ISSN 3042-0466

Conference Proceedings META 2024

meta

ISSN 3042-0466 (Online)

META 2024

The 9th Conference on Mathematics in Engineering: Theory and Applications

Conference Proceedings

May 31 - June 2, 2024
Novi Sad, Serbia

The 9th Conference on Mathematics in Engineering: Theory and Applications was organized by the Faculty of Technical Sciences, University of Novi Sad

Publisher: Faculty of Technical Sciences, University of Novi Sad

Editor in Chief: Professor Boris Dumnić, Ph.D.
Dean of the Faculty of Technical Sciences

Print: Grafički centar GRID, Trg Dositeja Obradovića 6, Novi Sad

The conference is supported by:

Ministry of Science, Technological Development and Innovation, Republic of Serbia

Faculty of Technical Sciences, University of Novi Sad

Predgovor

Pred vama je zbornik radova devetog izdanja konferencije Mathematics in Engineering: Theory and Applications - META 2024, koja je održana na Fakultetu tehničkih nauka Univerziteta u Novom Sadu od 31. maja do 2. juna 2024. Konferencija META od 2015. godine okuplja matematičare i inženjere koji izlažu naučne ili stručne radove iz svih teorijskih i primenjenih oblasti matematike, kao i iz metodike nastave matematike.

Ove godine, stiglo je dvadeset šest radova i svi pristigli radovi su prihvaćeni za izlaganje na konferenciji. Organizovane su tri sekcije usmenih izlaganja i jedna poster sekcija. Plenarna predavanja su održali dr Nebojša Gvozdenović, redovni profesor na Ekonomskom fakultetu u Subotici i dr Ljubo Nedović, vanredni profesor na Fakultetu tehničkih nauka u Novom Sadu.

Programski odbor su činili: dr Maja Nedović (predsednik), dr Jelena Ivetić, dr Ksenija Doroslovački, dr Sandra Buhmiler, dr Lidija Čomić, dr Slavica Medić, dr Filip Tomić, dr Ljubo Nedović i dr Biljana Carić. Organizacioni odbor su činili: dr Buda Bajić Papuga (predsednik), dr Zoran Ovcin, dr Vladimir Ilić, dr Ivan Prokić, dr Nataša Duraković, dr Jovana Dedeić, dr Srđan Milićević, dr Jelena Čolić Oravec, dr Marija Delić, mr Gabrijela Grujić, dr Jelena Stratijev, dr Jelena Đokić i dr Tijana Ostojić.

Zahvaljujemo se plenarnim predavačima i svim autorima koji svojim interesovanjem i učešćem omogućavaju održavanje konferencije. Zahvaljujemo se recenzentima, kao i članovima programskog i organizacionog odbora na stručnom i odgovornom radu. Zahvalni smo šefu Katedre za matematiku dr Kseniji Doroslovački, direktoru Departmana za opšte discipline u tehnici dr Nebojši Raleviću i sekretaru departmana Nataši Korpak na podršci, uloženom radu i trudu. Iskazujemo zahvalnost za pruženu finansijsku podršku Ministarstvu za nauku, tehnološki razvoj i inovacije Republike Srbije i Fakultetu tehničkih nauka u Novom Sadu.

Contents

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

CONSTRUCTION OF MEASURE-TYPE FUNCTIONS USING AGGREGATION FUNCTIONS[1]

Ljubo Nedović [2]
plenary talk

https://doi.org/10.24867/META.2024.00 Review article

Abstract. This paper presents the construction of a new measure-type functions of a certain type, by applying aggregation functions on the initial sequence of measure-type functions of the same type. The measure-type functions considered in this paper are distance functions, metrics, fuzzy metrics, and fuzzy measures. The properties of the constructed new measure-type functions depend on the properties of applied aggregation functions, as well as the properties of the initial functions on which the aggregation function is applied.

AMS Mathematics Subject Classification (2020): 28E10, 47S40, 68T37

Key words and phrases: aggregation function, distance function, metrics, fuzzy metrics, fuzzy measure

1. Introduction

In this review paper, some results of research conducted in the past ten years are presented, on the subject of the construction of new measure-type functions in the most wide sense, see [3, 4, 5]. The topic of the mentioned research is the methodology of construction of new measure-type functions by applying aggregation function on a sequence of initial functions of the same type. In this paper, considered measure-type functions are

(a) distance functions and metrics in the Section 3, see [4],

(b) fuzzy metrics in the Section 4, see [5],

(c) fuzzy measures in the Section 5, see [3].

The definition and some relevant properties of aggregation functions are presented in Section 2.

The idea of the mentioned methodology consists in the following. Let I be one of the intervals $I_1 = [0, 1]$, $I_2 = [0, \infty)$ or $I_3 = [0, \infty]$, and let $n \in \mathbb{N}$. Let X

[1]The author acknowledges the financial support of Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Improving the teaching process in the English language in fundamental disciplines".

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: nljubo@uns.ac.rs

be an arbitrary nonempty set measurable in a certain sense, and let $m_i : X \to I$, $i \in \{1, \ldots, n\}$ be a sequence of some measure-type functions of the same type. For n-ary aggregation function $A_{[n]} : I^n \to I$, function $m_{[n]} : X \to I$ defined with

$$m_{[n]}(x) = A_{[n]}\left(m_1(x), \ldots, m_n(x)\right), \quad x \in X \tag{1.1}$$

is also a measure-type function of the same type as the functions $m_i : X \to I$, $i \in \{1, \ldots, n\}$. Properties of the function m depend on the properties of the initial functions $m_i : X \to I$, $i \in \{1, \ldots, n\}$, as well as on the properties of applied aggregation functions $A_{[n]}$. Let I be one of the intervals $I_1 = [0, 1]$, $I_2 = [0, \infty)$ or $I_3 = [0, \infty]$ throughout the whole paper.

2. Aggregation functions

Aggregation functions are one type of fuzzy-operations that have found significant use in IT and other engineering disciplines, but also in other sciences. There are numerous types, classes, additional possible properties and methods of construction of aggregation functions, see [2]. Various sets of argument values and aggregation function values are discussed in the literature. In this article, the intervals I specified in Section 1 are considered as the mentioned sets.

Definition 2.1 (Aggregation function)**.** For each $n \in \mathbb{N}$, an *n-ary aggregation function* is a function $A_{[n]} : I^n \to I$ with the following properties.

(a01) *Boundary conditions* hold, which particulary means that

$$A_{[n]}(0, \ldots, 0) = 0, \tag{2.1}$$

and, depending on the observed cases for interval I,

[I1] for $I = I_1 = [0, 1]$,

$$A_{[n]}(1, \ldots, 1) = 1, \tag{2.2}$$

[I2] for $I = I_2 = [0, \infty)$,

$$\lim_{\forall i \in \{1, \ldots, n\},\ a_i \to \infty} A_{[n]}(a_1, \ldots, a_n) = \infty, \tag{2.3}$$

[I3] for $I = I_3 = [0, \infty]$,

$$A_{[n]}(\infty, \ldots, \infty) = \infty. \tag{2.4}$$

(a02) A function A is monotonically non-decreasing in each component, i.e., implication

$$\forall i \in \{1, \ldots, n\},\ a_i \leq b_i \ \Rightarrow\ A_{[n]}(a_1, \ldots, a_n) \leq A_{[n]}(b_1, \ldots, b_n) \tag{2.5}$$

hold for all $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in I^n$.

For $n = 1$, by definition is $A_{[1]}(x) = x$, $x \in I$.

Some additional properties that aggregation functions can have, can be of interest in the consideration of the construction of mentioned measure-type functions.

Definition 2.2. An n-ary aggregation function $A_{[n]} : I^n \to I$ can have some of the following properties.

(a03) Function $A_{[n]}$ is *continuous*.

(a04) Function $A_{[n]}$ is *symmetric in each component*, i.e., for each n-tuple $(a_1, \ldots, a_n) \in I^n$ and each permutation p of the set $\{1, \ldots, n\}$ hold

$$A_{[n]}(a_1, \ldots, a_n) = A_{[n]}(a_{p(1)}, \ldots, a_{p(n)}).$$

(a05) Function $A_{[n]}$ is *idempotent*, i.e., for each $a \in I$ holds $A_{[n]}(a, \ldots, a) = a$.

(a06) Function $A_{[n]}$ is *subadditive*, i.e., for all n-tuples $(a_1, \ldots, a_n) \in I^n$ and $(b_1, \ldots, b_n) \in I^n$ that satisfy condition $(a_1 + b_1, \ldots, a_n + b_n) \in I^n$ holds

$$A_{[n]}(a_1 + b_1, \ldots, a_n + b_n) \leq A_{[n]}(a_1, \ldots, a_n) + A_{[n]}(b_1, \ldots, b_n).$$

(a07) Function $A_{[n]}$ is *superadditive*, i.e., for all n-tuples $(a_1, \ldots, a_n) \in I^n$ and $(b_1, \ldots, b_n) \in I^n$ that satisfy condition $(a_1 + b_1, \ldots, a_n + b_n) \in I^n$ holds

$$A_{[n]}(a_1 + b_1, \ldots, a_n + b_n) \geq A_{[n]}(a_1, \ldots, a_n) + A_{[n]}(b_1, \ldots, b_n).$$

(a08) Function $A_{[n]}$ is *positively homogeneous*, i.e., for each $t \geq 0$ and all n-tuples $(a_1, \ldots, a_n) \in I^n$ that satisfy condition $(ta_1, \ldots, ta_n) \in I^n$ holds

$$A_{[n]}(ta_1, \ldots, ta_n) \leq tA_{[n]}(a_1, \ldots, a_n).$$

(a09) Depending on interval I, for $A_{[n]}$ holds one of following implications.

[I1] For $I = I_1 = [0, 1]$,
$$A_{[n]}(a_1, \ldots, a_n) < 1 \;\Rightarrow\; \forall i \in \{1, \ldots, n\},\ a_i < 1.$$

[I2] For $I = I_2 = [0, \infty)$ and $(b_1, \ldots, b_n) \in [0, \infty]^n$,
$$\lim_{\forall i \in \{1,\ldots,n\},\ a_i \to b_i} A_{[n]}(a_1, \ldots, a_n) < \infty \;\Rightarrow\; \forall i \in \{1, \ldots, n\},\ b_i < \infty.$$

[I3] For $I = I_3 = [0, \infty]$,
$$A_{[n]}(a_1, \ldots, a_n) < \infty \;\Rightarrow\; \forall i \in \{1, \ldots, n\},\ a_i < \infty.$$

(a10) Function $A_{[n]}$ satisfies

$$A_{[n]}(a_1, \ldots, a_n) = 0 \;\Rightarrow\; \exists i \in \{1, \ldots, n\},\ a_i = 0.$$

(a11) Function $A_{[n]}$ satisfies

$$A_{[n]}(a_1, \ldots, a_n) = 0 \;\Rightarrow\; \forall i \in \{1, \ldots, n\},\ a_i = 0.$$

3. Construction of distance functions

Distance functions and metrics have significant role and application in various scientific disciplines. Distance functions are functions that satisfy only two simple axioms, see [1].

Definition 3.1. Let X be an arbitrary nonempty set. *Distance function* on the set X is a function $d : X^2 \to [0, \infty)$ which has the following properties:

(d01) *reflexivity*, i.e.,

$$\forall x \in X, \;\; d(x, x) = 0,$$

(d02) *symmetry*, i.e.,

$$\forall x, y \in X, \;\; d(x, y) = d(y, x).$$

A *distance space* (X, d) is a set X equipped with a distance function d.

Some additional properties that distance functions can have can be of interest in various particular applications.

Definition 3.2. Let $X \neq \emptyset$, and let $d : X^2 \to [0, \infty)$ be a distance function on X. The function d may have some of the following additional properties.

(d03) *Identity of indiscernibles*, i.e.,

$$\forall x, y \in X, \;\; d(x, y) = 0 \;\Rightarrow\; x = y.$$

(d04) *Triangle inequality*, i.e.,

$$\forall x, y, z \in X, \;\; d(x, z) \leq d(x, y) + d(y, z).$$

(d05) *Ultrametric inequality*, i.e.,

$$\forall x, y, z \in X, \;\; d(x, z) \leq \max\{d(x, y), d(y, z)\}.$$

(d06) *C-triangle inequality*, i.e.,

$$\exists C \in [1, \infty), \; \forall x, y, z \in X, \;\; d(x, y) \leq C\,(d(x, z) + d(z, y)).$$

(d07) *Boundedness*, i.e.,

$$d : X^2 \to [0, 1] \;\; \vee \;\; \exists a > 0, \; d : X^2 \to [0, a].$$

Distance functions are one of measure-type function. In accordance with formula (1.1), starting from the initial distance functions, we can construct a new distance function as follows.

Let $n \geq 2$. Let $A_{[n]} : [0, 1]^n \to [0, 1]$ be an arbitrary n-ary aggregation function, and let $d_i : X^2 \to [0, 1]$, $i \in \{1, \ldots, n\}$ be a sequence of bounded distance functions on the nonempty set X. Let we observe the function $d_{[n]} : X^2 \to [0, 1]$ defined by

$$d_{[n]}(x, y) = A_{[n]}\left(d_1(x, y), \ldots, d_n(x, y)\right), \quad x, y \in X. \tag{3.1}$$

Theorem 3.3. *For functions $d_{[n]}$ defined with* (3.1)*, the following is valid.*

(a) *Function $d_{[n]}$ is a distance function.*

(b) *If for each of the distance functions* d_i, $i \in \{1, \ldots, n\}$ *the "identity of indiscernibles"* (d03) *holds and aggregation function* $A_{[n]}$ *has the property* (a10)*, then for function* $d_{[n]}$*, the "identity of indiscernibles"* (d03) *is also valid.*

(c) *If for at least one of the considered distance functions* d_i, $i \in \{1, \ldots, n\}$ *"identity of indiscernibles"* (d03) *holds and if the aggregation function* $A_{[n]}$ *has the property* (a11)*, then for the distance function* $d_{[n]}$ *"identity of indiscernibles"* (d03) *also holds.*

(d) *Let each of the distance functions* d_i, $i \in \{1, \ldots, n\}$ *be metric (satisfies* (d03) *and* (d04)*). Let* $A_{[n]} : [0, \infty)^n \to [0, \infty)$ *be subadditive function (satisfies* (a06)*) whose restriction on the set* $[0, 1]^n$ *is an aggregation function which also satisfies the property* (a10)*), then the distance function* $d_{[n]}$ *is a metric also.*

(e) *Assume that for each of the distance functions* d_i, $i \in \{1, \ldots, n\}$ *the C-triangle inequality* (d06) *holds. Let* $A_{[n]} : [0, \infty)^n \to [0, \infty)$ *be subadditive and positively homogeneous function (satisfies* (a06) *and* (a08) *properties) whose restriction on the set* $[0, 1]^n$ *is an aggregation function. Then for the distance function* $d_{[n]}$ *the C-triangle inequality* (d06) *also holds.*

4. Construction of fuzzy metrics

The definitions of fuzzy metrics are not unanimously accepted in the literature. For the purpose of presented results and terms in this paper see [5].

Definition 4.1. Let $X \neq \emptyset$.

(fS) For the continuous t-conorm $S : [0, 1]^2 \to [0, 1]$ and for the fuzzy set $s : X^2 \times (0, \infty) \to [0, 1]$, the ordered triple (X, s, S) is a *fuzzy S-metric space* and fuzzy set s *is a fuzzy S-metric* if for all $x, y, z \in X$ and all $\alpha, \beta > 0$ satisfies the following conditions.

(1) $s(x, y, \alpha) \in [0, 1)$,

(2) $s(x, y, \alpha) = 0 \Leftrightarrow x = y$,

(3) $s(x, y, \alpha) = s(y, x, \alpha)$,

(4) $S\left(s(x, y, \alpha), s(y, z, \beta)\right) \geq s(x, z, \alpha + \beta)$,

(5) Function $\overline{s} : (0, \infty) \to [0, 1]$, $\overline{s}(t) = s(x, y, t)$ is a continuous function.

(fT) For the continuous t-norm $T : [0, 1]^2 \to [0, 1]$ and for the fuzzy set $t : X^2 \times (0, \infty) \to [0, 1]$, the ordered triple (X, t, T) is a *fuzzy T-metric space* and fuzzy set t *is a fuzzy T-metric* if for all $x, y, z \in X$ and all $\alpha, \beta > 0$ satisfies the following conditions.

(1) $t(x, y, \alpha) \in (0, 1]$,

(2) $t(x, y, \alpha) = 1 \Leftrightarrow x = y$,

(3) $t(x, y, \alpha) = t(y, x, \alpha)$,

(4) $T\left(t(x, y, \alpha), t(y, z, \beta)\right) \leq t(x, z, \alpha + \beta)$,

(5) Function $\bar{t} : (0, \infty) \to [0, 1]$, $\bar{t}(t) = t(x, y, t)$ is a continuous function.

In the following theorem, according the formula (1.1) is presented a method for a fuzzy S-pseudo metric and fuzzy T-pseudo metric construction by application of aggregation function on the sequence of initial fuzzy S-pseudo metrics, i.e., initial fuzzy T-pseudo metrics, see [5]. For a definitions of aggregation function which is continuously compatible with continuous t-conorms $S_1, \ldots, S_n$ with respect to a continuous t-conorm S, and aggregation function which is continuously compatible with continuous t-norms $T_1, \ldots, T_n$ with respect to a continuous t-norm T see [5].

Theorem 4.2. *Let $A_{[n]}$ be a continuous n-ary aggregation function.*

(afS) *Let aggregation function $A_{[n]}$ be continuously compatible with continuous t-conorms $S_1, \ldots, S_n$ with respect to a continuous t-conorm S. If the functions $s_i : X_i^2 \times (0, \infty) \to [0, 1]$, $i \in \{1, \ldots, n\}$ are fuzzy S-pseudo metrics on $X_i \neq \emptyset$, $i \in \{1, \ldots, n\}$ with respect to the t-conorms S_i, $i \in \{1, \ldots, n\}$ respectively, then for $X = X_1 \times \cdots \times X_n$, function $s : X^2 \times (0, \infty) \to [0, 1]$ defined with*

$$s(x, y, \alpha) = A_{[n]}\left(s_1(x_1, y_1, \alpha), \ldots, s_n(x_n, y_n, \alpha)\right), \tag{4.1}$$

$x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in X$, $\alpha > 0$ is the fuzzy S-pseudo metric in a broader sense with respect to t-conorm S.

(afT) *Let aggregation function $A_{[n]}$ be continuously compatible with continuous t-norms $T_1, \ldots, T_n$ with respect to a continuous t-norm T. If the functions $t_i : X_i^2 \times (0, \infty) \to [0, 1]$, $i \in \{1, \ldots, n\}$ are fuzzy T-pseudo metrics on $X_i \neq \emptyset$, $i \in \{1, \ldots, n\}$ with respect to the t-norms T_i, $i \in \{1, \ldots, n\}$ respectively, then for $X = X_1 \times \cdots \times X_n$, function $t : X^2 \times (0, \infty) \to [0, 1]$ defined with*

$$t(x, y, \alpha) = A_{[n]}\left(t_1(x_1, y_1, \alpha), \ldots, t_n(x_n, y_n, \alpha)\right), \tag{4.2}$$

$x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in X$, $\alpha > 0$ is the fuzzy T-pseudo metric in a broader sense with respect to t-norm T.

5. Construction of fuzzy measures

There are many definitions and types of fuzzy measures and non-additive measures, see [6]. In this paper, some types of fuzzy measures and their construction will be considered, see [3].

Definition 5.1. Let $\mathcal{A}$ be a σ-algebra on the set $X \neq \emptyset$. Function $m : \mathcal{A} \to I$ is a *fuzzy measure* on $\mathcal{A}$ if

(fm1) $m(\emptyset) = 0$,

(fm2) $\forall A, B \in \mathcal{A}, \;\; A \subseteq B \;\Rightarrow\; m(A) \leq m(B)$.

Fuzzy measures can have many other important properties.

Definition 5.2. Let $m : \mathcal{A} \to I$ be a fuzzy measure defined on σ-algebra $\mathcal{A}$ on the set X.

(fm3) m is *continuous from below* if for every family of sets $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ nested as $A_1 \subseteq A_2 \subseteq \dots$ holds

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} m(A_i). \tag{5.1}$$

(fm4) m is *continuous from above* if for every family of sets $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ nested as $A_1 \supseteq A_2 \supseteq \dots$ and satisfying that there exists $n_0 \in \mathbb{N}$ such that $m(A_{n_0}) < \infty$ holds

$$m\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} m(A_i). \tag{5.2}$$

(fm5) m is *subadditive* if for each disjoint pair of sets $A, B \in \mathcal{A}$ holds

$$m(A \cup B) \leq m(A) + m(B). \tag{5.3}$$

(fm6) m is *superadditive* if for each disjoint pair of sets $A, B \in \mathcal{A}$ holds

$$m(A \cup B) \geq m(A) + m(B). \tag{5.4}$$

For $n \geq 2$, let $A_{[n]} : I^n \to I$ be an n-ary aggregation function, and let $m_i : \mathcal{A} \to I$, $i \in \{1, \dots, n\}$ be a sequence of fuzzy measures on $\mathcal{A}$. In accordance with formula (1.1), let we observe the function $m_{[n]} : \mathcal{A} \to I$ defined by

$$m_{[n]}(A) = A_{[n]}\left(m_1(A), \dots, m_n(A)\right), \quad A \in \mathcal{A}. \tag{5.5}$$

Properties of function $m_{[n]}$ depends on properties of aggregation function $A_{[n]}$ and properties of initial fuzzy measures m_i.

Theorem 5.3. *Function $m_{[n]}$ defined by* (5.5) *is a fuzzy measure on $\mathcal{A}$. Additionally, the following statements hold.*

(a) *Let all fuzzy measures m_i, $i \in \{1, \dots, n\}$ be continuous from below. If $A_{[n]}$ is an continuous aggregation function, then fuzzy measure $m_{[n]}$ is also continuous from below.*

(b) *Let all fuzzy measures m_i, $i \in \{1, \dots, n\}$ be continuous from above. If $A_{[n]}$ be a continuous aggregation function and, in the case $I = [0, \infty]$, let $A_{[n]}$ additionally have property* (a09)*. Then fuzzy measure $m_{[n]}$ is also continuous from above.*

(c) *Let all fuzzy measures m_i, $i \in \{1, \dots, n\}$ be subadditive. If $A_{[n]}$ is a subadditive aggregation function, then fuzzy measure $m_{[n]}$ is also subadditive.*

(d) *Let all fuzzy measures m_i, $i \in \{1, \dots, n\}$ be superadditive. If $A_{[n]}$ is a superadditive aggregation function, then fuzzy measure $m_{[n]}$ is also superadditive.*

6. Conclusion

As for distance functions, fuzzy metrics, and fuzzy measures, construction proposed by (1.1) can be applied on other measure-type functions.

References

[1] M. M. Deza and E. Deza, *Encyclopedia of distances*, *SpringerLink : Bücher*, Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2012.

[2] M. Grabisch, J.L. Marichal, R. Mesiar and E. Pap, *Aggregation functions*, vol. 127 of the *Encyclopedia of mathematics and its applications*, Cambridge: Cambridge University Press, 2009.

[3] Lj. Nedović and E. Pap, "Aggregation of sequence of fuzzy measures", *Iranian Journal of Fuzzy Systems*, vol. 17, no. 6, pp. 39-55, 2020.

[4] Lj. Nedović, N. M. Ralević and I. Pavkov, "Aggregated distance functions and their application in image processing", *Soft Computing*, vol. 22, no. 14, pp. 4723-4739, July 2018.

[5] N. M. Ralević, M. Delić and Lj. Nedović, "Aggregation of fuzzy metrics and its application in image segmentation", *Iranian Journal of Fuzzy Systems*, vol. 19, no. 3, pp. 19-37, 2022.

[6] V. Torra, Y. Narukawa and M. Sugeno, *Non-additive measures: theory and applications*, Switzerland: Springer International Publishing, 2014.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

ON THE COMPUTATION OF THE ORTHOGONAL HULL OF SIMPLE RECTILINEAR POLYGONS

Andrija Blesić[1], **and Lidija Čomić**[2]

https://doi.org/10.24867/META.2024.01 Review article

Abstract. We briefly review the algorithm for determining the orthogonal hull of a set of simple rectilinear polygons, proposed by Nicholl et al., based on determining their maximal vertices.

AMS Mathematics Subject Classification (2020): 68U05, 65D18

Key words and phrases: rectilinear polygons, orthogonal hull, orthogonal convexity

1. Introduction and preliminaries

A *simple polygon* is usually defined as a closed, connected plane figure determined by a cyclic sequence of points, called *(polygonal) vertices*, each one joined to the previous and the next point with straight line segments, and no other points of intersection exist between the line segments. The line segments themselves are then referred to as *(polygonal) edges*, and they form what is called the *(polygonal) boundary*, and outline a bounded region called the *(polygonal) area* or *(polygonal) interior*. See, for example, [9].

In [7], the authors consider only simple, rectilinear polygons (polygons whose interior angles are all 90° or 270°) , which they call $X-Y$ polygons. They also refer to the two classes of edges such polygons possess as *vertical* or *horizontal*, and we adopt the same terminology. In what follows, we consider only simple rectilinear polygons, which we refer to simply as "polygons" for brevity.

Definition 1.1. A polygon is *h-convex* (*v-convex*) if each non-empty intersection of the polygonal interior with a horizontal (vertical) line is connected. If the polygon is both h-convex and v-convex, it is called *orthogonally convex* (or hv-convex).

Definition 1.2. [7, 8] The *orthogonal hull* (or $X-Y$ *convex hull*) of a set of polygons, if it exists, is an orthogonally convex polygon that contains the set of polygons and has minimal area.

[1]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: andrija.blesic@uns.ac.rs

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: comic@uns.ac.rs

Note that the orthogonal hull of a set of polygons need not exist, as the polygon that contains the entire set and is of minimal area may not be simple at all, see Figure 1 for an example.

Figure 1: When considering a set of multiple polygons, the rectilinear polygon that contains all of them and is of minimal area: a) may be simple, and therefore the orthogonal hull, b) may not be simple, and is thus not the orthogonal hull.

In computer graphics, it is common to consider polygons as primitives and store the vertex coordinates and information regarding the connectivity of the edges [5]. Thus, for each polygon, we consider its vertices to be known and (their corresponding x- and y-coordinates) provided as the input for the algorithm.

Definition 1.3. [4, 10] Given a finite set of vertices, a point $p = (p_x, p_y)$ dominates (from the north-east direction) a distinct point $r = (r_x, r_y)$ if $r_x \leq p_x$ and $r_y \leq p_y$. A vertex p is maximal (from the north-east direction) if there is no other vertex from the set which dominates p.

The maximal vertices from the other three diagonal directions (north-west, south-east, south-west) are obtained in an analogous manner, see Figure 2.

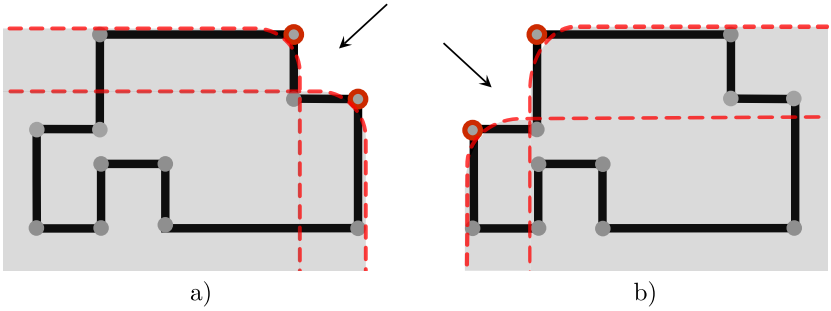


Figure 2: A polygon whose vertices are shown in grey, and its maximal vertices from a) north-east and b) north-west, shown outlined in red. Each maximal vertex dominates the vertices in its corresponding region (outlined with a red dashed line).

2. Algorithm by Nicholl et al.

In [7], the authors use as input a list of the polygonal vertices V, sorted in an order in which these vertices would have been visited during an oriented

(clockwise) traversal of the polygonal boundary. We first examine the case when a single polygon's orthogonal hull is considered.

2.1. Determining the orthogonal hull for a single polygon

The goal of the algorithm is to find all maximal vertices among the polygonal vertices, as the orthogonal hull is uniquely determined by such vertices. Obviously, the four extremal edges (i.e., the topmost horizontal, leftmost horizontal, bottommost vertical and rightmost vertical edges) necessarily all belong to the orthogonal hull. They determine at most eight distinct maximal vertices. The rest of the maximal vertices can be found between:

1. the right topmost and top rightmost vertices (from north-east),
2. the bottom rightmost and right bottommost vertices (from south-east),
3. the left bottommost and bottom leftmost vertices (from south-west),
4. the top leftmost and left topmost vertices (from north-west).

Note that the coordinates of these eight maximal vertices can easily be obtained by simply going through the input list of vertices and checking their coordinates.

For simplicity, we describe the algorithm as if we were determining non-trivial maximal vertices between the right topmost and top rightmost vertices (i.e. from the north-east direction). The other three cases are analogous.

The algorithm uses the fact that, for the orthogonal hull to be traversed (clockwise) between the right topmost and top rightmost vertices, the only allowed directions of traversal are downwards and to the right, which determines a sequence of alternately vertical and horizontal edges, forming what the authors in [7] refer to as a *staircase*, see Figure 3.

Note, as well, that a clockwise traversal reaches these maximal vertices from the left, via horizontal edge, and continues traversal downwards from them via vertical edge. The authors in [7] refer to all such vertices (not only maximal ones) as *right-down convex*, see Figure 3.

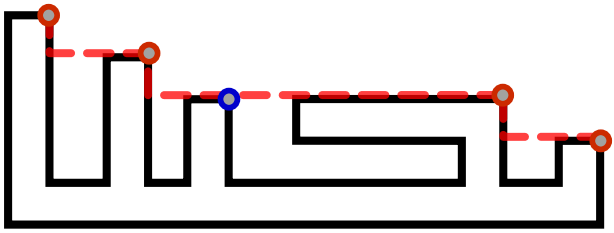


Figure 3: Maximal vertices (from the north-east direction), shown outlined in red, determine the part of the orthogonal hull from the right topmost to the top rightmost vertex of a given polygon (shown dashed in red). The alternately vertical and horizontal line segments connecting these vertices belong to the orthogonal hull. The only right-down convex vertex which is not maximal is shown outlined in blue.

The sorted list V_m of maximal vertices can be obtained as follows:

- At first, add the right topmost vertex to list V_m. It is necessarily a maximal (and right-down convex) vertex.
- After adding element v_i to V_m, find the first next right-down convex vertex v_{i+1} in V. Then:
  - if v_i dominates v_{i+1} (from north-east), examine the next right-down convex vertex in the same manner and do not add v_{i+1} to V_m;
  - if v_{i+1} does not dominate v_i, add v_{i+1} to V_m;
  - if v_{i+1} dominates v_i, remove all recent entries in V_m which are dominated by v_{i+1} (including v_i), and add v_{i+1} to V_m.
- If v_{i+1} is the top rightmost vertex, end the procedure. See Figure 4.

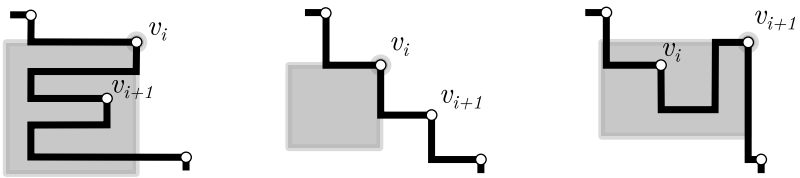


Figure 4: An example of traversal from the right topmost to the top rightmost vertex, with right-down convex vertices labeled v_i and v_{i+1}. Left: v_i dominates v_{i+1} (from north-east), center: v_{i+1} does not dominate v_i, right: v_{i+1} dominates v_i.

The output of the procedure described above returns a list of maximal vertices, which determine a part of the orthogonal hull. The maximal vertices in the remaining three directions may be found in an analogous manner. If a vertex is deleted at any point from the list of maximal vertices, it is not encountered again, nor added later. Thus, the total time complexity is $O(n)$, where n is the number of polygonal vertices, i.e. the size of the list V.

2.2. Determining the orthogonal hull for multiple polygons

The authors also describe how this algorithm could be adapted for the case of multiple polygons, provided they satisfy the condition of not having an $X-Y$ separation:

Definition 2.1. [7] The *x-extent* of a set of polygons is the open interval (x_{min}, x_{max}) where x_{min} and x_{max} are, respectively, the minimum and maximum x-coordinates among all vertices of the set. The *y-extent* is analogously defined.

Definition 2.2. [7] A set of polygons has an $X - Y$ separation if the set can be partitioned into disjoint nonempty subsets (of polygons) which have disjoint x-extents and y-extents.

Figure 1b) depicts a set of two polygons that has an $X - Y$ separation.

It is proven in [7] that the nonexistence of an $X - Y$ separation implies the existence of the orthogonal hull of a set of polygons. Thus, for a given set of k polygons which have no $X - Y$ separation, first the algorithm above is applied for each individual polygon, in order to find all maximal vertices. Then the maximal vertices forming, for example, the staircases connecting the right topmost to the top rightmost vertex of each individual polygon are also the candidates for forming the appropriate staircase of the entire orthogonal hull of the set.

These maximal vertices are then all lexicographically sorted, so that (x_1, y_1) comes before (x_2, y_2) if either: a) $x_1 > x_2$ or b) $x_1 = x_2$ and $y_1 > y_2$, and the sorted list is then traversed element by element. If an encountered element is dominated by the previous element in the sorted list, it is removed from the list. What remains is the list of maximal vertices that determine the required section of the orthogonal hull, see Figure 5. An analogous approach is used to find the remaining maximal vertices from the three other directions. The time complexity is $O(n \log k)$, where k is the number of polygons in the set, and n the total number of all polygonal vertices.

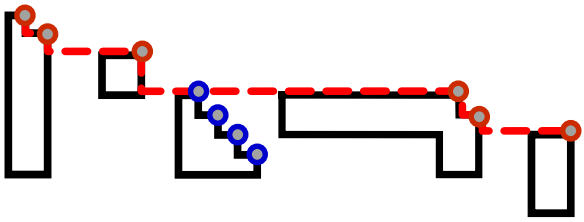


Figure 5: Maximal vertices (from the north-east direction), shown outlined in red and blue, for each individual polygon. Those outlined in blue will be removed, leaving only those which determine the required section of the orthogonal hull of the set (shown dashed in red).

3. Conclusion

The orthogonal hull of a digital object represents an important notion in computational geometry. It has found number of interesting applications tied to image processing and computer aided design, operations research, database concurrency control, fault-tolerant routing in mesh-connected multicomputers, as well as very large-scale integrated (VLSI) circuit layout design, see [1, 2, 3, 4, 6, 11], for instance. In this paper, we review the algorithm proposed by Nicholl et al. [7], which addresses the problem of determining the orthogonal hull for a single simple, rectilinear polygon, or a set of such polygons, and gives the condition for the existence of such a hull in the latter case.

