. Application

In this section, an application of the obtained results is presented. More precisely, a new formula for $\rho(A)$ is stated; its derivation is similar to that of [8, (79)]. First, known formulas for this quantity are recapitulated.

Known formulas for the spectral radius of $A \in \mathbb{C}^{n \times n}$

One formula is given by

$$\rho(A) = \lim_{n \to \infty} \|A^n\|^{\frac{1}{n}}, \tag{6.1}$$

see [4, Chapter I, p.27], where in (6.1) the spectral radius $\rho(A)$ is independent of the used operator norm $\|\cdot\|$.
Another representation is

$$\rho(A) = \max_{j=1,\cdots,n} |\lambda_j(A)|, \tag{6.2}$$

cf. [4, Chapter I,(5.12), p.38].

New formula for the spectral radius of $A \in \mathbb{C}^{n \times n}$

Let the conditions $(C1')$-$(C4')$ be fulfilled. Then, from Theorem 4.1, as Application, we deduce the new formula

$$\rho(A) = \max_{\substack{(u,v) \gg 0 \\ u \in N_{p,\mathbb{R}}, v \in N_{v^*,\mathbb{R}}}} \frac{|(Au,v)|}{(u,v)}. \tag{6.3}$$

7. New Generalized Numerical Ranges

In this section, a series of known numerical ranges are recapitulated and new numerical ranges of a general matrix are defined.

Known numerical range of $A \in \mathbb{C}^{n \times n}$ with respect to the full space $\mathbb{C}^n$

According to [12, Section 5.4,(5)], the *numerical range* of $A \in \mathbb{C}^{n \times n}$ with respect to the full space $\mathbb{C}^n$ is defined by

$$W_{\mathbb{C}^n}(A) = \left\{ z \in \mathbb{C} \,|\, z = \frac{(Au,u)}{(u,u)}, 0 \neq u \in \mathbb{C}^n \right\}, \tag{7.1}$$

which is a convex subset of $\mathbb{C}$. Employing this definition to A^*A instead of A, we obtain

$$W_{\mathbb{C}^n}(A^*A) = \left\{ x \in \mathbb{R}_0^+ \,|\, x = \frac{(A^*Au,u)}{(u,u)} = \frac{(Au,Au)}{(u,u)}, 0 \neq u \in \mathbb{C}^n \right\}, \tag{7.2}$$

which is a convex subset of $\mathbb{R}_0^+$. One has

$$W_{\mathbb{C}^n}(A^*A) = \left[\min_{j=1,\cdots,n} \lambda_j(A^*A), \max_{j=1,\cdots,n} \lambda_j(A^*A)\right] = \left[\frac{1}{\|A^{-1}\|_2^2}, \|A\|_2^2\right] \tag{7.3}$$

where $\frac{1}{\|A^{-1}\|_2^2}$ has to be interpreted as zero if A is singular.

The following four definitions of generalized numerical ranges are new.

Generalized numerical range of $A \in \mathbb{C}^{n\times n}$ with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$

Let the conditions $(C1')$-$(C4')$ be fulfilled. Then, we define the *generalized numerical range* of A with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$ by

$$W_{N_{p,\mathbb{R}},N_{v^*,\mathbb{R}},gen.}(A) = \left\{ z \in \mathbb{C} \,|\, z = \frac{(Au,v)}{(u,v)},\, (u,v) \gg 0,\, u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}} \right\}. \tag{7.4}$$

Real part of the generalized numerical range of $A \in \mathbb{C}^{n\times n}$ with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$

Let the conditions $(C1')$-$(C4')$ be fulfilled. Then, we define the *real part of the generalized numerical range* of A with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{u^*,\mathbb{R}}$ by

$$Re[W_{N_{p,\mathbb{R}},N_{v^*,\mathbb{R}},gen.}(A)] = \left\{ x \in \mathbb{R} \,|\, x = \frac{Re(Au,v)}{(u,v)},\, (u,v) \gg 0,\, u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}} \right\}. \tag{7.5}$$

Imaginary part of the generalized numerical range of $A \in \mathbb{C}^{n\times n}$ with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$

Let the conditions $(C1')$-$(C4')$ be fulfilled. Then, we define the *imaginary part of the generalized numerical range* of A with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$ by

$$Im[W_{N_{p,\mathbb{R}},N_{v^*,\mathbb{R}},gen.}(A)] = \left\{ x \in \mathbb{R} \,|\, x = \frac{Im(Au,v)}{(u,v)},\, (u,v) \gg 0,\, u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}} \right\}. \tag{7.6}$$

Modulus of the generalized numerical range of $A \in \mathbb{C}^{n\times n}$ with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$

Let the conditions $(C1')$-$(C4')$ be fulfilled. Then, we define the *modulus of the generalized numerical range* of A with respect to the subspaces $N_{p,\mathbb{R}}$ and $N_{v^*,\mathbb{R}}$ by

$$|[W_{N_{p,\mathbb{R}},N_{v^*,\mathbb{R}},gen.}(A)]| = \left\{ x \in \mathbb{R}_0^+ \,|\, x = \frac{|(Au,v)|}{(u,v)},\, (u,v) \gg 0,\, u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}} \right\}. \tag{7.7}$$

8. Numerical example

In this section, we check some of the formulas of Section 2 on an example from the theory of linear dynamical systems.

8.1. A two-mass vibration model

We take up the multi-mass vibration model of [5], shown in *Fig.8.1*

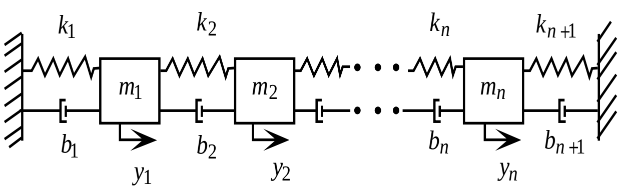

Fig.8.1: Multi-mass vibration model

and study the case $\mathbf{n} = \mathbf{2}$ as in [11]. For the sake of completeness, we give again the details. The associated initial value problem is given by

$$M\ddot{y} + B\dot{y} + Ky = 0,\ y(0) = y_0,\ \dot{y}(0) = \dot{y}_0,$$

where $y = [y_1, y_2]^T$ and

$$M = \left[\begin{array}{c|c} m_1 & 0 \\ \hline 0 & m_2 \end{array}\right],$$

$$B = \left[\begin{array}{c|c} b_1 + b_2 & -b_2 \\ \hline -b_2 & b_2 + b_3 \end{array}\right],$$

$$K = \left[\begin{array}{c|c} k_1+k_2 & -k_2 \\ \hline -k_2 & k_2+k_3 \end{array}\right],$$

with the *mass, damping,* and *stiffness matrices* M, B, and K, as the case may be, and the *displacement vector* y. In *state-space description*, this problem takes the form

$$\dot{x} = Ax,\, t \geq 0,\, x(0) = x_0, \tag{8.1}$$

where $x = [y^T, z^T]^T$, $z = \dot{y}$, and where the *system matrix* A is given by

$$A = \left[\begin{array}{c|c} 0 & E \\ \hline -M^{-1}K & -M^{-1}B \end{array}\right].$$

Like in [11], as numerical values for the quantities not yet specified, we choose $b_1 = 1/4$, $k_2 = 2^3 = 8$. On the whole, this delivers the following data:

$$m_1 = m_2 = 1;\; b_1 = 1/4, b_2 = 0, b_3 = 1/4;\; k_1 = 1/64 = 1/2^4, k_2 = 8, k_3 = 1/64 = 1/2^4,$$

which leads to

$$\begin{aligned} M &= \left[\begin{array}{c|c} m_1 & 0 \\ \hline 0 & m_2 \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & 1 \end{array}\right], \\ B &= \left[\begin{array}{c|c} b_1+b_2 & -b_2 \\ \hline -b_2 & b_2+b_3 \end{array}\right] = \left[\begin{array}{c|c} 0.25 & 0 \\ \hline 0 & 0.25 \end{array}\right], \\ K &= \left[\begin{array}{c|c} k_1+k_2 & -k_2 \\ \hline -k_2 & k_2+k_3 \end{array}\right] = \left[\begin{array}{c|c} 1/64+8 & -1/2 \\ \hline -1/2 & 8+1/64 \end{array}\right] = \left[\begin{array}{c|c} 8.015625 & -0.5 \\ \hline -0.5 & 8.015625 \end{array}\right]. \end{aligned}$$

Further, we choose

$$t_0 = 0$$

as well as

$$y_0 = [-1, 1]^T$$

and

$$\dot{y}_0 = [-1, -1]^T,$$

but y_0 and $\dot{y}_0$ are not needed here.

8.2. Computation of important quantities

Using the Matlab routine *jordan*, one obtains

$$\begin{aligned} \lambda_1(A) &= -0.1250 + 4.0000i, \\ \lambda_2(A) &= -0.1250 - 4.0000i, \\ \lambda_3(A) &= -0.1250, \\ \lambda_4(A) &= \lambda_3(A). \end{aligned} \tag{8.2}$$

The pertinent eigenvectors and principal vectors are

$$\left[p_1^{(1)}, p_1^{(2)}, p_1^{(3)}, p_2^{(3)}\right] = [p_1, p_2, p_3, p_4.]$$

They are unnormed. The algebraic multiplicities are thus $m_1 = m_2 = 1$ and $m_3 = 2$. So, here, $r = 3$.
For the adjoint matrix A^*, we obtain

$$\begin{aligned} \lambda_1(A^*) &= -0.1250 - 4.0000i, \\ \lambda_2(A^*) &= -0.1250 + 4.0000i, \\ \lambda_3(A^*) &= -0.1250, \\ \lambda_4(A^*) &= \lambda_3(A^*). \end{aligned} \tag{8.3}$$

The associated eigenvectors and principal vectors are

$$\left[u_1^{(1)*}, u_1^{(2)*}, u_1^{(3)*}, u_2^{(3)*}\right] = [u_1^*, u_2^*, u_3^*, u_4^*.]$$

They are also unnormed.
In [11], we biorthogonalized these vectors based on Theorem 1 such that the relations

$$(p_k^{(i)}, u_l^{(i)*}) = \begin{cases} 1, & l = m_i - k + 1 \\ 0, & l \neq m_i - k + 1 \end{cases} \tag{8.4}$$

and

$$(p_k^{(i)}, u_l^{(j)^*}) = 0,\ i \neq j. \tag{8.5}$$

So, with

$$v_l^{(i)^*} = u_{m_i-l+1}^{(i)}{}^*, \tag{8.6}$$

one has then the biorthogonality relations

$$(p_k^{(i)}, v_l^{(j)^*}) = \delta_{kl}\delta_{ij}. \tag{8.7}$$

The details of the biorthogonalization can be found in [11]. Define

$$\begin{aligned} v_1^* &= v_1^{(1)^*} = u_1^{(1)^*} = u_1^*, \\ v_2^* &= v_1^{(2)^*} = u_1^{(2)^*} = u_2^*, \\ v_3^* &= v_1^{(3)^*} = u_2^{(3)^*} = u_4^*, \\ v_4^* &= v_2^{(3)^*} = u_1^{(3)^*} = u_3^*. \end{aligned}$$

Herewith,

$$(p_i, v_j^*) = \delta_{ij},\ i,j = 1,\cdots,4 \tag{8.8}$$

where

$$p_1 = \begin{bmatrix} 0.364602 \\ -0.364602 \\ -0.045575+1.458408i \\ 0.045575-1.458408i \end{bmatrix}, \qquad v_1^* = \begin{bmatrix} 0.685679+0.021427i \\ -0.685679-0.021427i \\ 0+0.171420i \\ 0-0.171420i \end{bmatrix},$$

$$p_2 = \begin{bmatrix} 0.364602 \\ -0.364602 \\ -0.045575-1.458408i \\ 0.045575+1.458408i \end{bmatrix}, \qquad v_2^* = \begin{bmatrix} 0.685679-0.021427i \\ -0.685679+0.021427i \\ 0-0.171420i \\ 0+0.171420i \end{bmatrix},$$

$$p_3 = \begin{bmatrix} 0.707107 \\ 0.707107 \\ -0.088388 \\ -0.088388 \end{bmatrix}, \qquad v_3^* = \begin{bmatrix} 0.707107 \\ 0.707107 \\ 0 \\ 0 \end{bmatrix},$$

$$p_4 = \begin{bmatrix} 0 \\ 0 \\ 0.712610 \\ 0.712610 \end{bmatrix}, \qquad v_4^* = \begin{bmatrix} 0.087706 \\ 0.087706 \\ 0.701646 \\ 0.701646 \end{bmatrix}.$$

As in [11], we add the followings remarks.

Remark 8.1. *The vector $p_2^{(3)}$ is a principal vector of stage 2 for matrix A. But, since it is normed such that $(p_2^{(3)}, v_2^{(3)^*}) = 1$ instead of $\|p_2^{(3)}\|_2 = 1$, we have* **not** *$Ap_2^{(3)} = \lambda_3 p_2^{(3)} + p_1^{(3)}$, but instead, $Ap_2^{(3)} = \lambda_3 p_2^{(3)} + \gamma_1^{(3)} p_1^{(3)}$ with a factor $\gamma_1^{(3)} \neq 0$, $\gamma_1^{(3)} \neq 1$. Similarly, $u_2^{(3)^*}$ is principal vector of stage 2 for A^*. But, due to the biorthogonalization process, we have* **not** *$A^* u_2^{(3)^*} = \overline{\lambda_3} u_2^{(3)^*} + u_1^{(3)^*}$, but instead, $A^* u_2^{(3)^*} = \overline{\lambda_3} u_2^{(3)^*} + \delta_1^{(3)} u_1^{(3)^*}$ with a factor $\delta_1^{(3)} \neq 0$, $\delta_1^{(3)} \neq 1$. We leave it to the reader to check this numerically on our example.*

Remark 8.2. *Due to the foregoing remark, Formula (2.17) looks somewhat different. But, Formula (2.18) remains valid which is the important point since the subsequent findings are based on Formula (2.18), not on Formula (2.17).*

8.3. Numerical check of Theorems 2.6 and 2.9

From Theorem 2.6, Formula (2.30) and Theorem 2.9, Formula (2.33), we conclude

$$\min_{j=1,2,3} Re\lambda_j(A) \leq \frac{Re(Au,v)}{(u,v)} \leq \max_{j=1,2,3} Re\lambda_j(A),$$

$(u,v) \gg 0$, $u \in N_{p,\mathbb{R}}$, $v \in N_{v^*,\mathbb{R}}$ by setting $k = 1$, there. This can also be written as

$$\frac{Re(Au,v)}{(u,v)} \in \left[\min_{j=1,2,3} Re\lambda_j(A), \max_{j=1,2,3} Re\lambda_j(A)\right],$$

$(u,v) \gg 0,\ u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}}$. We check this for a series of vectors. One has

$$\left[\min_{j=1,2,3} Re\lambda_j(A),\ \max_{j=1,2,3} Re\lambda_j(A)\right] = [-0.1250, -0.1250];$$

in other words

$$\frac{Re(Au,v)}{(u,v)} = -0.1250,\ (u,v) \gg 0,\ u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}}.$$

Let

$$\begin{aligned} u_1 &= -5p_1 + 3p_3, \\ v_1 &= -4v_1^* + 2v_3^*. \end{aligned}$$

Then, $u_1 \in N_{p,\mathbb{R}}$ and $v_1 \in N_{v^*,\mathbb{R}}$ as well as $(u_1, v_1) \gg 0$, and one obtains

$$u_1 = \begin{bmatrix} 0.298311 \\ 3.944330 \\ -0.037289 - 7.292039i \\ -0.493041 + 7.292039i \end{bmatrix}, \quad v_1 = \begin{bmatrix} -1.328504 - 0.085710i \\ 4.156931 + 0.085710i \\ 0 - 0.685679i \\ 0 + 0.685679i \end{bmatrix},$$

$$\begin{aligned} (Au_1, v_1) &= -3.250000000000000 + 80.000000000000000i, \\ (u_1, v_1) &= 26.000000000000004, \end{aligned}$$

and thus

$$\frac{Re(Au_1,v_1)}{(u_1,v_1)} \doteq -0.125000000000000.$$

Let

$$\begin{aligned} u_2 &= 3p_2, \\ v_2 &= -4v_1^* + 2v_2^*. \end{aligned}$$

Then, $u_2 \in N_{p,\mathbb{R}}$ and $v_2 \in N_{v^*,\mathbb{R}}$ as well as $(u_2, v_2) \gg 0$, and one obtains

$$u_2 = \begin{bmatrix} 1.093806 \\ -1.093806 \\ -0.136726 - 4.375223i \\ 0.136726 + 4.375223i \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1.371359 - 0.128565i \\ 1.371359 + 0.128565i \\ 0 - 1.028519i \\ 0 + 1.028519i \end{bmatrix},$$

$$\begin{aligned} (Au_2, v_2) &= -0.750000000000000 - 24.000000000000000i, \\ (u_2, v_2) &= 6, \end{aligned}$$

and thus

$$\frac{Re(Au_2,v_2)}{(u_2,v_2)} \doteq -0.125000000000000.$$

Let

$$\begin{aligned} u_3 &= -5p_1 + 3p_2 - 4p_3, \\ v_3 &= -4v_1^* + 2v_2^* - 2v_3^*. \end{aligned}$$

Then, $u_3 \in N_{p,\mathbb{R}}$ and $v_3 \in N_{v^*,\mathbb{R}}$ as well as $(u_3, v_3) \gg 0$, and one obtains

$$u_3 = \begin{bmatrix} -3.557631 \\ -2.099223 \\ 0.444704 - 11.667262i \\ 0.262403 + 11.667262i \end{bmatrix}, \quad v_3 = \begin{bmatrix} -2.785572 - 0.128565i \\ -0.042855 + 0.128565i \\ 0 - 1.028519i \\ 0 + 1.028519i \end{bmatrix},$$

$$\begin{aligned} (Au_3, v_3) &= -4.250000000000000 + 56.000000000000000i, \\ (u_3, v_3) &= 34, \end{aligned}$$

and thus

$$\frac{Re(Au_3,v_3)}{(u_3,v_3)} \doteq -0.125000000000000.$$

Let

$$\begin{aligned} u_4 &= -5p_1 + 3p_2 + 6p_3 - 4p_4, \\ v_4 &= -2v_1^* + 4v_2^* + 2v_3^* - 3v_4^*. \end{aligned}$$

Then, $(u_4, v_4) \gg 0$, and one obtains

$$u_4 = \begin{bmatrix} 3.513437 \\ 4.971845 \\ -3.289618 - 11.667262i \\ -3.471919 + 11.667262i \end{bmatrix}, \qquad v_4 = \begin{bmatrix} 2.522455 - 0.128565i \\ -0.220262 + 0.128565i \\ -2.104939 - 1.028519i \\ -2.104939 + 1.028519i \end{bmatrix},$$

$$\begin{aligned} (Au_4, v_4) &= -13.812257748298542 - 7.999999999999989i, \\ (u_4, v_4) &= 46, \end{aligned}$$

and thus

$$\begin{aligned} \frac{Re(Au_4, v_4)}{(u_4, v_4)} &= -0.709454874462791 \\ &\neq -0.1250 \end{aligned}$$

which is not surprising since $u_4 \notin N_{p,\mathbb{R}} = [p_1, p_2, p_3]$ and $v_4 \notin N_{v^*,\mathbb{R}} = [v_1^*, v_2^*, v_3^*]$. Recall at this point that $r = 3$ and $v_j^* = u_{m_j}^{(j)^*}$, $j = 1,2,3$ are the principal vectors of maximum stage associated with the eigenvalues $\lambda_j(A^*), j = 1,2,3$, as the case may be. With $m_1 = m_2 = 1$ and $m_3 = 2$, therefore $v_1^* = u_1^{(1)^*}$ and $v_2^* = u_1^{(2)^*}$ are eigenvectors and $v_3^* = u_2^{(3)^*}$ is a principal vector of stage 2 whereas $v_4^* = u_1^{(3)^*}$ is an eigenvector and thus not a principal vector of maximum stage.
Let

$$\begin{aligned} u_5 &= [1,2,3,4]^T \in \mathbb{R}^4, \\ v_5 &= [4,3,2,1]^T \in \mathbb{R}^4. \end{aligned}$$

Here, one obtains

$$\begin{aligned} (Au_5, v_5) &= 29.437500000000000, \\ (u_5, v_5) &= 20, \end{aligned}$$

and thus

$$\frac{Re(Au_5, v_5)}{(u_5, v_5)} = 1.471875000000000 \neq -0.1250$$

which is neither surprising since $(u_5, v_5) \not\gg 0$ due to

$$\alpha^{(5)} := (\alpha_k^{(5)})_{k=1,\cdots,4} = ((u_5, v_k^*))_{k=1,\cdots,4} = \begin{bmatrix} -0.685679 + 0.192847i \\ -0.685679 - 0.192847i \\ 2.121320 \\ 5.174642 \end{bmatrix}$$

and

$$\beta^{(5)} := (\beta_k^{(5)})_{k=1,\cdots,4} = ((p_k, v_5))_{k=1,\cdots,4} = \begin{bmatrix} 0.319027 + 1.458408i \\ 0.319027 - 1.458408i \\ 4.684582 \\ 2.137829 \end{bmatrix}.$$

8.4. Computational aspects

In this subsection, we say something about the used computer equipment and the computation times.
(i) As to the *computer equipment*, the following hardware was available: an Intel Core2 Duo Processor at 3166 GHz, a 500 GB mass storage facility, and two 2048 MB high-speed memories. As software package for the computations, we used MATLAB, Version 7.11.
(ii) The *computation time t* of an operation was determined by the command sequence *t1=clock; operation; t=etime(clock,t1)*. It is put out in seconds, rounded to four decimal places. For the computation of the eigenvalues of matrix A in Subsection 8.2, we used the command [XA,DA]=eig(A); the pertinent computation time was less than 0.0001 s.