Acknowledgement

This paper has been supported by the project "Improving the teaching process in the English language in fundamental disciplines", Faculty of Technical Sciences, University of Novi Sad.

References

[1] T. Biedl, S. Durocher and J. Snoeyink, "Reconstructing polygons from scanner data", in *Theoretical Computer Science*, vol. 412, pp. 4161-4172, July 2011.

[2] M. Fredman and B. Weide, "On the complexity of computing the measure of $\cup[a_i, b_i]$", in *Communications of the ACM*, vol. 21, pp. 540-544, July 1978.

[3] B. Genç, C. Evrendilek, and B. Hnich, "Covering points with orthogonally convex polygons", *Computational Geometry*, vol. 44, pp. 249-264, 2011.

[4] R. Karlsson and M. Overmars, "Scanline algorithms on a grid", in *BIT Numerical Mathematics*, vol. 28, pp. 227-241, June 1988.

[5] J. Kessenich, G. Sellers, and D. Shreiner, *OpenGL Programming Guide*. Boston: Addison-Wesley Professional, 2016.

[6] W. Lipski and F. Preparata, "Segments, rectangles, contours", in *Journal of Algorithms*, vol. 2, pp. 63-78, March 1981.

[7] T. Nicholl, D. Lee, Y. Liao and C. Wong, "On the X-Y Convex Hull of a Set of X-Y Polygons", in *BIT Numerical Mathematics*, vol. 23, pp. 456-471, December 1983.

[8] T. Ottmann, E. Soisalon-Soininen and D. Wood,"On the definition and computation of Rectilinear Convex Hulls", in *Information Sciences*, vol. 33, pp. 157-171, September 1984.

[9] C. Pierce, *Algebra and Geometry*, vol. 2 of *The New Elements of Mathematics*. Paris: Mouton Publishers, 1976.

[10] F. Preparata and M. Shamos, *Computational Geometry: An Introduction*, in *Texts and Monographs in Computer Science*. Berlin: Springer, 1985.

[11] J. Wu and Z. Jiang, "On constructing the minimum orthogonal convex polygon for the fault-tolerant routing in 2-D faulty meshes", *IEEE Transactions on Reliability*, vol. 54, no. 3, pp. 449-458, 2005.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

A MODIFIED DURAN-SHISHKIN MESH FOR A SINGULARLY PERTURBED THIRD ORDER BOUNDARY VALUE PROBLEM[1]

Mirjana Brdar [2], Goran Radojev [3] and Ljiljana Teofanov [4]

https://doi.org/10.24867/META.2024.02 Original scientific paper

Abstract. In this article, a third order singularly perturbed problem with a weak layer is considered. To obtain approximation to the solution of this problem, a standard difference scheme on a new modification of the Duran-Shishkin mesh is used. Our modification has several advantages and provides numerical solution with better accuracy then the standard Shishkin mesh, which is confirmed in numerical experiments.

AMS Mathematics Subject Classification (2020): 65L10, 65L12, 65L50

Key words and phrases: singularly perturbed problem, layer-adapted mesh, Duran-Shishkin mesh, finite difference scheme, uniform convergence

1. Introduction

We consider the following third order singularly perturbed problem:

$$\begin{aligned} \varepsilon u'''(x) + a(x)u''(x) + b(x)u'(x) + c(x)u(x) &= f(x) \quad \text{in} \quad \Omega \\ u(0) = u(1) = u'(0) &= 0, \end{aligned} \tag{1.1}$$

where $\Omega = (0,1)$, ε is a small positive parameter, $a(x) > \alpha > 0$, and the coefficients are infinitely differentiable. This problem has a weak layer near $x = 0$ and from [6] we have

$$|u^{(k)}(x)| < C(1 + \varepsilon^{-k+1}e^{-\alpha x/\varepsilon}), \quad k = 0, 1, 2,$$

where C is some constant independent of ε.

There are only few papers dealing with third order singularly perturbed problems, see for example [3],[4],[6]-[8]. These problems are more difficult than second or fourth order singularly perturbed problem, which are studied widely.

[1]The third author's work has been financially supported by Department of Fundamentals Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Improving the teaching process in the English language in fundamental disciplines".

[2]Faculty of Technology Novi Sad, University of Novi Sad, e-mail: mirjana.brdar@uns.ac.rs
[3]Faculty of Sciences, University of Novi Sad, e-mail: goran.radojev@dmi.uns.ac.rs
[4]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: ljiljap@uns.ac.rs

2. Discretization

For a given arbitrary mesh $0 = x_0 < x_1 < \ldots < x_n = 1$, the following notation is used

$$h_i = x_{i+1} - x_i, \quad i = 0, 1, \ldots, n-1, \qquad h = \max_{i=0,1,\ldots,n} h_i,$$

and the divided difference operator

$$D^k u_i^h := k! u^h[x_i, \ldots, x_{i+k}], \quad k = 1, 2, 3$$

for a grid function u^h. The following discretization

(2.1)

$$\varepsilon D^3 u_i^h + a_i D^2 u_{i+1}^h + b_i D^1 u_{i+1}^h + c_i u_{i+1}^h = f_i, \quad \text{for} \quad i = 0, 1, \ldots, n-3,$$
$$u_0^h = u_n^h = 0, \quad D^1 u_0^h = 0.$$

is used to obtain approximate solution to problem (1.1), [6]. Also, to measure the error of the approximation, we introduce the following discrete norms

$$\|u^h\|_{h,\varepsilon} := \max\{\|u^h\|_{h,\infty}, \|D^1 u^h\|_{h,\infty}, \varepsilon\|D^2 u^h\|_{h,\infty}\},$$

$$\|D^k u^h\|_{h,\infty} := \max\{|D^k u_i^h| : i = 0, 1, \ldots, n-3\}, \quad k = 0, 1, 2.$$

Finally, $\|\cdot\|_{h,\infty}$ is the standard maximum discrete norm.

3. Duran-Shishkin mesh

To capture boundary layers for singularly perturbed problems, the layer-adapted meshes are most often used. One type of those meshes is the Duran mesh introduced in [1] and it is obtained as a simplified version of Gartland-type mesh from [2]. In [7] the authors presented a version of the Duran mesh, the so-called Duran-Shishkin (D-S) mesh which was used to capture the layers for a third order singularly perturbed problem.

Let $0 < \mu < 1$ be a given parameter and N some chosen even integer. Mesh from [7] tailored for problem (1.1) is defined as follows:

$$\begin{aligned}
&x_0 = 0,\\
&x_1 = \mu\varepsilon,\\
&x_i = x_{i-1} + \mu x_{i-1} = \mu(1+\mu)^{i-1}\varepsilon, && i = 2, \ldots, M,\\
&x_{M+i} = \bar{\tau} + 2iN^{-1}(1-\bar{\tau}), && i = 1, \ldots, N/2,\\
&x_{M+N/2} = 1,
\end{aligned}$$

where $x_M = \bar{\tau}$, and M is the smallest integer such that

$$\bar{\tau} = \mu(1+\mu)^{M-1}\varepsilon \geq \tau := \min\left\{\frac{1}{2}, \frac{2}{\alpha}\varepsilon \ln N\right\}.$$

This mesh has two issues. The fist one is a transition point $\bar{\tau}$ which is not the standard transition point of the Shishkin mesh τ. The second issue is related

to the number of mesh points, which is $M + N/2 + 1$. This number is not a priori known since it depends on the mesh parameter μ. Also, it is not clear what is the ratio between M - the number of mesh points in the layer, and $N/2$ - the number of mesh points outside of it. If μ is large, then this ratio could be too small (M is very small). If μ is small, than the ratio could be too large (M is very large number).

The main goal here is to modify this mesh, in order to solve both issues given the above. For a fixed N, our modified mesh is defined as follows

$$
\begin{aligned}
&x_0 = 0, \\
&x_1 = \mu\varepsilon, \\
&x_i = x_{i-1} + \mu x_{i-1} = \mu(1+\mu)^{i-1}\varepsilon, \quad && i = 2, \ldots, M, \\
&x_{N/2+j} = \tau + 2iN^{-1}(1-2\tau), && i = 1, \ldots, N/2,
\end{aligned}
\tag{3.1}
$$

where $M = N/2$, and μ is calculated so that

$$
x_M = \mu(1+\mu)^{M-1}\varepsilon = \tau. \tag{3.2}
$$

Now, the transition point of such a mesh is exactly τ and it has the same number of mesh points inside and outside of boundary layer. Precisely, there are $N/2$ points in the interval $(0, \tau)$ and $N/2$ points in the interval $(\tau, 1)$.

On Figure 1 the position of mesh points (for different values of ε) is shown. With decreasing of parameter ε, the number of mesh points in the neighborhood of $x = 0$ increases.

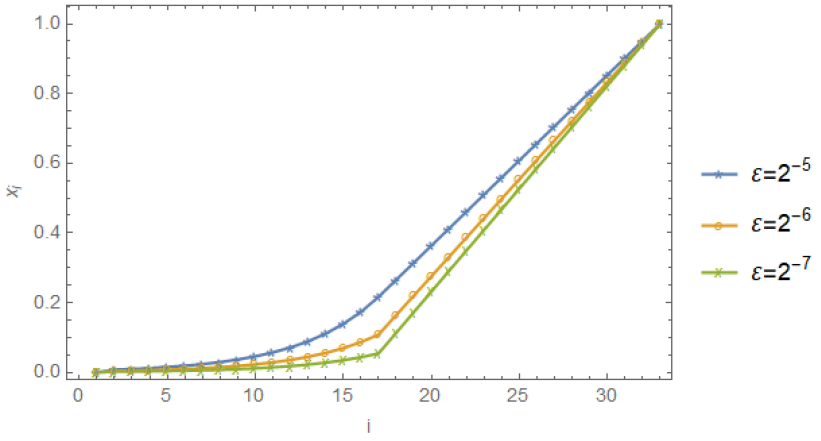


Figure 1: The modified D-S mesh (3.1) for $N = 32$.

The parameter μ in (3.2) can be calculated numerically. Table 1 shows the values of μ, for fixed ε and various N. Using these values, the obtained meshes are unique. In other words, for a chosen N, there exists only one modified D-S mesh, which is unlikely for the original D-S mesh [5], Duran mesh, as well for the one from [7].

For the method (2.1) applied on modified mesh (3.1), numerical solution converges to the exact solution of (1.1), when $N \to \infty$. Theoretical results

N	μ
2^5	0.2484575
2^6	0.1406596
2^7	0.0792910
2^8	0.0444235
2^9	0.0247098
2^{10}	0.0136418
2^{11}	0.0074769
2^{12}	0.0040706

Table 1: Values of parameter μ for $\varepsilon = 10^{-6}$

for different numerical methods on some kind of D-S mesh can be found in [6] and [7]. In this paper we want to compare numerical results for method (2.1) on mesh (3.1) with the results obtained by the same method on the standard piecewise uniform Shishkin mesh, which is defined by

$$x_i = \begin{cases} i\dfrac{2}{N}\tau, & i = 0, 1, \ldots, N/2 \\ \tau + \left(i - \dfrac{N}{2}\right)(1-\tau)\dfrac{2}{N}, & i = N/2+1, N/2+2, \ldots, N. \end{cases}$$

4. Numerical results

The following test problem is used to compare efficiency of method (1.1) on two mentioned

$$\begin{aligned} \varepsilon u'''(x) + u''(x) + \varepsilon u'(x) + u(x) &= x \quad \text{in} \quad \Omega = (0,1), \\ u(0) = u(1) &= u'(0) = 0, \end{aligned}$$

The exact solution of this problem is given by

$$u(x) = \frac{\left(e^{\frac{1}{\varepsilon}}(\varepsilon - 1) - \varepsilon\right)\sin x}{\varepsilon + e^{\frac{1}{\varepsilon}}(\sin 1 - \varepsilon\cos 1)} + \frac{\varepsilon\left(e^{\frac{1}{\varepsilon}}(1-\varepsilon) + \varepsilon\right)\cos x}{\varepsilon + e^{\frac{1}{\varepsilon}}(\sin 1 - \varepsilon\cos 1)} + \frac{\varepsilon e^{\frac{1}{\varepsilon} - \frac{x}{\varepsilon}}(\varepsilon(1-\cos 1) - 1 + \sin 1)}{\varepsilon + e^{\frac{1}{\varepsilon}}(\sin 1 - \varepsilon\cos 1)} + x - \varepsilon.$$

The errors and convergence rates

$$\chi_N = \|u - u^h\|_{h,\varepsilon}, \qquad p_N = \frac{\ln \chi_N - \ln \chi_{2N}}{\ln 2}$$

on modified D-S and Shishkin meshes are presented in Table 2. The rate of convergence on mesh (3.1) is slightly larger than on the Shishkin mesh. It can be observed that numerical results on the D-S mesh outperform results on the Shishkin mesh in the sense that errors are smaller on mesh (3.1).

Table 3 shows uniform convergence of previously described method on the D-S mesh – the number of mesh points is fixed ($N = 512$) while parameter ε takes various values.

	modified D-S mesh		Shishkin mesh	
N	χ_N	p_N	χ_N	p_N
2^5	6.73e-02	0.872	6.70e-02	0.834
2^6	3.68e-02	0.911	3.76e-02	0.860
2^7	1.96e-02	0.931	2.07e-02	0.878
2^8	1.03e-02	0.941	1.13e-02	0.890
2^9	5.35e-03	0.946	6.08e-03	0.900
2^{10}	2.78e-03	0.949	3.26e-03	0.907
2^{11}	1.44e-03	0.951	1.74e-03	0.912
2^{12}	7.44e-04	–	9.23e-04	–

Table 2: Comparison of results on D-S and Shishkin meshes for $\varepsilon = 10^{-6}$.

ε	χ_N
10^{-2}	2.23e-03
10^{-3}	2.70e-03
10^{-4}	2.77e-03
10^{-5}	2.78e-03
10^{-6}	2.78e-03
10^{-7}	2.78e-03
10^{-8}	2.78e-03

Table 3: Uniform convergence for $N = 512$.

Acknowledgement

The first and second authors gratefully acknowledge the financial support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Grants No. 451-03-66/2024-03/200134, 451-03-65/2024-03/200134 and No. 451-03-66/2024-03/200125, 451-03-65/2024-3/200125 respectively).

References

[1] R. G. Durán, A. Lombardi, "Finite Element Approximation of Convection Diffusion Problems using Graded Meshes", *Appl. Numer. Math.*, vol. 56, pp. 1314-1325, 2006.

[2] E. C. Gartland, "Graded-mesh difference schemes for singularly perturbed two-point boundary value problems", *Math. Comp.*, vol. 51, pp. 631-657, 1988.

[3] F. A. Howes, "Nonlinear dissipative systems: theory and examples", *Stud. Appl. Math.*, vol. 69, pp. 75–97, 1983.

[4] F. A. Howes, "The asymptotic solution of a class of third-order boundary value problems arising in the theory of thin film flows", *SIAM J. Appl. Math.*, vol. 43, pp. 993–1004, 1983.

[5] H. G. Roos, Lj. Teofanov and Z. Uzelac, "Graded meshes for higher order FEM" *J. Comput. Math.*, vol. 33 (1), pp. 1-16, 2015.

[6] H. G. Roos, Lj. Teofanov and Z. Uzelac, "Uniformly convergent difference schemes for a singularly perturbed third order boundary value problem", *Appl. Numer. Math.*, vol. 96, pp. 108-117, 2015.

[7] L. Yan and Y. Cheng, "Local discontinuous Galerkin method on graded meshes for a third-order singularly perturbed problem", *Z Angew Math Mech.*, https://doi.org/10.1002/zamm.202300238, 2023.

[8] H. Zarin, H. G. Roos and Lj. Teofanov, "A continuous interior penalty finite element method for a third-order singularly perturbed boundary value problem", *Comp. Appl. Math.*, vol. 37, pp. 175-190, 2018.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

PERCOLATION ON A TRIANGULAR LATTICE UNDER ANISOTROPIC CONDITIONS

Ljuba Budinski Petković [1] **and Ivana Lončarević** [2]

https://doi.org/10.24867/META.2024.03 Original scientific paper

Abstract. The properties of percolation of objects of various shapes on a two-dimensional triangular lattice is studied by means of Monte Carlo simulations. Depositing objects of various shapes and sizes are made by directed self-avoiding walks on the lattice. Anisotropy is introduced by positing unequal probabilities for orientation of depositing objects along different directions of the lattice. This probability is equal p or $(1-p)/2$, depending on whether the randomly chosen orientation is horizontal or not, respectively. It is found that the percolation threshold θ_p increases with the degree of anisotropy, having the maximum values for fully oriented objects.

AMS Mathematics Subject Classification (2020): 06, 65

Key words and phrases: percolation, RSA, triangular lattice

1. Introduction

Random sequential adsorption of extended objects at different surfaces is of considerable interest for a wide range of applications in biology, nanotechnology, device physics, physical chemistry, and materials science [1]. In the RSA model objects of a specified shape are randomly and sequentially deposited onto a substrate without overlapping each other. The adsorbed particles are permanently fixed at their spatial positions and they affect the geometry of all later placements so the jamming coverage θ_{jam} is less than in close packing. The kinetic properties of a deposition process are described by the time evolution of the coverage $\theta(t)$, which is the fraction of the substrate area occupied by the adsorbed particles. For discrete substrates the late time kinetics of the process is described by the time dependence:

$$\theta(t) = \theta_{jam} - Ae^{-t/\tau}, \tag{1}$$

where A and τ are parameters that depend on the shape, orientational freedom of the objects, and on the substrate dimensionality and heterogeneity.

During the process of irreversible deposition, coverage increases causing the growth of clusters of occupied sites. Percolation assumes the formation

[1]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: ljupka@uns.ac.rs

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: ivanalon@uns.ac.rs

of a large cluster that connects two opposite sides of the substrate [2]. In order to describe the inhomogenuous surfaces in the RSA model, anisotropy in the deposition procedure can be imposed [3]. Namely, the probability for deposition is different along different directions of the underlying lattice. This simple modification introduces preferential direction in the deposition process and causes a specific "patterning" of the deposited layer.

The aim of this research was to investigate the percolation properties in irreversible deposition of objects of various shapes under anisotropic conditions, with various probabilities p for depositions in a certain direction, i.e. for various values of the order parameter.

2. Definition of the model and the simulation method

Anisotropic irreversible deposition of extended objects that are modeled by self-avoiding walks on a triangular lattice (Table I) is studied by Monte Carlo simulations. Simulations are performed for k-mers (denoted as (A), angled objects (B), and triangles (C) up to the length $\ell = 20$, and for rhombuses (D) less then $\ell = 15$ because the percolation cannot be reached for larger ones.

shape	$A\ell$	shape	$B\ell$	shape	$C\ell$	shape	$D\ell$
	$\ell = 1$		$\ell = 2$		$\ell = 2$		$\ell = 3$
	$\ell = 2$		$\ell = 3$		$\ell = 5$		$\ell = 8$
...	...	...	...	...	...	...	...
...	$\ell = 10$	...	$\ell = 20$	...	$\ell = 20$	...	$\ell = 24$

TABLE I: Illustration of the construction of objects larger than basic ones. Simulation are performed for line segments (k-mers) Al; angled objects of sizes Bl; triangles Cl and rhombuses Dl.

Anisotropy is introduced by imposing different probabilities of deposition in the three possible directions. The choice of the horizontal direction occurs with probability p and for each of the other two directions with probability $(1 - p)/2$. Hence, the value of $p = 1/3$ corresponds to the isotropic case. The probability p actually stands for the order parameter characterizing the degree of anisotropy.

At each Monte Carlo step a lattice site is selected at random. If the selected site is unoccupied, one of the six possible orientations is chosen with the corresponding probability and deposition of the object is tried in that orientation. We fix the beginning of the walk that makes the shape at the selected site and search whether all successive ℓ sites are unoccupied. If so, we occupy these $\ell + 1$ sites and place the object. If the attempt fails, a new site is selected at random. The jamming limit is reached when the object of the specified shape can't be placed in any position on the lattice. The coverage of the surface is increased in the RSA process up to the percolation threshold, when there appears a cluster that extends through the whole system - from the left to

the right side of the lattice. The tree-based union/find algorithm is used to determine the percolation threshold [4].

The Monte Carlo simulations are performed on a triangular lattice of size up to $L = 3200$. Periodic boundary conditions are used in all directions. The time is counted by the number of attempts to select a lattice site and scaled by the total number of lattice sites. In all the simulations the data are averaged over 500 independent runs.

3. Results and discussion

Jamming coverages and percolation thresholds are determined for a large variety of objects shown in Table I. The effective percolation threshold θ_p, measured for a finite lattice, approaches the asymptotic value θ_p^* ($L \to \infty$) via the power law:

$$\theta_p - \theta_p^* \propto L^{-1/\nu} \, . \tag{2}$$

The theoretical value for the critical exponent is $\nu = 4/3$ for two-dimensional systems. The validity of the finite-size scaling is confirmed in the whole range of parameter p. Moreover, the asymptotic value of the percolation threshold θ_p^* coincides with the value of θ_p obtained for the largest lattice, within the limits of the statistical error. Scaling of the standard deviation σ_p, according to the relation:

$$\sigma_p \propto L^{-1/\nu} \, , \tag{3}$$

was plotted in a way that the values of σ_p are shown vs. L on a log-log scale and they lie on parallel straight lines. The slope of these lines corresponds to the exponent $1/\nu = 0.75 \pm 0.01$, so the results of simulations confirmed the theoretical value of critical exponent.

Results for jamming densities and percolation thresholds are obtained for various values of the order parameter p ranging from $p = 0$ to $p = 1$. Dependence of the percolation threshold θ_p on the order parameter p are presented in Figure 1 for various sizes of the basic objects from Table I. It can be seen that the lowest values of θ_p are obtained for the isotropic case ($p = 1/3$). Percolation threshold increases with the degree of anisotropy, having the largest values for fully oriented objects in one direction ($p = 1$). This property is most pronounced for k-mers (Fig. 1 a)). On the other hand, for triangles (C), neither jamming nor percolation are affected by the anisotropy (Fig. 1 c)). The relative increase of the percolation threshold due to the complete alighnement (maximum anisotropy) defined as:

$$R = \frac{\theta_p(p = 1) - \theta_p(p = 1/3)}{\theta_p(p = 1/3)} \, , \tag{4}$$

is largest for the k-mers of length $\ell = 11$ ($A11$), and its value is 28.6%. Simulations has shown that the impact of anisotropy on the percolation properties is largest for the elongated objects (A) on the contrary to the compact rhombuses (D), that are less affected by the anisotropic conditions. With the exception of fully symmetrical objects, the increase in the anisotropy always results in

higher percolation thresholds. It should be emphasized that the statistical errors are typically of the order of 10^{-3}, and in all the figures the error bars are smaller than the symbol size.

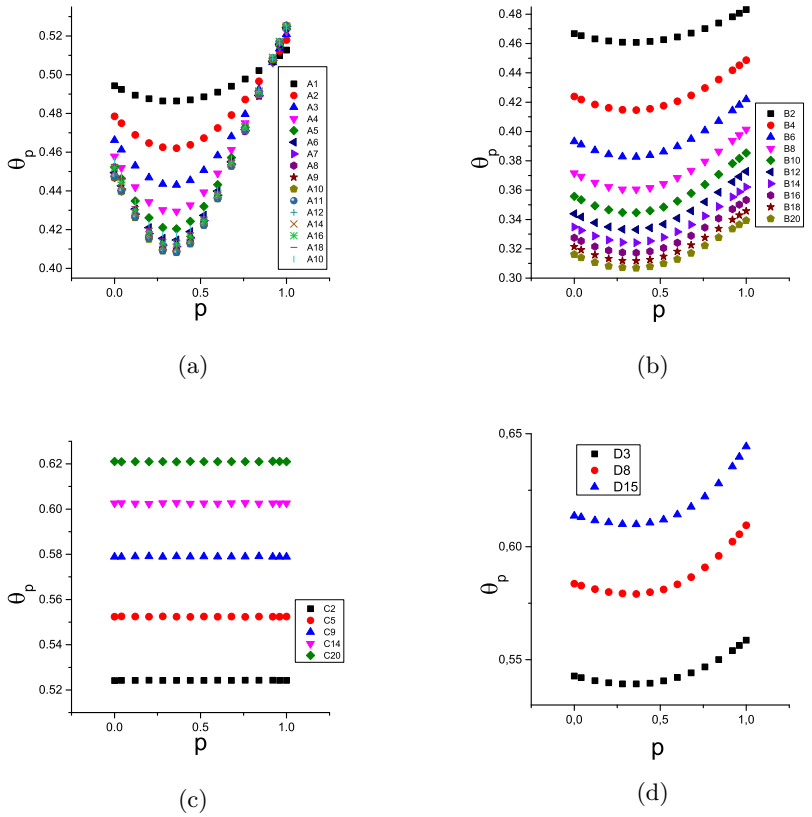


FIG. 1: Dependence of the percolation threshold θ_p on the probability p for deposition in the horizontal direction, i.e. on the order parameter, for various basic objects from Table I and for the larger sizes of these shapes: a) (A); b) (B); c) (C); d) (D)

In the isotropic case, percolation threshold decreases with ℓ for shorter k-mers, reaches a smooth minimum for $\ell \simeq 11$, and slightly increases for longer k-mers [5]. Dependence of the percolation threshold on the length of various objects from Table I is shown in Figure 2 for the values of the order parameter: $p =$ 0; 0.12; 0.28; 0.44; 0.60; 0.76; 0.92 and 1. Introducing the anisotropy shifts the minimum towards lower k-mer lengths. For highly anisotropic conditions a qualitatively different behavior is obtained - θ_p increases with the k-mer length, reaches a maximum, and decreases for longer k-mers (Fig. 2 a)). For the angled objects (B) θ_p decreases with ℓ (Fig. 2 b)), but for the triangles (C) increases with the object size (Fig. 2 c)). There is an essential difference between deposition of elongated objects and the compact ones. This feature

is connected with difference in the geometry exclusion effects. Blocking of the substrate area is enhanced by the growth of the k-mer length, making the surface more porous. The porosity of the surface is also responsible for the decrease of θ_p with the length of the angled objects (B) from Table I. On the other hand, for compact objects, such as triangles (C) and rhombuses (D), percolation threshold increases with their size. This is the consequence of a low connectivity of these objects.

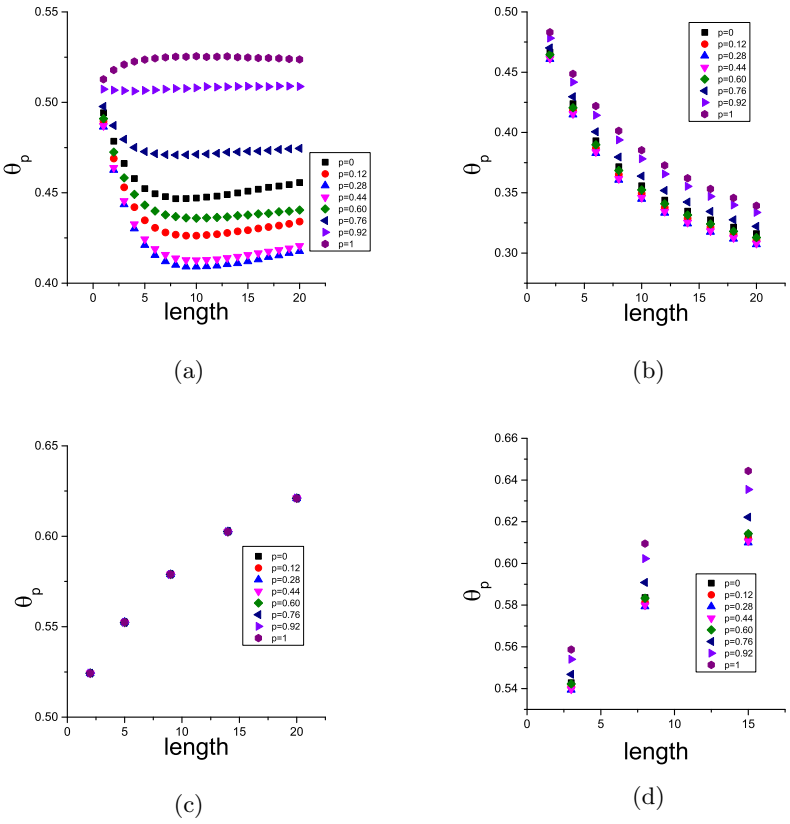


FIG. 2: Dependence of the percolation threshold θ_p on the length of various objects from Table I:
a) (A); b) (B); c) (C); d) (D); for the values of the order parameter:
$p = 0;\ 0.12;\ 0.28;\ 0.44;\ 0.60;\ 0.76;\ 0.92$ and 1.

4. Concluding remarks

Percolation properties in irreversible deposition under anisotropic conditions substrates have been investigated. Objects of various shapes were examined and percolation thresholds were determined for numerous degrees of deposition anisotropy characterized by the order parameter p taking values from $p = 0$ to $p = 1$. It was found that the percolation threshold increases

with the degree of anisotropy and have the maximum values for fully oriented objects in one direction. The relative increase of the percolation threshold for the maximum anisotropy ($p = 1$), compared to the isotropic case ($p = 1/3$), is largest for k-mers of length $\ell = 11$. k-mers are these that give the lowest value of θ_p in the isotropic case. On the contrary, percolation of the triangles with the symmetry axis of third order, is not affected by the anisotropy of the underlying lattice.

Crucial difference in the percolation properties of elongated and compact objects was also found. High porosity of the deposit and the high connectivity of elongated objects result in low percolation thresholds for the isotropic, as well as for the anisotropic deposition. θ_p decreases with the size for these objects. On the other hand, low connectivity of the compact objects, like triangles and rhombuses, results in higher percolation thresholds, while θ_p increases with the object size. (It should be noted that the k-mers show a more complex behavior. For the isotropic case θ_p decreases with ℓ for shorter k-mers, reaches a minimum, and increases for longer k-mers. In the presence of anisotropy the minimum is shifted towards shorter k-mers. For highly anisotropic deposition, percolation threshold practically does not depend on the k-mer length.)

Acknowledgement

This work was supported by the Department of Fundamental Sciences, Faculty of Technical Sciences through the project named "Improving the teaching process in the English language in fundamental disciplines".

References

[1] A. Dabrovski, "Adsorption from theory to practice" *Adv. Colloid Interface Sci.*, vol. 93, pp. 135-224, 2001.

[2] G. Condrat and A. Pekalski, "Percolation and jamming in random sequential adsorption of linear segments on a square lattice", *Phys. Rev. E*, vol. 63, pp. 051108, 2001.

[3] Lj. Budinski Petković, I. Lončarević, and Z. M. Jakšić, S. B. Vrhovac and N. Švrakić, "Simulation study of anisotropic random sequential adsorption of extended objects on a triangular lattice", *Phys. Rev. E*, vol. 84, pp. 051601, 2011.

[4] D. Dujak, A. Karač, Lj. Budinski Petković, I. Lončarević, and Z. M. Jakšić and S. B. Vrhovac, "Percolation in random sequential adsorption of mixtures on a triangular lattice", *J. Stat. Mech.*, pp. 113210, 2019.

[5] N. I. Lebovka, N. N. Karmazina, Y. Y. Tarasevich and V. V. Laptev, "Random sequential adsorption of partially oriented k-mers on a square lattice", *Phys. Rev. E*, vol. 84, pp. 061603, 2011.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

PROCESS CALCULI COMPARISONS: A JOURNEY THROUGH ENCODINGS [1]

Jovana Dedeić [2]

https://doi.org/10.24867/META.2024.04 Professional paper

Abstract. This paper provides a comprehensive literature review on the comparison of process calculi, specifically focusing on the encodings. It explores systematic methods for evaluating expressive power through encodings or proofs of their absence. The survey includes commonly used encodability criteria, general frameworks for assessing encoding quality, and methods for comparing these criteria. The insights gained from this review contribute to a better understanding of the concepts of encodings.

AMS Mathematics Subject Classification (2020): 03B30, 68V20

Key words and phrases: concurrent systems, process calculi, encodings, encodability criteria, expressiveness

1. Introduction

The complexity of programs necessitates advanced models for analysis, which highlights the use of formal methods to analyze the properties of complex systems. Formal methods rely on mathematical and logical frameworks for specifying and verifying intricate systems. They utilize modeling languages with precise mathematical syntax and semantics, enabling the demonstration of system properties and verification through mathematical proofs. Examples of formal methods for concurrency include Petri nets, communicating state machines, and process calculi.

In the following, this paper offers an extensive survey of existing literature concerning the evaluation of process calculi, with a specific emphasis on encoding techniques.

2. Expressiveness of Concurrent Calculi

The concept of expressive power in programming languages traces back to the late 1960s, notably with Landin's [12] work on a unified framework for describing language families. Felleisen's [7] framework in the 1990s contributed to studying relative expressiveness, emphasizing eliminable syntactic symbols and definitional extensions between languages. Mitchell [15] and Riecke [25] in 1993

[1]The author acknowledge the financial support of Department of Fundamentals Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Improving the teaching process in the English language in fundamental disciplines".

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: radenovicj@uns.ac.rs

analyze abstraction-preserving reductions in functional languages, focusing on their impact on program abstraction and expressive power.

Shapiro [27] in 1989 was the first to study expressiveness issues for concurrent languages. He suggested using embedding as a method to compare concurrent logic programming languages that are relatively similar, enabling a focused examination of their distinctions.

The introduction of the π-calculus in the early 1990s significantly advanced the exploration of expressiveness issues within process calculi. This simplicity and adaptability of name-passing, as demonstrated in the π-calculus, spurred numerous works proposing variants or extensions of it. Expressiveness studies for the π-calculus were essential for understanding its fundamental properties, identifying its inherent sources of expressiveness, and exploring relationships between its variants. Examples include studies on polyadic-to-monadic π-calculus translation properties ([24, 30]), point-to-point versus broadcasting communication ([6]), different choice operators ([16, 17]), mechanisms for synchronous and asynchronous communication ([1, 21]), and λ-calculus into the π-calculus [14]. Also, there are comparing various subcalculi of the π-calculus [9], comparing different process calculi and some separation results [2, 3, 4, 11, 13, 22, 23, 26, 28]. Due to varied motivations, each work proposed its own definition of encoding based on specific working intuitions or necessities.

Calculi undergo evaluation, with expressiveness being a key criterion for their assessment. Yet, the theory of concurrency lacks a formal definition of language expressiveness. Given the diversity of concurrency models, a unified theory that encompasses them all is unlikely [10].

In the study of expressiveness, the *concept of encoding* plays a crucial role, as outlined in [**?**]. This encoding, represented as $[\![\cdot]\!]$, translates terms from a *source calculus* to terms in a *target calculus*, adhering to correctness criteria that encompass both *structural* and *semantic* dimensions of the function $[\![\cdot]\!]$. The challenge lies in defining these criteria due to diverse practical needs, which hinders the development of a unified theory for language comparison. Various works in the literature, including those in [8, 19, 20, 21], emphasize the lack of consensus regarding a standardized set of criteria for meaningful encoding. These criteria are often customized to specific analysis requirements. Furthermore, in [20], the author underscores the significance of systematically comparing the increasing number of process models, highlighting a pivotal area of research in the field.

The concept of expressiveness raises questions about its purpose. Expressiveness studies typically focus on two key aspects: *encodability* and *non-encodability*. Encodability investigates the presence of an encoding, while non-encodability deals with the absence of such an encoding. Consider two languages $\mathcal{L}_1$ and $\mathcal{L}_2$. To establish that $\mathcal{L}_1$ is more expressive than $\mathcal{L}_2$, both encodability and non-encodability results must be provided. This means presenting/proving an encoding $[\![\cdot]\!] : \mathcal{L}_2 \longrightarrow \mathcal{L}_1$ and simultaneously demonstrating that an encoding $[\![\cdot]\!] : \mathcal{L}_1 \longrightarrow \mathcal{L}_2$ does not exist.

Another classification in the literature involves *absolute* and *relative expressiveness*. Absolute expressiveness focuses on a single process calculus, leading to either positive or negative absolute results depending on the ability to solve

specific problems. Conversely, relative expressiveness compares two languages, determining whether they have the same expressive power or evaluating the impact of specific operators on their expressiveness.

2.1. The Notation of Encoding

As stated in [5], an encoding function maps processes from a source calculus to a target calculus, indicating that the target language is as expressive as the source, or vice versa if no encoding exists. This approach, combining positive and negative encodability results, helps establish differences in expressivity between languages. Translations are often seen as syntax mappings between languages ($\mathcal{L}_{\mathbf{s}}$ to $\mathcal{L}_{\mathbf{t}}$), where $\mathcal{P}_{\mathbf{s}}$ and $\mathcal{P}_{\mathbf{t}}$ represent sets of process terms. Trivial mappings, like translating every process to inaction, are valid but don't reflect full expressive power. To evaluate encoding quality and avoid trivial mappings, encodings are assessed using a set of correctness criteria. References for further exploration of correctness criteria include [8, 18, 20, 29, 31].

It is common to relate the sources and target calculus through valid encodings (simply encodings). To define valid encodings, we adopt five correctness criteria formulated by Gorla [8]: (1) compositionality, (2) name invariance, (3) operational correspondence, (4) divergence reflection, (5) success sensitiveness. The first two criteria are structural criteria, while the other three are semantic criteria. Structural criteria describe the static structure of the encoding, whereas the semantic criteria describe its dynamics – how the behavior of encoded terms relates to that of source terms, and vice versa. As stated in [20], structural criteria are needed in order to measure the expressiveness of operators in contrast to expressiveness of terms. As for semantic criteria, operational correspondence is divided into *completeness* and *soundness* properties: the former ensures that the behavior of a source process is preserved by the translation in the target calculus; the latter ensures that the behavior of a translated (target) process corresponds to that of some source process. Divergence reflection ensures that a translation does not introduce spurious infinite computations, whereas success sensitiveness requires that source and translated terms behave in the same way with respect to some notion of *success*.

Following [4, 5, 8], we start by defining an abstract notion of calculus:

Definition 2.1 (Calculus [5])**.** We define a *calculus* as a triple $(\mathcal{P}, \longrightarrow, \approx)$, where: $\mathcal{P}$ is a set of processes; $\longrightarrow$ is its associated reduction semantics, which specifies how a process computes on its own; $\approx$ is an equality on processes, useful to describe the abstract behavior of a process, which is a congruence at least with respect to parallel composition.

We will further assume a countably infinite set of names, usually denoted $\mathcal{N}$. Accordingly, the abstract definition of encoding refers to those names.

Definition 2.2 (Encoding [5])**.** Let $\mathcal{N}_{\mathbf{s}}$ and $\mathcal{N}_{\mathbf{t}}$ be countably infinite sets of source and target names, respectively. An *encoding* of the source calculus $(\mathcal{P}_{\mathbf{s}}, \longrightarrow_{\mathbf{s}}, \approx_{\mathbf{s}})$ into the target calculus $(\mathcal{P}_{\mathbf{t}}, \longrightarrow_{\mathbf{t}}, \approx_{\mathbf{t}})$ is a tuple $(\llbracket\cdot\rrbracket, \varphi_{\llbracket\cdot\rrbracket})$ where $\llbracket\cdot\rrbracket : \mathcal{P}_{\mathbf{s}} \longrightarrow \mathcal{P}_{\mathbf{t}}$ denotes a *translation* and $\varphi_{\llbracket\cdot\rrbracket} : \mathcal{N}_{\mathbf{s}} \longrightarrow \mathcal{N}_{\mathbf{t}}$ denotes a *renaming policy* for $\llbracket\cdot\rrbracket$.

The renaming policy defines the way names from the source calculus are translated into the target calculus. A valid encoding cannot depend on the particular names involved in source processes.

As in [5], we shall use the following notations. We write $\longrightarrow^*$ to denote the reflexive, transitive closure of $\longrightarrow$. Also, given $k \geq 1$, we will write $P \longrightarrow^k P'$ to denote k consecutive reduction steps leading from P to P'. That is, $P_1 \longrightarrow^k P_{k+1}$ holds whenever there exist $P_2, \ldots, P_k$ such that $P_1 \longrightarrow P_2 \longrightarrow \cdots \longrightarrow P_k \longrightarrow P_{k+1}$.

We apply compositionality, as per [4], employing a context that combines translated subterms based on the source operator's combination of subterms.

In this paper, for definition of the operational correspondence we follow more strict criteria than Gorla [8]. We rely on [4, 5] form of operational completeness that explicitly describes the number of steps required to mimic a step in the source language. Also, for divergence reflection we use the following definition:

Definition 2.3 (Divergence [5])**.** A process P diverges, written $P \longrightarrow^\omega$, if there exists an infinite sequence of processes $\{P_i\}_{i\geq 0}$ such that $P = P_0$ and for any i, $P_i \longrightarrow P_{i+1}$.

To formulate success sensitiveness, as in [5], we assume that both source and target calculi contain the same success process $\checkmark$. Also, we assume that $\Downarrow$ is a predicate that asserts reducibility (in a "may" modality) to a process containing an unguarded occurrence of $\checkmark$.

Definition 2.4 (Success [5])**.** Let $(\mathcal{P}, \longrightarrow, \approx)$ be a calculus. A process $P \in \mathcal{P}$ (may)-succeeds, denoted $P \Downarrow$, if it is reducible to a process containing an unguarded occurrence of $\checkmark$, i.e., if $P \longrightarrow^* P'$ and $P' = C[\checkmark]$ for some P' and context $C[\bullet]$.

In the following definition, we formally present the *five criteria* for valid encoding:

Definition 2.5 (Valid Encoding [5])**.** Let $\mathcal{L}_\mathbf{s} = (\mathcal{P}_\mathbf{s}, \longrightarrow_\mathbf{s}, \approx_\mathbf{s})$ and $\mathcal{L}_\mathbf{t} = (\mathcal{P}_\mathbf{t}, \longrightarrow_\mathbf{t}, \approx_\mathbf{t})$ be source and target calculi, respectively, each with countably infinite sets of names $\mathcal{N}_\mathbf{s}$ and $\mathcal{N}_\mathbf{t}$. An encoding $(\llbracket\cdot\rrbracket, \varphi_{\llbracket\cdot\rrbracket})$, where $\llbracket\cdot\rrbracket : \mathcal{P}_\mathbf{s} \longrightarrow \mathcal{P}_\mathbf{t}$ and $\varphi_{\llbracket\cdot\rrbracket} : \mathcal{N}_\mathbf{s} \longrightarrow \mathcal{N}_\mathbf{t}$, is a *valid encoding* if it satisfies the following criteria:

(1) **Compositionality**: $\llbracket\cdot\rrbracket$ is *compositional* if for every n-ary ($n \geq 1$) operator $\mathtt{op}$ on $\mathcal{P}_s$ and for every set of names $\mathtt{N}$ there is an n-adic context $C^{\mathtt{N}}_{\mathtt{op}}[\bullet_1, \ldots, \bullet_n]$ such that, for all $P_1, \ldots, P_n$ with $\mathtt{fn}(P_1, \ldots, P_n) \subseteq \mathtt{N}$ it holds that $\llbracket\mathtt{op}(P_1, \ldots, P_n)\rrbracket = C^{\mathtt{N}}_{\mathtt{op}}[\llbracket P_1\rrbracket, \ldots, \llbracket P_n\rrbracket]$.

(2) **Name invariance**: $\llbracket\cdot\rrbracket$ is *name invariant* if for every substitution $\sigma : \mathcal{N}_\mathbf{s} \longrightarrow \mathcal{N}_\mathbf{s}$ there is a substitution $\sigma' : \mathcal{N}_\mathbf{t} \longrightarrow \mathcal{N}_\mathbf{t}$ such that (i) for every $a \in \mathcal{N}_\mathbf{s} : \varphi_{\llbracket\cdot\rrbracket}(\sigma(a)) = \sigma'(\varphi_{\llbracket\cdot\rrbracket}(a))$ and (ii) $\llbracket\sigma(P)\rrbracket = \sigma'(\llbracket P\rrbracket)$.

(3) **Operational correspondence**: $\llbracket\cdot\rrbracket$ is *operational corresponding* if it satisfies the two requirements:

   a) **Completeness**: If $P \longrightarrow_\mathbf{s} Q$ then there exists k such that $\llbracket P\rrbracket \longrightarrow^k_\mathbf{t} \approx_\mathbf{t} \llbracket Q\rrbracket$.

   b) **Soundness**: If $\llbracket P\rrbracket \longrightarrow^*_\mathbf{t} R$ then there exists P' such that $P \longrightarrow^*_\mathbf{s} P'$ and $R \longrightarrow^*_\mathbf{t} \approx_\mathbf{t} \llbracket P'\rrbracket$.

(4) **Divergence reflection**: $[\![\cdot]\!]$ *reflects divergence* if, for every P such that $[\![P]\!] \longrightarrow^{\omega}_{\mathbf{t}}$, it holds that $P \longrightarrow^{\omega}_{\mathbf{s}}$.

(5) **Success sensitiveness**: $[\![\cdot]\!]$ is *success sensitive* if, for every $P \in \mathcal{P}_{\mathbf{s}}$, it holds that $P \Downarrow$ if and only if $[\![P]\!] \Downarrow$.

Interested readers are encouraged to refer to [4, 5] to see how the authors defined another criterion for their encodings, known as "efficiency". There, the authors consider the number of reduction steps required in the target language to mimic the behavior of the source language.

References

[1] D. Cacciagrano, F. Corradini, and C. Palamidessi, "Separation of synchronous and asynchronous communication via testing", *Theor. Comput. Sci.* 386, 3, pp. 218–235, 2007.

[2] M. Carbone and S. Maffeis, "On the expressive power of polyadic synchronisation in pi-calculus", *Nordic Journal of Computing*, 10, 2, pp. 70–98, 2003.

[3] O. Dardha, E. Giachino, and D. Sangiorgi, "Session types revisited", *Information and Computation*, 256, pp. 253–286, 2017.

[4] J. Dedeić, J. Pantović, and J. A. Pérez, "On primitives for compensation handling as adaptable processes", *Journal of Logical and Algebraic Methods in Programming*, p. 100675, 2021.

[5] J. Dedeić, *Relative Expressiveness of Process Calculi with Dynamic Update and Runtime Adaptation*, PhD dissertation, Faculty of Technical Sciences, University of Novi Sad, 2022.

[6] C. Ene and T. Muntean, "Expressiveness of point-to-point versus broadcast communications", *In Proc. of FCT. Lecture Notes in Computer Science*, vol. 1684, Springer, pp. 258–268, 1999.

[7] M. Felleisen, "On the expressive power of programming languages", *Sci. Comput. Program.*, 17, 1-3, pp. 35–75, 1991.

[8] D. Gorla, "Towards a unified approach to encodability and separation results for process calculi", *Inf. Comput.*, 208, 9, pp. 1031–1053, 2010.

[9] K. Honda and M. Tokoro, "An object calculus for asynchronous communication", *In Proc. of ECOOP'91, Geneva, Switzerland, July 15-19, 1991, Proceedings*, vol. 512 of Lecture Notes in Computer Science, pp. 133–147. Springer, 1991.

[10] J. A. Pérez, *Higher-Order Concurrency: Expressiveness and Decidability Results*, PhD dissertation,, University of Bologna, Department of Computer Science, 2010.

[11] I. Lanese, J. A. Pérez, D. Sangiorgi, and A. Schmitt, "On the expressiveness and decidability of higher-order process calculi", *Information and Computation*, 209, 2, pp. 198–226, 2011.

[12] P. J. Landin, "The Next 700 Programming Languages", *Communications of the ACM*, 9, 3, pp. 157–166, 1966.

[13] D. Medić and C. A. Mezzina, "Static VS dynamic reversibility in CCS", *8th International Conference, RC 2016, Bologna, Italy, July 7-8, 2016, Proceedings*, vol. 9720 of Lecture Notes in Computer Science, pp. 36–51. Springer, 2016.

[14] R. Milner, "Functions as processes", *Mathematical Structures in Computer Science*, 2, 2, pp. 119–141, 1992.

[15] J. C. Mittchell, "On abstraction and the expressive power of programming languages", *Sci. Comput. Program.*, 21, 2, pp. 141–163, 1993.

[16] U. Nestmann, "What is a "good" encoding of guarded choice?", *Inf. Comput.*, 156, 1-2, pp. 287–319, 2000.

[17] U. Nestmann and B. C. Pierce, "Decoding choice encodings", *Inf. Comput.*, 163, 1, pp. 1–59, 2000.

[18] U. Nestmann, *On Determinacy and Nondeterminacy in Concurrent Programming*, PhD dissertation, Universität Erlangen-Nürnberg, 1996.

[19] U. Nestmann, "Welcome to the jungle: A subjective guide to mobile process calculi", *In Proceedings of the 17th International Conference on Concurrency Theory*, CONCUR'06, pp. 52–63, Berlin, Heidelberg, 2006.

[20] J. Parrow, "Expressiveness of process algebras", *Electronic Notes in Theoretical Computer Science*, 209, pp. 173–186, 2008.

[21] C. Palamidessi, "Comparing the expressive power of the synchronous and asynchronous pi-calculi", *Mathematical Structures in Computer Science*, 13, 5, pp. 685–719, 2003.

[22] C. Palamidessi, V. A. Saraswat, F. D. Valencia, and B. Victor, "On the expressiveness of linearity vs persistence in the asynchronous π-calculus", *In Proc. of the 21st IEEE Symposium on Logic in Computer Science (LICS 2006)*, Seattle, WA, USA, August 12-15, 2006, pp. 59–68. IEEE Computer Society, 2006.

[23] I. Prokic and H. T. Vieira, "The C_π-calculus: A model for confidential name passing", *Journal of Logical and Algebraic Methods in Programming*, 119, p. 100622, 2021.

[24] P. Quaglia and D. Walker, "Types and full abstraction for polyadic pi-calculus", *Inf. Comput.*, 200, 2, pp. 215–246, 2005.

[25] J. G. Riecke, "Fully abstract translations between functional languages", *Mathematical Structures in Computer Science*, 3, 4, pp. 387–415, 1993.

[26] D. Sangiorgi, *Expressing Mobility in Process Algebras: First-Order and Higher-Order Paradigms*, PhD thesis, University of Edinburgh, UK, 1993.

[27] E. Y. Shapiro, "The family of concurrent logic programming languages", *ACM Comput. Surv.*, 21, 3, pp. 413–510, 1989.

[28] M. G. Vigliotti, I. Phillips, and C. Palamidessi, "Tutorial on separation results in process calculi via leader election problems", *Theoretical Computer Science*, 388, 1, pp. 267–289, 2007.

[29] N. Yoshida, "Minimality and separation results on asynchronous mobile processes - representability theorems by concurrent combinators", *Theoretical Computer Science*, 274, 1, pp. 231–276, 2002.

[30] N. Yoshida, "Graph types for monadic mobile processes", *In Proc. of FSTTCS. Lecture Notes in Computer Science*, vol. 1180, Springer, pp. 371–386, 1996.

[31] K. Peters, *Translational expressiveness: comparing process calculi using encodings*, PhD dissertation, Technische Universität Berlin, Fakultät IV - Elektrotechnik und Informatik, Berlin, 2012.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

ISTORIJSKI PREGLED RAZVOJA TEORIJE VEROVATNOĆE: OD BACANJA KOCKICA DO MODERNE DEFINICIJE I TEORIJE ODLUČIVANJA

Nataša Duraković [1] **i Marko Uščebrka** [2]

https://doi.org/10.24867/META.2024.05 Review article

Sažetak. U ovom preglednom radu je dat istorijski prikaz razvoja teorije verovatnoće kroz vekove, kao i moderna aksiomatska definicija verovatnoće koja se danas koristi. Takođe, date su neke od modifikacija verovatnoće u opštoj teoriji mere i pseudo-analizi i osvrt na teoriju odlučivanja, kao granu primenjene verovatnoće.

AMS klasifikacija (2020): 01A40-01A61, 60-03

Ključne reči: verovatnoća, neaditivne mere, pseudoverovatnoća, teorija odlučivanja.

1. Igre na sreću i verovatnoća

Verovatnoća, slučajnost i šanse su pojmovi koji su poznati od davnina. Oko 1200. godine pre nove ere igrala se drevna igra sa četiri kosti kopitara, koje bi bile oblikovane u kocke, na kojima bi se pravila mala udubljenja (Slika 1.). Odatle je i došla ideja o tačkama na kockicama koje danas koristimo. Poznato je i da su Rimljani uživali u igrama sa kockicama (Slika 2.). Car Klaudije je u svojoj kočiji imao sto kako bi mogao da igra igre sa kockicama dok se vozi. Poznato je da je napisao knjigu u kojoj je razmatrao kako pobediti u kockama (lat. De arte aleae), koja je izgubljena. Međutim, kako u to vreme većina ljudi nije verovala u slučajnost, jer sve što bi se dogodilo bi se pripisivalo bogovima, ideja o verovatnoći se nije mogla razvijati ([2]).

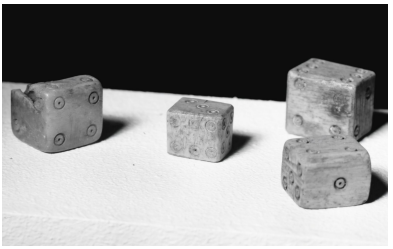
Problem u igri na sreću, u istoriji poznat kao „Problem nedovršene igre“, se pojavljuje 1654. godine u prepisci francuskih matematičara Pjera de Fermaa i Bleza Paskala i smatra se prvim koracima u začetku verovatnoće kao matematičke teorije ([10]).

Dva igrača, A i B, igraju jednu za drugom igre i u svakoj igri jedan od igrača koji pobedi, osvaja jedan poen. Ukupni pobednik je onaj igrač koji prvi osvoji tri poena. Svaki od igrača je uložio 32 novčica, međutim, igra je prekinuta nakon što je igrač A osvojio dva poena, a igrač B jedan poen. Postavlja se pitanje: kako raspodeliti igračima A i B uloženih 64 novčića tako da raspodela bude fer?

[1]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: natasa.durakovic@uns.ac.rs

[2]Muzej Vojvodine, Novi Sad, e-mail: marko.uscebrka@muzejvojvodine.org.rs

Slika 1: Kockice od kostiju



Slika 2: Rimske kockice

Oba matematičara su predložili rešenje problema sa istim konačnim rešenjem: igrač A bi trebalo da dobije 48 novčića, a igrač B 16 novčića.

Ferma je utvrdio da su dovoljne još dve igre da bi se znao ukupni pobednik. Naime, on je svoj odgovor predstavio preko šansi, tj. verovatnoće, jer postoje četiri moguća ishoda u dve odigrane igre, pri čemu je svaki podjednako verovatan. Ishodi su: igrač A može pobediti dva puta, ili prvo da pobedi igrač A pa igrač B, ili prvo da pobedi igrač B pa igrač A, ili da igrač B pobedi dva puta. Kako bi od četiri moguća ishoda samo poslednji rezultirao ukupnom pobedom igrača B, šansa da ukupni pobednik bude igrač A je $3 : 1$. Stoga, fer raspodela 64 novčica bi bila 48 novčica za igrača A i 16 novčica za igrača B.

Paskal je smatrao Fermaovo rešenje nezgrapnim i predložio je da se problem reši ne u smislu šansi, već u smislu količine koju on naziva „očekivanje“. Ako se pretpostavi da je igrač B već pobedio u sledećoj igri, tada su pozicije oba igrača izjednačene, svaki bi dobio po dva poena i svaki bi imao pravo na 32 novčića. Međutim, kako 32 novčića koja pripadaju igraču B zavise od pretpostavke da je pobedio u sledećoj odigranoj igri, ta igra se može tretirati kao fer igra sa ulogom od 32 novčića od kojih svaki od igrača ima očekivanje od 16 novčića. Stoga, igrač A dobija još 16 novčića na već osvojenih 32, a igrač B 16 novčića.

Ferma i Paskal nisu bili prvi koji su dali matematička rešenja za probleme poput ovog. Više od jednog veka ranije, italijanski matematičar, lekar i kockar Đirolamo Kardano istraživao je šanse za izvlačenje asova iz špila karata i bacanja sedmice sa dve kocke. On je prvi koji je shvatio da postoji jednaka šansa da se baci 1, 3 ili 5 kao i da se baci 2, 4 ili 6. Prvi je došao do ideje za izračunavanje šansi za dobitak na igrama na sreću tako što o prebrojavao povoljne ishode i upoređivao sa ukupnim brojem slučajeva i želeo je da verovatnoći ishoda dodeli broj od 0 do 1. Svoja zapažanja je zabeležio u knjizi koja je, nažalost, objavljena tek 1663. godine, kada su elementi teorije šansi već bili dobro poznati matematičarima u Evropi.

Jakob Bernuli, švajcarski matematičar je 1705. godine formulisao i primenio zakon velikih brojeva. U svom rasuđivanju je naveo da ako se izračuna proporcija ishoda eksperimenta nakon velikog broja ponavljanja tog eksperimenta, onda će ta proporcija biti tačan prikaz prave teorijske verovatnoće tog ishoda. Rad je objavljen 1713. godine, nakon njegove smrti. Dva veka kasnije, Bernulijevu teoriju su testirala tri čoveka. Dok je sedeo u nemačkom zatvoru tokom Drugog svetskog rata, grof Bufon je bacio novčić 4040 puta i dobio grb 2048 puta. Južnoafrički matematičar, Džon Kerič, bacio je novčić 10.000 puta, a grb

se pojavio 5067 puta. Engleski statističar Karl Pirson, bacio je novčić 24.000 puta i grb je pao 12.012 puta.

Jedna od velikih prekretnica u razvoju teorije verovatnoće je otkriće normalne raspodele. Francuski matematičar Abraham de Moavr je primetio da kada se broj bacanja novčića povećava, oblik binomne raspodele se približava veoma glatkoj krivoj, koja je danas poznata kao normalna (Gausova) kriva. Važnost normalne krive prvenstveno proizilazi iz činjenice da su raspodele mnogih prirodnih pojava približno jednake normalnoj raspodeli. Jedna od prvih primena normalne distribucije bila je analiza grešaka merenja u astronomskim posmatranjima, koje su nastajale zbog nesavršenih instrumenata. Galilej je u 17. veku primetio da su ove greške simetrične i da su se male greške dešavale češće od velikih grešaka. Nemački matematičar Karl Fridrih Gaus je 1809. razvio formulu za normalnu raspodelu i pokazao da se greške dobro uklapaju u ovu raspodelu.

Francuski matematičar Pjer-Simon de Laplas je početkom 19. veka nazvao verovatnoću „dobrim razumom svedenim na proračun“. Ova ideja nije bila toliko naučna kao što se možda čini. Naime, postojali su neki slučajevi u kojima je direktna primena matematičkih proračuna verovatnoće dovela do rezultata za koje se činilo da prkose racionalnosti. Problem poznat kao "Peterburški paradoks", koji je predložio Nikolas Bernuli, uključivao je opkladu sa eksponencijalno rastućom isplatom ([2]).

Igrač A baca ispravan novčić dok prvi put ne padne grb. Ako grb padne u prvom bacanju, igrač A dobija 2 dukata. Ako prvi put grb padne u drugom bacanju, igrač A dobija 4 dukata, a ako grb prvi put padne u n-tom bacanju, igrač A dobija 2^n dukata. Postavlja se pitanje: koliko dukata bi trebalo pripremiti kao početni ulog, da bi igra mogla da se odvija?

Bernuli je došao do zaključka da je za rešavanje ovog problema potrebna beskonačna količina dukata [3].

2. Formalna aksiomatska definicija verovatnoće

Jedna od poteškoća u razvoju verovatnoće kao matematičke teorije je ta što je trebalo dati definiciju verovatnoće koja je dovoljno precizna za upotrebu u matematici, ali dovoljno sveobuhvatna da može biti primenljiva na širok spektar pojava. Traganje za široko prihvatljivom definicijom je trajalo skoro tri veka. 1933. godine, ruski matematičar Andrej Nikolajevič Kolmogorov je izložio aksiomatski pristup definisanju verovatnoće koji predstavlja osnovu za savremenu teoriju verovatnoće.

Kada je započeo doktorske studije 1925. godine, Kolmogorov je objavio svoj prvi rad o teoriji verovatnoće, a 1929. godine, kada je završio doktorat imao je već 18 objavljenih radova među kojima su i verzije jakog zakona velikih brojeva. Kolmogorov je dao rigoroznu definiciju uslovnog očekivanja koje je kasnije

[3]Kako je verovatnoća da grb prvi put padne u prvom bacanju $\frac{1}{2}$, da prvi put padne u drugom bacanju $\frac{1}{4}$, sledi da je verovatnoća da prvi put padne u n-tom bacanju $\frac{1}{2^n}$. Tada je očekivana vrednost uloga $2\frac{1}{2} + 4\frac{1}{4} + \cdots + 2^n\frac{1}{2^n} + \cdots = \infty$.

postalo fundamentalno za definisanje Braunovskog kretanja, stohastičku integraciju i matematiku finansija. Postavio je temelje za izučavanje Markovljevih procesa ([5]).

U nastavku je prevod prve formalne definicije verovatnoće, objavljene u knjizi [6] "Grundbegriffe der Wahrscheinlichkeitsrechnung", autora Kolmogorova iz 1933. godine.

Definicija 2.1. (**Definincija Kolmogorova**)([6, 7])
Neka je E familija elemenata $\xi, \eta, \zeta, \ldots$, koji se nazivaju *elementarni događaji*, i neka je $\mathcal{F}$ familija podskupova od E. Elementi skupa $\mathcal{F}$ se nazivaju *slučajni događaji.*

I. $\mathcal{F}$ je *polje*[4] skupova.

II. $\mathcal{F}$ sadrži skup E.

III. Svakom skupu A iz $\mathcal{F}$ se dodeljuje nenegativan realan broj $P(A)$. Broj $P(A)$ se naziva *verovatnoća događaja* A.

IV. $P(E)$ je jednako 1.

V. Ako A i B nemaju zajedničkih elemenata, tada je

$$P(A + B) = P(A) + P(B).$$

Sistem skupova $\mathcal{F}$, zajedno sa definisanim dodeljivanjem brojeva $P(A)$ koje zadovoljava aksiome I-V, se naziva *prostor verovatnoće.*

U prethodnoj definiciji, svojstvo funkcije P dato u V. se naziva aditivnost.

3. Verovatnoća kao neaditivna mera

U klasičnoj teoriji mere, svojstvo aditivnosti verovatnoće kao mere je zadovoljeno. Međutim, u teoriji opštih mera se izučavaju neaditivne mere, tj. mere koje nemaju svojstvo aditivnosti. Neaditivne mere se dele na dve klase: klasu nula-aditivnih skupovnih funkcija i klasu fazi (monotonih) mera ([11, 12, 14]).

U nastavku su date definicije nekih fazi mera.

Neka je X neprazan skup i $\mathcal{P}(X)$ partitivni skup skupa X. Funkcija $m : \mathcal{P}(X) \to [0, 1]$ se naziva *osnovna dodela verovatnoće* (eng. *basic probability assignment*), ako ispunjava uslove:

(i) $m(\emptyset) = 0$,

(ii) $\sum_{A \in \mathcal{P}(X)} m(A) = 1$.

Definicija 3.1. (**Mera verovanja**)([12])
Neka je m osnovna dodela verovatnoće na $\mathcal{P}(X)$. *Mera verovanja* (eng. *belief*) $Bel : \mathcal{P}(X) \to [0, 1]$ indukovana sa m definiše se sa

$$Bel(A) = \sum_{B \subset A} m(B), \quad A \in \mathcal{P}(X).$$

[4]Familija skupova se naziva polje, ako je unija, proizvod i razlika dva skupa familije takođe element familije.

Za funkciju Bel i skupove $A, B \in \mathcal{P}(X)$ za koje je $A \cap B = \emptyset$ važi

$$Bel(A \cup B) \geq Bel(A) + Bel(B).$$

Prethodna osobina se naziva superaditivnost.

Definicija 3.2. (**Mera plauzabilnosti**)([12])
Neka je m osnovna dodela verovatnoće na $\mathcal{P}(X)$. *Mera plauzabilnosti* (eng. *plausability*) $Pl : \mathcal{P}(X) \to [0, 1]$ indukovana sa m definiše se sa

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad A \in \mathcal{P}(X).$$

Za funkciju Pl i skupove $A, B \in \mathcal{P}(X)$ za koje je $A \cap B = \emptyset$ važi

$$Pl(A \cup B) \leq Pl(A) + Pl(B).$$

Prethodna osobina se naziva subaditivnost.

Fazi mere kod kojih je operacija uobičajenog sabiranja zamenjena nekom opštijom realnom operacijom, u oznaci $\oplus$, se nazivaju $\oplus-$*dekompozabilne mere*. U slučaju kada je $\oplus = \sup$ dobijaju se tzv. maksitivne mere. Primeri maksitivnih mera su *mera mogućnosti* (eng. *possibility measure*) i *mera neophodnosti* (eng. *necessity measures*) ([11, 12, 14]).

U okviru pseudo-analize ([11, 12]) izučava se jos jedan primer dekompozabilne mere, tvz. *pseudoverovatnoća* (eng. *pseudo-probability measure*).

Neka je $(I, \oplus, \odot)$ poluprsten (više o poluprstenima se može naći u [11] i [12]) i Ω neprazan skup.

Definicija 3.3. (**Pseudoverovatnoća**)([13])
Neka je Σ $\sigma-$algebra podskupova od Ω. Funkcija $\mathbf{P} : \Sigma \to I$ sa osobinama

(a) $\mathbf{P}(\emptyset) = \mathbf{0}$ i $\mathbf{P}(\Omega) = \mathbf{1}$,

(b) $\mathbf{P}\left(\bigcup\limits_{\mathbf{i=1}}^{\infty} \mathbf{A_i}\right) = \bigoplus\limits_{i=1}^{\infty} \mathbf{P}(A_i)$, za svaki niz $\{A_i\}_{i \in \mathbf{N}}$ u parovima disjunktnih podskupova od Σ,

se naziva *pseudoverovatnoća*.

Više o pseudoverovatnoći i njenim osobinama se može videti u [1, 3, 4, 8, 9, 13].

4. Teorija odlučivanja

Teorija odlučivanja (ili teorija izbora) ([5, 11, 12]) je grana teorije verovatnoće koja se bavi teorijom donošenja odluka zasnovanih na dodeljivanju verovatnoće različitim faktorima i pripisivanju numeričkih vrednosti ishodima. Prvi ju je uveo Herbert A. Simon, dobitnik Nobelove nagrade za ekonomiju 1978. godine.

Osnov teorije odlučivanja je teorija korisnosti koja se bavi relacijama preferiranja, kojima se modelira ponašanje donosioca odluke. Teorija očekivane korisnosti se zasniva na meri verovatnoće i očekivanim vrednostima.

Razumevanje načina na koji se donose odluke je važno za mnoge druge nauke osim matematike. Teorija odlučivanja ima veliku ulogu u psihologiji, filozofiji, politici, ekonomiji i marketingu.

Zahvalnica

Rad je finansijski podržani od strane projekta Departmana za opšte discipline u tehnici, Fakulteta tehničkih nauka, Univerziteta u Novom Sadu "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama".

Literatura

[1] A. Chateauneuf, Đecomposable capacities, distorted probabilities and concave capacities", *Mathematical Social Sciences*, 31: pp. 19–37, 1996.

[2] L.Debnath and K. Basu, "A short history of probability theory and its applications", *International Journal of Mathematical Education in Science and Technology*, 46:1, pp. 13-39, 2015.

[3] N. Duraković, S. Medić, T. Grbić, A. Perović, and Lj. Nedović, "Generalization of portmanteau theorem for a sequence of interval-valued pseudo-probability measures", *Fuzzy Sets and Systems*, 364: pp. 96 – 110, 2019.

[4] T. Grbić, S. Medić, N. Duraković, S. Dumnić, and T. Gavrilov, "Weak convergence of sequences of distorted probabilities", In 2015 IEEE 13th International Symposium on Intelligent Systems and Informatics (SISY), Proceedings, pp. 307–312, 2015.

[5] E. T. Jaynes, "Probability Theory: The Logic of Science", Cambridge University Press, 2003.

[6] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin Verlag von Julius Springer, 1933.

[7] A. Kolmogoroff, *Foundations of the Theory of Probability*, Second English Edition, Chelsea Publishing Company, 1956.

[8] Lj. Nedović and T. Grbić, "The Pseudo Probability", *Journal of Electrical Engineering*, 53(12/s): pp. 27–30, 2002.

[9] Lj. Nedović, B. Mihailović, and N.M. Ralević, Šome properties of pseudo-measures and pseudo-probability", In 2007 IEEE 5th International Symposium on Intelligent Systems and Informatics (SISY), Proceedings, pp. 155–159, 2007.

[10] O. Ore, "Pascal and the Invention of Probability Theory", *The American Mathematical Monthly*, 67:5, pp. 409-419, 1960.

[11] E. Pap, *Null-Additive Set Functions*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1995.

[12] E. Pap, *Fazi mere i njihova primena*, Univerzitet u Novom Sadu, Prirodno-matematicki fakultet, Novi Sad, 1999.

[13] N. M. Ralević and Lj. M. Nedović, "The probability defined on semirings", *Bulletins for Applied and Computing Mathematics* (BAM), pp. 7–14, 1999.

[14] Z. Wang and G. J. Klir, *Generalized measure theory*, Springer, 2009.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

GARCH MODELI ZA PROCENU VOLATILNOSTI VREMENSKIH SERIJA

Jelena Erdeljan [1] **i Jelena Ivetić** [2]

https://doi.org/10.24867/META.2024.06 Professional paper

Sažetak. U radu su predstavljeni ARCH i GARCH modeli koji se koriste za procenu volatilnosti vremenskih serija, kao i određene modifikacije GARCH modela. Praktičan deo rada prikazuje primenu različitih GARCH modela na procenu volatilnosti relativnih dnevnih prinosa berzanskog indeksa BELEX15.

AMS klasifikacija (2020): 91B84, 62M10

Ključne reči: Vremenske serije, volatilnost, ARCH model, GARCH model

1. Uvod

Modelovanje i predviđanje volatilnosti privuklo je veliku pažnju poslednjih godina, uglavnom motivisano njenim značajem na finansijskim tržištima. Mnogi modeli za određivanje cene finansijskih instrumenata koriste procene volatilnosti kao jednostavnu meru rizika [5].