9. Numerical Example 2

In this section, we proceed in a similar way as in Section 8. Here, we present an example of a real non-diagonalizable matrix A with real eigenvalues.

9.1. The matrix A and its eigenvalues and principal vectors

We take the matrix A from [1, Example 5.2, p.82]. So, let

$$A = \begin{bmatrix} 4 & 1 & 1 \\ 2 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}.$$

In [1], the eigenvalues are given as

$$\begin{aligned} \lambda_1 &= 6, \\ \lambda_2 &= 3, \\ \lambda_3 &= \lambda_2, \end{aligned}$$

where the numbering is such that $\lambda_1 \geq \lambda_2$. According to [1], the associated right eigenvectors are given as

$$p_1 = p_1^{(1)} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}, \; p_2 = p_1^{(2)} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix};$$

they are unnormed. Here, $m_1 = 1$ and $m_2 = 2$. So, here, $r = 2$. A corresponding principal vector $p_3 = p_2^{(2)}$ is not given in [1].
Further, there, the vectors u_1^* and u_2^* of $A^* = A^T$ are given as

$$u_1^* = u_1^{(1)^*} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \; u_2^* = u_1^{(2)^*} = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix};$$

they are also unnormed. Again, the principal vector $u_3^* = u_2^{(2)^*}$ is not given in [1].
From

$$Ap_2^{(2)} = \lambda_2(A)p_2^{(2)} + p_1^{(2)},$$

one can determine a principal vector $p_2^{(2)}$, and from

$$A^* u_2^{(2)^*} = \lambda_2(A^*)u_2^{(2)^*} + u_1^{(2)^*},$$

a principal vector $u_2^{(2)^*}$. By hand calculation, we obtain

$$p_2^{(2)} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \; u_2^{(2)^*} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}$$

if we choose the third component of $p_2^{(2)}$ and the first component of $u_2^{(2)^*}$ as zero. At this point, we remind that these principal vectors of stage 2 are only determined up to an associated eigenvector.

9.2. Auxiliary computational results

Using the Matlab routine *eig.m*, we obtain

$$\begin{aligned} \lambda_1 &= 6, \\ \lambda_2 &= 3, \\ \lambda_3 &= \lambda_2. \end{aligned}$$

The pertinent computed biorthonormal right eigenvectors and principal vectors

$$\left[p_1^{(1)}, p_1^{(2)}, p_2^{(2)}\right] = [p_1, p_2, p_3]$$

are unnormed. The algebraic multiplicities are thus $m_1 = 1$ and $m_2 = 2$.
For the adjoint matrix A^*, we obtain

$$\begin{aligned} \lambda_1(A^*) &= \lambda_1(A), \\ \lambda_2(A^*) &= \lambda_2(A), \\ \lambda_3(A^*) &= \lambda_2(A). \end{aligned} \tag{9.1}$$

The associated eigenvectors and principal vectors are

$$\left[u_1^{(1)^*}, u_1^{(2)^*}, u_2^{(2)^*}\right] = [u_1^*, u_2^*, u_3^*]$$

are also unnormed.

As in [11], we biorthogonalized these vectors based on Theorem 2.1 such that the relations (8.4) - (8.7) hold. The details of the biorthogonalization can be found in [11].
Define

$$v_1^* = v_1^{(1)^*} = u_1^{(1)^*} = u_1^*,$$
$$v_2^* = v_1^{(2)^*} = u_2^{(2)^*} = u_3^*,$$
$$v_3^* = v_1^{(3)^*} = u_1^{(2)^*} = u_2^*.$$

Herewith,

$$(p_i, v_j^*) = \delta_{ij},\ i,j = 1,\cdots,3 \tag{9.2}$$

where

$$p_1 = \begin{bmatrix} -0.577350269189626 \\ -0.769800358919501 \\ -0.384900179459750 \end{bmatrix}, \qquad v_1^* = \begin{bmatrix} -0.577350269189626 \\ -0.577350269189626 \\ -0.577350269189626 \end{bmatrix},$$

$$p_2 = \begin{bmatrix} 0 \\ -1.427248064296125 \\ 1.427248064296125 \end{bmatrix}, \qquad v_2^* = \begin{bmatrix} -0.622799155329218 \\ 0.077849894416152 \\ 0.778498944161523 \end{bmatrix},$$

$$p_3 = \begin{bmatrix} -0.816496580927726 \\ 1.632993161855452 \\ -0.816496580927726 \end{bmatrix}, \qquad v_3^* = \begin{bmatrix} -0.816496580927726 \\ 0.408248290463863 \\ 0.408248290463863 \end{bmatrix}.$$

These results are based on the eigenvectors and principal vectors computed by using the Matlab routine *jordan*. We leave it to the reader to compute these vectors by starting with the unnormed vectors stated in Section 9.1. The result is somewhat different. This outcome is not surprising since the principal vectors are determined only up to eigenvectors for the treated matrix A.
We conclude this section by mentioning that, here, similar remarks hold to those at the end of Section 8.2.

9.3. Numerical check of Corollaries 5.2 and 5.5

From Corollary 5.2, Formula (5.3) and Corollary 5.5, Formula (5.7), we conclude

$$\min_{j=1,2} \lambda_j(A) \le \frac{(Au,v)}{(u,v)} \le \max_{j=1,2} \lambda_j(A),$$

$(u,v) \gg 0,\ u \in N_{p,\mathbb{R}} = [p_1, p_2]_{\mathbb{R}},\ v \in N_{v^*,\mathbb{R}} = [v_1^*, v_2^*]_{\mathbb{R}}$ by setting $k=1$, there. This can also be written as

$$\frac{(Au,v)}{(u,v)} \in \left[\min_{j=1,2} \lambda_j(A), \max_{j=1,2} \lambda_j(A)\right],$$

$(u,v) \gg 0,\ u \in N_{p,\mathbb{R}},\ v \in N_{v^*,\mathbb{R}}$. We check this for a series of vectors. One has

$$\left[\min_{j=1,2} \lambda_j(A), \max_{j=1,2} \lambda_j(A)\right] = [3;6].$$

Let

$$u_1 = -5p_1 + 3p_2,$$
$$v_1 = -4v_1^* + 2v_2^*.$$

Then, $u_1 \in N_{p,\mathbb{R}}$ and $v_1 \in N_{v^*,\mathbb{R}}$ as well as $(u_1, v_1) \gg 0$, and one obtains

$$u_1 = \begin{bmatrix} 2.886751345948128 \\ -0.432742398290872 \\ 6.206245090187128 \end{bmatrix}, \qquad v_1 = \begin{bmatrix} 1.063802766100067 \\ 2.465100865590808 \\ 3.866398965081549 \end{bmatrix},$$

$$(Au_1, v_1) = 138.0000000000000,$$
$$(u_1, v_1) = 25.999999999999996,$$

and thus

$$\frac{(Au_1, v_1)}{(u_1, v_1)} \doteq 5.307692307692308 \in [3;6].$$

Let

$$u_2 = 3p_2,$$
$$v_2 = -4v_1^* + 2v_2^*.$$

Then, $u_2 \in N_{p,\mathbb{R}}$ and $v_2 \in N_{v^*,\mathbb{R}}$ as well as $(u_2, v_2) \gg 0$, and one obtains

$$u_2 = \begin{bmatrix} 0 \\ -4.281744192888376 \\ 4.281744192888376 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 1.063802766100067 \\ 2.465100865590808 \\ 3.866398965081549 \end{bmatrix},$$

$$\begin{aligned} (Au_2, v_2) &= 17.999999999999996, \\ (u_2, v_2) &= 5.999999999999998, \end{aligned}$$

and thus

$$\frac{(Au_2, v_2)}{(u_2, v_2)} \doteq 3.000000000000000 \in [3;6].$$

Let

$$\begin{aligned} u_3 &= -2p_1 + 2p_2, \\ v_3 &= -2v_1^* + 2v_2^*. \end{aligned}$$

Then, $u_3 \in N_{p,\mathbb{R}}$ and $v_3 \in N_{v^*,\mathbb{R}}$ as well as $(u_3, v_3) \gg 0$, and one obtains

$$u_3 = \begin{bmatrix} 1.154700538379251 \\ -1.314895410753249 \\ 3.624296487511752 \end{bmatrix}, \qquad v_3 = \begin{bmatrix} -0.090897772279185 \\ 1.310400327211556 \\ 2.711698426702298 \end{bmatrix},$$

$$\begin{aligned} (Au_3, v_3) &= 36, \\ (u_3, v_3) &= 7.999999999999999, \end{aligned}$$

and thus

$$\frac{(Au_3, v_3)}{(u_3, v_3)} \doteq 4.500000000000001 \in [3;6].$$

Let

$$\begin{aligned} u_4 &= -5p_1 + 3p_2 + 6p_3, \\ v_4 &= -2v_1^* + 4v_2^* + 2v_3^*. \end{aligned}$$

Then, $(u_4, v_4) \gg 0$, and one obtains

$$u_4 = \begin{bmatrix} -2.012228139618228 \\ 9.365216572841840 \\ 1.307265604620772 \end{bmatrix}, \qquad v_4 = \begin{bmatrix} -2.969489244793074 \\ 2.282596696971587 \\ 5.085192895953069 \end{bmatrix},$$

$$\begin{aligned} (Au_4, v_4) &= 145.7298612851365, \\ (u_4, v_4) &= 34, \end{aligned}$$

and thus

$$\frac{(Au, v_4)}{(u_4, v_4)} = 4.286172390739309 \in [3;6]$$

even though $u_4 \notin N_{p,\mathbb{R}} = [p_1, p_2]_{\mathbb{R}}$ and $v_4 \notin N_{v^*,\mathbb{R}} = [v_1^*, v_2^*]_{\mathbb{R}}$.
Let

$$\begin{aligned} u_5 &= [1,2,3]^T \in \mathbb{R}^3, \\ v_5 &= [3,2,1]^T \in \mathbb{R}^3. \end{aligned}$$

Here, one obtains

$$\begin{aligned} (Au_5, v_5) &= 67, \\ (u_5, v_5) &= 10, \end{aligned}$$

and thus

$$\frac{(Au_5, v_5)}{(u_5, v_5)} = 6.700000000000000 \notin [3;6]$$

which is not surprising since $(u_5, v_5) \not\gg 0$ due to

$$\alpha^{(5)} := (\alpha_k^{(5)})_{k=1,2,3} = ((u_5, v_k^*))_{k=1,2,3} = \begin{bmatrix} -3.464101615137755 \\ 1.868397465987655 \\ 1.224744871391589 \end{bmatrix}$$

and

$$\beta^{(5)} := (\beta_k^{(5)})_{k=1,2,3} = ((p_k, v_5))_{k=1,2,3} = \begin{bmatrix} -3.656551704867629 \\ -1.427248064296125 \\ 0.000000000000000 \end{bmatrix}.$$

10. Conclusion

It has been shown that there exist generalized Rayleigh-quotient representations of the real and imaginary parts of the eigenvalues of general matrices that parallel those for the eigenvalues of diagonalizable matrices. As in that case, for the moduli, only a max-representation could be stated. The special case of general matrices with real eigenvalues has also been considered. The main difference to the case of diagonalizable matrices is that the space $N_{u^*,\mathbb{R}} = [u_1^*, \cdots, u_n^*]_{\mathbb{R}}$ is replaced by the space $N_{v^*,\mathbb{R}} = [v_1^*, \cdots, u_r^*]_{\mathbb{R}}$ where $v_j^* = v_1^{(j)^*} = u_{m_j}^{(j)^*}$ are the principal vectors of largest stage m_j pertinent to the eigenvalues $\lambda_j(A^*) = \overline{\lambda_j(A)}$ for $j = 1, \cdots, r$. As application, a new formula for the spectral radius $\rho(A)$ for general matrices is obtained. On a numerical example from the theory of linear dynamical systems with non-diagonalizable system matrix A (Example 1), we check that $\frac{Re(Au,v)}{(u,v)} \in [\min_{j=1,\cdots,r} Re\lambda_j(A), \max_{j=1,\cdots,r} Re\lambda_j(A)]$, $(u,v) \gg 0$, $u \in N_{p,\mathbb{R}}$, $v \in N_{v^*,\mathbb{R}}$. On a further example (Example 2), also for a non-diagonalizable matrix A, this time with real eigenvalues, we check numerically that $\frac{(Au,v)}{(u,v)} \in [\min_{j=1,\cdots,r} \lambda_j(A), \max_{j=1,\cdots,r} \lambda_j(A)]$, $(u,v) \gg 0$, $u \in N_{p,\mathbb{R}}$, $v \in N_{v^*,\mathbb{R}}$. We mention that, in the case of diagonalizable matrices, the results of [10] are obtained back since then $r = n$ and $m_j = 1$ for $j = 1, \cdots, r$ so that then $N_{v^*,\mathbb{R}} = N_{u^*,\mathbb{R}}$ and $N_{\sigma(A)} = \mathbb{C}^n$. The paper is of interest on its own in the areas of Linear Algebra and Numerical Analysis. Beyond this, it could be of value to mathematicians and engineers, in general.

References

[1] R.T. Gregory, D.L. Karney, *A Collection of Matrices for Testing Computational Algorithms*, Wiley-Interscience, New York, London Sydney Toronto, 1969.
[2] R.A. Horn, Ch.R. Johnson, *Matrix Analysis*, University Press, Cambridge, 22nd printing, 2009.
[3] R.A. Horn, Ch.R. Johnson, *Topics in Matrix Analysis*, University Press, Cambridge, 10th printing, 2008.
[4] T. Kato, *Perturbation Theory for Linear Operators*, Springer, New York, 1966.
[5] L. Kohaupt, *Construction of a biorthogonal system of principal vectors of the matrices A and A^* with applications to the initial value problem* $\dot{x} = Ax$, $x(t_0) = x_0$, J. Comput. Math. Optim., **3**(3) (2007),163-192.
[6] L. Kohaupt, *Biorthogonalization of the principal vectors for the matrices A and A^* with application to the computation of the explicit representation of the solution $x(t)$ of $\dot{x} = Ax$, $x(t_0) = x_0$*, Appl. Math. Sci., **2**(20) (2008), 961-974.
[7] L. Kohaupt, *Solution of the vibration problem $M\ddot{y} + B\dot{y} + Ky = 0$, $y(t_0) = y_0$, $\dot{y}(t_0) = \dot{y}_0$* without *the hypothesis $BM^{-1}K = KM^{-1}B$ or $B = \alpha M + \beta K$*, Appl. Math. Sci., **2**(41) (2008)1989-2024.
[8] L. Kohaupt, *Generalized Rayleigh-quotient representation of the eigenvalues of self-adjoint matrices*, J. Algebra Appl. Math., **14**(1) (2016)1-26.
[9] L. Kohaupt, *Rayleigh-quotient representation of the real parts, imaginary parts, and moduli of the eigenvalues of diagonalizable matrices*, J. Math. Sci. Model., **2**(2) (2019), 82-98.
[10] L. Kohaupt, *Generalized Rayleigh-quotient representation of the real parts, imaginary parts, and moduli of the eigenvalues of diagonalizable matrices*, J. Adv. Math., **14**(2) (2018), 7702-7728.
[11] L. Kohaupt, *Rayleigh-quotient representation of the real parts, imaginary parts, and moduli of the eigenvalues of general matrices*, J. Math. Sci. Model., **3**(2) (2020), 55-75.
[12] F. Stummel, K. Hainer, *Introduction to Numerical Analysis*, Scottish Academic Press, Edinburgh, 1980.
[13] L.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.
[14] R. Zurmühl, S. Falk, *Matrizen und ihre Anwendungen, Teil 1: Grundlagen (Matrices and Their Applications, Part 1: Foundations)*, Springer-Verlag, Berlin Heidelberg New York Tokyo, 1984.

Appendix

As in [10], with a minor additional hypothesis, the generalized min-, min-max-, and max-min-representaions for the moduli of eigenvalues can be proven. In this Appendix, we show this, but restrict ourselves to the min-max-representation. The minor additional hypothesis is $p_j \in M_p$ and $v_j^* \in M_{v^*}$. A further advantage of this additional hypothesis is that the proofs simplify. But, we omit the proof since it is similar to that in the case when matrix A is diagonalizable in [10].
We have the following theorem.

Theorem 10.1. *Let the conditions $(C1')$-$(C4')$ be fulfilled. Further, let the eigenvalues of A be arranged according to (4.1).*
Then, for every $j = 1, \cdots, n$ and every subspace $M_p \subset N_{p,\mathbb{R}}$ and $M_{v^*} \subset N_{v^*,\mathbb{R}}$ with $dim M_p = dim M_{v^*} = m = n+1-j$ where **additionally** $\mathbf{p_j} \in \mathbf{M_p}$ *and* $\mathbf{v_j^*} \in \mathbf{M_{v^*}}$*, the following inequalities are valid:*

$$|\lambda_j(A)| \leq \max_{\substack{(u,v)\gg 0 \\ u\in M_p, v\in M_{v^*}}} \frac{|(Au,v)|}{(u,v)} \leq |\lambda_1(A)|,$$

and the following representation formulas hold:

$$|\lambda_j(A)| = \min_{\substack{dim M_p = m, p_j \in M_p \\ dim M_{v^*} = m, v_j^* \in M_{v^*}}} \max_{\substack{(u,v)\gg 0 \\ u\in M_p, v\in M_{v^*}}} \frac{|(Au,v)|}{(u,v)}.$$

Remark 10.2. *We mention that, with the above additional hypotheses, the proofs of Theorems 2.6 - 2.10, Theorems 3.2 - 3.5, Theorem 4.1, and Corollaries 5.1 - 5.5 get also simpler.*

UJMA

Universal Journal of Mathematics and Applications, 4 (1) (2021) 26-32
Research paper

Universal Journal of Mathematics and Applications

Journal Homepage: www.dergipark.gov.tr/ujma
ISSN 2619-9653
DOI: https://doi.org/10.32323/ujma.818071

On the Resolution of the Acceleration Vector According to Bishop Frame[†]

Kahraman Esen Özen[1*] and Murat Tosun[2]

[1] *Sakarya, Turkey*
[2] *Department of Mathematics, Faculty of Science and Arts, Sakarya University, Sakarya, Turkey*
[*] *Corresponding author*

Article Info

Keywords: *Kinematics of a particle, Plane and space curves, Bishop frame, Siacci.*
2010 AMS: *70B05, 14H50.*
Received: *29 September 2020*
Accepted: *14 February 2021*
Available online: *22 March 2021*

Abstract

In the second half of the 19th century, Siacci investigated the motion of a particle in space under the influence of any forces (Atti R Accad Sci. Torino **14**(1879)). In this study, Siacci obtained a resolution of the acceleration vector which is very useful when the angular momentum is conserved. On the other hand, Bishop introduced the Bishop frame which is well defined for every curves and so very convenient for mathematical researches in the third quarter of the 20th century (Am Math Monthly **82**(1975)). In this study, we discuss the Siacci's resolution of the acceleration vector according to Bishop frame of the trajectory of the moving particle. Also, we provide an illustrative example for the obtained results.