U tom kontekstu, prvo se pojavljuju autoregresivni uslovno heteroskedastični (eng. Autoregressive conditional heteroskedasticity - ARCH) modeli, koje je uveo Robert F. Engle (1982), a zatim i generalizovani oblik ovog modela odnosno generalizovani autoregresivni uslovno heteroskedastični (eng. Generalized autoregressive conditional heteroskedasticity - GARCH) modeli. GARCH modeli su doveli do fundamentalne promene pristupa u finansijskoj kvantitativnoj analizi, kroz efikasno modeliranje volatilnosti cena finansijskih sredstava. Predložene su brojne klase različitih modela vremenskih serija, ali nijedan od njih nije izazvao interesovanje uporedivo sa GARCH modelima [4].

Volatilnost, kao finansijski pojam, opisuje nepredvidive promene cene nekog finansijskog instrumenta u određenom vremenskom periodu. U matematičkom smislu, volatilnost se predstavlja kao standardna devijacija slučajne promenljive. Informacije o prošlim vrednostima slučajne promenljive daju najbolje rezultate za ocenjivanje volatilnosti.

Ovaj rad predstavlja oblast modeliranja i predviđanja volatilnosti u finansijskim vremenskim serijama, fokusirajući se na GARCH modele. Primarni cilj GARCH modela je da uhvati i opiše vremenski promenljivu prirodu volatilnosti u finansijskim podacima. Eksplicitnim modeliranjem heteroskedastičnosti GARCH modeli pružaju moćan okvir za razumevanje i predviđanje obrazaca volatilnosti.

[1]Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: jelenae0501@gmail.com
[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: jelenaivetic@uns.ac.rs

2. Modeli volatilnosti

ARCH i GARCH modeli se koriste za konstruisanje i predviđanje promenljive volatilnosti u vremenskim serijama, posebno kada su fluktuacije cena i/ili prinosa često promenljive tokom vremena. Razvijeni su kako bi objasnili i modelirali heteroskedastičnost, koja se javlja kada standardna devijacija promenljive koja se modelira nije konstantna.

2.1. ARCH model

Autoregresivni uslovno heteroskedastični model (ARCH) je istorijski posmatrano prvi model volatilnosti vremenskih serija. Robert F. Engle je uslovnu heteroskedastičnost varijanse predstavio kao linearnu funkciju kvadrata ranijih grešaka [7].

Definicija 2.1 (ARCH(p) model [7])**.** Neka je $p \in \mathbb{N}$ i neka $t \in [1, \infty)$. ARCH(p) se definiše sa:

$$r_t = \mu_t + a_t$$

$$\mu_t = \mu$$

$$a_t = \sigma_t \xi_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_p a_{t-p}^2$$

$$\xi_t \sim N(0,1) \quad \text{i.i.d.}$$

U definisanom modelu:

- r_t predstavlja prinos u trenutku t,
- a_t predstavlja šok (inovaciju) u trenutku t,
- μ_t je srednja vrednost za r_t, gde μ_t ne zavisi od vremena t,
- σ_t^2 je volatilnost za r_t,
- ξ_t je greška koja se pravi tokom linearne regresije, predstavljena i.i.d. nizom,
- α_i, za $i = 1, \ldots, p$, predstavlja parametar modela,
- p predstavlja broj parametara ARCH modela,
- i.i.d. predstavlja niz nezavisnih, identično raspoređenih slučajnih promenljivih sa srednjom vrednošću 0 i varijansom 1.

Kako bi model bio dobro definisan i kako bi uslovna varijansa bila pozitivna, uvode se ograničenja [7]: $\alpha_0 > 0$ i $\alpha_i \geq 0$, za $i > 0$.

2.2. GARCH model

Tim Bollerslev je 1986. godine u [1] uveo generalizaciju ARCH modela koja je poznata kao generalizovani ARCH model (GARCH).

Definicija 2.2 (GARCH(p,q) model [7])**.** Neka su $p, q \in \mathbb{N}$ i $t \in [1, \infty)$. GARCH(p,q) je definisan sa:

$$r_t = \mu_t + a_t$$

$$\mu_t = \mu$$

$$a_t = \sigma_t \xi_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

$$\xi_t \sim N(0,1) \quad \text{i.i.d.}$$

U modelu važe sledeća ograničenja:

- $\alpha_0 > 0$, $\alpha_i \geq 0$ za $i = 1, \ldots, p$
- $\beta_j \geq 0$ za $j = 1, \ldots, q$
- $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$

Vidimo da se kod GARCH modela uslovna volatilnost opisuje preko grešaka iz prošlosti (kao i kod ARCH modela), ali i preko prošlih varijansi, što predstavlja generalizaciju. ARCH model traži veliki broj parametara, dok je kod GARCH modela često dovoljan GARCH(1,1) da bi se opisao veliki broj finansijskih vremenskih serija na tačan način.

2.3. Modifikacije GARCH modela

Tokom vremena su uvedene mnoge modifikacije osnovnog GARCH modela.

2.3.1. GARCH-M model

GARCH-M model (engl. GARCH in Mean) se koristi za analizu vremenskih serija finansijskih podataka gde se volatilnost promenljive proučava u kontekstu njenih efekata na srednju vrednost [2]. U ovom modelu GARCH komponenta se kombinuje sa srednjom vrednošću promenljive kako bi se modelovala volatilnost i istovremeno uključili i efekti na srednju vrednost.

Definicija 2.3 (GARCH-M(p,q) model [7])**.** Neka su $p, q \in \mathbb{N}$ i neka su date konstante μ i c. GARCH-M(1,1) se definiše sa:

$$r_t = \mu + c\sigma_t^2 + a_t$$

$$a_t = \sigma_t \xi_t$$

$$\sigma_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

gde navedeni parametri imaju isto značenje kao u prethodnim definicijama.

Ako je c pozitivno i statistički značajno, onda povećani rizik dovodi do porasta srednje vrednosti prinosa [7]. Takođe, pozitivno c ukazuje na to da je prinos pozitivno povezan sa svojom prethodnom volatilnošću [2].

2.3.2. EGARCH model

Eksponencijalni GARCH (EGARCH) model je omogućio asimetrične efekte između pozitivnih i negativnih prinosa [7]. Nelson je 1991. godine uveo novu funkciju definisanu na sledeći način:

$$g(\xi_t) = \theta\xi_t + \gamma[|\xi_t| - E(|\xi_t|)]$$

gde θ i γ predstavljaju realne konstante, $\xi_t = \frac{a_t}{\sigma_t}$, a ξ_t i $|\xi_t| - E(|\xi_t|)$ predstavljaju i.i.d. nizove sa srednjom vrednošću 0 i neprekidnom raspodelom. Bitno je napomenuti da je γ parametar asimetričnog odgovora, tako da samo negativni šok povećava buduću volatilnost. Ovo je u suprotnosti sa standardnim GARCH modelom gde šokovi iste magnitude, bilo pozitivni ili negativni, imaju isti efekat na buduću volatilnost.

Definicija 2.4 (EGARCH(p, q) model) [7])**.** Neka je α_0 konstanta, neka $p, q \in \mathbb{N}$ i neka B predstavlja operator pomeranja unazad (ili zaostajanja) takav da važi $Bg(\xi_t) = g(\xi_{t-1})$, i neka su $1 + \beta_1 B + \ldots + \beta_s B^s$ i $1 - \alpha_1 B - \ldots - \alpha_m B^m$ polinomi sa nulama izvan jediničnog kruga koji nemaju zajedničkih faktora.

Tada se EGARCH(p, q) model definiše sa:

$$a_t = \sigma_t \xi_t$$

$$\ln(\sigma_t^2) = \alpha_0 + \frac{(1 + \beta_1 B + \ldots + \beta_q B^q)}{(1 - \alpha_1 B - \ldots - \alpha_p B^p)} g(\xi_{t-1}).$$

Kako $\ln(\sigma_t^2)$ može biti negativan, nema potrebe da se parametri ograničavaju da bi desna strana prethodne jednačine ostala nenegativna kao kod osnovnog GARCH modela [6].

3. Modeliranje volatilnosti indeksa BELEX15

Konstruisanje GARCH modela je sprovedeno na osnovu realnog skupa podataka relativnih dnevnih promena berzanskog indeksa BELEX15, koji predstavlja najlikvidnijih 15 akcija na Beogradskoj berzi, posmatranog u periodu od 08.01.2020. do 31.12.2020 [8]. Period obuhvata uzorak od 252 radnih dana trgovanja, u kome je prosečna vrednost prinosa iznosila -0.022 ± 0.976, minimalna registrovana vrednost je -6, a maksimalna 3.53. Podaci su analizirani uz pomoć NumXL softverskog paketa.

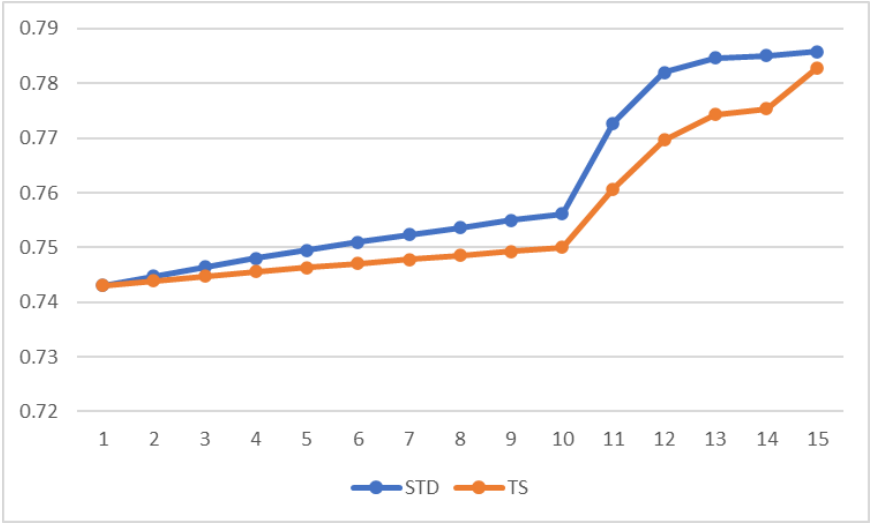
Prvi korak u analizi jeste testiranje uzorka na ARCH efekte, kako bi se proverilo da li posmatrana vremenska serija ima nekonstantu volatilnost, odnosno da li su GARCH modeli adekvatan model za posmatranu vremensku seriju. Analiza pokazuje da je ARCH test je statistički značajan na nivou značajnosti 0.05, pa sledi da je ova vremenska serija pogodna za primenu GARCH modela.

Za posmatranu vremensku seriju je potom konstruisano 12 modela: GARCH, GARCH-M i EGARCH modeli sa različitim brojem parametara (p,q = 1,2).Za svaki od primenjenih modela su izračunate prvo inicijalne vrednosti parametara koje nisu optimalne, zatim je model kalibrisan i procenjene su optimalne vrednosti parametara modela na osnovu dostupnih podataka. Cilj

kalibracije je prilagođavanje GARCH modela podacima kako bi se optimalno opisala i predvidela varijabilnost u vremenskoj seriji. Zbog ograničenog prostora ovde neće biti prikazani rezultati modeliranja, a čitaoca upućujemo na rad [3] u kome su navedeni svi detalji ovog, ali i nekoliko drugih primera.

Naredni korak je odabir najpogodnijeg modela za prognoziranje volatilnosti. Ovo je moguće utvrditi analizom kalibrisanih modela, metodom logaritam maksimalne verodostojnosti (skr. LLF) na osnovu maksimalne vrednosti, ili prema informacionom kriterijumu (skr. AIC) na osnovu najmanje vrednosti kriterijuma modela. Na osnovu LLF metode najbolji od konstruisanih modela model je EGARCH(2,1) ($LLF = -269.86$), dok je prema AIC kriterijumu najbolji model EGARCH(1,1) ($AIC = 553.55$).

Konačno, na osnovu odabranog optimalnog modela se vrši prognoza volatilnosti za određeni broj koraka unapred. Za svaki korak se računa standardna devijacija (STD) i struktura terma (TS). TS predstavlja prosečnu volatilnost od kraja posmatranog uzorka sve do kraja željenog perioda, dok STD predstavlja lokalnu volatilnost u određenom koraku. Rezultati predikcije za 15 koraka unapred na osnovu EGARCH(1,1) modela su predstavljeni na Slici 1.



Slika 1: Predikcija volatilnosti BELEX15 za 15 dana unapred primenom EGARCH(1,1) modela

4. Zaključak

Analiza finansijskih vremenskih serija predstavlja kompleksan proces koji zahteva primenu naprednih statističkih metoda. Razumevanje i modelovanje volatilnosti je neophodno za predviđanje kretanja na finansijskim tržištima, kao i za procenu rizika. U tom kontekstu, GARCH modeli su postali neizostavan alat u savremenoj ekonometrijskoj analizi.

GARCH modeli omogućavaju modelovanje heteroskedastičnosti tj. promenljive volatilnosti u vremenskim serijama. Njihova prednost u odnosu na ARCH modele se nalazi u tome što uzimaju u obzir prethodne informacije o volatilnosti, i što zahtevaju manji broj parametara.

Tokom vremena došlo je do razvoja različitih modifikacija GARCH modela. Ove modifikacije su korisne pri modeliranju specifičnih karakteristika finansijskih vremenskih serija, kao što su asimetrija, upornije promene u aktivnosti, efekat poluge i druge.

Na realnom skupu podataka vremenske serije relativnih dnevnih prinosa berzanskog indeksa BELEX15, prikazano je kako se volatilnost može modelirati i predviđati primenom GARCH modela.

Zahvalnica

Drugi autor je finansijski podržan od strane projekta Departmana za opšte discipline u tehnici, Fakulteta tehničkih nauka, Univerziteta u Novom Sadu ″Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama″.

Literatura

[1] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, vol. 31:3, pp. 307-327, 1986.

[2] C. Brooks, *Introductory Econometrics for Finance, 2nd edition.* Cambridge University Press, 2008.

[3] J. Erdeljan, *Uslovni heteroskedastični modeli za procenu volatilnosti vremenskih serija.* Master rad, Fakultet tehničkih nauka, Novi Sad, 2023.

[4] C. Francq and J-M. Zakoian, *GARCH models – Structure, Statistical Inference and Financial Applications.* John Wiley and Sons, 2010.

[5] J. Knight and S. Satchell, *Forecasting Volatility in the Financial Markets, 3rd edition.* Butterworth–Heinemann, 2007.

[6] D. Ruppert, *Statistics and Finance.* Springer-Verlag, 2004.

[7] R.S. Tsay, *Analysis of Financial Time Series, 3rd edition.* John Wiley and Sons, 2010.

[8] https://www.belex.rs/trgovanje/indeksi/belex15 (pristup 25.03.2023.)

The 7th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31–June 2nd 2024

REVERZIBILNI PROSTORI

Aleksandar Janjoš [1]

https://doi.org/10.24867/META.2024.07 Review article

Sažetak. Cilj ovoga rada je da prikaže neke osobine reverzibilnih topoloških prostora, kao i neke osnovne primere reverzibilnih i nereverzibilnih prostora koje su u svom radu istakli autori M. Rajagopalan i A. Wilansky [5]. Radi jasnijeg prikaza, uvešćemo i osnovne definicije i pojmove iz teorije topoloških prostora.

AMS klasifikacija (2020): 54C99

Ključne reči: Reverzibilnost, Homeomorfizam, Topološki prostori, Baza topologije, Povezanost

1. Definicije i opšti pojmovi

Definicija 1.1 ([2])**.** Neka je X neprazan skup. Kolekcija $\mathcal{O}$ podskupova skupa X je kolekcija otvorenih skupova ako važe sledeća tri uslova:

($O1$) $\emptyset$ i skup X su otvoreni, tj. $\emptyset, X \in \mathcal{O}$,

($O2$) Presek svaka dva otvorena skupa je otvoren skup, tj. za svako $O_1, O_2 \in \mathcal{O}$ važi $O_1 \cap O_2 \in \mathcal{O}$,

($O3$) Unija proizvoljno mnogo otvorenih skupova je otvoren skup, tj. za svaku familiju $\{O_i |\ i \in I\} \subset \mathcal{O}$ važi $\bigcup_{i \in I} O_i \in \mathcal{O}$.

Za familiju $\mathcal{O}$ kažemo i da je topologija na skupu X, dok uređeni par $(X, \mathcal{O})$ kažemo da je topološki prostor, a elemente skupa X nazivamo tačkama. Dalje, za skup $F \subseteq X$ kažemo da je zatvoren ako i samo ako je njegov komplement $X \setminus F$ otvoren skup.

Primer 1.2 ([1])**.** Neka je $\mathbf{R}$ skup realnih brojeva i neka je $\mathcal{O}$ familija koju čine svi skupovi $U \subset \mathbf{R}$ takvi da za svako $x \in U$ postoji $\epsilon > 0$ takvo da je $(x-\epsilon, x+\epsilon) \subseteq U$. Tada familija $\mathcal{O}$ zadovoljava osobine iz definicije 1.1. Familiju $\mathcal{O}$ nazivamo uobičajena topologija, a prostor $(\mathbf{R}, \mathcal{O})$ kraće označavamo samo sa $\mathbf{R}$.

Definicija 1.3 ([1])**.** Neka su $(X, \mathcal{O}_1)$ i $(X, \mathcal{O}_2)$ topološki prostori. Ako važi $\mathcal{O}_2 \subset \mathcal{O}_1$ kažemo da je toplogija $\mathcal{O}_1$ striktno finija, a topologija $\mathcal{O}_2$ striktno grublja. Diskretna topologija je uvek najfinija, a anti-diskretna je uvek najgrublja.

[1]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: aleksandar.janjos@uns.ac.rs

Definicija 1.4 ([4])**.** Neka je $(\mathbf{X}, \mathcal{O})$ topološki prostor i neka je $\mathcal{O} \subset \mathcal{O}'$. $\mathcal{O}'$ nazivamo jednostavnom ekstenzijom topologije $\mathcal{O}$ ako i samo ako postoji $A \in \mathcal{O}$ takav da je $\mathcal{O}' = \{O \cup (O' \cap A) : O, O' \in \mathcal{O}\}$

Definicija 1.5 ([1])**.** Za topološki prostor $(X, \mathcal{O})$ kažemo da je:

- T_0-prostor ako i samo ako za svake dve različite tačke $x, y \in X$ postoji otvoreni skup O koji sadrži tačno jednu od njih.
- T_1-prostor ako i samo ako za svaki par različitih tačaka $x, y \in X$ postoji otvoren skup O takav da je $x \in O \not\ni y$.
- T_2-prostor ili Hauzdorfov prostor ako i samo ako za svake dve različite tačke $x, y \in X$ postoje disjunktni otvoreni skupovi O_1 i O_2 da je $x \in O_1$ i $y \in O_2$.

Dalje, za aksiome separacije važi sledeći niz implikacija:

$$T_2 \implies T_1 \implies T_0.$$

Definicija 1.6 ([1])**.** Neka je $(X, \mathcal{O})$ topološki prostor i $M \subseteq X$. Tada familija $\mathcal{O}'$ koju čine svi skupovi oblika $U \cap M$ za $U \in \mathcal{O}$ je nasledna ili indukovana topologija potprostora $(M, \mathcal{O}')$.

Kada god je jasno o kojoj topologiji na skupu je reč prostor ćemo kratko označavati oznakom skupa.

Definicija 1.7 ([1])**.** Neka su $(X, \mathcal{O})$ i $(Y, \mathcal{O}')$ dva topološka prostora. Preslikavanje $f : X \to Y$ je nepekidno ako je $f^{-1}(U) \in \mathcal{O}$ za sve $U \in \mathcal{O}'$.

Definicija 1.8 ([1])**.** Neprekidno preslikavanje $f : X \to Y$ iz prostora $(X, \mathcal{O})$ u prostor $(Y, \mathcal{O}')$ je homeomorfizam ako je bijekcija i f^{-1} iz Y u X je neprekidno.

Definicija 1.9 ([1])**.** Neka je $\{(X_s, \mathcal{O}_s)\}_{s \in S}$ familija topoloških prostora gde su X_s, $s \in S$ po parovima disjunktni. Na skupu $X = \bigcup_{s \in S} X_s$ familija $\mathcal{O}$ svih skupova $U \subseteq X$ takvih da je $U \cap X_s$ otvoren u X_s za svako $s \in S$ čini topologiju na skupu X. Skup X sa ovom topologijom naziva se suma prostora $\{X_s\}_{s \in S}$.

Definicija 1.10 ([1])**.** Neka je $\{(X_s, \mathcal{O}_s)\}_{s \in S}$ familija topoloških prostora. Na skupu $X = \Pi_{s \in S} X_s$ neka je familija preslikavanja $\{p_s\}_{s \in S}$ definisana sa: svakoj tački $x = \{x_s\} \in X$ preslikavanje p_s dodeljuje $s-$tu koordinatu $x_s \in X_s$. Skup $X = \Pi_{s \in S} X_s$ zajedno sa topologijom koju na njemu generiše familija preslikavanja $\{p_s\}_{s \in S}$ naziva se Kartezijanski proizvod topoloških prostora $\{(X_s, \mathcal{O}_s)\}_{s \in S}$, a odgovarajuća topologija naziva se topologija Tihonova.

Teorema 1.11 (Bouwer-ova teorema invarijante domena)**.** *Ako je U otvoren u $\mathbf{R}^n$ i $f : U \to \mathbf{R}^n$ neprekidno injektivno preslikavanje onda je $V = f[U]$ otvoren u $\mathbf{R}^n$ i f je homeomorfizam između U i V.*

Definicija 1.12 ([1])**.** Topološki prostor X je povezan ako se ne može predstaviti kao suma bilo koja svoja dva disjunktna potprostora.

Teorema 1.13 ([1])**.** *Povezanost je invarijanta neprekidnih preslikavanja.*

2. Reverzibilni prostori. Osobine i primeri

Definicija 2.1 ([5])**.** Topološki prostor $(X, \mathcal{O})$ zovemo reverzibilnim ako nema striktno grublje topologije $\mathcal{O}'$ na skupu X takve da su $(X, \mathcal{O})$ i $(X, \mathcal{O}')$ homeomorfni, ekvivalentno, ako nema striktno finije topologije $\mathcal{O}''$ takve da su $(X, \mathcal{O})$ i $(X, \mathcal{O}'')$ homeomorfni.

Lema 2.2 ([5])**.** *Topološki prostor je reverzibilan akko je svaka neprekidna bijekcija iz prostora u njega samoga homeomorfizam.*

Dokaz. Pretpostavimo da je $(X, \mathcal{O})$ reverzibilan topološki prostor i neka je $f : X \to X$ neprekidna bijekcija. Definišimo $\mathcal{O}' = \{G \subseteq X : f[G] \in T\}$, što je nova topologija na skupu X. Tada je $\mathcal{O}'$ grublja topologija od topologije $\mathcal{O}$ jer za svako $G \in \mathcal{O}'$ imamo da je $G = f^{-1}(f[G])$. Time je jasno da imamo $f : (X, \mathcal{O}') \to (X, \mathcal{O})$ homeomorfizam. Kako smo pretpostavili reverzibilnost prostora $(X, \mathcal{O})$ dobijamo da je $\mathcal{O} = \mathcal{O}'$, što znači i da je $f : (X, \mathcal{O}) \to (X, \mathcal{O})$ homeomorfizam, što nam daje smer ($\Longrightarrow$). Obrnut smer se dokazuje analogno. □

Primer 2.3. Posmatrajmo skup realnih brojeva $\mathbf{R}$ gde će pozitivni brojevi i nula imati standardnu topologiju otvorenih intervala, a negativni brojevi diskretnu topologiju. Preslikavanje $f(x) = x + 1$ je neprekidna bijekcija, ali nije homeomorfizam, jer kako je $f^{-1}(x) = x - 1$ imamo da je $f^{-1}(-1) = 0$, gde je $\{-1\}$ otvoren skup, a $\{0\}$ nije. Potpuno isto funkcioniše i dokaz na skupu racionalnih brojeva. Time smo dali primer nereverzibilnog prostora koji je prebrojiv metrički prostor.

Definicija 2.4 ([3])**.** Topološki prostor $(X, \mathcal{O})$ je lokalno Euklidski ako postoji prirodan broj n takav da svaka tačka $x \in X$ ima okolinu homeomorfnu sa $\mathbf{R}^n$, gde prostor $\mathbf{R}^n$ posmatramo sa topologijom Tihonova.

Teorema 2.5 ([5])**.** *Svaki lokalno Euklidski topološki prostor je reverzibilan.*

Dokaz. Neka je f bijekcija iz lokalno Euklidskog prostora $(X, \mathcal{O})$ u njega samoga. Neka je $x \in X$, $y = f(x)$ i neka je V okolina od y takva da je $\psi : V \to W$ izomorfizam sa nekim otvorenim potprostorom W Euklindskog $n-$ prostora. Tada $f^{-1}[V]$ sadrži okolinu U tačke x sa $\gamma : U \to Z$ homeomorfizmom na otvoren potprostor Z Euklidskog $n-$ prostora. Uzmimo $g = \psi \circ f \circ \gamma^{-1}$. Preslikavanje g je bijekcija kao kompozicija bijekcija, pa je po Brouwer-ovoj teoremi 1.11 g homeomorfizam iz Z na otvorenu okolinu $\psi(y)$ i $f^{-1} = \gamma \circ g^{-1} \circ \psi^{-1}$ je neprekinda na okolini y. Kako je tačka x bila proizvoljna, a ovim je pokazano da proizvoljna neprekidna bijekcija ima inverznu funkciju koja je neprekinda bijekcija na okolini tačke $y = f(x)$, dobijamo da je f homeomorfizam. □

Teorema 2.6. *Prostor $(\mathbf{I}, \mathcal{O})$ iracionalnih brojeva sa naslednom topologijom prostora $\mathbf{R}$ nije reverzibilan.*

Dokaz. U ovom dokazu standardnom oznakom za interval (a, b) označavaćemo samo intervale iracionalnih brojeva, čak i da su vrednosti a i b racionalni brojevi. Dovoljno je pokazati da interval $I = (-1, 1)$ nije reverzibilan prostor. Izaberimo iracionalan broj $u \in I$ i posmatrajmo prostor I' čija je topologija

jednostavna ekstenzija topologije prostora I za skup $A = [u, 1)$ 1.4. Neka je $J = (0, 1) \cup (2, 4)$. J je homeomorfno sa I, ali mi ćemo pokazati da je homeomorfno i sa I'. Posmatrajmo $I' = (-1, u) \cup [u, 1)$, gde je jasno da su $(-1, u)$ i $(0, 1)$ homeomorfni, pa ostaje samo da napravimo homeomorfizam između $[u, 1)$ i $(2, 4)$. Neka su: niz $\{a_n\}_{n\in\mathbf{N}}$ rastći niz racionalnih brojeva koji konvergira ka $\sqrt{8}$ i $a_0 = 2$, $\{b_n\}_{n\in\mathbf{N}}$ opadajući niz racionalnih brojeva koji konvergira ka $\sqrt{8}$ sa $b_0 = 4$ i $\{c_n\}_{n\in\mathbf{N}}$ opadajući niz racionalnih brojeva koji konvergira ga u sa $c_0 = 1$. Za $n = 1, 2, ...$ definišimo f_n homeomorfizme iz (c_n, c_{n-1}) na $(a_{\frac{1}{2}(n-1)}, a_{\frac{1}{2}(n+1)})$ za neparne n, ili na $(b_{\frac{1}{2}n}, b_{\frac{1}{2}(n-2)})$ za parne n. Traženi homeomorfizam iz $[u, 1)$ na $(2, 4,)$ je definisan sa $f(u) = \sqrt{8}$ i $f|_{(c_n,c_{n-1})} = f_n$. □

Kako se prethodni dokaz analogno može sprovesti i za $\mathbf{Q}$, imamo da je prostor $\mathbf{R}$ direktna suma dva prostora koji nisu reverzibilni ($\mathbf{Q}$ i $\mathbf{I}$), ali on sam je reverzibilan kao lokalno Euklidski za $n = 1$. Time dobijamo sledeću teoremu kao posledicu teorema 2.4 i 2.6.

Teorema 2.7. *Postoji reverzibilni prostor koji je unija dva disjunktna prostora koji nisu reverzibilni.*

Teorema 2.8 ([5])**.** *Svaki potprostor koji je u isto vreme i otvoren i zatvoren reverzibilnog prostora je reverzibilan.*

Dokaz teoreme 2.8 je posledica činjenice da se svako neprekidno preslikavanje iz potprostora u njega samoga može neprekidno proširiti na ceo prostor (što možemo videti u [1]).

Teorema 2.9 ([5])**.** *Neka je $(X, \mathcal{O})$ topološki prostor sa konačno mnogo komponenti. Tada X je reverzibilan akko je svaka komponenta reverzibilna.*

Dokaz. Neka je $X = \bigcup\{C_i | i = 1, 2, 3, ..., n\}$ gde je C_i za svako i komponenta prostora X, koje su sve reverzibilne. Neka je $f : X \to X$ neprekidna bijekcija. Tada, kako $f[C_i]$ mora biti povezan potprostor za svako i, on će biti jednak nekom C_j za neko $j \in \{1, 2, 3, ..., n\}$. To znači da f permutuje komponente, te će za svako i postojati k, $1 \leq k \leq n$ takvo da je $f^k[C_i] = C_i$. Kako je f neprekidna bijekcija to je onda i $f^k|_{C_i}$, pa iz pretpostavke da su sve komponente reverzibilne imamo da je f^k homeomorfizam iz C_i na samoga sebe. Tada je i $f^{-1} = f^{k-1} \circ (f^k)^{-1}$ neprekidno na C_i. Iz činjenice da komponenti ima konačno mnogo znamo da su one i otvoreni i zatvoreni skupovi istovremeno čime dobijamo da je f^{-1} u neprekidno preslikavenje iz X u X, što je dovoljno da f bude homeomorfizam. Time dobijamo ($\Longleftarrow$), dok je ($\Longrightarrow$) posledica teoreme 2.8 i činjenice da su sve komponente i otvoreni i zatvoreni skupovi kada ih ima konačno mnogo. □

Sledeće tvrđenje je direktna posledica teoreme 2.9, jer će komponente prostora $D \times Y$ biti oblika $\{x\} \times Y$ za $x \in D$.

Tvrđenje 2.10 ([5])**.** *Neka je D konačan diskretan prostor i Y reverzibilan povezan prostor, tada je i $D \times Y$ reverzibilan.*

Definicija 2.11 ([5])**.** Za prostor X kažemo da je fisilni ako se može predstaviti kao disjunktna unija dve svoje homeomorfne kopije.

Primer fisilnog prostora je bilo koji poluotvoren interval realnih brojeva $(a, b]$. Sa druge strane, niti jedan prostor $\mathbf{R}^n$ za bilo koji prirodan broj n nije fisilan.

Primer 2.12 ([5])**.** Posmatrajmo skup prirodnih brojeva sa diskretnom topologijom, u oznaci $\mathbf{N}$, i Y neka je povezan fisilan prostor gde je $Y = A \cup B$, pri čemu su A i B njegove disjunktne homeomorfne kopije. Prostor $D \times Y$ nije reverzibilan jer za preslikavanje f koje $1 \times Y$ preslikava u $1 \times A$, $2 \times Y$ u $1 \times B$ i $n \times Y$ u $(n-1) \times Y$ za $n = 3, 4, 5, ...$ Ovim smo dobili neprekidnu bijekciju, ali inverzno preslikavanje nije neprekidno na osnovu 1.13 jer ono potprostor $1 \times Y$, koji je povezan, preslikava u nepovezan prostor $(1 \times Y) \cup (2 \times Y)$. Jedan od primera na kojem možemo ponoviti ovu konstrukciju je $\mathbf{N} \times \beta\mathbf{N}$, jer je $\mathbf{N}$ fisilan (skupovi parnih i neparnih brojeva), gde sa $\beta\mathbf{N}$ označavamo Stone-Čeh kompaktifijaciju prostora $\mathbf{N}$, o čemu se više može pročitati u [1].

Zahvalnica

Autor je finansijski podržan od strane projekta Departmana za opšte discipline u tehnici, Fakulteta tehničkih nauka, Univerziteta u Novom Sadu "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama".

Literatura

[1] R. Engelking, *General Topology,* Heldermann Verlag Berlin 1989.

[2] M. Kurilić, *Osnovi opšte topologije,* Prirodno-matematički fakultet, Univerzitet u Novom Sadu, Novi Sad 1998.

[3] J. M. Lee, *Introduction to Topological Manifolds,* Springer, Science and Bussunes Media, 2016.

[4] N. Levine, *Simple extensions of topologies,* American Mathematical Monthly 71, 1964.

[5] M Rajagopalan, A. Wilansky, *Reversible topological spaces,* Journal of the Australian Mathematical Society, 1966.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

BICAPACITIES ON BOUNDED LATTICES – BASIC PROPERTIES

Juraj Kalafut [1] **and Martin Kalina** [2]

https://doi.org/10.24867/META.2024.08 Preliminary communication

Abstract. The paper is a preliminary announcement of studying bipolar capacities on bounded lattices. An algebraic structure will be constructed where it is possible to define a bipolar capacity and one special case will be shown when the bipolar capacity is additive.

AMS Mathematics Subject Classification (2020): 03E72, 28E10

Key words and phrases: bicapacity, selfdual bicapacity, dual lattice, bounded lattice

1. Introduction

Bipolarity in solving decision-making problems has been used already for about 20 years. The theory of bipolar capacities and integral with respect to such capacities was introduced by Grabisch et al. [6, 7]. In some cases, a single value is not sufficient as a result. Interval-valued fuzzy sets or Atanasov's intuitionistic fuzzy sets (see [1]) are broadly used in fuzzy decision-making. Both, interval-valued fuzzy sets as well as Atanasov's intuitionistic fuzzy sets, are special cases of lattice-valued fuzzy sets, introduced by Goguen [5]. This paper contains some preliminary considerations in examining possible lattice-valued fuzzy sets where it is possible to define bipolar capacities. Readers are assumed to be familiar with basics of lattices. For details on the lattice theory readers are referred to the monograph [2]. The paper is organized as follows. Section 2 is devoted to recalling known results and notions that will be needed in authors' considerations. In Section 3 the main results will be formulated. Conclusions will be formulated in Section 4.

2. Preliminaries

In this section basic notions and results on bipolarity and also some types of lattices will be provided.

An important notion will be that of a complemented lattice.

[1]Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, e-mail: juraj.kalafut@stuba.sk

[2]Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, e-mail: martin.kalina@stuba.sk

Definition 2.1 ([2])**.** Let $(L, \vee, \wedge, 0, 1, ^c)$ be a bounded lattice and $\cdot^c$ a decreasing function such that for every $x \in L$ there exists a uniquely given element x^c with

$$x \wedge x^c = 0, \qquad x \vee x^c = 1. \tag{2.1}$$

Then the lattice L is said to be *complemented* and the element x^c the *complement of* x.