1. Introduction

In kinematics, the change in velocity of a moving particle in 3-dimensional Euclidean space with respect to the time parameter gives the acceleration. Since the force acting on a particle is concerned with its acceleration through the equation $\mathbf{F} = m\mathbf{a}$, the acceleration vector has an important place in kinematics and Newtonian physics.
The acceleration vector is usually written as the sum of its tangent and normal components. This writing style is useful in many applications. But we can not say this in movements where angular momentum is conserved. In this case, it is more useful to write the acceleration vector as the sum of its tangent and radial components. The success of obtaining the acceleration vector along tangent and radial components belongs to the Italian mathematician Francesco Siacci. The acceleration vector is stated in the aforementioned form by Siacci in the study [1]. In this study performed by Siacci, the motion of the particle is restricted to the plane. Also, Siacci performed a similar study for a moving particle in space [2].
Siacci's theorem has been studied widely by many authors. Whittaker was the first person to deal with this issue after Siacci. Whittaker proved the Siacci's theorem in the plane geometrically in his work carried out in 1937 [3]. Grossman succeeded in providing a more modern proof than Whittaker's in 1996 [4]. Afterwards, Casey discussed the Siacci's theorem in space which is based on the Serret-Frenet formulas to simplify the mathematical expressions in the theorem [5]. One of the most recent studies has been carried out by Kucukarslan et al [6]. The authors expressed and proved the Siacci's theorem for the curves lying on the Finsler manifold in this study. Then, Ozen studied on the Siacci's theorem for Bishop and Type-2 Bishop frames in his master's thesis [7] under the supervision of M. Tosun (The present article is derived from this master's thesis). Also Ozen et al. [8] researched the Siacci's theorem in view of the Darboux frame for the motion of a particle along the regular surface curve. Afterwards, Ozen et al [9] discussed the Siacci's theorem in the space endowed with the modified orthogonal frame. Finally, Ozen expressed and proved the Siacci's theorem for Frenet curves in 3-dimensional Minkowski space [10].
In the theory of curves, Serret-Frenet frame is a moving frame which is very useful and has an important place. To ride along a curve and illustrate the typical properties of this curve, e.g. the curvatures is possible thanks to this frame. But this frame has a disadvantage. For the curves which have vanishing second derivatives, it is not well defined. Hence an alternative frame, that is more convenient for mathematical investigations, was required. The discovery of Bishop frame finished this requirement in 1975 [11].

[†]This article is derived from the master's thesis titled "Siacci's theorem for the first and the second Bishop frame" which was carried out by K. E. Özen under the supervision of M. Tosun at Sakarya University / Graduate School of Natural and Applied Sciences.

***Email addresses and ORCID numbers*: kahraman.ozen1@ogr.sakarya.edu.tr, 0000-0002-3299-6709 (K. E. Özen), tosun@sakarya.edu.tr, 0000-0002-4888-1412 (M. Tosun)**

This frame is well defined for every curves. As a result of this, it has been studied by a lot of researchers to deal various concepts. Today, the studies on the Bishop frame have expanded to areas such as Biology and Computer graphics. Bishop's framework is used in predicting the structural information of DNA in biology and controlling virtual cameras in the field of Computer Graphics. The readers are referred to the studies [12–18] which are related to Bishop frame.

This article is organized as follows. In Section 2, we have given a short knowledge on the fundamental concepts to ensure understanding the ensuing sections. In Section 3, for a moving particle in space, we give Siacci's theorem in terms of Bishop elements of the trajectory. Moreover, an illustrative example is given for the aforementioned theorem.

2. Preliminaries

Let us consider the 3−dimensional Euclidean space E^3 with the standard scalar product:

$$\langle \mathbf{Q}, \mathbf{R} \rangle = q_1 r_1 + q_2 r_2 + q_3 r_3, \tag{2.1}$$

where $\mathbf{Q} = (q_1, q_2, q_3)$, $\mathbf{R} = (r_1, r_2, r_3)$ are arbitrary vectors in this space. The norm of the vector $\mathbf{Q}$ is given by $\|\mathbf{Q}\| = \sqrt{\langle \mathbf{Q}, \mathbf{Q} \rangle}$. If a curve $\sigma = \sigma(s) : I \subset \mathbb{R} \to E^3$ satisfies the equality $\left\| \frac{d\sigma}{ds} \right\| = 1$ for all $s \in I$, this curve is said to be a unit speed curve and s is said to be arc-length parameter of this unit speed curve.

The moving Serret-Frenet frame of $\sigma(s)$ is showed with $\{\mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s)\}$. In this frame, the vector $\mathbf{T}(s)$ is called the unit tangent vector, the vector $\mathbf{N}(s)$ is called the unit principal normal vector. Also, the vector $\mathbf{B}(s)$ is called the unit binormal vector and it is obtained by vectorial product of $\mathbf{T}(s)$ and $\mathbf{N}(s)$. Another thing that can be of importance is that this frame satisfies the following formulas:

$$\begin{aligned} \frac{d\mathbf{T}}{ds} &= \kappa \mathbf{N} \\ \frac{d\mathbf{N}}{ds} &= -\kappa \mathbf{T} + \tau \mathbf{B} \\ \frac{d\mathbf{B}}{ds} &= -\tau \mathbf{N}, \end{aligned} \tag{2.2}$$

where $\kappa = \left\| \frac{d\mathbf{T}}{ds} \right\|$ and $\tau = -\left\langle \frac{d\mathbf{B}}{ds}, \mathbf{N}(s) \right\rangle$ represent the curvature function and the torsion function, respectively [19].

We know that the unit tangent vector $\mathbf{T}(s)$ of a given curve is determined uniquely. The Bishop frame of this given curve comprises the unique tangent vector $\mathbf{T}(s)$ and two normal vectors $\mathbf{N_1}(s)$ and $\mathbf{N_2}(s)$, that are obtained by applying the circular rotation to the vectors $\mathbf{N}(s)$ and $\mathbf{B}(s)$ in the instantaneous normal plane $\mathbf{T}(s)^{\perp}$ such that $\mathbf{N_1}'(s)$ and $\mathbf{N_2}'(s)$ are collinear with $\mathbf{T}(s)$ [11]. Consequently, we have the Bishop frame $\{\mathbf{T}, \mathbf{N_1}, \mathbf{N_2}\}$ which satisfies the derivative formulas:

$$\begin{aligned} \frac{d\mathbf{T}}{ds} &= k_1 \mathbf{N_1} + k_2 \mathbf{N_2} \\ \frac{d\mathbf{N_1}}{ds} &= -k_1 \mathbf{T} \\ \frac{d\mathbf{N_2}}{ds} &= -k_2 \mathbf{T}, \end{aligned} \tag{2.3}$$

where k_1 and k_2 indicate the Bishop curvatures. As a result of the aforementioned circular rotation, there is a relation between the Serret-Frenet frame and Bishop frame as follows:

$$\begin{aligned} \mathbf{T} &= \mathbf{T} \\ \mathbf{N_1} &= \cos\varphi\, \mathbf{N} - \sin\varphi\, \mathbf{B} \\ \mathbf{N_2} &= \sin\varphi\, \mathbf{N} + \cos\varphi\, \mathbf{B}, \end{aligned} \tag{2.4}$$

where φ represents the aforementioned rotation angle. On the other hand, the equalities

$$\begin{aligned} \varphi(s) &= \arctan \frac{k_2(s)}{k_1(s)} \\ \kappa(s) &= \sqrt{k_1{}^2(s) + k_2{}^2(s)} \\ k_1(s) &= \sqrt{k_1{}^2(s) + k_2{}^2(s)} \cos\varphi(s) \\ k_2(s) &= \sqrt{k_1{}^2(s) + k_2{}^2(s)} \sin\varphi(s) \\ \tau(s) &= \frac{d\varphi}{ds} \end{aligned} \tag{2.5}$$

hold [16, 20].

In E^3, let us assume that a particle P moves along a curve γ endowed with the Bishop frame. At time t, let us show the position vector of P relative to the origin O with $\mathbf{x}$. Denote by s the arc-length parameter of γ which is a function of the time t. Then the equality

$$\mathbf{T} = \frac{d\mathbf{x}}{ds}$$

is immediately obtained. This equality yields the velocity of P as follows [5]:

$$\begin{aligned}\mathbf{v} &= \frac{d\mathbf{x}}{dt} \\ &= \frac{d\mathbf{x}}{ds}\frac{ds}{dt} \\ &= \frac{ds}{dt}\mathbf{T}.\end{aligned}$$

Similarly above, the acceleration

$$\begin{aligned}\mathbf{a} &= \frac{d\mathbf{v}}{dt} \\ &= \frac{d}{dt}\left(\frac{ds}{dt}\mathbf{T}\right) \\ &= \frac{d}{dt}\left(\frac{ds}{dt}\right)\mathbf{T} + \frac{ds}{dt}\frac{d\mathbf{T}}{dt} \\ &= \frac{d^2s}{dt^2}\mathbf{T} + \frac{ds}{dt}\frac{d\mathbf{T}}{ds}\frac{ds}{dt} \\ &= \frac{d^2s}{dt^2}\mathbf{T} + \left(\frac{ds}{dt}\right)^2 k_1\mathbf{N_1} + \left(\frac{ds}{dt}\right)^2 k_2\mathbf{N_2}\end{aligned}$$

is found. With the help of (2.5), $\mathbf{a}$ can be written as in the following form:

$$\mathbf{a} = \frac{d^2s}{dt^2}\mathbf{T} + \sqrt{{k_1}^2 + {k_2}^2}\left(\frac{ds}{dt}\right)^2 (\cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}). \tag{2.6}$$

Then we conclude that the acceleration vector lies in the instantaneous plane $S_\mathbf{p}\{\mathbf{T}, \cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}\}$. The instantaneous vector $-\sin\varphi\,\mathbf{N_1} + \cos\varphi\,\mathbf{N_2}$ is the normal vector of this instantaneous plane and the system $\{\mathbf{T}, \cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}, -\sin\varphi\,\mathbf{N_1} + \cos\varphi\,\mathbf{N_2}\}$ is a right-handed orthonormal system [7].

3. Alternative Resolution of Acceleration Vector According to Bishop Frame

In this section, we express Siacci's theorem according to Bishop Frame and give an example for the application of this theorem (see [7] for more details). We continue to take into account of the aforementioned particle P.
Suppose that the position vector of the particle P is resolved as

$$\mathbf{x} = a\mathbf{T} - b(\cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}) + c(-\sin\varphi\,\mathbf{N_1} + \cos\varphi\,\mathbf{N_2}), \tag{3.1}$$

where

$$\begin{aligned}a &= \langle \mathbf{x}, \mathbf{T}\rangle \\ b &= \langle \mathbf{x}, -\cos\varphi\,\mathbf{N_1} - \sin\varphi\,\mathbf{N_2}\rangle \\ c &= \langle \mathbf{x}, -\sin\varphi\,\mathbf{N_1} + \cos\varphi\,\mathbf{N_2}\rangle.\end{aligned} \tag{3.2}$$

Denote by $\mathbf{r}$ the vector

$$\mathbf{r} = a\mathbf{T} - b(\cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}), \tag{3.3}$$

lying in the instantaneous plane $Sp\{\mathbf{T}, \cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}\}$. Where r symbolizes the length of $\mathbf{r}$

$$r^2 = \langle \mathbf{r}, \mathbf{r}\rangle = a^2 + b^2 \tag{3.4}$$

can be written easily (see Figure 3.1).
On the other hand, the angular momentum vector of P about the origin O is obtained as

$$\mathbf{H}^O = mc\frac{ds}{dt}(\cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}) + mb\frac{ds}{dt}(-\sin\varphi\,\mathbf{N_1} + \cos\varphi\,\mathbf{N_2}) \tag{3.5}$$

by vector product of $\mathbf{x}$ and $m\frac{ds}{dt}\mathbf{T}$.
Now we try to resolve the acceleration vector in (2.6) along the radial direction BP and tangential direction in the instantaneous plane $Sp\{\mathbf{T}, \cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}\}$. To do so, let us state the vector $\cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2}$ in terms of $\mathbf{r}$ and $\mathbf{T}$. Due to (3.3), that can be possible when $b \neq 0$. If we assume that the component of angular momentum along the vector $-\sin\varphi\,\mathbf{N_1} + \cos\varphi\,\mathbf{N_2}$ never vanishes, we can ensure that b never equals to zero. Considering this assumption, we can write the following equalities

$$\cos\varphi\,\mathbf{N_1} + \sin\varphi\,\mathbf{N_2} = \frac{1}{b}(-\mathbf{r} + a\mathbf{T}) \tag{3.6}$$

and

$$\mathbf{e_r} = \frac{1}{r}\mathbf{r}. \tag{3.7}$$

Here $\mathbf{e_r}$ indicates the unit vector in direction of $\mathbf{r}$. By means of (3.7), we get

$$\cos\varphi\,\mathbf{N_1}+\sin\varphi\,\mathbf{N_2}=\frac{1}{b}\left(-r\,\mathbf{e_r}+a\mathbf{T}\right). \tag{3.8}$$

Considering (3.8) in (2.6), we obtain the fundamental form of the acceleration vector as in the following:

$$\mathbf{a}=\left[\frac{d^2s}{dt^2}+\frac{a}{b}\sqrt{{k_1}^2+{k_2}^2}\left(\frac{ds}{dt}\right)^2\right]\mathbf{T}+\left[-\frac{r}{b}\sqrt{{k_1}^2+{k_2}^2}\left(\frac{ds}{dt}\right)^2\right]\mathbf{e_r}=S_t\,\mathbf{T}+S_r\,\mathbf{e_r}. \tag{3.9}$$

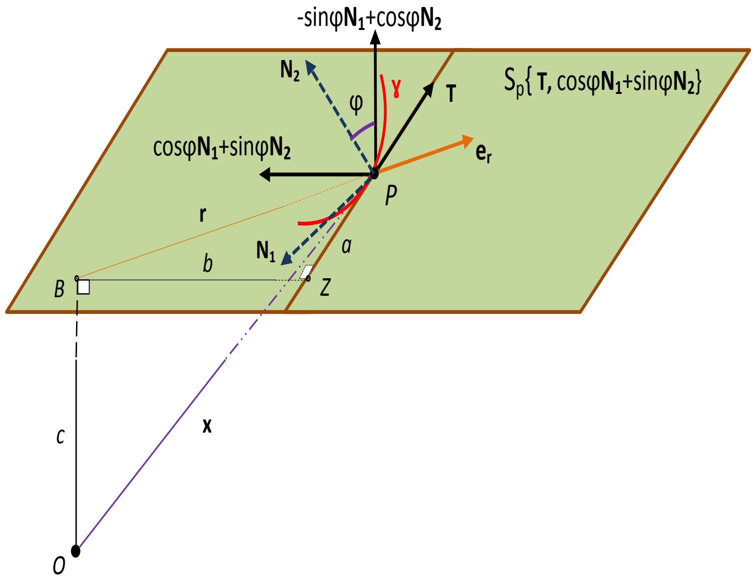

Figure 3.1: An illustration for explaining the motion of the moving particle P and the components of its acceleration vector.

Now we want to obtain the acceleration components S_t and S_r of the particle P in various forms. Because of that, we need some preparation. Let us differentiate the right and left side of (3.1) with respect to s. Then we obtain

$$\begin{aligned}\mathbf{T}=&\left(\frac{da}{ds}+k_1c\sin\varphi+k_1b\cos\varphi-k_2c\cos\varphi+k_2b\sin\varphi\right)\mathbf{T}\\&+\left(k_1a-\frac{dc}{ds}\sin\varphi-\frac{d\varphi}{ds}c\cos\varphi-\frac{db}{ds}\cos\varphi+\frac{d\varphi}{ds}b\sin\varphi\right)\mathbf{N_1}\\&+\left(k_2a+\frac{dc}{ds}\cos\varphi-\frac{d\varphi}{ds}c\sin\varphi-\frac{db}{ds}\sin\varphi-\frac{d\varphi}{ds}b\cos\varphi\right)\mathbf{N_2}.\end{aligned} \tag{3.10}$$

Because the vectors $\mathbf{T}$, $\mathbf{N_1}$ and $\mathbf{N_2}$ compose an orthonormal system,

$$\begin{aligned}1&=\frac{da}{ds}+k_1c\sin\varphi+k_1b\cos\varphi-k_2c\cos\varphi+k_2b\sin\varphi\\0&=k_1a-\frac{dc}{ds}\sin\varphi-\frac{d\varphi}{ds}c\cos\varphi-\frac{db}{ds}\cos\varphi+\frac{d\varphi}{ds}b\sin\varphi\\0&=k_2a+\frac{dc}{ds}\cos\varphi-\frac{d\varphi}{ds}c\sin\varphi-\frac{db}{ds}\sin\varphi-\frac{d\varphi}{ds}b\cos\varphi\end{aligned} \tag{3.11}$$

can be written. By keeping the equalities $k_1=\sqrt{{k_1}^2+{k_2}^2}\cos\varphi$ and $k_2=\sqrt{{k_1}^2+{k_2}^2}\sin\varphi$ in mind, we get

$$\begin{aligned}\frac{da}{ds}&=1-b\sqrt{{k_1}^2+{k_2}^2}\\\frac{db}{ds}&=a\sqrt{{k_1}^2+{k_2}^2}-c\frac{d\varphi}{ds}\\\frac{dc}{ds}&=b\frac{d\varphi}{ds}.\end{aligned} \tag{3.12}$$

If we differentiate (3.4) and use (3.12), it is not difficult to see the followings:

$$\begin{aligned}r\frac{dr}{ds}&=a-cb\frac{d\varphi}{ds}\\r\frac{dr}{ds}&=a-c\frac{dc}{ds}.\end{aligned} \tag{3.13}$$

Let us use the notation

$$h = b\frac{ds}{dt}. \tag{3.14}$$

Then, we obtain S_r as in the following form:

$$S_r = -\frac{rh^2\sqrt{{k_1}^2+{k_2}^2}}{b^3}. \tag{3.15}$$

If (3.12) and (3.14) are taken into consideration, S_t can be written as

$$S_t = \frac{1}{2b^2}\left(\frac{d}{ds}\left(h^2\right)+\frac{h^2}{b^2}\frac{d}{ds}\left(c^2\right)\right). \tag{3.16}$$

Similar to above, we can easily get

$$S_t = \frac{1}{2}\frac{d}{ds}\left(\left(\frac{ds}{dt}\right)^2\right)+\sqrt{{k_1}^2+{k_2}^2}\left(\frac{ds}{dt}\right)^2\left(\frac{1}{2b}\frac{d}{ds}\left(r^2\right)+\frac{d\varphi}{ds}c\right) \tag{3.17}$$

by using the first equality in (3.13).
Finally, it is very easy to see the following:

$$S_t = \frac{1}{2}\frac{d}{ds}\left(\left(\frac{ds}{dt}\right)^2\right)+\frac{1}{2b}\left(\frac{ds}{dt}\right)^2\sqrt{{k_1}^2+{k_2}^2}\frac{d}{ds}\left(\langle\mathbf{x},\mathbf{x}\rangle\right) \tag{3.18}$$

from the second equality in (3.13), since $\langle\mathbf{x},\mathbf{x}\rangle = a^2+b^2+c^2 = r^2+c^2$.
Consequently, if we consider the above derivation, we can state the following theorem and corollary for a particle moving along a space curve endowed with the Bishop frame.

Theorem 3.1 (Siacci's Theorem According to Bishop Frame)**.** *([7]) In E^3, let P be a particle moving on a curve γ endowed with Bishop frame. Suppose that the component of the angular momentum of P along the unit vector $-\sin\varphi\mathbf{N_1}+\cos\varphi\mathbf{N_2}$ never equals to zero. Then, the acceleration vector $\mathbf{a}$ of the particle P can be expressed as in (3.9). The component S_t, given in (3.9), lies along the tangent line of γ. The component S_r, given in (3.9), lies along the line which passes through P and the foot of the perpendicular that is from O to the instantaneous plane $Sp\{\mathbf{T},\cos\varphi\mathbf{N_1}+\sin\varphi\mathbf{N_2}\}$.*

Corollary 3.2. *([7]) S_t can be given as in (3.16), (3.17) and (3.18) except for the fundamental form, while S_r can be given as in (3.15) except for the fundamental form.*

Remark 3.3. *([7]) In Euclidean 3-space E^3, let the trajectory γ be restricted to the fixed plane $Sp\{\mathbf{T},\cos\varphi\mathbf{N_1}+\sin\varphi\mathbf{N_2}\}$ containing or not containing O. Then, it is obvious that the unit vector $-\sin\varphi\mathbf{N_1}+\cos\varphi\mathbf{N_2}$, that is the unit normal vector of this fixed plane, is constant along γ. This means that its derivative $\frac{d}{ds}(-\sin\varphi\mathbf{N_1}+\cos\varphi\mathbf{N_2})$ equals to zero for all s values of the parameter. If this derivative is calculated, one can easily conclude that $\frac{d\varphi}{ds}=0$. So, for this case, (3.16) and (3.17) reduce to*

$$S_t = \frac{1}{2b^2}\frac{d}{ds}\left(h^2\right) \tag{3.19}$$

and

$$S_t = \frac{1}{2}\frac{d}{ds}\left(\left(\frac{ds}{dt}\right)^2\right)+\frac{1}{2b}\sqrt{{k_1}^2+{k_2}^2}\left(\frac{ds}{dt}\right)^2\frac{d}{ds}\left(r^2\right), \tag{3.20}$$

respectively.

Example 3.4. *Assume that the helix curve $\delta(t) = \left(8\cos\frac{t}{17}, 8\sin\frac{t}{17}, 15\frac{t}{17}\right)$ is the trajectory of the moving particle P. Then we can easily write*

$$\mathbf{x} = \left(8\cos\frac{s}{17}, 8\sin\frac{s}{17}, 15\frac{s}{17}\right). \tag{3.21}$$

Firstly, let us note that

$$\langle\mathbf{x},\mathbf{x}\rangle = \left\langle\left(8\cos\frac{s}{17}, 8\sin\frac{s}{17}, 15\frac{s}{17}\right), \left(8\cos\frac{s}{17}, 8\sin\frac{s}{17}, 15\frac{s}{17}\right)\right\rangle = 64+\frac{225}{289}s^2.$$

By differentiating (3.21) twice with respect to time t, we get

$$\mathbf{a} = \left(-\frac{8}{289}\left(\frac{ds}{dt}\right)^2\cos\frac{s}{17}-\frac{8}{17}\frac{d^2s}{dt^2}\sin\frac{s}{17}, -\frac{8}{289}\left(\frac{ds}{dt}\right)^2\sin\frac{s}{17}+\frac{8}{17}\frac{d^2s}{dt^2}\cos\frac{s}{17}, \frac{15}{17}\frac{d^2s}{dt^2}\right).$$

Since δ is a unit speed curve, it is obvious that

$$\frac{ds}{dt} = 1$$
$$\frac{d^2s}{dt^2} = 0.$$

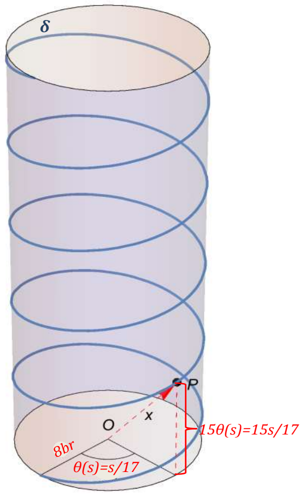

Figure 3.2: An illustration for the helix curve given in Example 3.4

On the other hand, the following equalities hold:

$$\begin{aligned} \mathbf{T}(s) &= \left(-\frac{8}{17}\sin\frac{s}{17}, \frac{8}{17}\cos\frac{s}{17}, \frac{15}{17}\right) \\ \mathbf{N}(s) &= \left(-\cos\frac{s}{17}, -\sin\frac{s}{17}, 0\right) \\ \mathbf{B}(s) &= \left(\frac{15}{17}\sin\frac{s}{17}, -\frac{15}{17}\cos\frac{s}{17}, \frac{8}{17}\right). \end{aligned} \tag{3.22}$$

From here, we obtain

$$\sqrt{{k_1}^2+{k_2}^2} = \left\|\frac{d\mathbf{T}}{ds}\right\| = \left\|\left(-\frac{8}{289}\cos\frac{s}{17}, -\frac{8}{289}\sin\frac{s}{17}, 0\right)\right\| = \frac{8}{289}.$$

By means of (2.4) *and* (3.22)*, the second and third Bishop bases are given by*

$$\mathbf{N_1} = \left(-\cos\varphi\cos\frac{s}{17} - \frac{15}{17}\sin\varphi\sin\frac{s}{17}, -\cos\varphi\sin\frac{s}{17} + \frac{15}{17}\sin\varphi\cos\frac{s}{17}, -\frac{8}{17}\sin\varphi\right) \tag{3.23}$$

and

$$\mathbf{N_2} = \left(-\sin\varphi\cos\frac{s}{17} + \frac{15}{17}\cos\varphi\sin\frac{s}{17}, -\sin\varphi\sin\frac{s}{17} - \frac{15}{17}\cos\varphi\cos\frac{s}{17}, \frac{8}{17}\cos\varphi\right). \tag{3.24}$$

Then we can write

$$\begin{aligned} a &= \langle \mathbf{x}, \mathbf{T}\rangle = \frac{225}{289}s \\ b &= \langle \mathbf{x}, -\cos\varphi\mathbf{N_1} - \sin\varphi\mathbf{N_2}\rangle = 8. \end{aligned}$$

So, we get

$$\begin{aligned} h &= 8 \\ r &= \sqrt{\frac{50625}{83521}s^2+64}. \end{aligned}$$

Substituting the obtained values of b, r, h, $\sqrt{{k_1}^2+{k_2}^2}$, $\frac{ds}{dt}$ *and* $\langle \mathbf{x}, \mathbf{x}\rangle$ *into* (3.15) *and* (3.18) *gives us the followings:*

$$S_t = \frac{225}{83521}s \tag{3.25}$$

and

$$S_r = -\frac{1}{289}\sqrt{\frac{50625}{83521}s^2+64}. \tag{3.26}$$

Finally, we must note that one can easily find the same solutions by means of the other options that are given in (3.9), (3.16) *and* (3.17).

References

[1] F. Siacci, *Moto per una linea piana*, Atti R Accad Sci. Torino, **14** (1879), 750-760.
[2] F. Siacci, *Moto per una linea gobba*, Atti R Accad Sci. Torino, **14** (1879), 946-951.
[3] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, Cambridge. Dover, New York, 1944.
[4] N. Grossman, *The Sheer Joy of Celestial Mechanics*, Birkhäuser, Basel, 1996.
[5] J. Casey, *Siacci's resolution of the acceleration vector for a space curve*, Meccanica, **46** (2011), 471-476.
[6] Z. Küçükarslan, M. Y. Yılmaz, M. Bektaş, *Siacci's theorem for curves in Finsler manifold* $\boldsymbol{F}^3$, Turkish J. Sci. Technol., **7** (2012), 181-185.
[7] K. E. Özen, *Siacci's theorem for the first and the second Bishop frame*, Master's Thesis, Sakarya University, 2015.
[8] K. E. Özen, M. Tosun, M. Akyiğit, *Siacci's theorem according to Darboux frame*, Analele Stiintifice ale Univ. Ovidius Constanta, Ser. Mat., **25** (2017), 155-165.
[9] K. E. Özen, M. Güner, M. Tosun, *A note on the acceleration and jerk in motion along a space curve*, Analele Stiintifice ale Univ. Ovidius Constanta, Ser. Mat., **28** (2020), 151-164.
[10] K. E. Özen, *Siacci's theorem for Frenet curves in Minkowski 3-space*, Math. Sci. Appl. E-Notes, **8** (2020), 159-167.
[11] R. L. Bishop, *There is more than one way to frame a curve*, Am. Math. Mon., **82** (1975), 246-251.
[12] B. Bükcü, M. K. Karacan. *The slant helices according to Bishop frame*, Int. J. Comput. Math. Sci., **3** (2009), 67-70.
[13] K. Eren.: *Geometry of coupled dispersionless equations by using Bishop frames*. In: Hvedri, I. (ed.) TBILISI-MATHEMATICS, pp 38-47. Sciendo, Berlin (2020).
[14] F. Doğan, Y. Yaylı, *On the curvatures of tubular surface with Bishop frame*, Commun. Fac. Sci. Univ. Ank. Series A1, **60** (2011), 59-69.
[15] T. Körpınar, M. T. Sarıydın, E. Turhan, *Associated curves according to Bishop frame in Euclidean 3 space*, Adv. Model. Optim., **15** (2013), 713-717.
[16] A. J. Hanson, H. Ma, *Parallel transport approach to curve framing* Indiana University, Techreports-TR425, **11** (1995), 3-7.
[17] M. Masal, A. Z. Azak, *Ruled surfaces according to Bishop frame in the Euclidean 3-space*, Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, **89** (2019), 415-424.
[18] A. Kelleci, M. Bektaş, M. Ergüt, *The Hasimoto surface according to Bishop frame*, Adıyaman Üniversitesi Fen Bilimleri Dergisi, **9** (2019), 13-22.
[19] T. Shifrin, *Differential Geometry: A First Course in Curves and Surfaces*, University of Georgia, Preliminary Version, 2008.
[20] T. Körpınar, S. Baş, *On characterization of B-Focal curves in* E^3, Bol. Soc. Paran. Mat., **31** (2013), 175-178.

Universal Journal of Mathematics and Applications, 4 (1) (2021) 33-40
Research paper

Universal Journal of Mathematics and Applications

Journal Homepage: www.dergipark.gov.tr/ujma
ISSN 2619-9653
DOI: https://doi.org/10.32323/ujma.711881

Stability Behaviour in Functional Differential Equations of the Neutral Type

Ali Fuat Yeniçerioğlu[1*], Cüneyt Yazıcı[1] and Vildan Yazıcı[2]

[1] *Kocaeli University, Faculty of Education, Kocaeli, 41380, Turkey*
[2] *Kocaeli University, Faculty of Arts and Sciences , Kocaeli, 41380, Turkey*

Article Info

Keywords: *Asymptotic behavior, Functional differential equation, Neutral, Characteristic equation, Stability*
2010 AMS: *34K06, 34K20, 34K25, 34K38, 34K40.*
Received: *30 November 2020*
Accepted: *1 March 2021*
Available online: *xxxx March 2021*

Abstract

In this study, we examine the behavior of solutions of the neutral functional differential equations. Using a suitable real root of the corresponding characteristic equation, the asymptotic behavior of the solutions and the stability of the trivial solution are explained. Three examples are also provided to illustrate our results.

1. Introduction and Preliminaries

This paper aims to describe the stability behaviour of the solutions of the neutral-type linear functional differential equations

$$\frac{d}{dt}\left[x(t)-\int_{-1}^{0}x(t-\tau(\theta))dq(\theta)\right]=\int_{-1}^{0}x(t-r(\theta))dv(\theta),\quad t\geq 0 \tag{1.1}$$

where $x(t)\in\mathbb{R}$, $r(\theta)$ and $\tau(\theta)$ are nonnegative real continuous functions in $[-1,0]$, and $v(\theta)$ and $q(\theta)$ are real functions of bounded variation in $[-1,0]$. Riemann-Stieltjes integrals are used. It is assumed that v and q are non-constant in $[-1,0]$. Consider the value $R=\max\{\|\tau\|,\|r\|\}$, where $\|\tau\|=\max\{\tau(\theta):-1\leq\theta\leq 0\}$ and $\|r\|=\max\{r(\theta):-1\leq\theta\leq 0\}$. The initial condition for (1.1) is determined by a function

$$x(t)=\phi(t),\quad -R\leq t\leq 0. \tag{1.2}$$

A *solution* of (1.1) refers to a continuous function $x:[-R,+\infty)\to\mathbb{R}$ satisfying (1.2), such that

$$x(t)-\int_{-1}^{0}x(t-\tau(\theta))dq(\theta)$$

is differentiable in $[0,+\infty)$ and satisfies (1.1) for every $t\geq 0$.

For a solution of (1.1) in the form $x(t)=e^{\lambda t}$ for $t\in\mathbb{R}$, λ represents a root of the *characteristic equation*

$$\lambda\left(1-\int_{-1}^{0}e^{-\lambda\tau(\theta)}dq(\theta)\right)=\int_{-1}^{0}e^{-\lambda r(\theta)}dv(\theta). \tag{1.3}$$

Additionally, this is applied to the relevant class of differential - difference equation

$$\frac{d}{dt}\left[x(t)-\sum_{j=1}^{m}b_jx(t-\tau_j)\right]=\sum_{j=1}^{m}a_jx(t-r_j),$$

Email addresses and ORCID numbers: **fuatyenicerioglu@kocaeli.edu.tr, 0000-0002-1063-0538, (A.F. Yeniçerioğlu), cuneyt.yazici@kocaeli.edu.tr, 0000-0002-4535-510X, (C. Yazıcı), maths.vildan@gmail.com, 0000-0001-5974-0167, (V. Yazıcı)**

where $j=1,...,m$, $a_j,b_j \in \mathbb{R}$ and $\tau_j,r_j \in (0,\infty)$. As it is well-known, this equation can be obtained from (1.1), under the assumption that $q(\theta)$ and $v(\theta)$ are step functions with a number m of jump points (See, [5], [12], [22], [26] and references therein). Furthermore, the equation (1.1) for $\tau(\theta)=r(\theta)=-r\theta \quad (r>0)$ and $\theta \in [-1,0]$ is reduced to the class of the equation

$$\frac{d}{dt}\left[x(t)-\int_{-1}^{0} x(t+\theta)dq(\theta)\right]=\int_{-1}^{0} x(t+\theta)dv(\theta), \tag{1.4}$$

where $q(\theta)=q(\theta/r)$ is atomic at zero, and $v(\theta)=v(\theta/r)$. The authors in [23] and [25] obtained behavior and stability analysis of the solutions of the equation (1.4). In this article, we examine the stability of equation (1.1), which is more general than equation (1.4). In other words, we make preference relying on equation (1.1) considering the chance to understand the impact of the delays on the stability behaviour of the neutral functional differential equations more clearly. One may look at the references [2], [7], [17], [19]- [21], [27] for a special case of equation (1.1)

$$x^{'}(t)=\int_{-1}^{0} x(t-r(\theta))dv(\theta).$$

In addition, the references [4] and [11] may also be reviewed.
Ferreira and Pedro [6] established the oscillatory criteria of the equation (1.1). However, the article in [6] has no information about asymptotic behavior and exponential estimate of solutions. In this article, we obtain the stability analysis of the solutions of the equation (1.1). Namely, we obtained the asymptotic behavior of the solutions and then we created a useful exponential estimate for these solutions and finally provided a stability criterion in this article which is different from the article in [6]. These results are obtained with a real root of the characteristic equation. For this purpose, the applied techniques are generated from a mix of methods used in the references [12], [13], [22]- [25], [27]. Examples are also given in this article.
The stability theory of the delay and the differential equations of neutral type in recent two decades has received widespread attention, as one can see through the textbooks [1], [3], [8]- [10], [14], [15], [18] and the references therein. Additionally, there are equations similar to (1.1) in the book by Kolmanovskii and Nosov [9].
Throughout this paper, $V(q)$ and $V(v)$ are denoted for the total variation function of q and v, respectively, defined in the interval $[-1,0]$. Note that the functions $V(q)$ and $V(v)$ are greater than zero in the interval $[-1,0]$. Moreover, it must be noted that $V(q)$ and $V(v)$ are not identically zero in the interval $[-1,0]$. The reader should know about both theories of the bounded variation functions and the Riemann-Stieltjes integrations. It is assumed that the reader knows the theory of Riemann-Stieltjes integration and the theory of functions of bounded variation (see [8], [16]).
Finally, in this section, we will give three well-known definitions of stability (see, for example, [9]). The trivial solution of (1.1) is defined as *stable* if for every $\varepsilon>0$, there exists a number $\ell=\ell(\varepsilon)$ such that, for any initial function ϕ with

$$\|\phi\|=\max_{-R\le t\le 0}|\phi(t)|<\ell$$

the solution x of (1.1)-(1.2) satisfies

$$|x(t)|<\varepsilon, \quad \text{for all} \quad t\in[-R,\infty).$$

In another case, the trivial solution of (1.1) is considered to be *unstable*. Provided that it is stable in the above-mentioned concept, the trivial solution of (1.1) is also considered *asymptotically stable*, and additionally, there exists a number $\ell_0>0$ such that, for any initial function ϕ with $\|\phi\|<\ell_0$, the solution x of (1.1)-(1.2) satisfies

$$\lim_{t\to\infty} x(t)=0.$$

2. Statement of the Main Results and Comments

Theorem 2.1. *We assume that $\lambda_0\in\mathbb{R}$ is a root of characteristic equation (1.3) with the property*

$$\mu(\lambda_0)=\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}(1+|\lambda_0|\tau(\theta))\,dV(q)(\theta)+\int_{-1}^{0} e^{-\lambda_0 r(\theta)}r(\theta)dV(v)(\theta)<1. \tag{2.1}$$

Set

$$\beta(\lambda_0)=\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}(\lambda_0\tau(\theta)-1)dq(\theta)+\int_{-1}^{0} e^{-\lambda_0 r(\theta)}r(\theta)dv(\theta). \tag{2.2}$$

Then, for any function $\phi\in C([-R,0],\mathbb{R})$, the solution x of (1.1)-(1.2) satisfies

$$\lim_{t\to\infty}\left[e^{-\lambda_0 t}x(t)\right]=\frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)}, \tag{2.3}$$

where

$$L(\lambda_0;\phi)=\phi(0)-\int_{-1}^{0}\left[\phi(-\tau(\theta))-\lambda_0 e^{-\lambda_0\tau(\theta)}\int_{-\tau(\theta)}^{0} e^{-\lambda_0 u}\phi(u)du\right]dq(\theta)+\int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left[\int_{-r(\theta)}^{0} e^{-\lambda_0 u}\phi(u)du\right]dv(\theta).$$

Note: Property (2.1) guarantees that $1+\beta(\lambda_0)>0$.

Proof. Property (2.1) implies $0<\mu(\lambda_0)<1$. From (2.2) we obtain

$$|\beta(\lambda_0)| = \left|\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\tau(\theta)dq(\theta) - \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}r(\theta)dv(\theta)\right|$$
$$\leq |\lambda_0|\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\tau(\theta)dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}r(\theta)dV(v)(\theta).$$

In this case, $|\beta(\lambda_0)|\leq\mu(\lambda_0)$ is satisfied, so $|\beta(\lambda_0)|<1$. Then $1+\beta(\lambda_0)>0$ is the outcome.