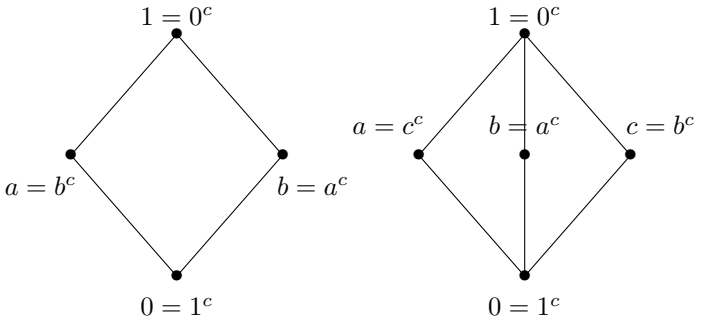


Figure 1: Examples of complemented lattices, left the lattice L_1, right L_2

Remark 2.2. Both of the lattice in Figure 1 are complemented, however, there is substantial difference between them. The lattice left has a so-called *involutive* complement, i.e., $(x^c)^c = x$ for all elements of L_1, while the complement in the lattice L_2 is not involutive.

Definition 2.3. Let X be a non-empty finite set and $\mathcal{P}(X)$ its powerset. A monotone set-function $\mu : \mathcal{P}(X) \to [0, 1]$ is said to be a *capacity* if $\mu(\emptyset) = 0$, $\mu(X) = 1$.

One of the main notions in this paper is that of a bicapacity. Its definition on a finite set X follows.

Definition 2.4 ([6])**.** Let X be a non-empty finite set and $\mathcal{P}(X)$ its powerset. Denote $\mathcal{Q}(X) = \{(A, B) \in \mathcal{P}(X) \times \mathcal{P}(X); A \cap B = \emptyset\}$. A function $h : \mathcal{Q}(X) \to [-1, 1]$ such that

1. h is increasing in the first variable,
2. h is decreasing in the second variable,
3. $h(\emptyset, \emptyset) = 0$, $h(X, \emptyset) = 1$, $h(\emptyset, X) = -1$,

is called a *bicapacity*.

Example 2.5. A typical example of a bicapacity on a non-empty finite set X is the following

$$h(A, B) = \mu(A) - \nu(B) \qquad \text{where } \mu \text{ and } \nu \text{ are capacities.} \tag{2.2}$$

Particularly, if $\mathrm{card}(X) = n$, one may have

$$h_{\mathrm{sym}}(A, B) = \frac{\mathrm{card}(A)}{n} - \frac{\mathrm{card}(B)}{n}, \tag{2.3}$$

where $(A, B) \in \mathcal{Q}(X)$ from Definition 2.4.

The bicapacity h_{sym} defined by formula (2.3), is called *symmetric* (see, e.g., [8, 9]).

An important notion will be also that of MV-algebra [4].

Definition 2.6 ([3, 4])**.** An *MV-algebra* is an algebra $(A, \oplus, \neg, 0)$ of type $(2, 1, 0)$, satisfying

(M1) $x \oplus y = y \oplus x$,

(M2) $x \oplus (y \oplus z) = (x \oplus y) \oplus z$,

(M3) $x \oplus 0 = x$,

(M4) $\neg\neg x = x$,

(M5) $x \oplus 1 = 1$ where $1 = \neg 0$,

(M6) $\neg(\neg x \oplus y) \oplus y = \neg(\neg y \oplus x) \oplus x$.

Remark 2.7. On any MV-algebra M, an order $\leq$ can be introduced in the following way ([3]):

$$x \leq y \quad \text{if and only if} \quad \neg x \oplus y = 1. \tag{2.4}$$

Moreover, the order $(M, \leq)$ can be organized into a bounded distributive lattice $(M, \vee, \wedge, 0, 1)$ by

$$x \vee y = \neg(\neg x \oplus y) \oplus y \quad \text{and} \quad x \wedge y = \neg(\neg x \vee \neg y). \tag{2.5}$$

The operation $\neg$ is not a complement in the sense of Definition 2.1.

3. Main results

Complemented lattices and MV-algebras are main structures that motivated this research. The lattice L_2 depicted in Fig. 1 right, is a complemented lattices, however, some technical problems might occur when constructing a bicapacity on that lattice because the complement is not involutive. However, it is possible to define an algebraic structure on the lattice L_2 skipping the lattice-theoretical complement.

Example 3.1. Consider the lattice L_2 from Fig. 1 right skipping the complement. Define a partial binary operation $\oplus$ and a unary operation $\neg$ such that $0 \oplus x = x \oplus 0 = x$ and $1 \oplus x = x \oplus 1 = 1$ for all $x \in L_2$ and $\neg 0 = 1$, $\neg 1 = 0$, and results for the set of inputs $\{a, b, c\}$ is given by Table 1. The algebraic structure $(L_2, \oplus, \neg, 0)$ is not an MV-algebra and neither $(L_2, \vee, \wedge, 1, 0)$ is a complemented lattice with $\neg$ as the complement.

$\oplus$	a	b	c
a	a	–	1
b	–	1	–
c	1	–	c

$\neg$	a	b	c
	c	b	a

Table 1: Operation $\oplus$ (left) and $\neg$ (right) on the set $\{a, b, c\}$

It is possible to define a dual operation to $\oplus$, $\odot$, defined for any pair $(x, y) \in L_2 \times L_2$ whenever $\oplus$ is defined for (x, y), by

$$x \odot y = \neg(\neg x \oplus \neg y). \tag{3.1}$$

Then the following holds for $\neg$:

$$x \oplus \neg x = 1, \quad x \odot \neg x = 0. \tag{3.2}$$

This means that formula (2.1) a lattice-theoretical complement is 'mimicked' by $\neg$ (cf. formula (3.2)) in the corresponding algebraic structure, just with respect to $\oplus$ and $\odot$.

Remark 3.2. From now on, mentioning an algebraic structure $(A, \oplus, \neg)$ it will be assumed that formula (3.2) is fulfilled, where $\odot$ is given by formula (3.1).

Definition 3.3. Let $(A, \oplus, \neg)$ be an algebraic structure. Then $\overline{(A, \oplus, \neg)}$ will denote the *dual algebraic structure*, i.e., the algebraic structure with the reverted order.

When no confusion may occur, A will denote also the algebraic structure itself and $\overline{A}$ its dual algebraic structure. By $\overline{x}$ an element of $\overline{A}$ will be denoted.

Definition 3.4. Let $(A, \oplus, \neg, 0)$ be an algebraic structure. A monotone function $\mu : A \to [0, 1]$ is said to be a capacity if $\mu(0) = 0$, $\mu(1) = 1$, where $1 = \neg 0$.

Definition 3.5. Denote $\mathcal{Q}(A) = \{(C_1, C_2) \in A \times A; C_2 \leq \neg C_1)\}$ for an algebraic structure A. A function $\mathcal{H} : \mathcal{Q}(A) \to A \cup \bar{A}$ will be called a *bipolar capacity* if the following properties are fulfilled

(B1) $\mathcal{H}$ is increasing in the first variable,

(B2) $\mathcal{H}$ is decreasing in the second variable,

(B3) $\mathcal{H}((1, 0) = 1$, $\mathcal{H}(0, 1) = \bar{1}$, $\mathcal{H}(0, 0) = 0$.

For a construction of a bipolar capacity on A, two capacities, $\mu : A \to [0, 1]$ and $\nu : A \to [0, 1]$ can be used.

Example 3.6. Let A be an algebraic structure and μ and ν two capacities on A. The following formula defines a bipolar capacity

$$\mathcal{H}(C_1, C_2) = \begin{cases} C_1 & \text{if } \mu(C_1) > \nu(C_2), \\ \overline{C_2} & \text{if } \nu(C_2) > \mu(C_1), \\ 0 & \text{if } \mu(C_1) = \nu(C_2). \end{cases}$$

Definition 3.7. A bipolar capacity $\overline{\mathcal{H}}$ is *dual* to $\mathcal{H}$ if the following holds for all $(C_1, C_2) \in \mathcal{Q}(A)$

$$\overline{\mathcal{H}}(C_2, C_1) = \mathcal{H}(C_1, C_2). \tag{3.3}$$

$\overline{\mathcal{H}}$ is said to be *self-dual* to $\mathcal{H}$ if $\overline{\mathcal{H}} = \mathcal{H}$.

Assume a bipolar capacity $\mathcal{H}$ is used in solving a decision-making problem. Then self-duality of $\mathcal{H}$ means that the preference for positive neither for negative evaluations is used. Particularly, the bipolar capacity $\mathcal{H}$ in Example 3.6 is self-dual if $\mu = \nu$.

Definition 3.8. Let A be an algebraic structure containing finitely many elements. A is said to be *atomic* if every element $x \in A$ can be decomposed as follows

$$x = \bigoplus_{i=1}^{k} a_i, \tag{3.4}$$

where a_i are elements such that $b \leq a_i$ implies $b = 0$ or $b = a_i$, Elements a_i are said to be *atoms*.

The expression of an element x as the sum of atoms is not necessarily unique.

Theorem 3.9. *Let A be an atomic algebraic structure. Let all atoms be enumerated by numbers from $N = \{1, 2, \ldots, n\}$. Assume that for all $x \in A$, if*

$$x = \bigoplus_{i-1}^{k} a_i = \bigoplus_{j=1}^{m} b_j \tag{3.5}$$

are two different decompositions into sets of atoms, then $k = m$. Then

$$\mathcal{H}(C_1, C_2) = \begin{cases} \bigoplus_{i=1}^{n_1 - n_2} a_i & \text{if } n_1 > n_2, \\ \bigoplus_{i=1}^{n_2 - n_1} \overline{a_i} & \text{if } n_1 < n_2, \\ 0 & \text{if } n_1 = n_2, \end{cases} \tag{3.6}$$

is a bipolar capacity, where n_1 and n_2 are numbers of atoms of decompositions of C_1 and C_2, respectively. The atoms a_i, $\overline{a_i}$ are chosen in such a way that always atoms with smaller enumeration numbers are taken. Moreover, the bipolar capacity $\mathcal{H}$ is additive with respect to $\oplus$.

4. Conclusions

The paper is an announcement of preliminary results of bipolar capacities on lattices (and algebraic structures). The bipolar capacities have been defined and a special case that led to additive bipolar capacities, was shown in Theorem 3.9.

Acknowledgement

The support for both authors by the VEGA grant agency, grant Nrs. 2/0128/24 and 1/0036/23 and by the bilateral project APVV SK-SRB-23-0044 is kindly announced.

References

[1] K. Atanassov, *Intuitionistic Fuzzy Sets*, Springer: Heidelberg, 1999.

[2] G. Birkhoff, *Lattice Theory*, American Mathematical Society Colloquium Publishers: Providence, RI. 1967.

[3] M. Botur, I. Chajda, R. Halaš, J. Kühr, J. Paseka, *Algebraic Methods in Quantum Logic*, Palacký University, Book - series Academic Publications: Olomouc, 2014.

[4] C. C. Chang, "Algebraic analysis of many valued logics", *Transactions of the American Mathematical Society*, vol. 88, pp. 467-490, 1958.

[5] J. A. Goguen, "L-fuzzy sets", *Journal of Mathematical Analysis and Applications*, vol. 18,1, pp. 145-174, 1967.

[6] M. Grabisch, Ch. Labreuche. "Bi-capacities I: definition, Möbius transform and intersection", *Fuzzy Sets and Systems*, vol. 151,2, pp. 211-236, 2005.

[7] M. Grabisch, Ch. Labreuche. "Bi-capacities II: Choquet integral", *Fuzzy Sets and Systems*, vol. 151,2, pp. 236-236, 2005.

[8] R. Mesiar, L. Jin, A. Stupňanová, "BIOWA operators", in *IPMU 2020*, pp. 419-425, Springer Cham, CCIS, vol. 1238, 2020.

[9] R. Mesiar, A. Stupňanová, L. Jin, "Bipolar ordered weighted averages: BIOWA operators", *Fuzzy Sets and Systems*, vol. 433, pp. 108-121, 2022.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

THE ORNESS MEASURE FOR OWA AND BIOWA OPERATORS [1]

Martin Kalina [2], Biljana Mihailović [3] and Mirjana Štrboja [4]

https://doi.org/10.24867/META.2024.09 Review article

Abstract. The OWA operators are useful tools in muticriteria decision-making. The BIOWA operators are a generalization of OWA operators. The orness measure for some operator gives information about the similarity of aggregation to OR (Max) operator. In this paper an overview on OWA and BIOWA operator and their orness measure are presented.

AMS Mathematics Subject Classification (2020): 28A25, 26E50

Key words and phrases: Aggregation function, OWA operator, BIOWA operator

1. Introduction

An aggregation function is a function that processes together several numerical values to obtain a single representative value [2]. These functions have numerous applications, in economy, finance, computer sciences, image processing, etc. The ordered weighted averages (OWA) introduced by Yager [13] form a special class of aggregation functions that includes the arithmetic average, minimum, maximum and median. The OWA operators are useful in muticriteria decision-making, and also in other fields where ranking is important ([15]). As a generalization of OWA operators, the bipolar ordered weighted averages (BIOWA), recently were introduced by Stupňanová and Jin [12] and later studied by Mesiar et al. in [9]. These new bipolar aggregation functions are based on the bipolar Choquet integral [1]. Besides the bipolar Choquet integral, other types of bipolar integrals were introduced [3, 4, 10, 11]. The

[1]The first author acknowledges the financial support of the project VEGA 2/0128/24 and 1/0036/23 and of the Science and Technology Assistance Agency under contract No. APVV SK-SRB-23-0044. The second and third authors acknowledge the financial support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, Serbian Slovak bilateral project: A Bipolar Approach in Mathematical Models of Decision-Making Processes, the second author by Grant No. 451-03-65/2024-14/200156, the third author by Grant No. 451-03-66/2024-03/ 200125 & 451-03-65/2024-03/200125.

[2]Department of Mathematics, Faculty of Civil Engineering, Slovak University of Technology, Slovakia, e-mail: martin.kalina@stuba.sk

[3]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: lica@uns.ac.rs

[4]Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, e-mail: mirjana.strboja@dmi.uns.ac.rs

notion of bipolarity in the framework of multi-criteria decision making was observed by Jin et al. in [5]. Recently, several other generalizations of OWA were studied ([6, 7, 8]).

The paper is organized as follows. In Section 2, we give an overview on the function aggregations and notations that is used in the rest of paper. In Section 3, we present the OWA operator with related the orness measure. The BIOWA operators and the measure of orness associated with BIOWA are given in Section 4.

2. Preliminaries

Let $\mathbb{I}$ be a nonempty subinterval of $[-\infty, \infty]$. A function of n variable $F : \mathbb{I}^n \to \mathbb{I}$ that is nondecreasing in each variable and that satisfies the boundary conditions

$$\inf_{(f_1,f_2,...,f_n)\in\mathbb{I}^n} F(f_1, f_2, ..., f_n) = \inf \mathbb{I} \text{ i } \sup_{(f_1,f_2,...,f_n)\in\mathbb{I}^n} F(f_1, f_2, ..., f_n) = \sup \mathbb{I}.$$

is an aggregation function on $\mathbb{I}$ [2]. Moreover, F is a symmetric aggregation function if

$$F(f_1, f_2, ..., f_n) = F\left(f_{\tau(1)}, f_{\tau(2)}, ..., f_{\tau(n)}\right),$$

for all $(f_1, f_2, ..., f_n) \in \mathbb{I}^n$ and any permutation of indexes τ.

In the rest of this paper we consider a non-empty finite set $X = \{x_1, x_2, \ldots, x_n\}$ with $\text{card}(X) = n$. Let f is a real-valued function on X, we denote $f(x_i) = f_i$, $i = 1, \ldots, n$.

3. OWA operator

Let us recall the definition of an OWA operator proposed by Yager [13].

Definition 3.1. *An* OWA *operator is a mapping* $\text{OWA}_{\mathbf{w}} : \mathbb{R}^n \to \mathbb{R}$ *that has an associated weighting vector* $\mathbf{w} = (w_1, \ldots, w_n)$ *with the following properties*

$$w_1 + w_2 + \cdots + w_n = 1 \quad \textit{and} \quad 0 \le w_i \le 1, \ i = 1, \ldots, n$$

given by

$$\text{OWA}_{\mathbf{w}}(f) = \sum_{i=1}^{n} w_i f_{\alpha(n-i+1)},$$

where $\alpha = (\alpha(1), \ldots \alpha(n))$ *is a permutation of indexes such that* $f_{\alpha(1)} \le f_{\alpha(2)} \le \cdots \le f_{\alpha(n)}$, *for all* i.

In the next example some notable OWA operators are given.

Example 3.2. *For*

(i) $\mathbf{w} = \mathbf{w}^*$ *where* $\mathbf{w}^* = (1, \ldots, 0)$ *we have* $\text{OWA}_{\mathbf{w}^*}(f) = \max\{f_i \mid x_i \in X\}$,

(ii) $\mathbf{w} = \mathbf{w}_*$ *where* $\mathbf{w}_* = (0, \ldots, 1)$ *we have* $\text{OWA}_{\mathbf{w}_*}(f) = \min\{f_i \mid x_i \in X\}$,

(iii) $\mathbf{w} = \mathbf{w}_{Ave}$ *where* $\mathbf{w}_{Ave} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ *we have* $\text{OWA}_{\mathbf{w}_{Ave}}(f) = \frac{1}{n}\sum_{i=1}^{n} f_i$.

It holds

$$\mathrm{OWA}_{\mathbf{w}_*}(f) \leq \mathrm{OWA}_{\mathbf{w}}(f) \leq \mathrm{OWA}_{\mathbf{w}^*}(f),$$

where $\mathbf{w}$ is an arbitrary weighting vector.

3.1. The orness measure for OWA operator

The orness measure for OWA operator was introduced by Yager [13].

Definition 3.3. *The orness measure for an* OWA *operator that has an associated weighting vector* $\mathbf{w}$ *is defined by*

$$\mathrm{orness}(\mathrm{OWA}_{\mathbf{w}}) = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)\,w_i.$$

Obviously, we have

$$\mathrm{orness}(\mathrm{OWA}_{\mathbf{w}^*}) = 1,\ \mathrm{orness}(\mathrm{OWA}_{\mathbf{w}_*}) = 0 \text{ and } \mathrm{orness}(\mathrm{OWA}_{\mathbf{w}_{Ave}}) = 0.5.$$

The orness measure gives information about closeness some OWA operator to the operator $\mathrm{OWA}_{\mathbf{w}^*}$. If $\mathrm{orness}(\mathrm{OWA}_{\mathbf{w}}) \leq 0.5$, then this operator is "orlike" operator, in the case that $\mathrm{orness}(\mathrm{OWA}_{\mathbf{w}}) \geq 0.5$ then it is "andlike" operator [14].

The measure of orness associated with a regular increasing quantifier (RIM), i.e., an increasing function $Q : [0,1] \to [0,1]$ that fulfills $Q(0) = 0$ and $Q(1) = 1$, also was observed in [14]. Hence, for $w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right)$ we have the operator in the following form

$$\mathrm{OWA}_{\mathbf{w}^Q}(f) = \sum_{i=1}^{n}\left(Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right)\right) f_{\alpha(n-i+1)}$$

and

$$\mathrm{orness}(\mathrm{OWA}_{\mathbf{w}^Q}) = \frac{1}{n-1}\sum_{i=1}^{n-1} Q\left(\frac{i}{n}\right).$$

The orness of a RIM quantifier Q is defined by

$$\mathrm{orness}(Q) = \lim_{n\to\infty}\frac{1}{n-1}\sum_{i=1}^{n-1} Q\left(\frac{i}{n}\right) = \int_0^1 Q(x)\,dx.$$

Example 3.4. *For*

(i) $Q(x) = Q^*(x)$ *where* $Q^*(x) = \begin{cases} 0, & x = 0 \\ 1, & x > 0 \end{cases}$ *we get weighting vector* $\mathbf{w}^*$ *and* $\mathrm{orness}(Q^*) = 1$,

(ii) $Q(x) = Q_*(x)$ *where* $Q_*(x) = \begin{cases} 0, & x < 0 \\ 1, & x = 0 \end{cases}$ *we get weighting vector* $\mathbf{w}_*$ *and* $\mathrm{orness}(Q_*) = 0$,

(iii) $Q(x) = Q_{Ave}(x)$ *where* $Q_{Ave}(x) = x$, $x \in [0,1]$, *we get weighting vector* $\mathbf{w}_{Ave}$ *and* $\mathrm{orness}(Q_{Ave}) = 0.5$.

If $Q(x) \geq x$ for all x, then Q is "orlike" and $\mathrm{orness}(Q) \geq 0.5$, and if $Q(x) \leq x$ for all x, then Q is "andlike" and $\mathrm{orness}(Q) \leq 0.5$ [14].

4. BIOWA operator

In this section BIOWA operator and related the orness measure are presented, according to [9]. Denote

$$\widetilde{N} = \{(i,j) \mid i,j \in \{0,1,\ldots,\mathrm{card}(X)\}, i+j \leq \mathrm{card}(X)\}.$$

Let $t : \widetilde{N} \to [-1,1]$ be a function such that the following conditions are satisfied:

(i) $t(n,0) = 1$, $t(0,0) = 0$, $t(0,n) = -1$,

(ii) t is increasing in the first coordinate,

(iii) t is decreasing in the second coordinate.

The class of functions $f : X \to \mathbb{R}$ is denoted by $\mathcal{S}$. For $f \in \mathcal{S}$, we use notations

$$X^{+0} = \{x_i \in X \mid f(x_i) \geq 0\},\ X^{-} = \{x_i \in X \mid f(x_i) < 0\}.$$

Observe a function $s : X \to \{(1,0),(0,1)\}$ defined by

$$s(x_i) = \begin{cases} (1,0), & x_{\alpha(i)} \in X^{+0}, \\ (0,1), & x_{\alpha(i)} \in X^{-}, \end{cases} \tag{4.1}$$

where α is any permutation of indexes such that $|f_{\alpha(1)}| \leq |f_{\alpha(2)}| \leq \cdots \leq |f_{\alpha(n)}|$. Denote

$$w_i^s = t\left(\sum_{j=i}^{n} s(x_j)\right) - t\left(\sum_{j=i+1}^{n} s(x_j)\right), \tag{4.2}$$

where $\sum_{j=n+1}^{n} s(x_j) = (0,0)$.

An BIOWA operator $\mathrm{BIOWA}_t : \mathbb{R}^n \to \mathbb{R}$ is given by

$$\mathrm{BIOWA}_t(f) = \sum_{i=1}^{n} w_i^s \cdot |f_{\alpha(i)}|,$$

where weights w_i^s and s are given by (4.1) and (4.2), respectively. The vector $\mathbf{w}_s = (w_1^s, .., w_n^s)$ is weighting vector, w_i^s are weights and function t is called a generating function. Any BIOWA operator is a symmetric aggregation function on $\mathbb{R}$.

Example 4.1. *For*

(i) $t(x,y) = t^*(x,y)$ *where* $t^*(x,y) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, y < n, \\ -1, & y = n, \end{cases}$ *we have* $\mathrm{BIOWA}_{t^*}(f) = \max\{f_i \mid x_i \in X\}$,

(ii) $t(x,y) = t_*(x,y)$ *where* $t_*(x,y) = \begin{cases} 1, & x = n, \\ 0, & x < n, y = 0, \\ -1, & y < n, \end{cases}$ *we have* $\mathrm{BIOWA}_{t_*}(f) = \min\{f_i \mid x_i \in X\}$,

(iii) $t(x,y) = t_{Ave}(x,y)$ *where* $t_{Ave}(x,y) = \frac{x-y}{n}$, *we have* $\mathrm{BIOWA}_{t_{Ave}}(f) = \frac{1}{n}\sum_{i=1}^{n} f_i$.

The BIOWA operator $\mathrm{BIOWA}_t : \mathbb{R}^n \to \mathbb{R}$ is OWA operator if and only if $t(x,y) = t_g(x,y) = g(x) - 1 + g(n-y)$, where $g : \{0,1,\ldots,n\} \to [0,1]$ is an increasing function that fulfills the conditions $g(0) = 0$ and $g(n) = 1$. In that case $w_i = g(n-i+1) - g(n-i)$ and

$$\mathrm{BIOWA}_{t_g}(f) = \sum_{i=1}^{n} w_i \cdot f_{\gamma(i)},$$

where γ is a permutation of indexes such that $f_{\gamma(1)} \leq f_{\gamma(2)} \leq \cdots \leq f_{\gamma(n)}$.

4.1. The orness measure for BIOWA operator

The orness measure for BIOWA operator was introduced in [9].

Definition 4.2. *The orness measure for an* BIOWA *operator* BIOWA_t *is defined by*

$$\mathrm{orness}(\mathrm{BIOWA}_t) = \frac{1}{(n-1)(n+2)} \sum_{(i,j)\in\widetilde{N}} t(i,j) + 0.5.$$

It holds:

$$\mathrm{orness}(\mathrm{BIOWA}_{t^*}) = 1,\ \mathrm{orness}(\mathrm{BIOWA}_{t_*}) = 0 \text{ and } \mathrm{orness}(\mathrm{BIOWA}_{t_{Ave}}) = 0.5.$$

In [9] the BIOWA-quantifiers were introduced and similar as the orness for RIM quantifiers, the orness for BIOWA-quantifiers was proposed.

Let us observe the lattice $(L, \leq_L)$ where $L = \left\{(u,v) \in [0,1]^2 \mid u + v \leq 1\right\}$ and $(u_1,v_1) \leq_L (u_2,v_2)$ iff $u_1 \leq u_2$ and $v_1 \leq v_2$. If $T : L \to [-1,1]$ is an order homomorphism between lattices $(L, \leq_L)$ and $([0,1], \leq)$ satisfying $T(0,0) = 0$, then T is called a BIOWA-quantifiers.

The orness of a BIOWA-quantifier T is defined by

$$\mathrm{orness}(T) = \iint_L T(u,v)\, dudv + 0.5.$$

Example 4.3. *For*

(i) $T(u,v) = T^*(u,v)$ *where* $T^*(u,v) = \begin{cases} 1, & u > 0, \\ 0, & u = 0, v < 1, \\ -1, & v = 1, \end{cases}$ *we have* $\mathrm{orness}(T^*) = 1$,

(ii) $T(u,v) = T_*(u,v)$ *where* $T_*(u,v) = \begin{cases} 1, & u = 1, \\ 0, & u < 1, v = 0, \\ -1, & v < 1, \end{cases}$ *we have* $\mathrm{orness}(T_*) = 0$,

(iii) $T(u,v) = T_{Ave}(u,v)$ *where* $T_{Ave}(u,v) = u - v$, *we have* $\mathrm{orness}(T_{Ave}) = 0.5$.

References

[1] M. Grabisch, Ch. Labreuche, "Bi-capacities II: Choquet integral", *Fuzzy Sets Syst.* 151, pp. 236-236, 2005.

[2] M. Grabisch, J.-L. Marichal, R. Mesiar, E. Pap, *Aggregation Functions*, Encyclopedia of Mathematics and its Applications, No 127, Cambridge University Press, 2009.

[3] S. Greco, F. Rindone, "Bipolar fuzzy integrals", *Fuzzy Sets Syst.* 220, pp. 21–33, 2013.

[4] S. Greco, R. Mesiar, F. Rindone, "Discrete bipolar universal integrals", *Fuzzy Sets Syst.* 252, pp. 55–65, 2014.

[5] L. Jin, Z-S. Chen, J-Y. Zhang, R. R. Yager, R. Mesiar, M. Kalina, H. Bustince, L. Martinez, "Bi-polar preference based weights allocation with incomplete fuzzy relations", *Information Sciences* 621, pp. 308–318, 2023.

[6] L. Jin, R. Mesiar, M. Kalina, R. R. Yager, "Canonical form of ordered weighted averaging operators", *Ann. Oper. Res.* 295(2), pp. 605–631, 2020.

[7] M. Kalina, "Continuous OWA Operators", *Proceedings of EUSFLAT 2021*, Atlantic Press, pp. 589–594, 2021.

[8] M. Kalina, B. Mihailović, M. Štrboja, "On induced bipolar OWA operators", Book of Abstracts of MMA2023, Jurmala, Latvia, 24-24, 2023.

[9] R. Mesiar, A. Stupňanová, L. Jin, "Bipolar ordered weighted averages: BIOWA operators", *Fuzzy Sets Syst.* 433, pp. 108–121, 2022.

[10] B. Mihailović, M. Štrboja, M. Todorov, "Jensen type inequality for the bipolar Shilkret, Sugeno and Choquet integrals", *Acta Polytechnica Hungarica* 18(9), pp. 9–25, 2021.

[11] M. Štrboja, E. Pap, B. Mihailović, "Discrete bipolar pseudo-integrals", *Inf. Sci.* 468, pp. 72–88, 2018.

[12] A. Stupňanová, L. Jin, "BIOWA Operators". In: Lesot MJ. et al. (eds) Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020. Communications in Computer and Information Science, vol 1238. Springer, Cham, pp. 419–425, 2020.

[13] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making", *IEEE Transactions on Systems, Man and Cybernetics* 18, pp. 183-190, 1988.

[14] R. R. Yager, "Families of OWA operators", *Fuzzy Sets Syst.* 59 (2), pp. 125–143, 1993.

[15] R. R. Yager, J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer: Norwell, MA, 1997.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

INTEGRACIJA MOODLE PLATFORME I MAPE UMA U NASTAVNI PROCES

Mikloš Kovač [1], **Goran Radojev** [2] **i Mirjana Brdar** [3]

https://doi.org/10.24867/META.2024.10 Original article

Sažetak. Glavni cilj rada je da predstavi mogućnost integracije Moodle platforme (Modular Object-Oriented Dynamic Learning Environment) u nastavni proces, pomoću zadataka koji su prezentovani učenicima u obliku mape uma. Ovaj rad pruža uvid u prednosti integracije platforme u nastavni proces i daje preporuku za njenu upotrebu u obrazovnom okruženju, kao i preporuku za primenu mapa uma tokom nastavnog procesa sa ciljem poboljšanja postignuća učenika iz matematike, dok u isto vreme predstavlja metodu izvođenja nastave pomoću koje će nastavnici moći učiniti nastavu matematike interesantnijom i zanimljivijom za učenike. Ovaj pristup bi trebalo da ima značajan pozitivan uticaj na motivaciju, angažovanje i dostignuća učenika.

AMS klasifikacija (2020): 97D40, 97G70

Ključne reči: Moodle platforma, mapa uma, nastavni proces

1. Prednosti Moodle platforme

U današnjem digitalnom dobu, tehnologija je postala neizostavan deo obrazovanja, omogućujući nastavnicima i učenicima da pristupe obrazovnom materijalu na inovativne načine. Moodle platforma predstavlja jednu od najkorišćenijih onlajn obrazovnih platformi, koja omogućava nastavnicima da kreiraju kurseve i njima upravljaju, interaktivnim zadacima i materijalima za učenje. Ona pruža brojne prednosti i može unaprediti proces učenja matematike za učenike na sveobuhvatan i interaktivan način. Pomoću Moodle platforme, nastavnici mogu da vrše interakciju sa učenicima kako unutar, tako i van učionice. Ova platforma pruža sve potrebne osnovne alate za dizajniranje interaktivnog okruženja kako bi podržala nastavu i učenje, čime se podstiče angažovanje i motivisanost učenika [2].

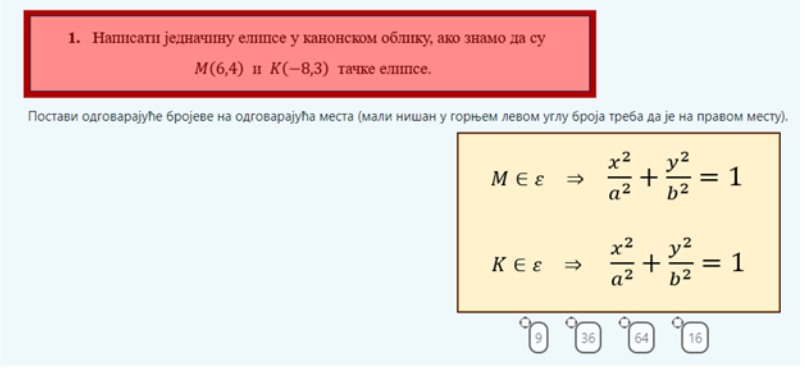
Pre svega, Moodle platforma omogućava pristup nastavnom materijalu u bilo koje vreme i sa bilo kog mesta, što znači da učenici imaju fleksibilnost u procesu učenja. Ovo je posebno važno u nastavi matematike, gde se nastavnik često suočava sa različitim individualnim potrebama učenika, kao i sa različitim tempom učenika potrebnim za savladavanje određene nastavne oblasti.

[1]Elektrotehnička škola "Mihajlo Pupin", Novi Sad, e-mail: kovac.miklos@gmail.com
[2]Departman za matematiku i informatiku, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, e-mail: radojev@dmi.uns.ac.rs
[3]Tehnološki fakultet Novi Sad, Univerzitet u Novom Sadu, e-mail: mirjana.brdar@uns.ac.rs

Korišćenjem Moodle platforme, učenici mogu pristupiti lekcijama, zadacima i dodatnim resursima (Slika 1) kad god im odgovara, što podstiče kontinuirano angažovanje i samostalno učenje.

Svaki učenik ima svoj jedinstveni stil učenja i različite nivoe razumevanja matematike. Ova platforma omogućava nastavnicima da prilagode nastavne materijale prema potrebama svakog učenika. Kroz interaktivne domaće zadatke i povratne informacije nakon svakog koraka, Moodle omogućava personalizovano učenje koje podstiče učenike na postizanje boljih rezultata.

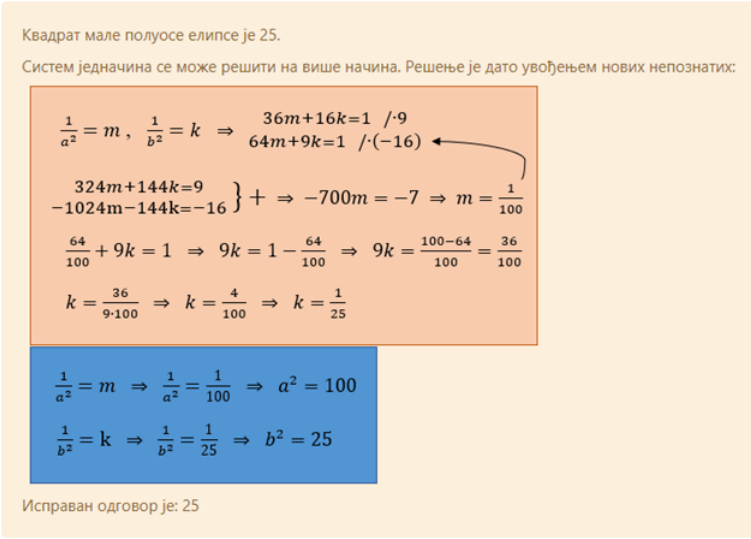
Tradicionalna nastava matematike je često nerazumljiva i neinspirativna za učenike. Tokom učenja, kontinuirano i sukcesivno bombardovanje našeg uma informacijama čini učenje dosadnim i zamornim [7]. Korišćenjem Moodle platforme, nastavnici mogu učiniti nastavu matematike dinamičnijom i interesantnijom kroz interaktivne zadatke i kvizove i na taj način da podstiču učenike na aktivno učešće u rešavanju problema, kao i da olakšaju razumevanje apstraktnih koncepata matematike. Platforma omogućava kreiranje različitih vrsta zadataka, uključujući pitanja sa više ponuđenih odgovora, kvizove, da/ne odgovore, povezivanje pojmova od više ponuđenih, a ovo je tek deo mogućnosti koje pruža Moodle. Ova raznovrsnost omogućava nastavnicima da prilagode zadatke različitim metodama učenja.



Slika 1: Primer pitanja na Moodle platformi

Jedna od ključnih prednosti Moodle platforme je i mogućnost praćenja napretka učenika i evaluacije njihovog rada u realnom vremenu. Nastavnici mogu lako pratiti koji su koncepti učenicima izazovni i pružiti im dodatnu podršku kad je to potrebno. Takođe, učenici mogu dobiti povratne informacije o svom radu odmah nakon svakog koraka, kao i na kraju zadatka, što im omogućava brže učenje i lakše razumevanje, obraćajući pažnju na načinjene greške koje se prezentuju učenicima odmah nakon svake etape zadatka.

Moodle omogućava nastavnicima da pokažu svoju kreativnost prilikom stvaranja materijala prilagođenih aktuelnoj temi. Osim toga, može da pruži pozitivne i motivacione povratne informacije učenicima u realnom vremenu (Slika 2). Pokazano je da primena Moodle platforme povećava angažovanost, uspeh i zadovoljstvo učenika, istovremeno unapređujući fleksibilnost i raznovrsnost njihovog okruženja za učenje [5].



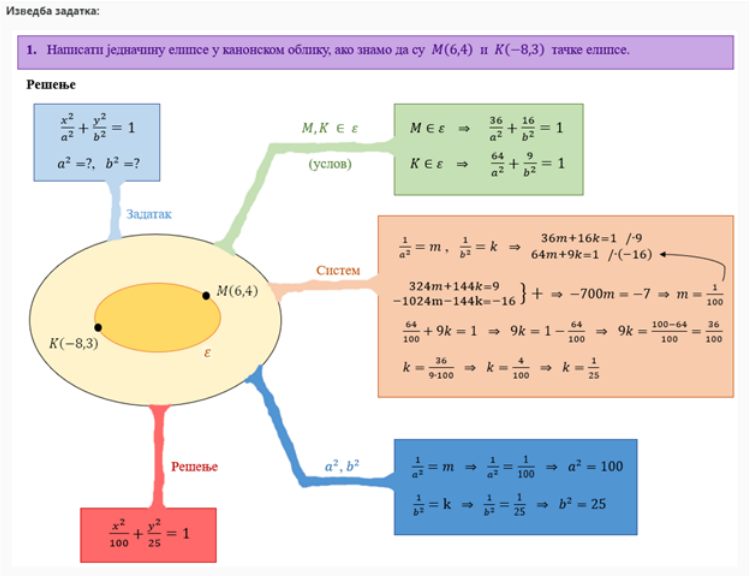
Slika 2: Primer izvedbe zadatka

2. Primena mape uma u nastavi matematike

Potreba za ažuriranjem, grupisanjem i sistematizacijom informacija je ključna, pogotovo u današnjem užurbanom informatičkom svetu. Bez adekvatnog rukovanja i upravljanja informacijama teško je zamisliti svakodnevni život, a školski uspeh učenika takođe se dovodi u pitanje. Bez organizovanja podataka, učenik neće biti u stanju da na adekvatan način usvaja i upravlja svojim znanjem. Mape uma (Slika 3) mogu poslužiti kao veoma efikasan alat u prevazilaženju tih poteškoća.

Ljudi su po prirodi vizuelna bića. To je glavni razlog zašto su mnogi prigrlili tehnike mape uma prilikom usvajanja novih znanja. Mape uma možemo opisati kao vizuelnu tehniku koja predstavlja znanje, ideje, pojmove i odnose među njima u mentalnoj konstrukciji pojedinca na dvodimenzionalnoj ravni [1]. Mapa uma ilustruje strukturu, hijerarhiju i odnos između pojmova, povećava efikasnost procesa usvajanja znanja i promoviše kreativnost [10]. Uz pomoć njih, pamćenje pojmova je efikasnije i trajnije. Učenici lakše povezuju pojmove, uviđaju njihov odnos, problemima prilaze sa više strana i na taj način ih uspešnije rešavaju [4]. Takođe pomažu u uočavanju nelogičnosti i suprotnosti, što motiviše učenike da postavljaju nova pitanja [10]. One povezuju maštu sa strukturom i slike sa logikom [9]. Imajući u vidu date definicije, možemo zaključiti da postoji nekoliko paralelnih mišljenja o tome šta su mape uma. Čini se da je mapa uma hijerarhijska struktura, grafički prikaz, grafički obrazac, nastavno sredstvo, tehnika vođenja beleški, instrument za učenje i metodologija uspostavljanja korelacija između informacija [8].

Mape uma nastaju prenošenjem određenih aspekata znanja u grafički oblik koji krajnji korisnici mogu lako razumeti. Ključna stvar u razvoju mape uma je lociranje bitnog znanja, a zatim organizovanje ovih informacija [4]. U procesu kreiranja mape uma dolazi do slobodnog, spontanog razmišljanja, a cilj je pro-

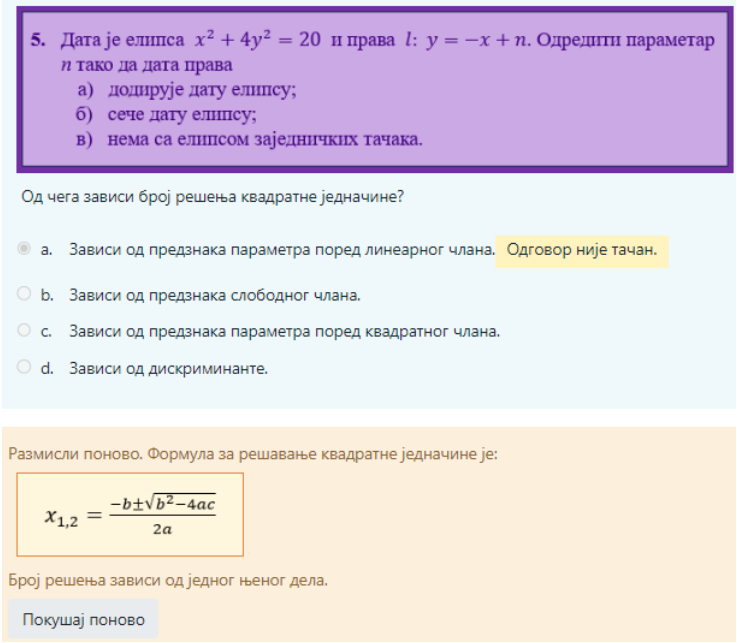


Slika 3: Rešenje zadatka pomoću tehnike mape uma

nalaženje kreativnih asocijacija između ideja. Zbog toga, mape uma se mogu smatrati i mapama asocijacija [3].

3. Primena Moodle platforme i mape uma u nastavnom procesu

Ovaj rad je fokusiran na primenu Moodle platforme u nastavnom procesu, pomoću koje se zadaju interaktivni domaći zadaci za samostalan rad učenika, a pomoć nakon svakog koraka, kao i rešenja zadataka su predstavljeni pomoću mape uma. Glavna nastavna jedinica koja se obrađivala na časovima matematike, uz upotrebu pomenute metode, je bila "Jednačina elipse". Postupak rešavanja zadataka predstavljen je preko mapa uma, sa jasno naznačenim međukoracima ključnim za uspešno rešavanje zadatka. Za vreme izvođenja nastave učenici su podsticani da postavljaju svoja pitanja i da na taj način, uz komunikaciju i diskusiju sa nastavnikom i među sobom, stvaraju nove grane mape uma. Na taj način aktivno su učestvovali u nastavi, donosili svoje zaključke i menjali ih ako su se pokazali netačnim. Neki od zadataka su kreirani pomoću Moodle platforme. Zadaci su zadati na vizuelno dopadljiv način u obliku kviza, primenjujući različite mogućnosti za zadavanje pitanja koje platforma nudi. Svako pitanje je obrađivalo određenu etapu zadatka. U slučaju netačnih odgovora, učenici bi dobili pomoć u obliku grafičke skice i/ili tekstualnog objašnjenja, kao u primeru sa Slike 4. Nakon svake faze zadatka, međukoraci su predstavljeni pomoću tehnike mape uma. Prilikom završetka zadatka, učenici bi dobili jasnu povratnu informaciju o uspešnosti, kao i izvedbu zadatka u obliku mape uma. Nekoliko zadataka postavljenih na Moodle su sa učenicima



Slika 4: Mogući vizuelni prikaz zadatka pomoću Moodle platforme

obrađeni na času, kada su učenici zadacima pristupali preko mobilnih telefona, a preostali zadaci su ostavljeni za domaći zadatak.

Primećeno je da su učenici bili značajno motivisaniji prilikom izrade domaćih zadataka, i u većoj meri su obraćali pažnju na detalje i uslove zadataka. Moodle platforma je omogućila interesantan i interaktivan pristup domaćim zadacima u obliku kviza, a mape uma su obezbedile vizuelno dopadljiv prikaz međukoraka i rešenja zadataka, kao i pomoć učenicima prilikom netačnih odgovora. Na narednim časovima matematike primećeno je bolje razumevanje gradiva, kao i efikasnije povezivanje novih saznanja sa prethodno naučenim lekcijama.

4. Zaključak

U poslednjoj deceniji primetno je da učenici sve teže uspostavljaju smislenu vezu između onoga što uče u školi i životnih problema. Učenici sa najboljim ocenama često imaju poteškoća u primeni znanja pri rešavanju realnih životnih pitanja. Dostupnost elektronskih uređaja kao što su mobilni telefoni omogućili su lak pristup informacijama svima. Međutim, mogućnost da se sve reši odmah i pritiskom na dugme dovelo je do toga da se pamćenje ne koristi u dovoljnoj meri. Moć povezivanja podataka i pojmova kod učenika se smanjuje, a motivacija se gubi. To je samo jedan od razloga zbog kojih konstruktivističke teorije izazivaju sve veću pažnju stručnjaka iz oblasti obrazovanja. Mogućnosti primene konstruktivističkih ideja u nastavnom procesu nisu male, a elementi konstruktivizma su snažni podsticajni i motivacioni faktori za efikasno učenje učenika.

Može se reći da kroz upotrebu Moodle platforme i mape uma kao tehnike koja je zasnovana na konstruktivizmu, otvaramo vrata novim mogućnostima za učenje u nastavi matematike ali i šire, prilagođavajući se potrebama učenika omogućavajući im da dostignu zadate ishode na efikasan i inspirativan način.

Možemo zaključiti da je primena Moodle platforme i tehnike mape uma u nastavnom procesu efektivnija u poređenju sa tradicionalnom, klasičnom metodom nastave.

Zahvalnica

Drugi i treći autor su podržani od strane Ministarstva nauke, tehnološkog razvoja i inovacija Republike Srbije, programi: 451-03-66/2024-03/200125, 451-03-65/2024-3/200125, 451-03-66/2024-03/200134 i 451-03-65/2024-03/200134.

Literatura

[1] A. G. Balım, "The effect of mind-mapping applications on upper primary students' success and inquiry-learning skills in science and environment education", *International Research in Geographical and Environmental Education*, 22(4), pp. 337–352, 2013. DOI: 10.1080/10382046.2013.826543

[2] S. Colombero and P. Dal Zotto, "Moodle: A platform that enables gamification", *Academy of Management Learning and Education*, Vol. 22(4), pp. 743–744, 2023. DOI: 10.5465/amle.2021.0453

[3] M. Davies, "Concept mapping, mind mapping and argument mapping: what are the differences and do they matter?", *Higher Education*, 62(3), pp. 279–301, 2010.

[4] M. Dunić, *Metodičko uputstvo za primenu mape znanja u nastavi informatike i računarstva*, Fakultet tehničkih nauka u Čačku, 2017.

[5] S. H. P. W. Gamage, J. R. Ayres, M. B. Behrend, "A systematic review on trends in using Moodle for teaching and learning", *International Journal of STEM Education*, 9, 2022. DOI: 10.1186/s40594-021-00323-x

[6] von Glasersfeld, E. "Radical Constructivism: A Way of Knowing and Learning", The Falmer Press, 1995.

[7] É. Gyarmathy, Gondolatok térképe. Taní-Tani, 18-19. (108-115. old), 2001.

[8] E. K. Pullu and A. Ü. Kan, "The effect of mind mapping on academic achievement of vocational school students", *International Journal of Eurasian Education and Culture*, 7(16), pp. 238–268, 2022.

[9] I. Svantesson, *Mind Mapping und Gedächtnistraining*, GABAL, Bremen, 1992.

[10] A. Vanko, "Preparation and Application of Mind Maps in Mathematics Teaching and Analysis of their Advantages in Relation to Classical Teaching Methods", *Ratio Mathematica*, Vol. 35, pp. 87-99, 2018. DOI: 10.23755/rm.v35i0.428

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

"EQUATION SOLVING" GENERALIZED INVERSES - WHAT ARE THEY?[1]

Biljana Mihailović [2]

https://doi.org/10.24867/META.2024.11 Short communication

Abstract. A new block representation theorem for the core inverse of square matrix with the index less or equal to 1 is presented. Two examples are given and some open problems are stated, having in mind an application of generalized inverses in solving fuzzy linear systems.

AMS Mathematics Subject Classification (2020): 15A06, 15A09, 15A10

Key words and phrases: system of linear equations, {1}-inverse, Drazin inverse, core inverse, BT inverse

1. Introduction

The most prominent generalized inverse of finite matrices (square or rectangle) in literature known as the Moore-Penrose inverse or "pseudo-inverse", was originally introduced by E.H. Moore (1920) and rediscovered by R. Penrose (1955). Moore's first publication on this unique inverse of matrices, its main properties and its application to linear equations, appeared in 1920, as an abstract of a talk given at a meeting of the American Mathematical Society. Later, his detailed results on the subject were published in 1935. However, some authors considered that Moore's discovery has been obtained much earlier, probably in 1906. Since only very dedicated readers could understand unnecessarily complicated Moore's notation, his work was sinking into oblivion. Penrose rediscovered the general reciprocal, nowadays called the Moore-Penrose inverse, 35 years after Moore's first publication. In 1951, Bjerhammar, who first rediscovered Moore's inverse, also noted the relationship of generalized inverses to solutions of linear systems, whereas, in 1955 Penrose sharpened and extended Bjerhammar's results on linear systems, and showed that Moore's inverse, for a given matrix F, is the unique matrix G satisfying the four equations, nowadays called Penrose's equations. Since 1955 thousands of papers on various aspects of generalized inverses and their applications have appeared.

For a given matrix F, generalized inverses satisfying the first Penrose's equation, e.g. $FGF = F$, are called {1}-inverses or inner inverses. Recently, in [1], Baksalary and Trenkler introduced the core inverse of square matrices with index less or equal to 1, whereas, in [2] they proposed BT inverses. The application of generalized inverses in solving fuzzy linear systems of Friedman et al.'s type was consequently studied by

[1]The author acknowledges the financial support of Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, the project "Improving the teaching process in the English language in fundamental disciplines".

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: lica@uns.ac.rs

numerous authors. In particular, recently, B. Mihailović et al. [7] showed the general algebraic form of fuzzy linear systems based on an {1}-inverse of its coefficient matrix. Very recently, the first straightforward method for solving dual fuzzy linear systems using the block structure of an arbitrary {1}-inverse was introduced by Dragić et al. in [4]. In [6], the core inverse was considered for solving fuzzy linear systems, whereas in [5] the method for solving fuzzy linear systems using the block structure of BT inverse was introduced.

The paper is organized as follows. In Section 1, some preliminaries related to generalized inverses are presented. In Section 2, a new block structure of the core inverse is presented and two open problems are stated.

2. Generalized inverses

Throughout this paper, we denote the set of all $m \times n$ real matrices by $\mathcal{M}^{m\times n}$. For $F \in \mathcal{M}^{m\times n}$ the symbols F^T, F^{-1}, $rank(F)$, $\mathcal{N}(F)$, and $\mathcal{R}(F)$ will stand for the transpose, the ordinary inverse ($m = n$), the rank, the kernel, and the range space of F, respectively. Moreover, let us denote with $\mathcal{M}^n$ the class of all square $n \times n$ real matrices, I_n denotes the identity matrix of order n, and O denotes the null matrix of order n. For $F \in \mathcal{M}^n$, the index of F, denoted by $Ind(F)$, is the smallest non-negative integer k such that $rank(F^{k+1}) = rank(F^k)$. For each $F \in \mathcal{M}^n$, define $F^0 = I_n$. For a non-singular matrix F it holds $Ind(F) = 0$, since $rank(F) = rank(F^0) = rank(I_n) = n$, while $Ind(O) = 1$, since $rank(O^2) = rank(O) = 0$ and $rank(O^0) = n$.

First, let us recall the system of four Penrose's equations for $F \in \mathcal{M}^n$:

(P1) $$FGF = F,$$

(P2) $$GFG = G,$$

(P3) $$(FG)^{\mathrm{T}} = FG,$$

(P4) $$(GF)^{\mathrm{T}} = GF,$$

where matrix $G \in \mathcal{M}^{n\times m}$ is unknown [3]. Let us consider the additional matrix equations as follows:

(P1') $$FGF^2 = F^2,$$

(P2') $$FG^2 = G,$$

(P4') $$(F^TGF^2)^T = F^TGF^2,$$

(P5) $$FG = GF,$$

(P5') $$GF^{k+1} = F^k.$$

Definition 2.1. ([3]) For any $F \in \mathcal{M}^{m\times n}$, let $\mathcal{H}\{i, j, \dots h\}$ denote the set of matrices $G \in \mathcal{M}^{n\times m}$ which fulfill equations $(Pi), (Pj), \dots, (Ph)$ among the equations (P1) to (P5'). A matrix $G \in \mathcal{H}\{i, j, \dots, h\}$ is called an $\{i, j, \dots, h\}$-inverse of F and it will be denoted by $F^{(i,j,\dots,h)}$.

According to [1, 2, 6], let us present the most prominent generalized inverse, the Moore-Penrose inverse, and some recently introduced inverses, the BT inverse and the core inverse.

Definition 2.2. Let $F \in \mathcal{M}^{m\times n}$.

(i) Let $G \in \mathcal{M}^{n\times m}$ be a matrix which fulfills the system of four matrix equations (P1)-(P4). This matrix G is called *the Moore-Penrose inverse* of F, and it is denoted by $F^\dagger$ or by $F^{(1,2,3,4)}$.

(ii) Let $F \in \mathcal{M}^n$ with $Ind(F) = k$, and let $G \in \mathcal{M}^n$ be a matrix which fulfills the system of three matrix equations (P2), (P5) and (P5'). This matrix G is called *the Drazin inverse* of F, and denoted by F^D or by $F^{(2,5,5')}$. Moreover, if $Ind(F) \leq 1$, the Drazin inverse G is called *the group inverse* of F, and denoted by $F^{\#}$ or by $F^{(1,2,5)}$.

(iii) Let $F \in \mathcal{M}^n$ with $Ind(F) = k$, and let $G \in \mathcal{M}^n$ be a matrix which fulfills the system of four matrix equations (P1'), (P2), (P3) and (P4'). This matrix G is called *the BT inverse* of F, and it is denoted by $F^{\diamond}$ or by $F^{(1',2,3,4')}$.

(iv) Let $F \in \mathcal{M}^n$ with $Ind(F) \leq 1$, and let $G \in \mathcal{M}^n$ be a matrix which fulfills the system of three matrix equations (P1), (P2') and (P3). This matrix G is called *the core inverse* of F, and denoted by $F^{\circledcirc\#}$ or by $F^{(1,2',3)}$.

Trivially, for each non-singular square matrix all those generalized inverses are equal to its ordinary inverse. The group inverse exists for all $F \in \mathcal{M}^n$, such that $Ind(F) \leq 1$ and it holds $F^{\#} = F^D$, i.e., in that case the group inverse of a matrix F is identical to its Drazin inverse. Obviously, (P1) implies (P1'), (P4) implies (P4'), but not vice versa. It holds $F^{\diamond} = (F^2F^{\dagger})^{\dagger}$ and if F is a singular matrix, with $Ind(F) = 1$, it holds $F^{\diamond} = F^{\circledcirc\#} = F^{\#}FF^{\dagger}$. Since the Moore-Penrose inverse and the group inverse are uniquely determined, the core inverse and the BT inverse of $F \in \mathcal{M}^n$ are uniquely determined. Recall, a matrix $F \in \mathcal{M}^n$ that satisfies $F^{\dagger}F = FF^{\dagger}$ is called an EP matrix.

For $F \in \mathcal{M}^n$, $F = [f_{ij}]$, we denote with $|F|$ the matrix whose entries are the absolute values of entries of F, i.e., $|F| = [|f_{ij}|]$, $|F| \in \mathcal{M}^n$. We say that F is non-negative matrix if $f_{ij} \geq 0$, for all i, j.

Example 2.3. Let us consider the next singular matrices:

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \qquad |A| = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

In Table 1, the group inverse, the Drazin inverse and the Moore-Penrose inverse, whereas in Table 2, the core inverse, the BT inverse and an arbitrary $\{1\}$-inverse of these matrices are obtained.

F	$F^{\#}$	F^D	$F^{\dagger}$
A, $Ind(A) = 2$	$--$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} \end{bmatrix}$
$\|A\|$, $Ind(\|A\|) = 1$	$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$

Table 1: Generalized inverses of A and $|A|$

Obviously, $|A|$ is a singular matrix with $Ind(|A|) = 1$, it is symmetric, therefore $|A|^{\#} = |A|^{\dagger}$, and it is an EP matrix. On the other hand, the matrix A is a nilpotent matrix satisfying $A^2 = O$, and its index equals 2. Let us solve the next linear system:

$$\begin{array}{ccccc} x_1 & - & x_2 & = & 5 \\ x_1 & - & x_2 & = & 5 \end{array}.$$

F	$F^{\textcircled{\#}}$	$F^{\diamond}$	$F^{(1)}$
$A,\ Ind(A)=2$	$--$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1-a+b+c & a \\ b & c \end{bmatrix},\ a,b,c\in\mathbb{R}$
$\lvert A\rvert,\ Ind(\lvert A\rvert)=1$	$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$	$\begin{bmatrix} 1-a-b-c & a \\ b & c \end{bmatrix},\ a,b,c\in\mathbb{R}$

Table 2: Generalized inverses of A and $|A|$

As it is well-known, the general solution is given in the next form (see, [3]):

$$X = A^{(1)}B + (I_n - A^{(1)}A)V, \tag{2.1}$$

where $A^{(1)}$ denotes any of $\{1\}$-inverses of A and $V = (v_1, \ldots, v_n)^T$ is arbitrary. Using e.g. $A^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, we compute, $A^{(1)}B = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$, where $B = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$, and therefore, $X = \begin{bmatrix} 5 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, for arbitrary $v_1, v_2 \in \mathbb{R}$. Finally, $X = \begin{bmatrix} 5+p \\ p \end{bmatrix},\ p \in \mathbb{R}, \quad SS = \{\,(x_1, x_2)\,|\,x_1 = 5+p, x_2 = p,\ p \in \mathbb{R}\,\}$, where SS stands for the solution set of this system of linear equations. Notice that neither $A^{\diamond}B = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a solution of this system, nor $A^{D}B$.

In general, the Drazin inverse and the BT inverse are not "equation solving" generalized inverses. In the case of square linear systems, when the index of the coefficient matrix is less or equal to 1, we can compute the group inverse and the core inverse that are both "equation solving" generalized inverses, i.e. they are $\{1\}$-inverses of A. However, consider the consistent system of linear equations $x_1 - x_2 = 5$, $x_1 - x_2 = 5$, $x_3 = 3$, with the coefficient matrix M, then $M^{\diamond}B_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$ and $(0, 0, 3)$ is not a solution, but according to Theorem 2.5 in [5], the system $x_1 - x_2 = 0$, $x_1 - x_2 = 0$, $x_3 = 3$, has a solution $M^{\diamond}B_2$ because $B_2 \in \mathcal{R}(M^2)$.

3. A block structure of generalized inverses

In this section, a new block representation theorem of the core inverse is presented. Also, some open problems related to nonnegativity and block structure of "equation solving" generalized inverses are stated.

First, recall that for any $F \in \mathcal{M}^{m\times n}$ of rank r there exist matrices $Q \in \mathcal{M}^{m\times m}$ and $P \in \mathcal{M}^{n\times n}$ such that:

$$QFP = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = E_r, \tag{3.1}$$

where $I_r \in \mathcal{M}^r$ is the identity matrix. A block representation of any $\{1\}$-inverse of F is presented in the next theorem from [7].

Theorem 3.1. *Let $F \in \mathcal{M}^{m\times n}$ be a matrix of rank r. Let Q and P be non-singular, square matrices which fulfill (3.1). A matrix $G \in \mathcal{M}^{n\times m}$ is a solution of the matrix equation $FGF = F$ if and only if*

$$G = P \cdot \begin{bmatrix} I_r & Z_1 \\ Z_2 & Z_3 \end{bmatrix} \cdot Q, \tag{3.2}$$

where $I_r \in \mathcal{M}^r$ is the identity matrix, and $Z_1 \in \mathcal{M}^{r\times(m-r)}$, $Z_2 \in \mathcal{M}^{(n-r)\times r}$ and $Z_3 \in \mathcal{M}^{(n-r)\times(m-r)}$ are arbitrarily chosen matrices.

We will present a new method for computing the core inverse. For each $F \in \mathcal{M}^n$, $rank(F) = r$, there exist non-singular matrices $P, Q \in \mathcal{M}^n$, such that (3.1) holds (for details, see [7]). Such matrices P and Q are not uniquely determined, and they can be obtained by making the same elementary operations by rows or by columns on a matrix F and the identity matrix $I_n \in \mathcal{M}^n$, which provides:

$$\begin{bmatrix} F & I_n \\ I_n & 0 \end{bmatrix} \sim \begin{bmatrix} E_r & Q \\ P & 0 \end{bmatrix}.$$

Further, the product of such matrices Q and P and the product $Q \cdot Q^T$ admit the next block structures:

$$Q \cdot Q^{\mathrm{T}} = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix} \text{ and } Q \cdot P = \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix}, \tag{3.3}$$

where $V_4 \in \mathcal{M}^{n-r}$ denotes a square sub-matrix of order $n-r$, and appropriate sub-matrices are $W_1 \in \mathcal{M}^r$, $W_2 \in \mathcal{M}^{r\times(n-r)}$, $W_3 \in \mathcal{M}^{(n-r)\times r}$, $W_4 \in \mathcal{M}^{n-r}$.

A new representation of the core inverse is presented in the next theorem.

Theorem 3.2. *(The core inverse) Let $F \in \mathcal{M}^n$, $rank(F) = r$ and $Ind(F) = 1$. Let Q and P be non-singular, square matrices which fulfill (3.1), such that the products $Q \cdot P$ and $Q \cdot Q^T$ admit (3.3). A matrix $G \in \mathcal{M}^n$ is the unique solution of the system of three matrix equations (P1), (P2') and (P3) if and only if*

$$G = P \cdot \begin{bmatrix} I_r & -W_2 \cdot W_4^{-1} \\ -V_4^{-1} \cdot V_3 & V_4^{-1} \cdot V_3 \cdot W_2 \cdot W_4^{-1} \end{bmatrix} \cdot Q, \tag{3.4}$$

where $V_4 \in \mathcal{M}^{n-r}$ is non-singular.

The next example and some directions for further investigations related to block structure of generalized inverses of the block matrix S_A, for $A \in \mathcal{M}^{m\times n}$ conclude this paper. Denote $A^+ = \frac{1}{2}(A + |A|)$, $A^- = \frac{1}{2}(|A| - A)$.

Example 3.3. Let $|A|$ be matrix considered in Example 2.3. Then for

$$S_{|A|} = \begin{bmatrix} |A|^+ & |A|^- \\ |A|^- & |A|^+ \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad S_{|A|}^{\dagger} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \end{bmatrix},$$

it holds $S_{|A|}^{\dagger} = S_{|A|}^{D} = S_{|A|}^{\#} = S_{|A|}^{\textcircled{\#}} = S_{|A|}^{\diamond}$ (by Theorem 4([7]), Theorem 5([7]), Theorem 3.2 and Example 2.3, since $Ind(S_{|A|}) = 1$ and $|A|$ is EP matrix).

By Theorem 3 from [7], $G_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -2 & 2 \end{bmatrix}$, $G_1 = S_{|A|}^{(1)}$, is obtained by

$|A|^{(1)} = \begin{bmatrix} 1 & 0 \\ -2 & 2 \end{bmatrix}$. However, one of $\{1\}$-inverses of $S_{|A|}$ is also:

$$G_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \; G_2 = S^{(1)}_{|A|}.$$

Open problem 1: A characterization of the class of matrices A such that each $\{1\}$-inverse of S_A admits the symmetric block structure.
Open problem 2: A characterization of the class of matrices A such that each $\{1\}$-inverse of S_A is non-negative.

References

[1] O.M. Baksalary, G. Trenkler, "Core inverse of matrices", *Linear and Multilinear Algebra,* 58(6), pp. 681–697, 2010.

[2] O.M. Baksalary, G. Trenkler, "On a generalized core inverse", *Applied Mathematics and Computation* 236, pp. 450–457, 2014.

[3] A. Ben-Israel, T. N. E. Greville, *Generalized Inverses: Theory and Applications*, Springer, New York, 2003.

[4] Đ. Dragić, B. Mihailović, Lj. Nedović, "The general algebraic solution of dual fuzzy linear systems and Stein fuzzy matrix equations", *Fuzzy Sets and Systems*, https://doi.org/10.1016/j.fss.2024.108997

[5] M. He, H. Jiang, X. Liu, "General strong fuzzy solutions of fuzzy Sylvester matrix equations involving the BT inverse", *Fuzzy Sets and Systems*, https://doi.org/10.1016/j.fss.2024.108862

[6] H. Jiang, H. Wang, X. Liu, "Solving fuzzy linear systems by a block representation of generalized inverse: the core inverse", *Computational and Applied Mathematics* 39(2), pp. 1–20, 2020.

[7] B. Mihailović, V. M. Jerković, B. Malešević, "Solving fuzzy linear systems using a block representation of generalized inverses: the group inverse", *Fuzzy Sets and Systems* 353, pp. 66–85, 2018.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, June 01-03th 2024

PRIMENA JAKOBIJEVOG POSTUPKA ZA REŠAVANJE FAZI LINEARNIH SISTEMA

Srđan Milićević[1], Biljana Mihailović[2] i Đorđe Dragić[3]

https://doi.org/10.24867/META.2024.12 Review article

Sažetak

U ovom radu predstavljen je algoritam za rešavanje kvadratnog, nesingularnog fazi linearnog sistema oblika $A\tilde{X} = \tilde{Y}$, gde je matrica koeficijenata A data realna matrica, $\tilde{Y}$ je poznat, a $\tilde{X}$ nepoznat vektor fazi brojeva. Algoritam je zasnovan na primeni Jakobijevog postupka i prikazano je njegovo izvođenje i teorema o njegovoj konvergenciji. Rezultati su ilustrovani kroz numerički primer. Rad predstavlja prikaz rada [2].

AMS klasifikacija (2010): 03E72, 93C05

Ključne reči: fazi broj, fazi linearni sistem, Jakobijev postupak

1 Uvod

Sistemi linearnih jednačina imaju značajnu primenu u rešavanju raznih problema u matematici, fizici, statistici, inženjerstvu, kao i mnogim drugim naučnim oblastima. Često se prilikom rešavanja takvih problema susrećemo sa parametrima sistema koji su umesto klasičnim, realnim brojevima, predstavljeni fazi brojevima. Stoga se javlja potreba za rešavanjem fazi linearnih sistema (FLS). U zavisnosti od matrice koeficijenata sistema $A \in \mathbb{R}^{m\times n}$, fazi linearne sisteme delimo na kvadratne ($m = n$) i pravougaone ($m \neq n$). Kvadratni fazi linearni sistemi po prirodi rešenja mogu se podeliti na singularne i nesingularne. Opšti metod za rešavanje kvadratnog nesingularnog FLS-a dat je 1998. godine u radu [7] (Friedman i drugi). Umesto kvadratnog fazi linearnog sistema formata $n \times n$ oblika $A\tilde{X} = \tilde{Y}$, posmatran je klasičan sistem linearnih jednačina čija je matrica koeficijenata formata $2n \times 2n$. Kasnije su se na temu rešavanja fazi linearnih sistema javili i brojni drugi radovi ([1], [2], [4]).

Struktura rada je organizovana na sledeći način. U sekciji 2 prikazana je teorijska osnova rada. Predstavljene su poznate osobine, definicije i teoreme iz oblasti fazi brojeva i fazi linearnih sistema. U sekciji 3 opisana je konstrukcija Jakobijevog postupka za rešavanje fazi linearnih sistema. Implementacija datog postupka na numeričkom primeru prikazana je u sekciji 4.

[1]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail:srdjan88@uns.ac.rs

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: lica@uns.ac.rs

[3]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail:djordje.dragic@uns.ac.rs

2 Metode za rešavanje fazi linearnih sistema

U ovom delu rada prikazani su osnovni pojmovi, definicije i teoreme vezane za fazi brojeve i rešavanje fazi linearnih sistema.

Definicija 2.1. Fazi broj dat u parametarskom obliku je uređeni par funkcija $\tilde{u} = (\underline{u}(r), \overline{u}(r)),\ r \in [0,1]$, koje zadovoljavaju sledeća svojstva:
1. $\underline{u}(r)$ je ograničena, neprekidna sa leve strane, neopadajuća funkcija na intervalu $[0,1]$;
2. $\overline{u}(r)$ je ograničena, neprekidna sa leve strane, nerastuća funkcija na intervalu $[0,1]$;
3. $\underline{u}(r) \leq \overline{u}(r),\ r \in [0,1]$.

Skup svih fazi brojeva označavamo sa $\mathcal{E}$.

Definicija 2.2. Za proizvoljne fazi brojeve $\tilde{u} = (\underline{u}(r), \overline{u}(r))$ i $\tilde{v} = (\underline{v}(r), \overline{v}(r))$ i realni broj k, za svako $r \in [0,1]$ definišemo:

1. *jednakost:* $\tilde{u} = \tilde{v} \Longleftrightarrow \underline{u}(r) = \underline{v}(r)$ i $\overline{u}(r) = \overline{v}(r)$;
2. *sabiranje:* $[\tilde{u} + \tilde{v}]_r = [\underline{u}(r) + \underline{v}(r), \overline{u}(r) + \overline{v}(r)]$;
3. *množenje skalarom:* $[k\tilde{u}]_r = \begin{cases} [k\underline{u}(r), k\overline{u}(r)], & k \geq 0 \\ [k\overline{u}(r), k\underline{u}(r)], & k < 0 \end{cases}$.