Let us define $y(t)=e^{-\lambda_0 t}x(t), t\in[-R,\infty)$. Then, by considering that λ_0 is a real root of characteristic equation (1.3), for every $t\geq 0$, we obtain

$$\left[x(t)-\int_{-1}^{0} x(t-\tau(\theta))dq(\theta)\right]' - \int_{-1}^{0} x(t-r(\theta))dv(\theta)$$
$$= e^{\lambda_0 t}\left\{\left[y(t)-\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta)\right]' + \lambda_0\left[y(t)-\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta)\right] - \int_{-1}^{0} e^{-\lambda_0 r(\theta)}y(t-r(\theta))dv(\theta)\right\}$$
$$= e^{\lambda_0 t}\left\{\left[y(t)-\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta)\right]' + \left[\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}dv(\theta)\right]y(t)\right.$$
$$\left. -\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta) - \int_{-1}^{0} e^{-\lambda_0 r(\theta)}y(t-r(\theta))dv(\theta)\right\}$$
$$= e^{\lambda_0 t}\left\{\left[y(t)-\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta)\right]' + \lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}[y(t)-y(t-\tau(\theta))]dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}[y(t)-y(t-r(\theta))]dv(\theta)\right\}.$$

Hence, x satisfies (1.1) for all $t\geq 0$, it follows that y satisfies

$$\left[y(t)-\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta)\right]' = -\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}[y(t)-y(t-\tau(\theta))]dq(\theta) - \int_{-1}^{0} e^{-\lambda_0 r(\theta)}[y(t)-y(t-r(\theta))]dv(\theta). \tag{2.4}$$

And then, the initial condition (1.2) becomes

$$y(t)=e^{-\lambda_0 t}\phi(t), \qquad t\in[-R,0]. \tag{2.5}$$

When equation (2.4) is integrated from 0 to t, the following equation is obtained

$$y(t) - \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta) = y(0) - \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(-\tau(\theta))dq(\theta) - \lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left(\int_{0}^{t}[y(s)-y(s-\tau(\theta))]ds\right)dq(\theta)$$
$$-\int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left(\int_{0}^{t}[y(s)-y(s-r(\theta))]ds\right)dv(\theta)$$
$$=\phi(0) - \int_{-1}^{0}\phi(-\tau(\theta))dq(\theta) - \lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left(\int_{t-\tau(\theta)}^{t} y(u)du\right)dq(\theta)$$
$$-\int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left(\int_{t-r(\theta)}^{t} y(u)du\right)dv(\theta) + \lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left(\int_{-\tau(\theta)}^{0} y(u)du\right)dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left(\int_{-r(\theta)}^{0} y(u)du\right)dv(\theta)$$
$$=-\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left(\int_{t-\tau(\theta)}^{t} y(u)du\right)dq(\theta) - \int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left(\int_{t-r(\theta)}^{t} y(u)du\right)dv(\theta) + L(\lambda_0;\phi).$$

Here, $L(\lambda_0;\phi)$ is given in Theorem 2.1. So that the following equation (2.6) is obtained

$$y(t) - \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}y(t-\tau(\theta))dq(\theta) = -\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left(\int_{t-\tau(\theta)}^{t} y(u)du\right)dq(\theta)$$
$$-\int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left(\int_{t-r(\theta)}^{t} y(u)du\right)dv(\theta) + L(\lambda_0;\phi), \quad t\geq 0. \tag{2.6}$$

This equation is equivalent to equation (1.1). Now, let us define following expression:

$$z(t)=y(t)-\frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)}, \quad t\geq -R.$$

If this definition applied to equation (2.6), we obtain

$$z(t) - \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}z(t-\tau(\theta))dq(\theta) = -\lambda_0\int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left(\int_{t-\tau(\theta)}^{t} z(u)du\right)dq(\theta)$$
$$-\int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left(\int_{t-r(\theta)}^{t} z(u)du\right)dv(\theta), \quad t\geq 0. \tag{2.7}$$

Moreover, the following expression is obtained when the initial condition (2.5) is obtained

$$z(t) = e^{-\lambda_0 t}\phi(t) - \frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)}, \quad t \in [-R,0]. \tag{2.8}$$

Here, $\beta(\lambda_0)$ is given by expression (2.2).

From the definitions of y and z, we have

$$\lim_{t\to\infty} z(t) = 0. \tag{2.9}$$

We will prove the statement (2.9) later. We define,

$$M(\lambda_0;\phi) = \max_{t\in[-R,0]} \left| e^{-\lambda_0 t}\phi(t) - \frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)} \right|. \tag{2.10}$$

In this case, from (2.8), the following expression is obtained:

$$|z(t)| \le M(\lambda_0;\phi), \quad -R \le t \le 0. \tag{2.11}$$

Now, let us show that the following inequality is satisfied in the interval $[-R,\infty)$

$$|z(t)| \le M(\lambda_0;\phi). \tag{2.12}$$

Consider an arbitrary number $\varepsilon > 0$. We claim that

$$|z(t)| < M(\lambda_0;\phi) + \varepsilon, \quad \text{for} \quad t \ge -R. \tag{2.13}$$

Let us assume that inequality (2.13) is not satisfied. In this case, because of (2.11), there exist a point $t_0 > 0$ such that

$$|z(t)| < M(\lambda_0;\phi) + \varepsilon, \quad -R \le t < t_0 \quad \text{and} \quad |z(t_0)| = M(\lambda_0;\phi) + \varepsilon.$$

Since $\mu(\lambda_0) < 1$, from equation (2.7) we obtain

$$\begin{aligned}
M(\lambda_0;\phi) + \varepsilon = |z(t_0)| &= \left| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} z(t_0-\tau(\theta))dq(\theta) - \lambda_0 \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \left(\int_{t_0-\tau(\theta)}^{t_0} z(u)du \right) dq(\theta) \right. \\
&\quad \left. - \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left(\int_{t_0-r(\theta)}^{t_0} z(u)du \right) dv(\theta) \right| \\
&\le \left| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} z(t_0-\tau(\theta))dq(\theta) \right| + |\lambda_0| \left| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \left(\int_{t_0-\tau(\theta)}^{t_0} z(u)du \right) dq(\theta) \right| \\
&\quad + \left| \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left(\int_{t_0-r(\theta)}^{t_0} z(u)du \right) dv(\theta) \right| \\
&\le \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} |z(t_0-\tau(\theta))| dV(q)(\theta) + |\lambda_0| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \left(\int_{t_0-\tau(\theta)}^{t_0} |z(u)|du \right) dV(q)(\theta) \\
&\quad + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left(\int_{t_0-r(\theta)}^{t_0} |z(u)|du \right) dV(v)(\theta) \\
&\le [M(\lambda_0;\phi)+\varepsilon] \left\{ \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} dV(q)(\theta) + |\lambda_0| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \tau(\theta) dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} r(\theta) dV(v)(\theta) \right\} \\
&= [M(\lambda_0;\phi)+\varepsilon] \left\{ \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} (1+|\lambda_0|\tau(\theta)) dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} r(\theta) dV(v)(\theta) \right\} \\
&= [M(\lambda_0;\phi)+\varepsilon]\mu(\lambda_0) < [M(\lambda_0;\phi)+\varepsilon],
\end{aligned}$$

which is a contradiction. So, inequality (2.13) must be true. Therefore, inequality (2.12) must also be true. Now, by virtue of (2.12), from (2.7) we obtain for $t \ge 0$

$$\begin{aligned}
|z(t)| &\le \left| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} z(t-\tau(\theta))dq(\theta) \right| + |\lambda_0| \left| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \left(\int_{t-\tau(\theta)}^{t} z(u)du \right) dq(\theta) \right| + \left| \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left(\int_{t-r(\theta)}^{t} z(u)du \right) dv(\theta) \right| \\
&\le \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} |z(t-\tau(\theta))| dV(q)(\theta) + |\lambda_0| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \left(\int_{t-\tau(\theta)}^{t} |z(u)|du \right) dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left(\int_{t-r(\theta)}^{t} |z(u)|du \right) dV(v)(\theta) \\
&\le M(\lambda_0;\phi) \left\{ \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} dV(q)(\theta) + |\lambda_0| \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} \tau(\theta) dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} r(\theta) dV(v)(\theta) \right\} \\
&= M(\lambda_0;\phi) \left\{ \int_{-1}^{0} e^{-\lambda_0\tau(\theta)} (1+|\lambda_0|\tau(\theta)) dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} r(\theta) dV(v)(\theta) \right\}.
\end{aligned}$$

Due to the definition of $\mu(\lambda_0)$, we have

$$|z(t)| \leq \mu(\lambda_0) M(\lambda_0;\phi), \quad \forall t \geq 0. \tag{2.14}$$

Using inequalities (2.12) and (2.14), by the induction method we can easily show

$$|z(t)| \leq (\mu(\lambda_0))^n M(\lambda_0;\phi), \quad t \geq nR - R \quad (n = 0,1,2,\ldots). \tag{2.15}$$

Here, due to $\lim_{n\to\infty}(\mu(\lambda_0))^n = 0$, from inequality (2.15) $\lim_{t\to\infty} z(t) = 0$ is obtained, that is, (2.9) is true. Hence, Theorem 2.1 is proven at all. □

A root of characteristic equation (1.3) is $\lambda = 0$ if and only if the following expressions hold:

$$\int_{-1}^{0} dv(\theta) = 0 \quad \text{and} \quad \int_{-1}^{0} dV(q)(\theta) + \int_{-1}^{0} r(\theta) dV(v)(\theta) < 1$$

or

$$v(0) = v(-1) \quad \text{and} \quad V(q)(0) - V(q)(-1) + \int_{-1}^{0} r(\theta) dV(v)(\theta) < 1. \tag{2.16}$$

So, an application of Theorem 2.1 with $\lambda = 0$ leads to the following corollary.

Corollary 2.2. *Let us satisfy the conditions of* (2.16)*. In this case, for any $\phi \in C([-R,0],\mathbb{R})$, the solution x of equation* (1.1)-(1.2) *is given as follows:*

$$\lim_{t\to\infty} x(t) = \frac{\phi(0) - \int_{-1}^{0} \phi(-\tau(\theta)) dq(\theta) + \int_{-1}^{0} \left[\int_{-r(\theta)}^{0} \phi(u) du\right] dv(\theta)}{1 - q(0) + q(-1) + \int_{-1}^{0} r(\theta) dv(\theta)}.$$

Note: Because of the second condition of (2.16), $1 - q(0) + q(-1) + \int_{-1}^{0} r(\theta) dv(\theta) > 0$ holds.

Theorem 2.3. *Let λ_0 be a real root of the characteristic equation* (1.3)*, and the condition* (2.1) *is provided for λ_0. Let us consider $\beta(\lambda_0)$ in Theorem 2.1. Then, for any $\phi \in C([-R,0],\mathbb{R})$, the solution x of* (1.1)-(1.2) *satisfies*

$$|x(t)| \leq \left[\frac{(1+\mu(\lambda_0))^2}{1+\beta(\lambda_0)} + \mu(\lambda_0)\right] N(\lambda_0;\phi) e^{\lambda_0 t}, \quad \forall t \geq 0$$

where

$$N(\lambda_0;\phi) = \max_{t\in[-R,0]} \left|e^{-\lambda_0 t}\phi(t)\right|. \tag{2.17}$$

Moreover, the trivial solution of equation (1.1) *is stable if $\lambda = 0$, asymptotically stable if $\lambda_0 < 0$ and unstable if $\lambda_0 > 0$.*

Proof. Let y and z be defined as in the proof of Theorem 2.1, i.e.

$$z(t) = y(t) - \frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)}, \quad t \geq -R,$$

where $L(\lambda_0;\phi))$ and $M(\lambda_0;\phi)$ are defined as in Theorem 2.1. Then, we can express the following for $t \geq 0$

$$y(t) \leq \frac{|L(\lambda_0;\phi)|}{1+\beta(\lambda_0)} + \mu(\lambda_0) M(\lambda_0;\phi). \tag{2.18}$$

From the definition of $L(\lambda_0;\phi)$ we get

$$\begin{aligned}
L(\lambda_0;\phi) =& \phi(0) - \int_{-1}^{0} \left[e^{\lambda_0\tau(\theta)}\phi(-\tau(\theta)) - \lambda_0 \int_{-\tau(\theta)}^{0} e^{-\lambda_0 u}\phi(u) du\right] e^{-\lambda_0\tau(\theta)} dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left[\int_{-r(\theta)}^{0} e^{-\lambda_0 u}\phi(u) du\right] dv(\theta) \\
\leq& |\phi(0)| + \int_{-1}^{0} \left[\left|e^{\lambda_0\tau(\theta)}\phi(-\tau(\theta))\right| + |\lambda_0| \int_{-\tau(\theta)}^{0} \left|e^{-\lambda_0 u}\phi(u)\right| du\right] e^{-\lambda_0\tau(\theta)} dV(q)(\theta) \\
&+ \int_{-1}^{0} e^{-\lambda_0 r(\theta)} \left[\int_{-r(\theta)}^{0} \left|e^{-\lambda_0 u}\phi(u)\right| du\right] dV(v)(\theta) \\
\leq& \left\{1 + \int_{-1}^{0} \left[1 + |\lambda_0|\tau(\theta)\right] e^{-\lambda_0\tau(\theta)} dV(q)(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)} r(\theta) dV(v)(\theta)\right\} N(\lambda_0;\phi) \\
=& (1+\mu(\lambda_0)) N(\lambda_0;\phi).
\end{aligned}$$

Furthermore, by the definition of $\mu(\lambda_0;\phi)$, we obtain

$$\begin{aligned}
M(\lambda_0;\phi) =& \max_{t\in[-R,0]} \left|e^{-\lambda_0 t}\phi(t) - \frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)}\right| \leq \max_{t\in[-R,0]} \left|e^{-\lambda_0 t}\phi(t)\right| + \frac{|L(\lambda_0;\phi)|}{1+\beta(\lambda_0)} \\
=& N(\lambda_0;\phi) + \frac{|L(\lambda_0;\phi)|}{1+\beta(\lambda_0)} \leq N(\lambda_0;\phi) + \frac{(1+\mu(\lambda_0)) N(\lambda_0;\phi)}{1+\beta(\lambda_0)} \\
=& \left(1 + \frac{1+\mu(\lambda_0)}{1+\beta(\lambda_0)}\right) N(\lambda_0;\phi).
\end{aligned}$$

So, from (2.18) we get

$$\begin{aligned}|y(t)| \leq & \frac{(1+\mu(\lambda_0))N(\lambda_0;\phi)}{1+\beta(\lambda_0)} + \mu(\lambda_0)\left(1+\frac{1+\mu(\lambda_0)}{1+\beta(\lambda_0)}\right)N(\lambda_0;\phi) \\ = & \left[\frac{(1+\mu(\lambda_0))^2}{1+\beta(\lambda_0)} + \mu(\lambda_0)\right]N(\lambda_0;\phi).\end{aligned}$$

Finally, by the definition of y, we obtain

$$|x(t)| \leq \left[\frac{(1+\mu(\lambda_0))^2}{1+\beta(\lambda_0)} + \mu(\lambda_0)\right]N(\lambda_0;\phi)e^{\lambda_0 t} \tag{2.19}$$

for all $t \geq 0$. So, the proof of the first part of the theorem is completed. Now let us show the stability criterion of the theorem.

Assume that $\lambda_0 \leq 0$, and let $\phi \in ([-R,0],\mathbb{R})$ be an arbitrary initial function. Then, from inequality (2.19) we get

$$|x(t)| \leq \left[\frac{(1+\mu(\lambda_0))^2}{1+\beta(\lambda_0)} + \mu(\lambda_0)\right]N(\lambda_0;\phi), \quad \forall t \geq 0.$$

Because $\frac{(1+\mu(\lambda_0))^2}{1+\beta(\lambda_0)} > 1$, we obtain

$$|x(t)| \leq \left[\frac{(1+\mu(\lambda_0))^2}{1+\beta(\lambda_0)} + \mu(\lambda_0)\right]N(\lambda_0;\phi), \quad \forall t \geq -R.$$

So, the trivial solution of equation (1.1) is stable. For $\lambda_0 < 0$, it is clear that, from inequality (2.19) it follows that

$$\lim_{t\to\infty} x(t) = 0.$$

Thus, the trivial solution is asymptotically stable. Finally, let $\lambda_0 > 0$. We want to show that this solution is unstable. Let us assume that it is stable. Then, there is a number $\delta > 0$ such that, for each $\phi \in C([-R,0],\mathbb{R})$ with $\|\phi\| < \delta$, the solution x of (1.1)-(1.2) satisfies

$$|x(t)| < 1, \quad t \geq -R.$$

Given the following,

$$\phi_0(t) = e^{\lambda_0 t}, \quad t \in [-R,0].$$

We see that $\phi_0 \in C([-R,0],\mathbb{R})$. From the definition of $L(\lambda_0;\phi)$ we obtain

$$\begin{aligned}L(\lambda_0;\phi_0) \equiv & \phi_0(0) - \int_{-1}^{0}\left[\phi_0(-\tau(\theta)) - \lambda_0 e^{-\lambda_0\tau(\theta)}\int_{-\tau(\theta)}^{0} e^{-\lambda_0 u}\phi_0(u)du\right]dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left[\int_{-r(\theta)}^{0} e^{-\lambda_0 u}\phi_0(u)du\right]dv(\theta) \\ = & 1 - \int_{-1}^{0}\left[e^{-\lambda_0\tau(\theta)} - \lambda_0 e^{-\lambda_0\tau(\theta)}\int_{-\tau(\theta)}^{0} e^{-\lambda_0 u}e^{\lambda_0 u}du\right]dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}\left[\int_{-r(\theta)}^{0} e^{-\lambda_0 u}e^{\lambda_0 u}du\right]dv(\theta) \\ = & 1 - \int_{-1}^{0} e^{-\lambda_0\tau(\theta)}\left[1-\lambda_0\tau(\theta)\right]dq(\theta) + \int_{-1}^{0} e^{-\lambda_0 r(\theta)}r(\theta)dv(\theta) \\ = & 1+\beta(\lambda_0) > 0.\end{aligned}$$

Now, we select a number δ_0 with $0 < \delta_0 < \delta$ and let

$$\phi = \frac{\delta_0}{\|\phi_0\|}\phi_0,$$

where it is clear that $\phi \in C([-R,0],\mathbb{R})$ and $\|\phi\| = \delta_0 < \delta$. So, we obtain

$$\lim_{t\to\infty}\left[e^{-\lambda_0 t}x(t)\right] = \frac{L(\lambda_0;\phi)}{1+\beta(\lambda_0)} = \frac{(\delta_0/\|\phi_0\|)L(\lambda_0;\phi_0)}{1+\beta(\lambda_0)} = \frac{\delta_0}{\|\phi_0\|} > 0.$$

However, due to $\lambda_0 > 0$, we get

$$\lim_{t\to\infty}\left[e^{-\lambda_0 t}x(t)\right] = 0.$$

We reach a contradiction, and hence Theorem 2.3 is proven. □

3. Examples

Example 3.1. Let $\tau(\theta)=-\frac{\theta}{4}$, $r(\theta)=\frac{\theta^2}{2}$, $q(\theta)=-\frac{\theta^2}{5}$ and $v(\theta)=\frac{3\theta^2}{5}$. In this case, the characteristic equation (1.3) is obtained as follows:

$$\lambda\left(1-\int_{-1}^{0} exp\Big(\frac{\lambda\theta}{4}\Big)d\Big(-\frac{\theta^2}{5}\Big)\right)=\int_{-1}^{0} exp\Big(-\frac{\lambda\theta^2}{2}\Big)d\Big(\frac{3\theta^2}{5}\Big). \tag{3.1}$$

It is seen that $\lambda_0\cong -1,03$ is a root of equation (3.1). Here, this root is one of the roots of (3.1) that we obtain with the MATLAB program. Thus, the condition of Theorem 2.3 is satisfied by using the root $\lambda_0=-1,03$. Namely, because the functions q and v are monotonous in $[-1,0]$, from expression (2.1) we obtain

$$\begin{aligned}\mu(-1,03)&=\int_{-1}^{0} e^{-\frac{1,03\theta}{4}}\left(1+|-1,03|\Big(-\frac{\theta}{4}\Big)\right)dV\Big(-\frac{\theta^2}{5}\Big)+\int_{-1}^{0} e^{\frac{1,03\theta^2}{2}}\left(\frac{\theta^2}{2}\right)dV\Big(\frac{3\theta^2}{5}\Big)\\&\le \max_{-1\le\theta\le0}\left|e^{-\frac{1,03\theta}{4}}\left(1-\frac{1,03\theta}{4}\right)\right|\left|-\frac{1}{5}\right|+\max_{-1\le\theta\le0}\left|e^{\frac{1,03\theta^2}{2}}\left(\frac{\theta^2}{2}\right)\right|\left|\frac{3}{5}\right|\\&=\frac{1}{5}e^{\frac{1,03}{4}}\left(1+\frac{1,03}{4}\right)+\frac{3}{5}\frac{e^{\frac{1,03}{2}}}{2}\cong 0,83<1.\end{aligned}$$

Since $\lambda_0=-1,03<0$, the solution $x(t)\equiv 0$ is asymptotically stable.