Definicija 2.3. Neka je dat vektor fazi brojeva $\tilde{Y} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n)^T$, $\tilde{y}_i \in \mathcal{E}$, $i \in \{1, 2, \ldots, n\}$ i matrica koeficijenata $A = [a_{ij}] \in \mathbb{R}^{n \times n}$. Linearni sistem u matričnom obliku

$$A\tilde{X} = \tilde{Y}, \tag{2.1}$$

gde je $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^T$, $\tilde{x}_j \in \mathcal{E}$, $j \in \{1, 2, \ldots, n\}$, nepoznati vektor fazi brojeva, naziva se fazi linearni sistem (FLS).

Definicija 2.4. Vektor fazi brojeva $\tilde{U} = (\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n)^T$ dat sa
$\tilde{u}_j = (\underline{u}_j(r), \overline{u}_j(r)),\ j \in \{1, 2, \ldots, n\},\ r \in [0,1]$, je rešenje FLS-a (2.1) ako važi:

$$\underline{\sum_{j=1}^{n} a_{ij}\tilde{u}_j} = \sum_{j=1}^{n} \underline{a_{ij}\tilde{u}_j} = \underline{y}_i, \quad \text{i} \quad \overline{\sum_{j=1}^{n} a_{ij}\tilde{u}_j} = \sum_{j=1}^{n} \overline{a_{ij}\tilde{u}_j} = \overline{y}_i,$$

za sve $i \in \{1, 2, \ldots, n\}$.

Kvadratni nesingularni fazi linearni sistem (2.1) formata $n \times n$ možemo posmatrati kao klasičan sistem linearnih jednačina formata $2n \times 2n$ čija je matrica koeficijenata $S \in \mathbb{R}^{2n \times 2n}$:

$$SX = Y, \tag{2.2}$$

pri čemu je S matrica pridružena matrici A oblika

$$S = \begin{bmatrix} S_1 & S_2 \\ S_2 & S_1 \end{bmatrix},$$

$$X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n)^T \text{ i}$$
$$Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n, \overline{y}_1, \overline{y}_2, \ldots, \overline{y}_n)^T.$$

Matrice S_1 i S_2 su kvadratne matrice reda n, $S_1 = [a_{ij}^+]$ i $S_2 = [a_{ij}^-]$, gde je

$$a_{ij}^+ = \begin{cases} a_{ij}, & a_{ij} > 0 \\ 0, & \text{inače} \end{cases} \text{ i } a_{ij}^- = \begin{cases} a_{ij}, & a_{ij} < 0 \\ 0, & \text{inače} \end{cases},$$

za sve $i, j \in \{1, 2, \ldots, n\}$. Primetimo da je $A = S_1 + S_2$.

Teorema 2.5. *(Theorem 2.1, [2]) Matrica S iz sistema* (2.2) *je nesingularna ako i samo ako su obe matrice $S_1 + S_2$ i $S_1 - S_2$ nesingularne.*

Ako postoji matrica S^{-1}, gde je matrica S iz sistema (2.2), ona mora imati istu strukturu kao i matrica S, tj. može se zapisati u obliku

$$S^{-1} = \begin{bmatrix} T_1 & T_2 \\ T_2 & T_1 \end{bmatrix},$$

gde su T_1 i T_2 kvadratne matrice reda n.

Teorema 2.6. *(Theorem 3.1, [2]) Ako je S^{-1} nenegativna, jedinstveno rešenje X sistema (2.2) za proizvoljno Y je $X = S^{-1}Y$ i tada je njemu pridruženi vektor fazi brojeva $\tilde{X}$ rešenje FLS (2.1).*

Definicija 2.7. Matrica $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ je strogo dijagonalno dominantna matrica (SDD) ako je $|a_{ii}| > \sum_{j=1,\ j \neq i}^{n} |a_{ij}|$, $i \in \{1, 2, ..., n\}$.

Teorema 2.8. *(Theorem 3.2, [2]) Matrica A u fazi sistemu (2.1) čiji su svi dijagonalni elementi pozitivni je strogo dijagonalno dominantna ako i samo ako je matrica S strogo dijagonalno dominantna.*

3 Jakobijev postupak za rešavanje fazi linearnih sistema

Jakobijev postupak je jedan od najpoznatijih postupaka za rešavanje sistema linearnih jednačina. Naziv je dobio po nemačkom matematičaru iz devetnaestog veka Karlu Gustavu Jakob Jakobiju. Jakobijev postupak za rešavanje fazi linearnih sistema prvobitno je predstavljen u radu [2]. Takođe, u literaturi su poznati i drugi postupci (Gaus-Zajdelov, relaksacioni postupci, itd.) koji obezbeđuju bržu konvergenciju ka rešenju. Zbog jednostavnosti primene, u ovom radu za rešavanje fazi linearnih sistema izabaran je Jakobijev postupak.

Neka je $S = D + L + U$ realna matrica formata $2n \times 2n$ sa pozitivnim dijagonalnim elementima gde je

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_1 \end{bmatrix},\ L = \begin{bmatrix} L_1 & 0 \\ S_2 & L_1 \end{bmatrix} \text{ i } U = \begin{bmatrix} U_1 & S_2 \\ 0 & U_1 \end{bmatrix},$$

pri čemu su L i L_1 donje trougaone matrice, a U i U_1 gornje trougaone matrice. Stoga sistem $SX = Y$ dobija sledeću strukturu:

$$\begin{bmatrix} D_1 & 0 \\ 0 & D_1 \end{bmatrix} \begin{bmatrix} \underline{X} \\ \overline{X} \end{bmatrix} + \begin{bmatrix} L_1 + U_1 & S_2 \\ S_2 & L_1 + U_1 \end{bmatrix} \begin{bmatrix} \underline{X} \\ \overline{X} \end{bmatrix} = \begin{bmatrix} \underline{Y} \\ \overline{Y} \end{bmatrix},$$

odakle je

$$\underline{X} = D_1^{-1}\underline{Y} - D_1^{-1}(L_1 + U_1)\underline{X} - D_1^{-1}S_2\overline{X}$$

i

$$\overline{X} = D_1^{-1}\overline{Y} - D_1^{-1}(L_1 + U_1)\overline{X} - D_1^{-1}S_2\underline{X}.$$

Jakobijev postupak se svodi na izračunavanje sledećih iteracija:

$$\underline{X}^{k+1} = D_1^{-1}\underline{Y} - D_1^{-1}(L_1 + U_1)\underline{X}^k - D_1^{-1}S_2\overline{X^k}$$

i

$$\overline{X^{k+1}} = D_1^{-1}\overline{Y} - D_1^{-1}(L_1 + U_1)\overline{X^k} - D_1^{-1}S_2\underline{X}^k,$$

$X^k = (\underline{X}^k, \overline{X^k})$, $k = 0, 1, 2, .., n-1$, gde je $n \in \mathbb{N}$ ukupan broj iteracija, a početna iteracija X^0 je unapred zadata. Ako je zadata tačnost $\epsilon > 0$, a tačno rešenje nam nije poznato, broj iteracija se može odrediti na osnovu uslova:

$$\frac{||\underline{X}^{k+1} - \underline{X}^k||}{||\underline{X}^{k+1}||} < \epsilon \quad \text{i} \quad \frac{||\overline{X^{k+1}} - \overline{X^k}||}{||\overline{X^{k+1}}||} < \epsilon,$$

gde je za $X = X(r) \in \mathbb{R}^n$ preslikavanje $||\ ||\ : \mathbb{R}^n \to [0, \infty)$ dato sa

$$||X(r)|| = \max_{r \in [0,1]} |x_k(r)|,\ k \in \{1, 2, ..., n\}.$$

Zapisan u matričnoj formi, Jakobijev postupak ima oblik $X^{k+1} = PX^k + C$, gde je:

$$P = \begin{bmatrix} -D_1^{-1}(L_1 + U_1) & -D_1^{-1}S_2 \\ -D_1^{-1}S_2 & -D_1^{-1}(L_1 + U_1) \end{bmatrix},\ C = \begin{bmatrix} D_1^{-1}\underline{Y} \\ D_1^{-1}\overline{Y} \end{bmatrix} \text{ i } X = \begin{bmatrix} \underline{X} \\ \overline{X} \end{bmatrix}.$$

Teorema 3.1. *([8], p. 120) Neka je matrica A u fazi sistemu (2.1) SDD matrica. Tada Jakobijev postupak konvergira ka rešenju sistema (2.2) za proizvoljno* $X^0 \in \mathbb{R}^{2n}$.

4 Numerički rezultati

Jakobijev postupak konvergira ako je matrica fazi sistema SDD matrica. Međutim, može se dogoditi da sistem konvergira i kada matrica sistema nije SDD, što nam pokazuje naredni primer.

Primer 4.1. Posmatrajmo sledeći 2×2 fazi linearni sistem:

$$\begin{array}{rcrcl} \tilde{x}_1 & - & \tilde{x}_2 & = & (r, 2-r) \\ \tilde{x}_1 & + & 3\tilde{x}_2 & = & (4+r, 7-2r) \end{array}.$$

Prvo ćemo rešenje tražiti primenom Jakobijevog postupka uzimajući za početni vektor $X^0 = \begin{bmatrix} r \\ r \\ 2-r \\ 2-r \end{bmatrix}$, tj. $\underline{X}^0 = \begin{bmatrix} r \\ r \end{bmatrix}$ i $\overline{X^0} = \begin{bmatrix} 2-r \\ 2-r \end{bmatrix}$.

Iz sistema nalazimo da je:

$$\underline{Y} = \begin{bmatrix} r \\ 4+r \end{bmatrix},\ \overline{Y} = \begin{bmatrix} 2-r \\ 7-2r \end{bmatrix},\ D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},\ L_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

$$U_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ i } S_2 = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}.$$

Koristeći jednačine

$$\underline{X}^{k+1} = D_1^{-1}\underline{Y} - D_1^{-1}(L_1+U_1)\underline{X}^k - D_1^{-1}S_2\overline{X^k}$$

i

$$\overline{X^{k+1}} = D_1^{-1}\overline{Y} - D_1^{-1}(L_1+U_1)\overline{X^k} - D_1^{-1}S_2\underline{X}^k,$$

dobijamo da je

$$\underline{X}^1 = \frac{1}{3}\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} r \\ 4+r \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\cdot\begin{bmatrix} r \\ r \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\cdot\begin{bmatrix} 2-r \\ 2-r \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{4}{3} \end{bmatrix} \text{ i}$$

$$\overline{X^1} = \frac{1}{3}\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 2-r \\ 7-2r \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\cdot\begin{bmatrix} 2-r \\ 2-r \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\cdot\begin{bmatrix} r \\ r \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{5}{3}-\frac{r}{3} \end{bmatrix}.$$

Stoga je $X^1 = \begin{bmatrix} 2 \\ \frac{4}{3} \\ 2 \\ \frac{5}{3}-\frac{r}{3} \end{bmatrix}$, a analogno dobijamo i naredne iteracije:

$$X^2 = \begin{bmatrix} \frac{5}{3}+\frac{2r}{3} \\ \frac{2}{3}+\frac{r}{3} \\ \frac{10}{3}-r \\ \frac{5}{3}-\frac{2r}{3} \end{bmatrix},\ X^3 = \begin{bmatrix} \frac{5}{3}+\frac{r}{3} \\ \frac{7}{9}+\frac{r}{9} \\ \frac{8}{3}-\frac{2r}{3} \\ \frac{11}{9}-\frac{r}{3} \end{bmatrix},\ X^4 = \begin{bmatrix} \frac{11}{9}+\frac{2r}{3} \\ \frac{7}{9}+\frac{2r}{9} \\ \frac{25}{9}-\frac{8r}{9} \\ \frac{13}{9}-\frac{4r}{9} \end{bmatrix},\ X^5 = \begin{bmatrix} \frac{13}{9}+\frac{5r}{9} \\ \frac{25}{27}+\frac{r}{9} \\ \frac{25}{9}-\frac{7r}{9} \\ \frac{38}{27}-\frac{10r}{27} \end{bmatrix},$$

$$X^6 = \begin{bmatrix} \frac{38}{27}+\frac{17r}{27} \\ \frac{23}{27}+\frac{4r}{27} \\ \frac{79}{27}-\frac{8r}{9} \\ \frac{38}{27}-\frac{11r}{27} \end{bmatrix},\ X^7 = \begin{bmatrix} \frac{38}{27}+\frac{16r}{27} \\ \frac{70}{81}+\frac{10r}{81} \\ \frac{77}{27}-\frac{23r}{27} \\ \frac{110}{81}-\frac{30r}{81} \end{bmatrix},\ X^8 = \begin{bmatrix} \frac{110}{81}+\frac{51r}{81} \\ \frac{70}{81}+\frac{11r}{81} \\ \frac{232}{81}-\frac{71r}{81} \\ \frac{112}{81}-\frac{31r}{81} \end{bmatrix} \text{ i}$$

$$X^9 = \begin{bmatrix} \frac{112}{81} + \frac{50r}{81} \\ \frac{214}{243} + \frac{10r}{81} \\ \frac{232}{81} - \frac{70r}{81} \\ \frac{335}{243} - \frac{91r}{243} \end{bmatrix}.$$

Uzimajući poslednje sračunatu iteraciju kao približno rešenje posmatranog fazi linearnog sistema dobijamo:
$\tilde{x}_1 \approx (1.383 + 0.617r,\ 2.864 - 0.864r)$ i $\tilde{x}_2 \approx (0.881 + 0.124r,\ 1.379 - 0.374r)$.
Tačno rešenje sistema je:

$$X = S^{-1}Y = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & 3 & 0 & 0 \\ 0 & -1 & 1 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}^{-1} \cdot \begin{bmatrix} r \\ 4+r \\ 2-r \\ 7-2r \end{bmatrix},$$

tj. $\tilde{x}_1 = (1.375 + 0.625r,\ 2.875 - 0.875r)$ i $\tilde{x}_2 = (0.875 + 0.125r,\ 1.375 - 0.375r)$.
Primetimo da je $\dfrac{||\underline{X} - \underline{X}^9||}{||\underline{X}||} < 0.01$ i $\dfrac{||\overline{X} - \overline{X^9}||}{||\overline{X}||} < 0.01$.

Zahvalnica

Autori su podržani od strane projekta ,,Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama" Departmana za opšte discipline u tehnici, Fakulteta tehničkih nauka, Univerziteta u Novom Sadu.

Literatura

[1] S. Abbasbandy and M. Alavi, "A method for solving fuzzy linear systems", *Iranian Journal of Fuzzy Systems*, vol. 2, pp. 37–43, 2005.

[2] T. Allahviranloo, "Numerical methods for fuzzy systems of linear equations", *Applied Mathematics and Computation*, vol. 155, pp. 493–502, 2004.

[3] T. Allahviranloo and M.A. Kermani, "Solution of a fuzzy system of linear equation", *Applied Mathematics and Computation*, vol. 175, pp. 519–531, 2006.

[4] T. Allahviranloo and M. Ghanbari, "On the algebraic solution of fuzzy linear systems based on interval theory", *Applied Mathematical Modelling*, vol. 36, 5360–5379, 2012.

[5] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

[6] R. Ezzati, "Solving fuzzy linear systems", *Soft. Comput.*, vol. 15, pp. 193–197, 2011.

[7] M. Friedman, M. Ming and A. Kandel, "Fuzzy linear systems", *Fuzzy Sets and Syst,* vol. 96, pp. 201–209, 1998.

[8] J.M. Ortega, *Numerical Analysis a Second Course*, Siam, 1990.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st - June 2nd 2024

BIPOLAR FUZZY LINEAR SYSTEMS WITH A UNIQUE SOLUTION[1]

Ljubo Nedović [2], **Biljana Mihailović** [3] **and Đorđe Dragić** [4]

https://doi.org/10.24867/META.2024.13 Short communication

Abstract. For a given bipolar fuzzy number vector Y and a real matrix A, the system of equations $AX = Y$ is called a bipolar fuzzy linear system, where X is an unknown bipolar fuzzy number vector. We present a new method for solving bipolar fuzzy linear systems with a unique solution and illustrate this new approach by an example.

AMS Mathematics Subject Classification (2020): 03E72, 15A09

Key words and phrases: bipolar fuzzy numbers, bipolar fuzzy linear systems, completely non-singular matrix

1. Introduction

Fuzzy linear systems (FLS), introduced by Friedman et al. in [5], arose as a generalization of linear systems. There are numerous papers devoted to FLS of Friedman et al.'s type. In [2] Allahviranloo and Ghanbari presented a new, efficient method for obtaining exact algebraic solutions of a square FLS whose coefficient matrix is non-singular. The application of inner inverses ($\{1\}$-inverses) in solving fuzzy linear systems was consequently studied by numerous authors. A general algebraic solution of fuzzy linear systems of Friedman et al.'s type was characterized for the first time by Mihailović et al. in [6]. Also, the algorithm for solving non-square FLS based on the Moore-Penrose inverse of its coefficient matrix was presented in this paper. Recently, in [8], a straightforward method for solving $m \times n$ FLS $A\tilde{X} = \tilde{Y}$ was introduced, as a generalization of the obtained results from [6, 7]. The first straightforward method for solving dual fuzzy linear systems using arbitrary $\{1\}$-inverses of its coefficient matrices was introduced by Dragić et al. in [3, 4].

Bipolar fuzzy linear systems (BFLS) were introduced in [1]. The main aim of this paper is to propose a new method for solving square BFLS with a unique

[1]The authors acknowledge the financial support of the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, Serbian Slovak bilateral project: A Bipolar Approach in Mathematical Models of Decision-Making Processes.

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: nljubo@uns.ac.rs

[3]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: lica@uns.ac.rs

[4]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: djordje.dragic@uns.ac.rs

solution. The paper is organized as follows. In Section 1, some preliminaries related to bipolar fuzzy numbers and bipolar fuzzy linear systems are presented. In Section 2, the general form of unique (strong) solutions of square BFLS is presented and an illustrative example is given.

2. Bipolar fuzzy linear systems

Recall some basic definitions and terminology ([1, 5, 7]).

Definition 2.1. [1] A bipolar fuzzy set (BFS) u in parametric form is a quadruple $\prec \underline{u}^P, \overline{u}^P, \underline{u}^N, \overline{u}^N \succ$ of the functions $\underline{u}^P(r)$, $\overline{u}^P(r)$, $\underline{u}^N(s)$, $\overline{u}^N(s)$; $0 \leq r \leq 1$, $-1 \leq s \leq 0$, satisfying the next conditions:

(i) $\underline{u}^P(r)$ is a bounded monotonically increasing (non-decreasing) left-continuous function on a set $(0, 1]$ and right-continuous at point 0,

(ii) $\overline{u}^P(r)$ is a bounded monotonically decreasing (non-increasing) left-continuous function on a set $(0, 1]$ and right-continuous at point 0,

(iii) $\underline{u}^N(s)$ is a bounded monotonically decreasing (non-increasing) left-continuous function on a set $(-1, 0]$ and right-continuous at point -1,

(iv) $\overline{u}^N(s)$ is a bounded monotonically increasing (non-decreasing) left-continuous function on a set $(-1, 0]$ and right-continuous at point -1,

(v) $\underline{u}^P(r) \leq \overline{u}^P(r)$,

(vi) $\underline{u}^N(s) \leq \overline{u}^N(s)$.

The set of all bipolar fuzzy numbers will be denoted by $\mathcal{B}$. For $u, v \in \mathcal{B}$, in parametric form $u = \prec \underline{u}^P, \overline{u}^P, \underline{u}^N, \overline{u}^N \succ$, $v = \prec \underline{v}^P, \overline{v}^P, \underline{v}^N, \overline{v}^N \succ$ any real number k, we define:

1. *Addition:* $u + v = \prec \underline{u}^P + \underline{v}^P, \overline{u}^P + \overline{v}^P, \underline{u}^N + \underline{v}^N, \overline{u}^N + \overline{v}^N \succ$,

2. *Scalar multiplication:* $$ku = \begin{cases} \prec k\underline{u}^P, k\overline{u}^P, k\underline{u}^N, k\overline{u}^N \succ, & k \geq 0, \\ \prec k\overline{u}^P, k\underline{u}^P, k\overline{u}^N, k\underline{u}^N \succ, & k < 0, \end{cases}$$

3. *Equality:* $u = v \Leftrightarrow \underline{u}^P = \underline{v}^P,\ \overline{u}^P = \overline{v}^P,\ \underline{u}^N = \underline{v}^N,\ \overline{u}^N = \overline{v}^N$.

We will use the next notation and terminology. Let $\mathcal{BV}_n$ denotes the class of all n-dimensional bipolar fuzzy number vectors. Let $X = (x_1, \ldots, x_n)^T$ denotes a bipolar fuzzy number vector, where $x_i \in \mathcal{B}$, $x_i = \prec \underline{x}_i^P, \overline{x}_i^P, \underline{x}_i^N, \overline{x}_i^N \succ$, for all $i = 1, \ldots, n$. For $X \in \mathcal{BV}_n$, the associated $2n \times 1$ classical functional vectors $X^P = (\underline{x}_1^P, \ldots, \underline{x}_n^P, -\overline{x}_1^P, \ldots, -\overline{x}_n^P)^T$, $X^N = (\underline{x}_1^N, \ldots, \underline{x}_n^N, -\overline{x}_1^N, \ldots, -\overline{x}_n^N)^T$ such that all its components are the unit interval functions, will be called the representative vectors for X. For any $X \in \mathcal{BV}_n$, the associated $n \times 1$ classical functional vectors $\underline{X}^P$ and $\underline{X}^N$, (resp. $\overline{X}^P$ and $\overline{X}^N$) with the lower (resp. upper) branches as their components, are $\underline{X}^P = (\underline{x}_1^P, \ldots, \underline{x}_n^P)^T$ and $\underline{X}^N = (\underline{x}_1^N, \ldots, \underline{x}_n^N)^T$ (resp. $\overline{X}^P = (\overline{x}_1^P, \ldots, \overline{x}_n^P)^T$ and $\overline{X}^N = (\overline{x}_1^N, \ldots, \overline{x}_n^N)^T$). We will use the notation $\overline{X}^P \geq \underline{X}^P$ if and only if for all $i = 1, \ldots, n$, and for each $r \in [0, 1]$, it holds $\overline{x}_i^P(r) \geq \underline{x}_i^P(r)$, and similarly $\overline{X}^N \geq \underline{X}^N$.

For a given $Y \in \mathcal{BV}_n$ and real square matrix A of order n, *the bipolar fuzzy linear system* in the matrix form is $AX = Y$, where $X \in \mathcal{BV}_n$ is unknown.

For each real square matrix A of order n, denote $A^+ = \frac{1}{2}(A + |A|)$, $A^- = \frac{1}{2}(|A| - A)$, where $|A|$ is a square real matrix of order n whose entries are the absolute values of entries of A. Let S_A be a square matrix of order $2n$ defined by

$$S_A = \begin{bmatrix} A^+ & A^- \\ A^- & A^+ \end{bmatrix}. \tag{2.1}$$

Definition 2.2. For a given $Y \in \mathcal{BV}_n$ and real square matrix A of order n, *a solution* of BFLS $AX = Y$ is any $U \in \mathcal{BV}_n$ such that for all $r \in [0,1]$ it holds $S_A U^P(r) = Y^P(r)$ and for all $s \in [-1,0]$ it holds $S_A U^N(s) = Y^N(s)$.

3. A unique solution of BFLS

As a new result, we present a necessary condition for the consistency of BFLS with completely non-singular coefficient matrix. According to [2, 5], recall that A is a completely non-singular matrix if both matrices A and $|A|$ are non-singular and S_A is non-singular if and only if A is completely non-singular.

Theorem 3.1. *Let $AX = Y$ be a BFLS, where A is a completely non-singular matrix of order n and $Y \in \mathcal{BV}_n$. Let $X^{*P} = S_M Y^P$ and $X^{*N} = S_M Y^N$, where $M = A^{-1}$ and S_M is given by (2.1). Let $\mathsf{R}^P = Y^P - S_A X^{*P}$ and $\mathsf{R}^N = Y^N - S_A X^{*N}$.*

If the bipolar fuzzy linear system $AX = Y$ is consistent, then there exist $\mathsf{V}^P = \begin{bmatrix} V^P \\ V^P \end{bmatrix}$, and $\mathsf{V}^N = \begin{bmatrix} V^N \\ V^N \end{bmatrix}$, where $V^P = (v_1^P(r), \ldots, v_n^P(r))^T$ and $V^N = (v_1^N(s), \ldots, v_n^N(s))^T$, such that for all $r \in [0,1]$ and $s \in [-1,0]$ it holds $S_A \mathsf{V}^P = \mathsf{R}^P$, $S_A \mathsf{V}^N = \mathsf{R}^N$, $2V^P \leq \overline{X}^{*P} - \underline{X}^{*P}$, and $2V^N \leq \overline{X}^{*N} - \underline{X}^{*N}$.

Moreover, a unique solution X of BFLS is determined by $X^P = X^{*P} + \mathsf{V}^P$ and $X^N = X^{*N} + \mathsf{V}^N$.

Example 3.2. Let us solve the next 2×2 bipolar fuzzy linear system:

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & \prec -2 + 2r, 5 - 5r, -2 - 4s, 13 + 11s \succ \\ x_1 & - & 2x_2 & = & \prec -6 + 3r, 3 - 6r, -17 - 4s, 1 + 14s \succ \end{array}.$$

Matrices A, A^+ and A^- of this BFLS are:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix},\ A^+ = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},\ A^- = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

Since both matrices A and $|A|$ are invertible, the matrix A is completely non-singular. The classical inverse of A is:

$$M = A^{-1} = -\frac{1}{3}\begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix}.$$

First, we compute $X^{*P} = S_M Y^P$, and $X^{*N} = S_M Y^N$, and obtain:

$$\begin{aligned} X^{*P} &= \left(-\frac{10}{3} + \frac{7}{3}r, -\frac{5}{3} + \frac{8}{3}r, -\frac{13}{3} + \frac{16}{3}r, -\frac{11}{3} + \frac{8}{3}r\right)^T, \\ X^{*N} &= (-7 - 4s, -1 - 6s, -9 - 12s, -10 - 5s)^T. \end{aligned}$$

We obtain $\mathsf{R}^P = Y^P - S_A X^{*P} = \left(3-3r, \frac{14}{3} - \frac{14}{3}r, 3-3r, \frac{14}{3} - \frac{14}{3}r\right)^T$ and $\mathsf{R}^N = Y^N - S_A X^{*N} = (6+6s, 10+10s, 6+6s, 10+10s)^T$. Further, the unique solutions of $S_A \mathsf{V}^P = \mathsf{R}^P$ and $S_A \mathsf{V}^N = \mathsf{R}^N$ are:

$$\begin{aligned} \mathsf{V}^P &= \left(\frac{4}{3}(1-r), \frac{5}{3}(1-r), \frac{4}{3}(1-r), \frac{5}{3}(1-r)\right)^T, \\ \mathsf{V}^N &= (2(1+s), 4(1+s), 2(1+s), 4(1+s))^T, \end{aligned}$$

respectively. In order to obtain a bipolar fuzzy number vector, i.e., in order to $\overline{x}_i^P(r) \geq \underline{x}_i^P(r)$ be fulfilled for $i = 1, 2$ and each $r \in [0, 1]$, V^P should to satisfy the next necessary condition $2v_i^P(r) \leq \overline{x_i}^{*P}(r) - \underline{x}_i^{*P}(r)$, $i = 1, 2$, for each $r \in [0, 1]$. In that sense, the obtained V^P is feasible, similarly, V^N is feasible, therefore, the unique solution of BFLS is $X = (x_1, x_2)^T$, given by:

$$\begin{aligned} x_1 &= \prec -2+r, 3-4r, -5-2s, 7+10s \succ, \\ x_2 &= \prec r, 2-r, 3-2s, 6+s \succ . \end{aligned}$$

References

[1] M. Akram, G. Muhammad, T. Allahviranloo, "Bipolar fuzzy linear system of equations", *Computational and Applied Mathematics* 38, article number 69, 2019.

[2] T. Allahviranloo, M. Ghanbari, "On the algebraic solution of fuzzy linear systems based on interval theory", *Applied Mathematical Modelling* 36, pp. 5360–5379, 2012.

[3] Đ. Dragić, B. Mihailović, Lj. Nedović, "The general algebraic solution of dual fuzzy linear systems", *Book of Abstracts of FSTA*, Liptovsky Jan, Slovak Republic, pp. 58–59, 2024.

[4] Đ. Dragić, B. Mihailović, Lj. Nedović, "The general algebraic solution of dual fuzzy linear systems and Stein fuzzy matrix equations", *Fuzzy Sets and Systems*, https://doi.org/10.1016/j.fss.2024.108997

[5] M. Friedman, M. Ming, A. Kandel, "Fuzzy linear systems", *Fuzzy Sets and Systems* 96, pp. 201–209, 1998.

[6] B. Mihailović, V. Miler Jerković, B. Malešević, "Solving fuzzy linear systems using a block representation of generalized inverses: The Moore-Penrose inverse", *Fuzzy Sets and Systems* 353, pp. 44–65, 2018.

[7] B. Mihailović, V. Miler Jerković, B. Malešević, "Solving fuzzy linear systems using a block representation of generalized inverses: The group inverse", *Fuzzy Sets and Systems* 353, pp. 66-85, 2018.

[8] V. Miler Jerković, B. Mihailović, B. Malešević, "The general algebraic solution of fuzzy linear systems based on a block representation of 1-inverses", *Iranian Journal of Fuzzy Systems* 20 (3), pp. 115-126, 2023.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

ON A SIMPLE SCALING CONDITION FOR $H-$MATRICES AND APPLICATIONS [1]

Maja Nedović [2] and Srđan Milićević [3]

https://doi.org/10.24867/META.2024.14 Original scientific paper

Abstract. In this paper we consider a simple scaling condition for a special subclass of $H-$matrices and we present applications of this condition in eigenvalue localization for Schur complements of matrices in this subclass. Spectra localizations for the Schur complement matrix are based on the construction of the diagonal scaling matrix for the given $H-$matrix and use only the entries of the original matrix.

AMS Mathematics Subject Classification (2020): 15A18, 15B99

Key words and phrases: $H-$matrices, Schur complement, Eigenvalue localization

1. Introduction

The theory of $H-$matrices, together with related knowledge on classes of $M-$matrices and $P-$matrices, represents a research area of interest for mathematicians as well as for researchers in the field of economy, ecology, engineering. It is well-known that investigation of some special subclasses of $H-$matrices brought to light different possibilities for localizing spectra of square complex matrices in general. Spectra localizations provide important information on stability of dynamical systems. Also, $H-$matrices and related classes play an important role in research on existence and uniqueness of solutions in linear complementarity problems, in construction of iterative procedures for solving these problems and in error analysis. Schur complement appears in block-Gaussian elimination and proved to be useful in reducing the dimension of the problem in solving linear systems of equations. When considering Schur complements of $H-$matrices, in recent years different authors provided results on closure properties of some special subclasses of $H-$matrices under Schur complement. Also, dominance degree of Schur complement was discussed and compared to the dominance degree of the original matrix.

[1]The authors acknowledge the financial support of Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama", "Improving the teaching process in the English language in fundamental disciplines".

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: maja.nedovic@uns.ac.rs

[3]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: srdjan88@uns.ac.rs

In this paper we deal with a special subclass of $H-$matrices defined by a simple condition based on scaling. We provide information on eigenvalues of Schur complements for matrices in this class.

2. Special $H-$matrices

In this section we recall results on SDD matrices and $H-$matrices and deal with a simple condition defining $SDD-$scal matrices.

A matrix $A = [a_{ij}] \in \mathbb{C}^{n,n}$ is a strictly diagonally dominant (SDD) matrix if

$$|a_{ii}| > r_i(A), \text{ for all } i \in N = \{1, 2, ..., n\},$$

with $r_i(A)$ being a deleted row sum defined as follows

$$r_i(A) = \sum_{j \in N \setminus \{i\}} |a_{ij}|.$$

This well-known class of non-singular matrices is the starting point for research on wider classes of special $H-$matrices. Its main advantage is that it is defined by a very simple condition, easily checkable, with low calculation cost.

Although there are many different ways to introduce non-singular $H-$matrices, see [1], the scaling characterization given by Fiedler and Pták, see [8], is the most revealing for the subject of this paper. According to [8], a given matrix $A = [a_{ij}] \in \mathbb{C}^{n,n}$ is an H-matrix if and only if there exists a diagonal non-singular matrix D such that AD is an SDD matrix. Moreover, we can always assume that D has only positive diagonal entries.

In literature, there are several subclasses of $H-$matrices that are introduced or further researched through a construction of special diagonal scaling matrices, see [3, 7, 17, 18]. We apply the type of scaling from [3] in the next section in order to obtain spectra information for Schur complements.

Let us consider a different type of row sums for the given complex square matrix with nonzero diagonal entries, as follows. If $a_{ii} \neq 0,\ i \in N$, define

$$R_i(A) = \sum_{k \in N \setminus \{i\}} \frac{r_k(A)}{|a_{kk}|} |a_{ik}|.$$

Let $A \in \mathbb{C}^{n,n},\ n \geq 2$, be a matrix with nonzero diagonal entries and let

$$r_i(A) > R_i(A),$$

for all $i \in N$. We call matrices satisfying this condition $SDD-$scal matrices. It is easy to see that any $SDD-$scal matrix is an $H-$matrix. Namely, define a non-singular diagonal matrix

$$D = diag(d_k),\ k = 1, ..., n,$$

with

$$d_k = \frac{r_k(A)}{|a_{kk}|},\ k = 1, ..., n.$$

Let us consider the matrix AD. It holds that

$$(AD)_{ii} = r_i(A),\ i = 1, ..., n,$$

$$r_i(AD) = \sum_{k \in N \setminus \{i\}} \frac{r_k(A)}{|a_{kk}|} |a_{ik}| = R_i(A),\ i = 1, ..., n.$$

Therefore, as A is an $SDD-$scal matrix, AD is an SDD matrix, implying that A can be scaled to SDD matrix by a non-singular diagonal matrix D. This proves that A is an $H-$matrix.

Although this condition represents a simple modification of SDD condition, it is easy to see that the class of $SDD-$scal matrices is not a subclass of SDD, nor SDD class is a subclass of $SDD-$scal. For example, if we consider any SDD matrix with at least one deleted row sum equal to zero, it is easy to notice that such matrix is not $SDD-$scal. On the other hand, there are matrices that belong to $SDD-$scal class, but do not belong to SDD, such as the matrix B, while the intersection of these classes is not empty, as there exist matrices that belong both to SDD and $SDD-$scal, such as the matrix C,

$$B = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 2 & 1 & 0 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & 2 \end{bmatrix}.$$

Therefore, classes SDD and $SDD-$scal stand in a general relation. If we consider the recently introduced class of SDD_1 matrices, see [14, 16], that contains SDD class, it is easy to see that B does not belong to SDD_1, meaning that $SDD-$scal and SDD_1 also stand in a general relation.

Recall that the famous Geršgorin theorem, that provides an eigenvalue localization set for an arbitrary complex square matrix, is equivalent to the statement that every SDD matrix is non-singular. Geršgorin localization set is the union of n Geršgorin disks, one for each row of the given matrix, defined as $\Gamma_i(A) = \{z \in \mathbb{C} \,|\, |z - a_{ii}| \leq r_i(A)\},\ i = 1, 2, ..., n$. For more details on eigenvalue localization sets see [18]. In the next section, we apply Geršgorin localization set and the construction of diagonal scaling matrices for $SDD-$scal matrices in order to define areas that include (or exclude) spectra of Schur complements.