Example 3.2. Let $\tau(\theta)=\theta^2$, $r(\theta)=-\frac{\theta}{2}$, $q(\theta)=-\frac{\theta^2}{4}$ and $v(\theta)=4\theta^3$. In this case, the characteristic equation (1.3) is obtained as follows:

$$\lambda\left(1-\int_{-1}^{0} exp(-\lambda\theta^2)d\Big(-\frac{\theta^2}{4}\Big)\right)=\int_{-1}^{0} exp\Big(\frac{\lambda\theta}{2}\Big)d(4\theta^3). \tag{3.2}$$

We see $\lambda_0\cong 2,08745$ is a root of equation (3.2). Here, this root is one of the roots of (3.2) that we obtain with the MATLAB program. Thus, the condition of Theorem 2.3 is satisfied for the root. Namely, because the functions q and v are monotonous in $[-1,0]$, from (2.1) we obtain

$$\begin{aligned}\mu(2,08745)&=\int_{-1}^{0} e^{-2,08745\theta^2}\left(1+2,08745\theta^2\right)dV\Big(-\frac{\theta^2}{4}\Big)+\int_{-1}^{0} e^{\frac{2,08745\theta}{2}}\Big(-\frac{\theta}{2}\Big)dV\left(4\theta^3\right)\\&\le \max_{-1\le\theta\le0}\left|e^{-2,08745\theta^2}\left(1+2,08745\theta^2\right)\right|\frac{1}{4}+\max_{-1\le\theta\le0}\left|e^{\frac{2,08745\theta}{2}}\Big(-\frac{\theta}{2}\Big)\right|4\\&=e^0(1+0)\frac{1}{4}+\frac{e^{-\frac{2,08745}{2}}}{2}4=0,954<1.\end{aligned}$$

Since $\lambda_0>0$, the solution $x(t)\equiv 0$ is unstable.

Example 3.3. Let $\tau(\theta)=-2\theta$, $r(\theta)=-\frac{\theta}{2}$, $v(\theta)=\frac{\theta^2+\theta}{2}$ and $q(\theta)=\frac{\theta}{4}$. In this case, the characteristic equation (1.3) is obtained as follows:

$$\lambda\left(1-\int_{-1}^{0} exp(2\lambda\theta)d\Big(\frac{\theta}{4}\Big)\right)=\int_{-1}^{0} exp\Big(\frac{\lambda\theta}{2}\Big)d\left(\frac{\theta^2+\theta}{2}\right). \tag{3.3}$$

We see $\lambda_0=0$ is a root of equation (3.3). The following expressions is satisfied since the function $v(\theta)$ is decreasing in the interval $[-1,-\frac{1}{2}]$ and increasing in the interval $[-\frac{1}{2},0]$, whereas $q(\theta)$ is also increasing in the interval $[-1,0]$:

$$\begin{aligned}\mu(0)&=\int_{-1}^{0} dV\left(\frac{\theta}{4}\right)+\int_{-1}^{0}\Big(-\frac{\theta}{2}\Big)dV\left(\frac{\theta^2+\theta}{2}\right)\\&=\int_{-1}^{0}\frac{1}{4}dV(\theta)+\int_{-1}^{-\frac{1}{2}}\Big(-\frac{\theta}{2}\Big)dV\left(\frac{\theta^2+\theta}{2}\right)+\int_{-\frac{1}{2}}^{0}\Big(-\frac{\theta}{2}\Big)dV\left(\frac{\theta^2+\theta}{2}\right)\\&\le\frac{1}{4}+\max_{-1\le\theta\le0}\left\{\Big(-\frac{\theta}{2}\Big)\right\}\frac{1}{2}\left[\int_{-1}^{-\frac{1}{2}}dV\left(\theta^2+\theta\right)+\int_{-\frac{1}{2}}^{0}dV\left(\theta^2+\theta\right)\right]\\&=\frac{1}{4}+\frac{1}{4}\left(\left|-\frac{1}{4}\right|+\left|\frac{1}{4}\right|\right)=\frac{1}{4}+\frac{1}{4}\frac{1}{2}=\frac{3}{8}\\&<1.\end{aligned}$$

Hence, the solution $x(t)\equiv 0$ is stable for the root $\lambda_0=0$.

4. Conclusions

In this study, firstly, a basic asymptotic result for the solution of the equation (1.1) is proved. Secondly, we obtained a useful exponential boundary for solutions and the stability of trivial solutions were shown. These results were obtained using a suitable real root for the characteristic equation. Namely, this real root played an important role in establishing the results of the article. Finally, three examples were given for stability.

References

[1] T.A. Burton, *Volterra integral and differential equations*, Academic Press, New York, 1983.
[2] K.L. Cooke, J.M. Ferreira, *Stability conditions for linear retarded functional differential equations*, J. Math. Anal. Appl., **96** (1983), 480–504.
[3] C. Corduneanu, *Integral equations and applications*, Cambridge University Press, New York, 1991.
[4] J.M. Ferreira, I. Györİ, *Oscillatory behavior in linear retarded functional differential equations*, J. Math. Anal. Appl., **128** (1987), 332–346.
[5] J.M. Ferreira, A.M. Pedro, *Oscillations of differential-difference systems of neutral type*, J. Math. Anal. Appl., **253** (2001), 274–289.
[6] J.M. Ferreira, A.M. Pedro, *Oscillatory behaviour in functional differential systems of neutral type*, J. Math. Anal. Appl., **269** (2002), 533–555.
[7] N.J. Ford, Y. Yan, Md.A. Malique, *Numerical treatment of oscillatory functional differential equations*, J. of Comput. and Appl. Math., **234** (2010), 2757–2767.
[8] J.K. Hale, S.M. Verduyn Lunel, *Introduction to functional differential equations*, Springer, Berlin, Heidelberg, New York, 1993.
[9] V.B. Kolmanovskii, V.R. Nosov, *Stability of functional differential equations*, Academic Press, Inc., London, 1986.
[10] V.B. Kolmanovskii, A. Myshkis, *Applied theory of functional differential equations*, Kluver Academic, Dordrecht, 1992.
[11] Q. Kong, *Oscillation for systems of functional differential equations*, J. Math. Anal. Appl., **198** (1996), 608–619.
[12] I.G. E. Kordonis, N. T. Niyianni, Ch. G. Philos, *On the behavior of the solutions of scalar first order linear autonomous neutral delay differential equations*, Arch. Math. (Basel), **71** (1998), 454–464.
[13] I.G.E. Kordonis, Ch.G. Philos, *The behavior of solutions of linear integro-differential equations with unbounded delay*, Comput. Math. Appl., **38** (1999), 45–50.
[14] Y. Kuang, *Delay differential equations with applications in population dynamics*, Academic Press, San Diego, 1993.
[15] V. Lakshmikantham, L. Wen, B. Zhang, *Theory of differential equations with unbounded delay*, Kluwer Academic Publishers, London, 1994.
[16] K. Leffler, The Riemann-Stieltjes integral and some applications in complex analysis and probability theory. Umea University, Institutionen för matematik och matematisk statistik, 2014.
[17] Md. A. Malique, *Numerical treatment of oscillatory delay and mixed functional differential equations arising in modeling*, PhD thesis, University of Liverpool (University of Chester), 2012.
[18] S.I. Niculescu, *Delay effects on stability*, Springer-Verlag London Limited, 2001.
[19] A.M. Pedro, *Oscillation and nonoscillation criteria for retarded functional differential equations*, in: International Conference on Differential Equations (Proc. Internat. Sympos., Czecho-Slovak series, Bratislava, 2005), Comenius University Press, Bratislava, (2007), 353–362.
[20] A.M. Pedro, *Nonoscillation criteria for retarded functional differential systems*, Inter. J. of Pure and Appl. Math., **31** (2006), 47–71.
[21] A.M. Pedro, *Oscillation criteria for retarded functional differential systems*, Inter. J. of Pure and Appl. Math., **63** (2010), 75–84.
[22] Ch.G. Philos, I.K. Purnaras, *Periodic first order linear neutral delay differential equations*, Appl. Math. Comput., **117** (2001), 203–222.
[23] Ch.G. Philos, I.K. Purnaras, *Asymptotic properties, nonoscillation, and stability for scalar first order linear autonomous neutral delay differential equations*, Electron. J. Differential Equations, **03** (2004), 1–17.
[24] Ch.G. Philos, I.K. Purnaras, *More on the behavior of solutions to linear integro differantial equations with unbounded delay*, Funkcial. Ekvac., **48** (2005), 393–414.
[25] Ch.G. Philos, I.K. Purnaras, *A result on the behavior of the solutions for scalar first order linear autonomous neutral delay differential equations*, Math. Proc. Camb. Phil. Soc., **140** (2006), 349–358.
[26] U. Stroinski, *Order and oscillation in delay differential systems*, J. Math. Anal. Appl., **207** (1997), 158–171.
[27] A.F. Yeniçerioglu, C. Yazıcı, *Stability criteria for retarded functional differential equations*. Eskişehir Technical Univ. J. of Sci. and Tech. B – Theo.Sci., **8**(2), (2020), 212 – 222

Universal Journal of Mathematics and Applications, 4 (1) (2021) 41-49
Research paper

UJMA

Universal Journal of Mathematics and Applications

Journal Homepage: www.dergipark.org.tr/en/pub/ujma
ISSN 2619-9653
DOI: https://doi.org/10.32323/ujma.738463

Boolean Hypercubes: The Origin of a Tagged Recursive Logic and the Limits of Artificial Intelligence†

Ramon Carbó-Dorca[1]

[1]*Institut de Química Computacional i Catalisi Universitat de Girona, Girona 17005, Catalonia, Spain*

†*This paper is dedicated to the memory of Professor Mario Bunge (1919-2020), who enlightened the mind of many, in such a manner that the bright beam of his thought will never fade away.*

Article Info

Keywords: *Fuzzy sets, Fuzzy logic, Tagged sets, Boolean tags, Logical tags, Boolean tagged sets, Boolean hypercubes, Concatenation of Boolean hypercubes, Natural number recursion, Mersenne numbers, Multivalued logic, Gödel like properties of Boolean sets, Artificial intelligence.*
2010 AMS: *03-11, 68T01.*
Received: *16 May 2020*
Accepted: *12 March 2021*
Available online: *23 March 2021*

Abstract

Boolean and logical hypercubes are discussed as providers of tags to logical object sets, transforming them into logical tagged sets, a generalization of fuzzy sets. The equivalence of Boolean and logical sets permits to consider natural tags as an equivalent basis of logical tagged sets. Boolean hypercube concatenation easily allows studying how Boolean information is transmitted. From there a Gödel-like behavior of Boolean hypercubes and thus of logical object sets can be unveiled. Later, it is discussed the iterative building of natural numbers, considering Mersenne numbers as upper bounds of this kind of recursive construction. From there information acquisition, recursive logic, and artificial intelligence are also examined.

1. Introduction

This paper can be considered as another application of the versatile structure of Boolean hypercubes, which has been steadily developing in this laboratory since the first paper on a generalized form of fuzzy sets [1]. Reference [2] can be thought of as an application example of Boolean vertex structures to molecular similarity and design. Since these two works had been published, many developments in mathematical structures definition and applications in chemistry, biology, and physics were produced, as a condensed list see references [3]- [11]. On the other hand, the final prospection of a recent Bunge's book [13], has motivated the background where the present study is inspired and built. Bunge presented in an appendix of the previous reference an extremely interesting prospect to formulate the structure of logic.
The present paper will be developed using Boolean hypercubes as a tool but following essentially Bunge's directives. In this sense, the following working organization will be used. First, the basic ideas about fuzzy sets will be detailed. Then, the concept of Boolean tagged sets will be established. Boolean hypercubes and their transcription into natural numbers will be described next. Concatenation of Boolean hypercubes and the possibility to design a recursive construction of natural numbers will follow. Next, Mersenne numbers will be shown that play a leading role in this recursive construction. In a new section, a simple distance definition will permit in the continuing development of the paper to classify the truth and falsehood contents of available logical tags described at some Boolean hypercube dimension level. The Gödel-like structure of Boolean hypercubes tags associated with their concatenation will be later discussed. Finally, an analysis of the limits of artificial intelligence will close this presentation.

2. Binary Functions Over Object Sets: Basic Fuzzy Sets

One can classically define in general a set containing some sort of objects, an object set Ω. Then according to the ideas of fuzzy set definition by Zadeh [14], see also for extended information on fuzzy sets the reference [15], one can also describe a function f, the so-called membership function, having as domain the elements of such an object set. When applying the function to any object belonging to Ω, one can suppose that if such a function yields a binary outcome, then the resulting value of the function codomain may be associated with a

Email address and ORCID number: **ramoncarbodorca@gmail.com, 0000-0002-9219-0686, (R. Carbó-Dorca)**

Boolean pair of digits: $B = \{0,1\}$, which alternatively can be also taken as a logical pair: $L = \{F,T\}$, of false and true, values.
Initially, this scheme can be formally written as a classical set theory definition, providing information about some objects and some set, through:

$$\forall \varpi \notin \Omega \to f(\varpi) = 0 \wedge \forall \omega \in \Omega \to f(\omega) = 1. \tag{2.1}$$

So, one can use the Boolean digits $\mathbf{B}$ as the function codomain leading to the primary definition of a fuzzy set [1], [14] or substitute them with the logical values $\mathbf{L}$ and thus leading to a fuzzy logical set.
This previous fuzzy set definition is equivalent to use the results of the application of the membership function f over the objects' domain as the vertices of a one-dimensional Boolean hypercube: $\mathbf{H}_1 = \{(0),(1)\}$, where the parenthesis means one can take the two binary digits as one-dimensional Boolean vectors, the two vertices of the one-dimensional Boolean hypercube.
To attach fuzziness to the membership function codomain, one might proceed one step further and suppose that some objects cannot be considered as sharply belonging to the set Ω or as clearly being in the outside of the object set, but in kind of no-man's-land between the outside and inside of the object set. In this case, the simplest way is to use as function codomain the elements of a two-dimensional Boolean hypercube as follows:

$$\mathbf{H}_2 = \{(0,0),(0,1),(1,0),(1,1)\} = \{\langle\mathbf{h}_0|;\langle\mathbf{h}_1|;\langle\mathbf{h}_2|;\langle\mathbf{h}_3|\}. \tag{2.2}$$

A membership function outcome like the zero vertex: $\langle\mathbf{h}_0|$ might mean that the domain object does not belong to the set. However, any vertex of the pair: $\{\langle\mathbf{h}_1|;\langle\mathbf{h}_2|\}$ obtained as a membership function value within the codomain, will represent a fuzzy answer to the question about the object belongs or not to the object set. Finally, the outcome $\langle\mathbf{h}_3|$ can be interpreted as positively belonging to the object set. The whole number of possibilities can be described using as the object set a fuzzy set Φ instead defined as:

$$\begin{aligned}
&\forall\omega \in \Phi : f(\omega) = \langle\mathbf{h}_3|,\\
&\forall\varpi \notin \Phi : f(\varpi) = \langle\mathbf{h}_0|,\\
&\exists\varphi : f(\varphi) = \langle\mathbf{h}_2| \to \varphi\tilde{\in}\Phi,\\
&\exists\phi : f(\phi) = \langle\mathbf{h}_1| \to \phi\tilde{\notin}\Phi.
\end{aligned}$$

where the symbols $\tilde{\in}$ and $\tilde{\notin}$ mean an incomplete knowledge of the objects: $\{\varphi;\phi\}$ about belonging or not to the set Φ.
The use of the four vertices of the two-dimensional Boolean hypercube in the equation 2.2 as membership function outcome codomain permits, besides the positive and negative belonging issues, to add two nuances to the fact that the objects are located or not into the set.
A simple fuzzy set Φ can be defined in this manner, constituting an example of the most basic one, a part of the monodimensional case defined in the equation 2.1, which can be defined based on Boolean hypercubes.

3. Basic Boolean Tagged Sets

Nevertheless, the possibility of defining a fuzzy logic structure along with an object set can be easily generalized. Precisely taking as start up the two-dimensional Boolean hypercube as a platform to generalize fuzzy logical sets. From this initial scaffold, then using the properties of Boolean hypercubes of higher dimensions it can be in general created a fuzzy logical framework of arbitrary dimensions.
At the same time, to avoid the fuzzy set idea of defining a membership function, whose variable values in its domain are objects of some set, one can be aware of the fact that the same idea of fuzziness attached to an object set might be also generalized, using instead of the membership function the definition of a tagged set, see for instance references [1], [2].
In a primitive way, similar to the previous fuzzy set description, to the objects of some set one can connect tags, which for this study can be essentially characterized as the elements of the Boolean hypercube as defined in the equation 2.2.
Suppose an object set Ω and the two-dimensional Boolean hypercube $\mathbf{H}_2$ as defined in the equation 2.2. A Cartesian product can be described between the elements of the object set and the two-dimensional Boolean hypercube vertices: $T = \Omega \times \mathbf{H}_2$, such that the elements of the tagged set $\mathbf{T}$ can be defined by the ordered pairs:

$$\forall t \in \mathbf{T} : t = (\omega;\langle h_I|) \leftarrow \omega \in \Omega \wedge \langle h_I| \in \mathbf{H}_2. \tag{3.1}$$

The elements of the tagged set $\mathbf{T}$ permit to consider that every element ω of the object set Ω can be associated with a tag $\{\langle h_I\|I = 0,1,2,3\}$ which is one of the Boolean vertices belonging to $\mathbf{H}_2$. The elements of the object set can be identified with the vertices of the Boolean hypercube and thus classified into four kinds, coincident with the fuzzy set structure already discussed.
The object tags might be also replaced by the four natural numbers which coincide with the digital transcriptions of the four Boolean vertices of $\mathbf{H}_2$. In one way or another, the four tags will classify in the same manner the elements of the object set. Of course, instead of Boolean tags, constructed by the set $B = \{0;1\}$ of bits, when looking at the construction of a logic system, one can use the logical values contained in the set: $\mathbf{L} = \{F,T\}$. Then, four logical classes associated with the object set can be also easily described. That is, the logical equivalent of the two-dimensional Boolean hypercube in the equation 2.2 can be written like:

$$\mathbf{L}_2 = \{(F,F);(F,T);(T,F);(T,T)\} = \{\langle\mathbf{l}_0|;\langle\mathbf{l}_1|;\langle\mathbf{l}_2|;\langle\mathbf{l}_3|\}$$

and every logical vertex is used as a logical tag.
From now on, the Boolean framework will be used as a collection of Boolean vertex sets, which can be subjected to a trivial transformation, leading to a logical framework simply corresponding to a substitution of the Boolean bits: (0)bit–(1)bit, by logical (F)alse-(T)rue values.

4. Boolean Hypercubes

Because everything from now on will be based on the Boolean hypercubes structure, the building of such mathematical elements will be succinctly discussed next.
Any $N-$ dimensional Boolean hypercube $\mathbf{H}_N$ possess 2^N vertices, which can be expressed as binary row vectors or Boolean strings. [1]

$$\begin{aligned}&\mathbf{H}_N = \{\langle \mathbf{h}_I \| I = 0, 2^N - 1\}\\&\rightarrow \langle \mathbf{h}_I| = (\beta_{I,N}; \beta_{I,N-1}; ...\beta_{I,1}) \leftarrow \forall J = 1, N : \beta_{I,J} \in \{0,1\}.\end{aligned} \tag{4.1}$$

This notation obeys the fact that the binary strings $\langle \mathbf{h}_I|$ are supposed here to bear the most significant bits on the leftmost side.

5. Transcription of a Boolean Hypercube into a Natural Number Sequence

Every Boolean vertex corresponds to a decimal number within a well-defined natural number sequence. Constituting a relationship which can be defined as:

$$\forall I = 0, (2^N - 1) : \delta(\langle \mathbf{h}_I|) = \sum_{J=1}^{N} \beta_{IJ} 2^{(J-1)} = \upsilon_I \in \mathbf{S}_N \subset \mathbb{N} \tag{5.1}$$

where using as variables the Boolean hypercube vertices, then the multivariate Boolean function $\delta(\langle \mathbf{h}_I|) \in \mathbb{N}$ yields a natural number, including zero.
Thus, the natural number set $\mathbf{S}_N$ in equation 5.1 contains a sequence of 2^N natural numbers:

$$\mathbf{S}_N = \{0, 1, 2, ..., (2^N - 1)\} \tag{5.2}$$

where the last value corresponds to the so-called Mersenne number, which can be represented by:

$$\mu_2(N) = 2^N - 1.$$

6. Equivalence of Boolean and Logical Hypercube Vertices with a Natural Number Sequence

One can see without a problem that the elements of the natural sequence $\mathbf{S}_N$, described in the equation 5.2, can be used as a set of tags as such, accompanying any object set element. While the tagged set in the equation 3.1 can be described as a Boolean tagged set, substituting the Boolean hypercube vertices by elements of the natural number sequence of equation 5.2 results in a natural tagged set.
Natural tagged sets are even more identifiable with fuzzy sets because one can easily define the connection between the values of some membership function as elements of some natural number sequence, like $\mathbf{S}_N$, which can instead be employed as tags in an equivalent tagged set.
However, the connection of Boolean-logical hypercubes and natural numbers can be seen as an iterative way to build the hypercube vertices and the equivalent natural elements, but subject to the binary increase of elements both in $\mathbf{H}_N$ and in $\mathbf{S}_N$.
The natural elements of the sequence $\mathbf{S}_N$ are equivalent to the vertices of some Boolean hypercube $\mathbf{H}_N$. It has also been seen that these Boolean vertices can be transformed into logical ones. Thus, all the possible Boolean tags are transformed into the vertices of a logical hypercube: $\mathbf{L}_N$ and those into a natural number set: $\mathbf{S}_N$.
Next, one can admit that natural sequences, Boolean and logical hypercubes are the same structure described in three equivalent forms and therefore interchangeable.
In this manner the natural number sequence $\mathbf{S}_N$ in equation 5.2 might be well used as a set of tags instead of a set of logical hypercube vertices. Fuzzy logical sets and natural tagged sets are the same mathematical construct.

7. Specific and Complementary Vertices of a Boolean Hypercube

To deepen in the previously commented equivalence of the three possible kinds of tagged sets, the specific nature of some Boolean vertices will be presented now.
In any Boolean hypercube there are two well-defined vertices: The zero vertex:

$$\langle \mathbf{h}_0| \equiv \langle \mathbf{0}| = (0, 0, 0, ..., 0)$$

and the unity or Mersenne vertex:

$$\langle \mathbf{h}_{(2^N-1)}| \equiv \langle \mathbf{1}_N| = (1, 1, 1, ..., 1)$$

which for any dimension the unity vector corresponds to the binary representation of the corresponding so-called Mersenne number:

$$\delta(\langle \mathbf{1}_N|) = \mu_2(N).$$

Another well-defined set of vertices corresponds to a set containing N vertices, which can be called the canonical basis set. Such a specific Boolean vertex set may be constructed associated with the vertices, which starting from the zero vector, then one 0-bit at a time is substituted by a 1-bit only. The canonical basis set is made by the Boolean vertices, which also correspond to the decimal sequence of the powers of two:

$$\mathbf{K}_N = \{\langle \mathbf{h}_{2^0}|; \langle \mathbf{h}_{2^1}|; \langle \mathbf{h}_{2^2}|; ...; \langle \mathbf{h}_{2^{N-1}}|\} \Rightarrow \delta(\mathbf{K}_N) = \{2^0; 2^1; 2^2; ...; 2^{N-1}\}.$$

[1] In the present paper, the Dirac notation is used to represent vectors. Thus, an $N-$dimensional row vector is written as: $\langle \mathbf{a}| = (a_1; a_2; ,,,; a_N)$ and its transpose which corresponds to a column vector is represented as: $|\mathbf{a}\rangle$.

The canonical basis set vertices can also be expressed with the usual symbols employed to describe the canonical basis set of any vector space:

$$\mathbf{E}_N = \{\langle \mathbf{e}_1|;\langle \mathbf{e}_2|;\langle \mathbf{e}_3|;\ldots;\langle \mathbf{e}_N|\} \rightarrow$$
$$\{|\mathbf{e}_1\rangle;|\mathbf{e}_2\rangle;|\mathbf{e}_3\rangle;\ldots;|\mathbf{e}_N\rangle\} \equiv \mathbf{I}_N = \{I_{IJ} = \delta_{IJ}|I,J = 1,N\}$$

which when considering them as columns and their zero-one elements as natural numbers, the final structure is the $(N \times N)$ unit matrix $\mathbf{I}_N = \{I_{IJ} = \delta_{IJ}\}$, being δ_{IJ} a Kronecker delta. However, turning again into the Boolean meaning side provides 1-bits in the whole diagonal and 0-bits elsewhere. The unity vertex is the complete sum of the canonical basis vertices:

$$\langle \mathbf{1}| = \sum_{I=1}^{N} \langle \mathbf{e}_I| \rightarrow \mu_2(N) = \sum_{I=0}^{N-1} 2^I = 2^N - 1.$$

Every Boolean hypercube vertex has a complementary vertex, where the 0-bits are substituted by 1-bits and vice versa, every 1-bit is substituted by a 0-bit. In particular, a conversion of this kind occurs when the zero and unity vector both are considered as complementary vertices.

The canonical basis vectors possess a complementary basis set made by substitution one of a time of the 1-bit of the unity vector by a 0-bit. As the canonical basis set can be arranged to construct the unit matrix, the complementary canonical basis set can be arranged into a matrix $\mathbf{C}_N$, defined with the aid of the unity matrix: $1_N = \{1_{IJ} = 1\}$, a matrix with all the elements equal to the 1-bit, minus the unit matrix, that is: $\mathbf{C}_N = \mathbf{1}_N - \mathbf{I}_N = \{C_{IJ} = \delta(I \neq J)\}$, where $\delta(I \neq J)$ is a logical Kronecker delta [2], which in this case yields 0-bits at the diagonal and 1-bits at every one of the off-diagonal elements.

8. Concatenation of Boolean Hypercubes

Moreover, Boolean hypercubes can be concatenated [9]. The symbol $\cup$ will be employed to signal the operation of vertex concatenation. For instance, one can simply write the concatenation:

$$\mathbf{H}_2 = \mathbf{H}_1 \cup \mathbf{H}_1 \rightarrow \mathbf{H}_2 = \{(0 \quad 0),(0 \quad 1),(1 \quad 0),(1 \quad 1)\} \tag{8.1}$$

which can be interpreted as concatenating the vertices of the right Boolean hypercube by each vertex of the one on the left. If necessary, one can construct higher dimensional Boolean hypercubes, see reference [9] for more details, just concatenating two or more Boolean hypercubes of lower-dimensional structures, for instance:

$$\mathbf{H}_{M+N} = \mathbf{H}_M \cup \mathbf{H}_N.$$

Following this hypercube building up, concatenation of the mono-dimensional Boolean hypercube $\mathbf{H}_1$ with any arbitrary $N-$ dimensional one: $\mathbf{H}_N$ produces:

$$\mathbf{H}_{N+1} = \mathbf{H}_1 \cup \mathbf{H}_N \tag{8.2}$$

meaning that the resultant Boolean hypercube $\mathbf{H}_{N+1}$ possesses a double number of vertices than the original Boolean hypercube $\mathbf{H}_N$. This is so because with the concatenation 8.2 the original Boolean hypercube vertices, as defined in the equation 5.1, transform into two kinds of vertices belonging to $\mathbf{H}_{N+1}$:

$$\begin{aligned} \forall I &= 0,2^N : \langle \mathbf{h}_{N;I}| \in \mathbf{H}_N : \\ &\rightarrow (0) \cup \langle \mathbf{h}_{N;I}| = \langle \mathbf{h}_{(N+1);I}| = \langle 0;\langle \mathbf{h}_{N;I}|| \in \mathbf{A}_{N+1} \\ &\rightarrow (1) \cup \langle \mathbf{h}_{N;I}| = \langle \mathbf{h}_{(N+1);(2^N+I)}| = \langle 1;\langle \mathbf{h}_{N;I}|| \in \mathbf{B}_{N+1} \end{aligned}$$

with the additional property, involving the two newly defined sets:

$$\mathbf{H}_{N+1} = \mathbf{A}_{N+1} \cup \mathbf{B}_{N+1} \wedge \mathbf{A}_{N+1} \cap \mathbf{B}_{N+1} = \varnothing.$$

Thus, in such a concatenation operation two new Boolean vertex sets are well-defined. The set $\mathbf{A}_{N+1}$ is equivalently made by the old Boolean vertices of $\mathbf{H}_N$, considering that they are one dimension larger, but possessing the most significative bit as the 0-bit; while the set $\mathbf{B}_{N+1}$ contains all the new vertices which conform to the Boolean hypercube $\mathbf{A}_{N+1}$ and cannot belong to the initial $\mathbf{H}_N$.

9. Recursive Construction of Boolean Hypercubes and Natural Numbers

As explained above, this construction indicates that starting from any Boolean hypercube one can obtain another one with the dimension augmented in one unit. An in a deep application can be found in reference [10].

Half of the vertices in $\mathbf{H}_{N+1}$ are those of $\mathbf{H}_N$, because whenever one writes the decimal representations of the involved vertices it is obtained: $\delta(\mathbf{H}_N) = \mathbf{S}_N$, but also it can be written: $\delta(\mathbf{A}_{N+1}) = \mathbf{S}_N$. The Boolean vertex set $\mathbf{B}_{N+1}$ corresponds to a decimal representation which corresponds to the very new nature of the extended dimension and can be easily written as:

$$\delta(\mathbf{B}_{N+1}) = 2^N \oplus \mathbf{S}_N = \{2^N;(2^N+1);(2^N+2);\ldots;(2^{N+1}-1)\} \tag{9.1}$$

meaning that the power of two constant 2^N is summed to every element of the natural set $\mathbf{S}_N$. The final element of this sequence 9.1 is the Mersenne number $\mu_2(N+1)$. Accordingly, the natural number set associated with $\mathbf{H}_{N+1}$ can be expressed as:

$$\mathbf{S}_{N+1} = \delta(\mathbf{A}_{N+1}) \cup \delta(\mathbf{B}_{N+1}) = \mathbf{S}_N \cup (2^N \oplus \mathbf{S}_N). \tag{9.2}$$

[2]A logical Kronecker delta is a symbol $\delta(E)$ where E is an expression, which if $E = .True.$ returns $\delta(T) = 1$ and if $E = .False.$, then $\delta(F) = 0$.

A result meaning that natural numbers can be constructed within a recursive framework, associated in turn with the building of Boolean hypercubes via an iterative concatenation as defined in the equation 8.2.
Moreover, in the equation, 9.2 the structure of the recursion involving the decimal representation of Boolean hypercubes suggests one can define a recursion operator R, such that:

$$\mathbf{S}_{N+1} = R[\mathbf{S}_N] = \mathbf{S}_N \cup (2^N \oplus \mathbf{S}_N).$$

Consequently, fuzzy logical arbitrarily complex structures can be iteratively constructed in the same way.

9.1. Mersenne Numbers and Their Twins

According to the equations 9.1 and 9.2, the lower and upper limits of the new natural numbers attached to a Boolean hypercube digital transcription, are respectively the Mersenne number and a power of 2, both associated with the appropriate dimension of the hypercube.
Furthermore, it is interesting to note the role of the Mersenne number: $\mu_2(N) = 2^N - 1$, contained in a natural number sequence similar to the equation 5.2, and another number contained as a second element of the sequence of the equation 9.1 $\upsilon_2(N) = 2^N + 1$, which is a member of a general sequence, which can be called Mersenne twins [10]. While any Mersenne number is translated into binary form as a 1-bit string of dimension N, a unity vector: $\langle \mathbf{1}_N| = (1,1,1,...,1,1)$, the Mersenne twins are well-defined bit strings too, being nearly a zero vector but having two 1-bits, one at the most significative left bit and the other at the rightmost position:$\langle \mathbf{v}_{N+1}| = (1,0,0,...,0,1)$. Between any Mersenne number and his twin there is the vertex corresponding to the power 2^N , which obviously enough corresponds to an elements of the canonical basis $\langle \mathbf{e}_N|$.

9.2. Complex Logical Object Sets

Coming back to a logical object set Ω, one can choose some function, which applied on an object yields a Boolean hypercube vertex, that is:

$$\forall \omega \in \Omega : f(\omega) = \langle \mathbf{h}_I| \in \mathbf{H}_N$$

or similarly as done in the equation 3.1, one can associate to each object belonging to the set Ω a Boolean hypercube vertex, forming a Boolean tagged set. That is, first forming a tagged Set using the vertices of a Boolean hypercube as tags:

$$\mathbf{T_H} = \Omega \times \mathbf{H}_N \rightarrow \forall t \in \mathbf{T_H} : t = (\omega; \langle \mathbf{h}_I|) \Leftarrow \omega \in \Omega \wedge \langle \mathbf{h}_I| \in \mathbf{H}_N. \tag{9.3}$$

Alternatively, one can use as tags the elements of the decimal transcription $\mathbf{S}_N$ of a given Boolean hypercube $\mathbf{H}_N$:

$$\mathbf{T_S} = \Omega \times \mathbf{S}_N \rightarrow \forall t \in \mathbf{T_S} : t = (\omega; s_I) \Leftarrow \omega \in \Omega \wedge s_I \in \mathbf{S}_N = \delta(\mathbf{H}_N). \tag{9.4}$$

In equations 9.3 and 9.4, it is supposed that $\mathbf{H}_N$ and what is the same, the natural transcription $\mathbf{S}_N$, contain sufficient information on the elements of Ω, which in turn can be considered in particular as logical objects, to leave almost all sensible knowledge covered up to some extent and experience, whose measure can be easily associated to the dimension N. Of course both Boolean hypercubes and the decimal transcript forming natural number sequences are equivalent to the logical hypercubes $\mathbf{L}_N$, whose vertices are made by logical elements. One can write the implication of interchange among the three descriptions of the same concept:

$$\mathbf{L}_N \Leftrightarrow \mathbf{H}_N \Leftrightarrow \mathbf{S}_N.$$

10. The Distance of a Boolean Hypercube Vertex from the $\langle \mathbf{0}|$ and $\langle \mathbf{1}|$ Vectors

While the extreme vectors $\langle \mathbf{0}|$ and $\langle \mathbf{1}|$ represent the most far away vertices of any Boolean hypercube $\mathbf{H}_N$, representing the natural numbers 0 and $(2^N - 1)$ respectively, the rest of Boolean-logical vertices can be also located concerning these extremal vertices belonging to any hypercube.
Boolean vertices can be classified in a complementary manner by the number of 0-bits or 1-bits they have, but this just describes some trivial classification of Boolean vertices. There are in a Boolean hypercube $\mathbf{H}_N$, $(N+1)$ vertex classes, which also can be complementary to another set of classes with the same number of elements. Depending on that, each class corresponds to having: $\{0,1,2,...,(N-1),N\}$ 1-bits, or in a complementary way, the vertex classes might be contemplated as possessing: $\{N,(N-1),...,2,1,0\}$ 0-bits.
Although there are several ways to measure the similarity between two-bit strings, see for instance references [17]- [22] as an assorted example, the distance of a precisely given Boolean vertex say $\langle \mathbf{v}_I|$ to both extreme vectors can be also used. Such a measure might be obtained first via its decimal transcription: $\delta[\langle \mathbf{v}_I|] = d_I$.
Once obtained the decimal transcription of a Boolean vertex, then two Minkowski-like distances: $D_0\big|\langle \mathbf{v}_I| - \langle \mathbf{0}|\big|$ and $D_1\big|\langle \mathbf{v}_I| - \langle \mathbf{1}|\big|$ can be constructed, taking directly the decimal transcription d_I for the first distance to the zero vertex, and using the absolute difference: $|d_I - (2^N - 1)| \equiv |d_I + 1 - 2^N| \equiv |2^N - (d_I + 1)|$ for the second one to the unity vertex.
Thus, every vertex could be associated with a pair of natural numbers: $(D_0; D_1)$, which will locate it nearest one or another of the extremal vertices or make it equidistant to both.
For example, one could be interested to locate from the zero and unity vertices respectively, in the way explained above, the 5-dimensional Boolean hypercube vertex: $\langle \mathbf{h}| = (1,0,1,0,1) \rightarrow \delta[\langle \mathbf{h}|] = 21$. Thus, the distance to the zero vector is just 21, while the distance to the Mersenne vertex, which in this case has a decimal value: $\mu_2(5) = 2^5 - 1 = 31$ is 10. One can conclude that this specific vertex example is nearby to the Mersenne vertex than to the zero vertex. Better, one can represent this situation with a two-dimensional vector, in this specific case: $(21,10)$. From the logical point of view, if the zero vertex represents absolute falsehood and the unity vertex absolute trueness, the vertex of the above example can be classified as nearer truth than from untruth.

10.1. Double distances and the falsehood-trueness content of logical objects

Such double distance classification, providing for every vertex a two-dimensional non-negative definite coordinate element, as explained above, might be interesting when using Boolean hypercube vertices as sequences of false and true associations, because even if every vertex will correspond to a non-trivial sequence of false and true elements, the double distances to the extremal hypercube vertices can locate an associated complicated logical object as being lying next to complete falsehood, represented by the zero vertex $\langle \mathbf{0}|$, or sitting nearby to the absolute truth, represented by the unity vector $\langle \mathbf{1}|$.
Another short example will provide the already mentioned possible use of the double distance to associate Boolean hypercube vertices to the extremal vertices for logical purposes. In the equation 8.1, the vertices of a two-dimensional Boolean hypercube are given. The natural number sequence associated with every one of these vertices is: $\mathbf{S}_2 = \{0,1,2,3\}$. Then the two middle vertices double distances are simply obtained as: $(1,2)$ and $(2,1)$ respectively. The meaning of this resultant pattern corresponds to associate the vertex $(0,1)$ nearby to the zero vector, while the vertex $(1,0)$, which is complementary to the former one, is located closer to the Mersenne vertex.
The above simple case, connected with the two-dimensional Boolean hypercube of the equation 8.1, can be also seen from a multidimensional logical point of view and to the three-valued logic system [23]. After substituting the 0-bit by a logical $F-$value and the 1-bit by a logical $T-$value, then immediately one can consider that the two vertices (F,T) and (T,F) can be classified as undecidable logical objects, from the point of view of absolute falsity-trueness values. However, using the defined double distance technique, the first vertex can be thought located nearby to the absolute falsehood vector (F,F), while the second one, could be seen lying nearest to the absolute truth vertex (T,T). This interpretation means that even undecidable logical objects, corresponding to a fuzzy or tagged logical frame, might be studied possessing extra logical nuances, allowing them to consider holding truth and false contents in different amounts when compared with respect Aristotelian complete falsehood and trueness vertices.

11. Gödel-Like Properties of Boolean Tagged Sets

The recursive structure of the natural numbers as shown in the previous pages, essentially in section 9, permit us to observe how every time a Boolean hypercube concatenates with a mono-dimensional one, then, as a result, the number of vertices or elements in the decimal transcription, is duplicated. What is the same, the tag set construct duplicates itself its elements every time a simple concatenation of this kind is performed, as it is also shown in the equation 9.2.
That is: one-bit concatenation is enough to duplicate the amount of information, which one can suppose is ready to be attached as Boolean or logical vertices or natural numbers tags to the elements of an object set. In case this extra information is not needed; even so, the new tags, which can be generated in this one-bit way, can be used to enrich the available information with minimal effort.
However, if the new extended tags are used, then the possibility appears to create this augmented information once again and thus new extended tags can be added to the initial tag set.
No end could be envisaged to stop this iterative process, which will reappear along every time one might need more information to be added as Boolean or logical vertex tags to an object set. Such characteristics shared by both Boolean and logical basis sets and the attached natural numbers can be taken as a Gödel-like theorem [34] property, a characteristic which was already discussed in reference [9]. With this, it is meant the intrinsic incompleteness and unprovability of any logical structure which can be based on a Boolean or logical basic description. One can advance that in general almost all, if not all, human knowledge might be translated into a Boolean-logical framework. Therefore, any human knowledge construct is incomplete and unprovable.
Such a recursive procedure illustrates the possibility to assemble a simple path leading to an increase in the information knowledge about a known object set, and the impossibility that any (in the present case: logical) theory becomes complete. A simple 1-bit reconstructs a completely new set of logical information structures, equal in number to the initial ones, which are also preserved with the addition of a complementary single 0-bit.
Describing this issue again in short: a unique 1-bit concatenated to any initial set of Boolean-logical tags can completely transform the tagged set information contents, while the initial original tags are preserved via a 0-bit concatenation.
Therefore, no Boolean-logical information about some object set can be considered final and complete, but always lying one step back, until the use of a 1-bit concatenation changes everything, doubling the potentially available information.
Of course, the same incompleteness occurs in the set of decimal transcription of a given Boolean-logical hypercube. That is, in the information duplication: $\mathbf{H}_N \rightarrow \mathbf{H}_{N+1}$, there is also a duplication of the sequence of the natural number transcription, where: $\mathbf{S}_N \rightarrow \mathbf{S}_{N+1}$ implies that: $\mathbf{S}_{N+1} = \mathbf{S}_N \cup 2^N \oplus \mathbf{S}_N$. Meaning, as before, that to every element of $\mathbf{S}_N$ the power: 2^N is added. Evidently, because of this information doubling, a similar expansion appears associated with the tags or fuzzy logical values in logical object sets.

12. The Limits of Artificial Intelligence

The issue of fuzzy Boolean tagged sets might not only be associated with a Gödel-like impossibility to get rid of incompleteness in any kind of theoretical or experimental structure, whenever such structure can be transformed into bit strings. It has immediate consequences in handling Artificial Intelligence (AI) devices (wherein this denomination also machine learning procedures are included to simplify the reasoning of the present discussion) constructed for any purpose and at any complexity level, the development of logical systems included. Something similar has been commented on machine learning not long ago [30].
What it seems not commented at all, as far as the present author knows, is the nature of any AI procedure, by extension of any computational procedure, which is or can be running in some computer, from small desktop to supercomputers or clusters of them. They undoubtedly run as programs written in any available programming language.
Any application program constructed for any purpose is, sorry for the redundancy, programmed by one or several programmers. Because of this, a program cannot perform anything else *outside* of the programmed code provided by the programmer(s). That is: not included by the programmer(s) within the code. This is not the *only* limit of any given computing procedure. In old and modern computers results are in need to be reproducible, and the digital machines are thus lacking the freedom to provide different results than these expected to be yielded by a given program.

That is, every time that the same input is fed up *with* a program, no matter how complex can be, the *same* output will be obtained. In this sense, programs running on digital computers are predictable. Certainly, one can simulate randomness through adequate programming techniques, but these techniques need a random seed, which repeats the pseudo-random series every time the same seed is introduced as input. In brief, computers' lack of freedom of will, is *deterministic*.

No one can predict if the announced quantum computers will continue to be fully deterministic or for some purposes can be programmed to act with some output freedom of choice. Certainly, for engineering purposes, say, even quantum computers must behave deterministically, otherwise, it will become impossible to compute any object with reasonable confidence.

However, with some easy thought, one can describe fuzzy Boolean hypercubes, which can possess even the ability to change the information content, depending on some parameter(s) which can act as a time-like dependence or evolution, see reference [35]. That is, with the aid of this kind of hypercubes not only the fuzziness of the involved (pseudo-)Boolean tags, which can be chosen as the hypercube vertices, can be present and attached to any kind of logical objects, but one can even envisage their time-like evolution depending on ad hoc parameter(s).

Such previous statements appear to be obvious if one analyzes the essence of an AI procedure. One can consider that they are based on a complex mathematical scheme, built on computational programming grounds, and previously trained by a finite object subset to become operative. An AI procedure is hopefully trained, so it achieves the stage of being able to recognize and handle new objects, which have been not previously used in the training or learning period stage.

It seems that the forest of methods where AI is grounded; among them logic, statistical learning, and classifiers, image recognition, ... could also be essentially based on a large variety of the so-called Neural Networks, see for an excellent resume reference [24].

A brief description example of the use and characteristics of AI within the neural network context will be done within the interesting background of medicinal chemistry. The use of neural networks in drug design is quite old, see for instance reference [25]. However, modern AI users in this field can be quite preposterous sometimes, offering publications that intend to astonish the naive readers. A good example of the idea of obtaining exhaustivity within the field of drug design can be well represented by the article of reference [26]. There, the authors describe an AI way to obtain molecular information.

Pretending to have produced a sample of all molecular structures bearing potential biological activity [3]. In the not very far years there is a large amount of paper, which in one way or another use AI to the aim of molecular design, it is too short space to describe and quote all of them. However, some examples are worth comment. Another at least not so pretending set of studies on the same AI application field was recently published [27], this time various authors were offering a more reasonable point of view to the public, though. But a preposterous and overwhelming very recent contribution, precludes a not particularly good perspective in the field of AI application to molecular engineering [28] via theoretically dubious docking techniques. The same can be told about another pair AI recently published papers, which claim to have solved old chemical problems. The first describes a solution of the protein folding problem (see the web site [30] for more information and references), while the basic background relies on a classical parametric structure of the computations. The second claims to provide a black box Schrödinger equation solution [31] of any quantum chemical problem but based on a Monte Carlo procedure to solve the wave function equation, but using simple ideas to construct the wave functions, as atomic centered basis sets. One can wonder, among several technical questions if a pseudo random procedure will appear reliable enough for large molecular systems.

This kind and other multiple applications of AI algorithms, in case one agrees that these processes (usually acting as black boxes) can be called in this way, try to simulate as much as possible a brain activity knowledge process. Similarly, as brain most of the development does, these procedures need a training initialization, a specific information gathering as a starting point for every purpose of AI deployment.

The resultant computational trained structure, according to the AI users, must be hopefully able to build new information concerning objects not used in the training period.

Now, one must be aware that in the present discussion the AI structure sits into a computer in the form of a program. Thus, a discussion about the limits of AI must consider such a program background, and the boundaries of such a situation are to be linked to the proper AI limits.

These points above are coupled to the fact that new information is generated from the initial one, which has been provided during the training step. Then, additionally, some important questions arise.

For instance, how pretended AI efficiency can happen? Whenever the AI system, must consist of a program written by some operator, and thus cannot generate new information within the program itself if left intact. Unless such an issue is provided from an external extra training step or by a program modifying source: included in the program or feed on it in some moment of the analysis?

Is any new source of information pre-programmed and based on the knowledge provided by every new object feed to the procedure, analyzed by the trained algorithm, and finally used to modify the initial information? If so, the AI will act on every new analysis as being in another stage of the training step. Therefore, entering a never-ending loop, taking for granted the intrinsic Gödel-like incompleteness of any information system [30].

Alternatively, the AI answer on a new object analysis is just given by an as sophisticated as one can imagine kind of interpolation (or extrapolation, which constitutes a shady workout, prone to failure) among the information of the training objects used?

After these preliminary discussions on the nature of computing and AI, one can also consider now the problem of causality, which is dubious that any AI-based procedure can include, at least with the actual level of programmed AI systems. Nobody can say how these issues will evolve with a new breed of quantum computers, which even might be able (difficult to think how this be reached, though) to escape in some circumstances the deterministic behavior of digital computers. To peruse an excellent discussion on the nature of the important causality topic by Bunge, reference [33] is recommended.

The acausal character of AI appears to be another issue and can be stored on the same shelf where statistical correlation remains. It is well-known that such a statistical relationship does not imply a causal connection between the variables involved. No doubt about this, because there is a shared set of similarities between statistical correlation and AI. Both might use a training set, and both provide acausal relationships on new objects and some parameters, both present Gödel-like incompleteness.

Perhaps the AI acausal background is a consequence of the fact that the information content of the related procedures also bears the paradox of apparently creating new information, starting from an a priori established set of computational structures. This formal appearance is certainly shared by AI with the well-known set of statistical correlation procedures.

Such a problem is not the only one that can be found in AI procedures. As it has been previously commented, unprovability is another one,

[3]The title of the quoted paper might be taken as representative of the authors' attitude: "Stochastic Voyages into Unchartered Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds". One can say that, not only the apparently complete advertised results seem a bit preposterous, considering of the Gödel-like properties of information manipulation, but even the wording at the beginning of the title statement has been duly criticized some years ago [28].

see reference [30] to have some information about this problem too, which can be linked to one of Gödel's theorems [34], and thus also to the discussion contained in the previous section.
Even more, another similar problem that also can be found within AI procedures will be discussed now. When setting up initial information, any AI might be considered that produces the equivalent of some bit string, provided that any information or logical string is constructed, bearing a dimension, say, as large as the nature of the problem requires.
Such an initial setup cannot provide new appearances of the same kind of bit strings showing out of nowhere, except information already contained into the original bit string appears. What could be conceivable to consider is the possibility to obtain Boolean vertex transformations belonging to the Boolean-logical hypercube containing the initial bit string. This might be allowed as a posterior step of information manipulation because, given a bit string of dimension N, the remaining $2^N - 1$ vertex components of the hypercube holding them are well-known and can be easily chosen as vertices of some Boolean hypercube $\mathbf{H}_N$ for any purposes.
That is, such information changes shall be found within the set of the initial well-defined dimension of the Boolean-logical bit structures, where belongs to the gathered training information.
To better visualize this situation, suppose that the AI training set has been able to construct information contained into some $N-$dimensional Boolean hypercube: $\mathbf{H}_N$. In principle, no limit must hinder the dimension N. Thus, theoretically one can suppose that the information generated after the training period might be contained in some Boolean vertex:$\langle \mathbf{h}_I | \in \mathbf{H}_N$.
What one can easily grasp is that without adding at all even a new 1-bit of information to the trained system represented by $\langle \mathbf{h}_I |$, any answers the AI procedure can provide are compelled to be located among the Mersenne number $\mu_2(N)$ of the remaining vertices of the original Boolean-logical hypercube, where the Boolean vertex representing the trained AI belongs.
It has already been previously discussed how by addition of one bit, the amount of information of any Boolean hypercube duplicates. If one bit is added to the initial Boolean-logical AI structure, then the training period might be invalidated or transformed into another completely different result.
Therefore, no new information content must be created by an AI post-training manipulation out of the vertices belonging to $\mathbf{H}_N$. The AI resulting outcome must be just the consequence of a shift from a Boolean-logical vertex to another, that is: $\langle \mathbf{h}_I | \rightarrow \langle \mathbf{h}_J |$ or any other manipulation among the known vertices of $\mathbf{H}_N$.
Such a program could be modified if and only if, with the analysis of a new object, the AI answer and the information of the newly tested object could be absorbed into the old pool of training objects. Then, the resulting algorithm issued from the training set period results in being necessarily modified. Subsequently, one could face the transformation of the original Boolean-logical hypercube $\mathbf{H}_N$ into another one of larger dimensions, say: $\mathbf{H}_N \rightarrow \mathbf{H}_{N+M}$.
In case this possibility appears to be feasible (an insecure question considering all possible problems to be treated), without the intervention of an external operator cooperating with the whole AI process, then it might appear the same as a digital machine can program itself. Nowadays, such self-programming ability still is a bit difficult to admit, without another program taking care of the initial one and doing the reprogramming job, say automatically. But such an exploit within a program's ability must be made with another program made by an external operator beforehand and somehow included in the whole procedure. Further reprogramming will initiate a never-ending process.
Hence, unless an external programmer re-programs the code or simpler: the AI algorithm acting with a re-programming issue, might re-educate itself accordingly.
An AI algorithm once trained means that one has fed the training stage with objects, described in some manner, and their attached tags. Once the AI is trained, the desideratum is such that the learned AI system in front of a new object, described in the same fashion as the ones previously used in the training period, could be able to generate part of some new unknown object tag.
If this previously unknown tag is not contained in the AI training phase, then that will be the same as new information has appeared from nowhere. In the case it is so, this freshly new information must be feed again as part of the training source. And the AI process will enter an unending loop or become corrupt at the end. The only possibility to avoid the appearance of such a loop is that the information generated because of the new object input corresponds to an interpolation of the information already known by the trained AI.
Therefore, it seems a bit out of question pretending that, after a correct training phase, a computational AI structure can generate by itself, even partially, new Boolean-logical structures, not contained in the Boolean-logical pool, the $N-$dimensional hypercube $\mathbf{H}_N$, organized by the initial training object set.
Nevertheless, even if some kind of procedure, generating new information not contained in the old experience, could be successfully implemented with the aid of new computing machines, and possessing additional self-programming abilities, one must be again aware that a unique bit, added to any old information structure, can duplicate the information content of the whole affair. This might revert an AI process into a never-ending learning procedure and drive the whole computational structure into a combinatorial information explosion.
For the moment, the best one can say about this problem is that any AI result appears forcibly deterministic, incomplete, and acausal. In agreement with the author of reference [30], one might learn and accept to become a bit humbler about the abilities of AI procedures.

13. Conclusion

A succinct revision of the fuzzy set definition, extending this mathematical structure into the tagged set construction, permits with the aid of Boolean hypercubes to create a general formalism involving Boolean vertices, fuzzy logical vertices, and natural number tags. This point of view results in an equivalent formalism, which permits to study of falsity-verity contents of logical object sets from three equivalent kinds of points of view, represented by the three kinds of different tags available. Because of the option to construct Boolean hypercubes recursively, then logical objects tags can be generated also recursively. This possibility causes the equivalent of Gödel's unprovability and incompleteness in fuzzy multivalued logical theories. Finally, such a result also permits the description of some limitations appearing in the AI procedures, which can be involved in the development of fuzzy logical systems based on computational grounds.

Conflict of Interest

The author declares no conflict of interest.

References

[1] R. Carbó-Dorca, *Fuzzy sets and boolean tagged sets*, J. Math. Chem., **22**(1997), 143-147.
[2] R. Carbó-Dorca,*Fuzzy sets and Boolean tagged sets, vector semispaces and convex sets, QSM and ASA density functions, diagonal vector spaces and quantum chemistry*, Adv. Molec. Simil., **2**(1998), 43-72.
[3] R. Carbó-Dorca, *N-dimensional Boolean hypercubes and the Goldbach conjecture*, J. Math. Chem., **54**(2016), 1213-1220.
[4] R. Carbó-Dorca, *Natural vector spaces, (inward power and Minkowski norm of a natural vector, natural Boolean hypercubes) and Fermat's last theorem*, J. Math. Chem., **55**(2017), 914-940.
[5] R. Carbó-Dorca, *Boolean hypercubes as time representation holders*, J. Math. Chem., **56**(2018), 1349-1352.
[6] R. Carbó-Dorca, Boolean hypercubes and the structure of vector spaces, J. Math. Sci. Model., **1**(2018), 1-14.
[7] R. Carbó-Dorca, *Transformation of Boolean hypercube vertices into unit interval elements: QSPR workout consequences*, J., Math. Chem., **57**(2019), 694-696.
[8] R. Carbó-Dorca, *Role of the structure of Boolean hypercubes when used as vectors in natural (Boolean) vector semispaces*, J. Math. Chem., **57**(2019), 697-700.
[9] R. Carbó-Dorca, *Cantor-like infinity sequences and Gödel-like incompleteness revealed by means of Mersenne infinite dimensional Boolean hypercube concatenation*, J. Math. Chem., **58**(2020), 1-5.
[10] R. Carbó-Dorca, *Boolean Hypercubes, Mersenne numbers and the Collatz conjecture* Research Gate Preprint DOI:10.13140/RG.2.2.28323.40482.
[11] R. Carbó-Dorca, T. Chakraborty, J. Comp. Chem., **40**(2019), 2653-2653.
[12] R. Carbó-Dorca, J. Math. Chem., **56**(2018), 1353-1356.
[13] M. Bunge, *Matter and Mind. A Philosophical Inquiry*, Boston Studies in the Philosophy of Science, Vol 287, Springer (Dordrecht) (2010).
[14] L.A. Zadeh, Information and Control, **8**(1965), 338-353.
[15] https://en.wikipedia.org/wiki/Fuzzy_set
[16] E. A. J. Sloane(Editor), *Online Encyclopedia of Integer Sequences*, https://oeis.org/wiki/Welcome, Sequence A083318.
[17] P.Jaccard, Bulletin de la Société Vaudoise des Sciences Naturelles, **37**(1901), 547-579.
[18] P.Jaccard, New Phytologist, **11**(1912), 37-50.
[19] T.T. Tanimoto, *An Elementary Mathematical theory of Classification and Prediction* ,Internal IBM Technical Report, (1958).
[20] D. J. Rogers, T. T. Tanimoto, Science, **132**(1960), 1115-1118.
[21] J. T. Tou, R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley Pub. Co. Reading Mass., USA (1974).
[22] A. Rácz A, D. Bajusz, Héberger, J. Cheminform., **10**(2018), 48, https://doi.org/10.1186/s13321-018-0302-y.
[23] S. Gottwald S, *Many-Valued Logic*, http://plato.stanford.edu/entries/logic-manyvalued/, Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University (2009).
[24] https://en.wikipedia.org/wiki/Artificial_intelligence#Tools
[25] J. Devillers (Editor), *Neural Networks in QSAR and Drug Design*, Vol 2 in the Series: Principles of QSAR and Drug Design, Academic Press, San Diego, (1996).
[26] A. M. Virshup, J. Contreras-García, P. Wipf P, W. Yang W, D. N. Beratan, J. Am. Chem. Soc., **135**(2013), 7296-7303.
[27] L. K. Boerner (Editor), *Artificial Intelligence in Drug Discovery*, Chemical and Engineering News Discovery Report, American Chemical Society (2019).
[28] A. Acharya, R. Agarwal, M. B. Baker, J. Baudry, D. Bhowmik, S. Boehm, K. G. Byler, S. Y. Chen, L. Coates, C. J. Cooper, O. Demerdash, I. Daidone, J. D. Eblen, S. Ellingson, S. Forli, J. Glaser, J. C. Gumbart, J. Gunnels, O. Hernandez, S. Irle, D. W. Kneller, A. Kovalevsky, J. Larkin, T. J. Lawrence, S. LeGrand, S.-H. Liu, J.C. Mitchell, G. Park, J.M. Parks, A. Pavlova, L. Petridis, D. Poole, L. Pouchard, A. Ramanathan, D. M. Rogers, D. Santos-Martins, A. Scheinberg, A. Sedova, Y. Shen, J. C. Smith, M. D. Smith, C. Soto, A. Tsaris, M. Thavappiragasam, A. F. Tillack, J. V. Vermaas, V. Q. Vuong, J. Yin, S. Yoo, M. Zahran, L. Zanetti-Polzi, *Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19*, J. Chem. Inf. Model., **60**(2020), 5832-5852.
[29] R. Carbó-Dorca, J. Math. Chem., **51**(2013), 413-419.
[30] https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology.
[31] J. Hermann, J. Schätzle, F. Noé, *Deep-Neural-Network Solution of the Electronic Schrödinger Equation*, Nature Chemistry, **12**(2020), 891-897.
[32] L. Reyzin, *Unprovability comes to machine learning*, Nature **65**(2019), 166-167.
[33] M. Bunge, *Causality and Modern Science*, Dover Publications (New York) 3rd Revised Edition, 2011.
[34] K. Gödel K, Monatsh. Math., **38**(1931),173-198.
[35] J. Chang, R. Carbó-Dorca, *Fuzzy Hypercubes and their Time-Like Evolution*, Research Gate Preprint: DOI: 10.13140/RG.2.2.26064.25605 (2020).