3. Eigenvalues of Schur complements of $SDD-$scal matrices

Spectra localizations and closure properties for Schur complements of some special matrices were researched in [4, 5, 6, 9, 10, 13, 15, 17].

The Schur complement of A with respect to a proper subset of N, α, is denoted by A/α and defined as

$$A(\overline{\alpha}) - A(\overline{\alpha}, \alpha)(A(\alpha))^{-1} A(\alpha, \overline{\alpha})$$

where $A(\alpha, \beta)$ stands for the submatrix of $A \in \mathbb{C}^{n,n}$ consisting of the rows indexed by α and the columns indexed by β, while $A(\alpha, \alpha)$ is abbreviated to $A(\alpha)$. We assume $A(\alpha)$ to be a nonsingular matrix. For details on applications of Schur complement matrices see [2, 19]. In the remainder of this section, we are interested in providing information on eigenvalues of Schur complements without calculating Schur complements. It is, off course, possible to

apply well-known results on localizing spectra once the Schur complement matrix is calculated using the entries of Schur complement matrix. However, in what follows, we provide preliminary information on the eigenvalues of Schur complements using only the entries of the original matrix. These preliminary bounds can be useful when solving large scale systems of linear equations via Schur-based iteration, as the concentration of eigenvalues could predict faster convergence.

In [12], the following result is obtained. If a matrix $A = [a_{ij}] \in \mathbb{C}^{n,n}$ is an SDD matrix with real diagonal entries, and α a proper subset of the index set, then, A/α and $A(\overline{\alpha})$ have the same number of eigenvalues whose real parts are greater (less) than $w(A)$ (resp. $-w(A)$), with

$$w(A) = \min_{j \in \overline{\alpha}} \left[|a_{jj}| - r_j(A) + \min_{i \in \alpha} \frac{|a_{ii}| - r_i(A)}{|a_{ii}|} \sum_{k \in \alpha} |a_{jk}| \right].$$

In [11], one can find another result on the dominant degree and the spectra localization for the Schur complement. If $A \in \mathbb{C}^{n,n}$ and $\alpha = \{i_1, i_2, \ldots, i_k\} \subseteq N_2(A) = \{i \in N : \ |a_{ii}| > r_i(A)\}$, $\overline{\alpha} = \{j_1, j_2, \ldots, j_l\}$, then, for every eigenvalue λ of A/α, there exists $1 \leq t \leq l$ such that

$$|\lambda - a_{j_t j_t}| \leq r_{j_t}(A).$$

This means that the localization area for eigenvalues of A/α can be obtained by taking the union of those Geršgorin disks, formed for the matrix A, whose indices are in $\overline{\alpha}$.

Now we present main results on eigenvalue localization for Schur complements of $SDD-$scal matrices.

Theorem 3.1. *Let $A = [a_{ij}] \in \mathbb{C}^{n,n}$ be an $SDD-$scal matrix with real diagonal entries and let α be a proper subset of the index set N. Then, A/α and $A(\overline{\alpha})$ have the same number of eigenvalues whose real parts are greater (less) than $w(D^{-1}AD)$ (resp. $-w(D^{-1}AD)$), where $D = diag(d_1, d_2, ..., d_n)$, $d_k = \dfrac{r_k(A)}{|a_{kk}|}$, $k = 1, ..., n$.*

Proof. Notice that $D^{-1}AD$ is an SDD matrix with real diagonal entries. For α being a proper subset of the index set N, it holds that

$$(D^{-1}AD)/\alpha = D^{-1}(\overline{\alpha})(A/\alpha)D(\overline{\alpha}),$$

which is similar to A/α. As $(D^{-1}AD)(\overline{\alpha})$ and $A(\overline{\alpha})$ are similar for any choice of α, their spectra coincide. Applying result from [12] to SDD matrix $D^{-1}AD$ we obtain that A/α and $A(\overline{\alpha})$ have the same number of eigenvalues whose real parts are greater (less) than $w(D^{-1}AD)$ (resp. $-w(D^{-1}AD)$). □

When real parts of all the eigenvalues of the matrix $A(\overline{\alpha})$ lie out of the interval $(-w(D^{-1}AD), w(D^{-1}AD))$, this vertical band represents the exclusion area for the spectrum of A/α.

The next result gives a preliminary localization for the spectrum of the Schur complement matrix through a modification of the Geršgorin set of the original matrix.

Theorem 3.2. *Let $A = [a_{ij}] \in \mathbb{C}^{n,n}$ be an $SDD-$scal matrix, let α be a subset of the index set N, and let $D = diag(d_1, d_2, ..., d_n)$, $d_k = \dfrac{r_k(A)}{|a_{kk}|}$, $k = 1, ..., n$. Then,*

$$\sigma(A/\alpha) = \sigma((D^{-1}AD)/\alpha) \subseteq \bigcup_{j\in\overline{\alpha}} \Gamma_j(D^{-1}AD).$$

Proof. It is easy to see that

$$(D^{-1}AD)/\alpha = D^{-1}(\overline{\alpha})(A/\alpha)D(\overline{\alpha}).$$

Therefore the matrices $(D^{-1}AD)/\alpha$ and A/α are similar. This implies that

$$\sigma(A/\alpha) = \sigma((D^{-1}AD)/\alpha).$$

As $D^{-1}AD$ is an SDD matrix, applying the result from [11] to the matrix $D^{-1}AD$, we prove our statement. □

The benefit of this result is that it allows us to construct a localization area for the spectrum of Schur complement matrix A/α by scaling those Geršgorin disks of the original matrix A that correspond to $\overline{\alpha}$. Notice that results of this type are obtained for $PH-$matrices, see [13], for Partition-Nekrasov matrices, see [17] and for SDD_1 matrices, see [14], only with different constructions of corresponding scaling matrices.

Acknowledgement

The authors acknowledge the financial support of Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama", "Improving the teaching process in the English language in fundamental disciplines".

References

[1] A. Berman and R. J. Plemmons, "Nonnegative Matrices in the Mathematical Sciences", *Philadelphia: Classics in Applied Mathematics* 9, SIAM, 1994.

[2] D. Carlson and T. Markham, "Schur complements of diagonally dominant matrices", *Czech. Math. J.* 29 (104), pp. 246-251, 1979.

[3] Lj. Cvetković and V. Kostić, "New criteria for identifying $H-$matrices", *J. Comput. Math.* 180, pp. 265-278, 2005.

[4] Lj. Cvetković, V. Kostić, M. Kovačević, and T. Szulc, "Further results on $H-$matrices and their Schur complements", *Appl. Math. Comput.* 198, pp. 506-510, 2008.

[5] Lj. Cvetković and M. Nedović, "Special $H-$matrices and their Schur and diagonal-Schur complements", *Appl. Math. Comput.* 208, pp. 225-230, 2009.

[6] Lj. Cvetković and M. Nedović, "Eigenvalue localization refinements for the Schur complement", *Appl. Math. Comput.* 218 (17), pp. 8341-8346, 2012.

[7] T.-B., Gan, T.-Z., Huang, "Simple criteria for nonsingular $H-$matrices", *Linear Algebra Appl.* 374, pp. 317-326, 2003.

[8] M. Fiedler and V. Pták, "On matrices with nonpositive off-diagonal elements and positive principal minors", *Czechoslovak Math. J.* 12, pp. 382-400, 1962.

[9] J. Liu, Y. Huang, and F. Zhang, "The Schur complements of generalized doubly diagonally dominant matrices", *Linear Algebra Appl.* 378, pp. 231-244, 2004.

[10] J. Liu and Y. Huang, "Some properties on Schur complements of $H-$matrices and diagonally dominant matrices", *Linear Algebra Appl.* 389, pp. 365-380, 2004.

[11] J. Liu, Y. Huang, and F. Zhang, "The dominant degree and the disc theorem for the Schur complement of matrix", *Appl. Math. Comput.* 215, pp. 4055-4066, 2010.

[12] J. Z. Liu and F. Z. Zhang, "Disc separation of the Schur complements of diagonally dominant matrices and determinantal bounds", *SIAM J. Matrix Anal. Appl.* 27 (3), pp. 665-674, 2005.

[13] M. Nedović and Lj. Cvetković, "The Schur complement of $PH-$matrices", *Applied Mathematics and Computation*, vol. 362, Article Number 124541, 2019.

[14] M. Nedović and I. Prodanović, "Spectra localization for Schur complements of SDD_1 matrices", *META 2023, The 8th Conference on Mathematics in Engineering: Theory and Applications, Novi Sad, Serbia, May 26-28*, pp. 31-36, 2023.

[15] J. M. Peña, "Eigenvalue localization and pivoting strategies for Gaussian elimination", *Appl. Math. Comput.* 219 (14), pp. 7725-7729, 2013.

[16] J. M. Peña, "Diagonal dominance, Schur complements and some classes of H-matrices and P-matrices", *Adv. Comput. Math* 35, pp. 357-373, 2011.

[17] T. Szulc, Lj. Cvetković and M. Nedović, "Scaling technique for Partition-Nekrasov matrices", *Appl.Math.Comput.* vol. 271C, pp. 201-208, 2015.

[18] R. S. Varga, "Geršgorin and his Circles", *Series in Computational Mathematics*, 36, Springer, 2004.

[19] F. Zhang (Ed.), *The Schur Complement and its Applications*, Springer, New York, 2005.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, June 01-03th 2024

STABILITY ANALYSIS OF A NANO BEAM WITH NONSYMMETRIC BOUNDARY CONDITIONS

Branislava N. Novaković [1] **and Nina Novaković-Marinković** [2]

https://doi.org/10.24867/META.2024.15 Original scientific paper

Abstract. In this paper we investigate stability of an axially loaded nano beam that is clamped at one end and elastically restrained against rotation on the other. We analyze elastically buckling nano beam based on Eringen's nonlocal elasticity theory. The Euler method of adjacent equilibrium configuration is used to derive the nonlinear governing equations. The critical axial force and postbuckling shape are obtained for the beam with the unit cross-sectional area. New numerical results are obtained. The numerical analysis includes the influence of the characteristic parameter of the small scale length on the critical load and the postbuckling shape.

AMS Mathematics Subject Classification (2020): 74G60

Key words and phrases: nano beam, stability, postbuckling shape

1. Introduction

Recently, there has been significant research attention on structures at very small length scales. Nano rods, nano beams, and nano plates have particularly attracted interest due to their technical applications as nanoactuators, nanosensors, and electrochemical sensors. The classical local continuum theory is not entirely applicable to microstructures, especially nanostructures. In particular in cases where the influence of small-scale effects becomes more significant and cannot be ignored. The nonlocal continuum theory was introduced by Eringen [1] to incorporate an internal length scale. According to this theory, the stress tensor at a reference point in an elastic continuum depends on the strain field at all points within the domain. Peddieson et al. [2] developed a nonlocal Euler-Bernoulli beam model in their research. The nonlocal continuum theory has been utilized in numerous studies to model nano beams ([3], [4], [5], [6], [7]).

Researchers have developed various nonlocal beam models to investigate the behavior of nano beams. Wang and Lee [8] formulated the nonlocal theory for both Timoshenko and Euler-Bernoulli beams. Their results provided the

[1]Department of Technical Mechanics, Faculty of Technical Sciences, University of Novi Sad, e-mail: nbrana@uns.ac.rs

[2]Institut für Physik and Astronomie, Universität Potsdam, Germany, Helmholtz-Zentrum Berlin für Materialien und Energie, 12489 Berlin, Germany, e-mail: nina.novakovic@helmholtz-berlin.de

first insights into the influence of nonlocal elasticity on the bending of nano beams, considering concentrated loads as Dirac delta functions. Additionally, Tuna and Kirca [9] derived the exact solution of the integral form of Eringen's nonlocal model for bending analysis of Euler-Bernoulli and Timoshenko beams. In [10], the static deflection and critical buckling load of nano beams were studied for different boundary conditions, taking into account varying nonlocal and material-distribution parameters.

The influence of nonlocal parameters on critical load levels, post-critical shapes, and the effect of various boundary conditions on the stability of nano beams under conservative loading was analyzed in studies conducted by [11].

The aim of this research is to analyze the stability and determine the post-critical shape of a nano beam with nonsymmetrical boundary conditions which have not been used before. The beam is clamped on one end and elastically restrained against rotation on the other end.

2. Mathematical formulation

Consider a straight nano beam of length L loaded by an axial force F with the action line coinciding with the x axis of a rectangular coordinate system $x - B - y$ (see Fig. 1). The beam is clamped at one end and elastically restrained against rotation on the other end (linear rotational spring), with the end C having the possibility of sliding along the x -axis. At the end C the beam is loaded by a compressive force F.

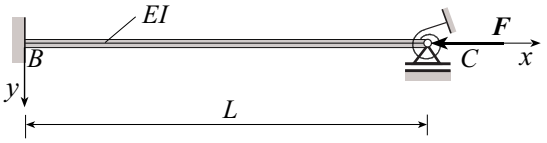


Figure 1: Coordinate system and load configuration

Equilibrium equations for the beam are (see [12])

$$\frac{dH}{dS} = 0, \quad \frac{dV}{dS} = 0, \quad \frac{dM}{dS} = -V\cos\theta + H\sin\theta, \tag{2.1}$$

where H and V are components of the contact force (i.e. the resultant force in an arbitrary cross-section) along x and y axes, respectively, M is the bending moment, θ is the angle between the tangent to the column axis and the x- axis of a rectangular Cartesian coordinate system $x - B - y$, S is the arc-length of the column axis measured from the origin of the coordinate system B. We adjoin to (2.1) the geometrical equations

$$\frac{d\bar{x}}{dS} = \cos\theta, \quad \frac{d\bar{y}}{dS} = \sin\theta, \tag{2.2}$$

and the constitutive equation for nonlocal beam theory

$$M - l^2\frac{d^2M}{dS^2} = EI\frac{d\theta}{dS}. \tag{2.3}$$

where l is an additional length scale specific to nonlocal constitutive law, that is, to a specific material. The value of l can be identified from the Born–Kármán model of lattice dynamics (see [1]). It can be observed from the experimental interpolations that $l \leq 2$ nm (nanometers). In (2.3) E is the modulus of elasticity and I is the moment of inertia of the cross section. In Equations (2.2) we used $\bar{x}$ and $\bar{y}$ to denote coordinates of an arbitrary point on the bear axis in the coordinate system $x - B - y$. If l=0 Equation (2.3) corresponds to the classical Bernoulli-Euler rod theory.

The boundary conditions for the column shown in Fig. 1 are

(2.4)
$$\bar{x}(0) = 0, \quad \bar{y}(0) = \bar{y}(L) = 0, \quad \theta(0) = 0, \quad M(L) = -c\theta(L), \quad H(L) = -F.$$

where c is a spring constant of the support and $c \neq \infty$. Solving $(2.1)_1$ and by using $(2.4)_5$, we obtain $H = -F$. By introducing the dimensionless quantities

(2.5)
$$t = \frac{S}{L},\ \zeta = \frac{\bar{x}}{L},\ \eta = \frac{\bar{y}}{L},\ k = \frac{l}{L},\ \lambda = \frac{FL^2}{EI},\ v = \frac{VL^2}{EI},\ m = \frac{ML}{EI},\ b = \frac{cL}{EI}$$

we obtain from (2.1)-(2.3)

$$(2.6)\quad \dot{v} = 0, \quad \dot{m} = -v\cos\theta - \lambda\sin\theta, \quad \dot{\zeta} = \cos\theta, \quad \dot{\eta} = \sin\theta, \quad \dot{\theta} = m - k^2\ddot{m},$$

subject to

$$(2.7)\qquad \eta(0) = \eta(1) = 0, \quad \theta(0) = 0, \quad m(1) = -b\theta(1),$$

where $\dot{(\cdot)} = \frac{d}{dt}(\cdot)$.

The trivial solution for the systems (2.6), (2.7) in which the axis of the rod remains straight for any value of and the dimensionless load parameter is

$$(2.8)\qquad \theta_0 = v_0 = \eta_0 = 0, \quad \zeta_0 = t.$$

Euler method is used to examine stability of the trivial configuration defined by Equations (2.8) (see [12]). In order to obtain nontrivial solution to (2.6), (2.7) and determine $\lambda \in \mathbb{R}$ for it, it is assumed that

$$(2.9)\qquad \theta = \theta_0 + \Delta\theta, \quad v = v_0 + \Delta v, \quad \eta = \eta_0 + \Delta\eta, \quad \zeta = \zeta_0 + \Delta\zeta,$$

where $\Delta\theta, ..., \Delta\zeta$ are perturbations. After substituting this in Equations (2.6) and (2.7) and by neglecting the higher order terms in perturbations, Equations (2.6) become (omitting Δ in front of variables) linearized equations describing relative equilibrium of the beam

$$(2.10)\qquad \dot{v} = 0, \quad \dot{m} = -v - \lambda\theta, \quad \dot{\zeta} = 1, \quad \dot{\eta} = \theta, \quad \dot{\theta} = m - k^2\ddot{m},$$

subject to (2.7).

3. Critical values of load parameter

In this section we will determine the critical loads (λ) of the beam for which the beam loses its stability, e.i. when system (2.6), (2.7) has nontrivial solution. A necessary condition for this to occur is that the linearized systems (2.10), (2.7) have a nontrivial solution.

The system (2.10) can be reduced, so it can be written as

$$\dddot{\eta} + \frac{\lambda}{1-k^2\lambda}\ddot{\eta} = 0 \tag{3.1}$$

Then the solution of Equations (3.1) is

$$\eta = C_1 \cos(\beta\zeta) + C_2 \sin(\beta\zeta) + C_3\zeta + C_4, \tag{3.2}$$

where C_j, $j = 1, 2, 3, 4$are arbitrary constants and

$$\beta = \sqrt{\frac{\lambda}{1-k^2\lambda}}. \tag{3.3}$$

The boundary conditions subjected to Equations (3.2) are

$$\eta(0) = \eta(1) = 0, \quad \dot{\eta}(0) = 0, \quad \ddot{\eta}(1) = -b\dot{\eta}(1). \tag{3.4}$$

By using boundary conditions (3.4) the following condition for the existence of non-trivial solutions is derived

$$2b - \left(\beta^2 + 2b\right)\cos\beta + (\beta - b\beta)\sin\beta = 0. \tag{3.5}$$

The necessary condition for the existence of solutions of the Equations (3.5) is $k^2\lambda \leq 1$.

4. Numerical results

The critical (smallest positive root of (3.5)) value of the axial force for several values of k and b is determined from Equations (3.5) and shown in Table 1. Case when parameter $b{=}\infty$ represents the nano beam that is clamped on the both ends.

Table 1. Critical values of λ for different values of k and parameter b

b	0	0.5	50	500
$k = 0$	20.19072856	21.65942569	29.57476388	39.32098463
$k = 0.05$	19.22053695	20.54684361	27.5386418	35.80159929
$k = 0.1$	16.79890687	17.80332726	22.82447831	28.22330372
$k = 0.2$	11.1697295	11.60506445	13.54782035	19.07257867

The value of critical axial force increases for the increasing value of spring constant b and constant value of k. When the nonlocal parameter (k) increases

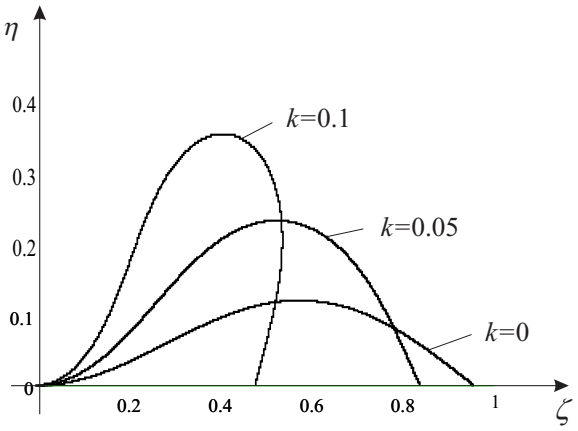


Figure 2: Postbuckling modes $\lambda = 22$, $b = 0.5$

than value of critical axial force also increases for constant value of b. For the case when $b = \infty$ and $k = 0$ (Bernoulli–Euler clamped beam) we have $\lambda = 4\pi^2 = 39.4784176$ (see [12]).

The postbuckling shapes of first modes for chosen $\lambda = 22$, $b = 0.5$ and several values of k are shown in Fig. 2. Equations (2.6) and (2.7) are used for determining postbuckling shapes.

5. Conclusions

In this paper we analyzed the stability for an elastic nano beam. The beam is clamped on one end and elastically restrained against rotation on the other end. The characteristic equation (3.5) that determins critical loads for the nano beam with constant cross-section is derived. By using the characteristic equation we determined the lowest value of λ for several values of parameters b and k (nonlocal parameter). Numerical analysis demonstrates that the value of spring constant and nonlocal parameters have an impact on critical value of axial force. The postbuckling shapes of first modes for chosen values of λ and b and several values of parametar k is determined.

Acknowledgement

This research is supported by Project No. 054-2024 of Department of Mechanics, Faculty of Technical Sciences, University of Novi Sad.

References

[1] A. C. Eringen, *Nonlocal Continuum Field Theories*. New York: Springer-Verlag, 2002

[2] J. Peddieson, R. Buchanan, and R. P. Miles, "Application of nonlocal continuum models to nanotechnology", *International Journal of Engineering Science*, vol. 41, pp. 305-312, March 2003.

[3] L. J. Sudak, "Column buckling of multiwalled carbon nanotubes using nonlocal continuum mechanics", *Journal of Applied Physics*, vol. 94, pp. 7281-7287, December 2003.

[4] J. N. Reddy, "Nonlocal theories for bending, buckling and vibration of beams" *International Journal of Engineering Science*, vol. 45, pp. 288-307, February-August 2007.

[5] J. N. Reddy, "Nonlocal nonlinear formulations for bending of classical and shear deformation theories of beams and plates" *International Journal of Engineering Science*, vol. 48, pp. 1507-1518, November 2010.

[6] A. C. Eringen, "On differential equations of nonlocal elasticity and solutions of screw dislocation and surface waves" *Journal of Applied Physics*, vol. 54, pp. 4703-4710, September 1983.

[7] N. Challamel, Z. Zhang, C. M. Wang, "Nonlocal equivalent continuum for the buckling and the vibrations of microstructured beams" *ASCE Journal of Nanomechanics and Micromechanics*, vol. 5, pp. A401 4004-1 (13), March 2015.

[8] Q. Wang, K. M. Liew, "Application of nonlocal continuum mechanics to static analysis of micro-and nano-structure" *Physics Letters A*, vol. 363, pp. 2364-242, March 2007.

[9] M. Tuna, M. Kirca, "Exact solution of Eringen's nonlocal integral model for bending of Euler-Bernoulli and Timoshenko beams" *International Journal of Engineering Science*, vol. 105, pp. 80-92, August 2016.

[10] M. A. Eltaher, S. A. Emam, F. F. Mahmoud, "Static and stability analysis of nonlocal functionally graded nanobeams" *Conposite Structures*, vol. 96, pp. 82-88, February 2013.

[11] C. M. Wang, N. Challamel, W. H. Duan "On boundary conditions for buckling and vibration of nonlocal beams" *Europian Journal of Mechanics A/Solid*, vol. 61, pp. 73-81, January-February 2017.

[12] T. M. Atanackovic, *Stability Theory of Elastic Rods*. World Scientific. River Edge. New Jersey. 1997

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

METODI NJUTNOVOG TIPA [1]

Tijana Ostojić [2] i Manojlo Vuković [3]

https://doi.org/10.24867/META.2024.16 Review article

Sažetak. Njutnov metod je jedan od osnovnih i najpoznatijih alata u numeričkoj analizi, operacionim istraživanjima i optimizaciji. Pored dobrih teoretskih osobina, kao što je lokalna kvadratna konvergencija, ovaj iterativni metod ima brojne primene u različitim naukama, gde se pokazao kao efikasan za rešavanje problema numeričke optimizacije. Međutim, postoji značajan skup praktičnih problema u kojima implementacija Njutnovog metoda zahteva visoke računske troškove, što nas dovodi do metoda Njutnovog tipa, koji čine temu ovog preglednog rada. Cilj rada je da se da prikaz ključnh ideja metoda Njutnovog tipa u istorijskoj perspektivi, kao i savremena istraživanja uz odgovarajuće reference.

AMS klasifikacija (2020): 90C30, 90C53

Ključne reči: Njutnov metod, metodi Njutnovog tipa, kvazi-Njutnovi metodi, spektralni gradijentni metod, pravac pretrage

1. Uvod

Posmatramo problem optimizacije bez ograničenja gde je potrebno pronaći optimalno rešenje x^* tako da važi

$$f(x^*) = \min_{x \in \mathbb{R}^n} f(x). \tag{1.1}$$

Problem (1.1) se može rešiti iterativno tako što se odabere početna tačka x_0, a potom se generiše niz iterativnih tačaka $\{x_k\}_{k\in\mathbb{N}}$, pri čemu se naredna tačka određuje na osnovu prethodne. Cilj je da niz $\{x_k\}_{k\in\mathbb{N}}$ bude generisan tako da teži ka rešenju x^* posmatranog problema. U okviru ovog rada posmatraćemo metode linijskog pretraživanja [7], pa je u nastavku formalno navedeno pravilo po kom se ažurira iteracija u ovom metodu

$$x_{k+1} = x_k + \alpha_k p_k, \tag{1.2}$$

gde $\alpha_k > 0$ predstavlja dužinu koraka i p_k pravac pretraživanja. Osnovna ideja ovog metoda leži u tome da se pronađe nova iteracija x_{k+1} sa manjom vrednosti

[1]Ovo istraživanje je podržano od strane Fakulteta tehničkih nauka Univerziteta u Novom Sadu kroz projekat "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama".

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: tijana.ostojic@uns.ac.rs

[3]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: manojlo.vukovic@uns.ac.rs

funkcije f. Većina metoda linijskog pretraživanja zahteva da p_k bude opadajući pravac kako bi se ostvarilo poboljšanje.

Definicija 1.1. Za datu tačku $x_k \in \mathbb{R}^n$ pravac $p_k \in \mathbb{R}^n$ je opadajući, ako postoji $\overline{\alpha}$ tako da važi

$$f(x_k + \alpha p_k) < f(x_k),\ \forall \alpha \in (0, \overline{\alpha}).$$

Ako je funkcija f neprekidno diferencijabilna, može se pokazati da je pravac p_k opadajući u x_k ako je sledeći uslov zadovoljen

$$\nabla f(x_k)^T p_k < 0. \tag{1.3}$$

Ova osobina, koja garantuje da se vrednost funkcije f može smanjiti duž pravca p_k, je direktna posledica razvoja prvog reda funkcije f Tejlorovom formulom u okolini tačke x_k

$$f(x_k + \alpha p_k) = f(x_k) + \alpha p_k^T \nabla f(x_k) + O(\alpha^2).$$

Primetimo, za dovoljno mali korak α sledeća nejednakost će uvek biti zadovoljena

$$f(x_k + \alpha p_k) < f(x_k),$$

što je zapravo i glavni cilj u metodama optimizacije-da se u svakoj iteraciji dobije tačka koja je bolja od trenutne, tj. tačka sa manjom vrednosti funkcije cilja.

2. Metodi Njutnovog tipa

Optimizacioni metod koji za pravac pretraživanja koristi

$$p_k = -B_k^{-1} \nabla f(x_k),$$

gde je B_k simetrična nesingularna matrica, zove se *metod Njutnovog tipa.*

Lema 2.1. *(Opadajući pravac): Ako je matrica B_k pozitivno definitna, tj. ako važi $B_k \succ 0$, tada je $p_k = -B_k^{-1} \nabla f(x_k)$ opadajući pravac.*

Primetimo da prethodna lema važi jer za pravac $p_k = -B_k^{-1} \nabla f(x_k)$ važi da je

$$\nabla f(x_k)^T p_k = -\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) < 0,$$

odnosno uslov (1.3) je zadovoljen.

Najjednostavniji metod, *gradijentni metod*, dobija se za izbor $B_k = I$, tj. opadajući pravac definisan je sa $p_k = -\nabla f(x_k)$. Ovaj metod se još naziva i *metod najbržeg pada* zato što se duž tog pravca vrednost funkcije najbrže smanjuje. Ukoliko se pretpostavi da je dužina koraka fiksna za sve iteracije, važi sledeći rezultat.

Teorema 2.2. *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ neprekidno diferencijabilna konveksna funkcija i neka je ∇f $L-$Lipšic neprekidno. Neka je $\{x_k\}$ niz generisan sa* (1.2) *za pravac pretrage $p_k = -\nabla f(x_k)$ i dužinu koraka $\alpha_k = \alpha$ za svako k. Ako je $\alpha \leq 1/L$, tada x_k konvergira linearno ka rešenju problema* (1.1).

Gradijentni metod sa fiksnom dužinom koraka je veoma primenjiv i široko korišćen zbog svoje jednostavnosti i niskih računskih troškova, pošto zahteva samo izvode prvog reda i nema dodatnih izračunavanja za izbor dužine koraka. Međutim, glavna mana ovog metoda je stopa konvergencije koja je najviše linearna. Prema tome, gradijentni metod može biti veoma spor i može zahtevati mnogo iteracija kako bi se pronašlo rešenje sa dobrom tačnošću.

Ukoliko pretpostavimo da je funkcija dva puta neprekidno diferencijabilna možemo dobiti sofisticiranije izbore za pravac pretrage p_k. Funkcija cilja $f \in C^2(\mathbb{R}^n)$ može biti aproksimirana u okolini trenutne iteracije x_k korišćenjem drugog reda Tejlorovog razvoja:

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T \nabla^2 f(x_k) p. \tag{2.1}$$

Ako označimo desnu stranu (2.1) sa $m_k(p)$, cilj je izračunati pravac p u iteraciji k minimizirajući kvadratnu funkciju m_k. Ako pretpostavimo da je $\nabla^2 f(x_k) \succ 0$, tada funkcija m_k ima jedinstveni minimizator oblika

$$p_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k).$$

Drugim rečima, za $B_k = \nabla^2 f(x_k)$ dobija se *Njutnov metod*. Ovaj metod postiže lokalnu kvadratnu konvergenciju pod odgovarajućim pretpostavkama o funkciji.

Teorema 2.3. *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ dva puta neprekidno diferencijabilna funkcija i neka je Hesijan $\nabla^2 f(x)$ Lipšic neprekidan u okolini rešenja x^* gde su zadovoljeni dovoljni uslovi optimalnosti. Neka je $\{x_k\}$ niz generisan Njutnovim metodom pri čemu je u svakoj iteraciji linijskog pretraživanja* (1.2) *$\alpha = 1$. Tada važi:*

i) Ako je početna tačka x_0 dovoljno blizu x^*, niz iteracija $\{x_k\}$ konvergira ka x^*;

ii) Ako metod konvergira, stopa konvergencije niza $\{x_k\}$ je kvadratna;

iii) Ako metod konvergira, niz normi gradijenata $||\nabla f(x_k)||$ kvadratno konvergira ka nuli.

Kao što je već pomenuto, da bi se obezbedilo da je p_k opadajući pravac, uslov iz Leme 2.1 treba da bude zadovoljen. Tačnije, matrica $\nabla^2 f(x_k)$ mora biti pozitivno definitna u svakoj iteraciji.

Važno je istaći da uprkos dobrim teoretskim osobinama, u praksi se pojavljuju i izvesni nedostaci. Svaka iteracija Njutnovog metoda zahteva određivanje matrice Hesijana, što podrazumeva izračunavanje izvoda drugog reda u svakoj iteraciji, što samo po sebi može biti izuzetno skupo. Metod, takođe, čak i ako je $\nabla^2 f(x_k) \succ 0$, može postati nestabilan u slučaju loše uslovljenog $\nabla^2 f(x_k)$ u nekoj iteraciji.

Zbog ovih nedostataka, pojavile su se brojne varijante i modifikacije Njutnovog metoda koji imitiraju Njutnovu ideju i u čijoj osnovi je aproksimacija

matrice Hesijana. U ovim metodama, nazvanim *metode Njutnovog tipa* ili *kvazi-Njutnove metode (QN)*, matrica Hesijana je zamenjena matricom $B_k \in \mathbb{R}^{n\times n}$, takvom da je

$$B_k \approx \nabla^2 f(x_k)$$

dobra aproksimacija prave matrice Hesijana sa nižim troškovima izračunavanja jer je B_k bazirana samo na informacijama prvog reda. Red konvergencije u ovim metodama je najviše superlinearan, pa je konvergencija sporija u poređenju sa Njutnovim metodom, ali s druge strane, trošak je značajno manji. Ideja na kojoj se zasnivaju QN metode je da dve uzastopne iteracije x_k i x_{k+1} zajedno sa gradijentima $\nabla f_k := \nabla f(x_k)$ i $\nabla f_{k+1} := \nabla f(x_{k+1})$ sadrže informacije o krivini (tj. Hesijanu).

U literaturi su predložene različite strategije za izračunavanje matrice B_k zasnovane na uslovima koje bi B_{k+1} trebalo da zadovolji. Glavni uslov poznat je kao jednačina sečice

$$B_{k+1}s_k = y_k, \tag{2.2}$$

gde je razlika između dve iteracije i između gradijenata u dve susedne iteracije data sa

$$s_k = x_{k+1} - x_k \text{ and } y_k = \nabla f_{k+1} - \nabla f_k.$$

Međutim, jedinstveno rešenje za B_{k+1} nije obezbeđeno iz uslova (2.2). Zato se nameću dodatni zahtevi na B_{k+1}, kao što su simetrčnost i ograničenje da razlika između uzastopnih aproksimacija B_k i B_{k+1} ima niski rang. Drugim rečima, B_{k+1} je rešenje sledećeg problema

$$\min ||B - B_k||_* \tag{2.3}$$

$$\text{tako da } B^T = B, Bs_k = y_k.$$

Različite formule za ažuriranje matrice B_{k+1} dobijaju se rešavanjem problema (2.3) u zavisnosti od matrične norme $||\cdot||_*$, [7].

Jedna od najčešće korišćenih formula za ažuriranje ovog tipa je DFP formula predložena od strane Dejvidsona, Flečera i Pauela. Ona se dobija korišćenjem ponderisane Frobeniusove norme i definisana je sa

$$B_{k+1} = (I - \frac{y_k s_k^T}{y_k^T s_k})B_k(I - \frac{s_k y_k^T}{y_k^T s_k}) + \frac{y_k y_k^T}{y_k^T s_k}.$$

Zatim, Brojden, Flečer, Goldfarb i Šenoa predložili su BFGS formulu oblika

$$H_{k+1} = (I - \frac{s_k y_k^T}{y_k^T s_k})H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k},$$

gde H_{k+1} predstavlja aproksimaciju inverzne matrice Hesijana, tj. $H_k = B_k^{-1}$ ($H_{k+1}y_k = s_k$). Početna aproksimacija H_0 se bira od strane korisnika i često je definisana kao $H_0 = \gamma I, \gamma > 0$. Bitno je istaći da BFGS formula očuvava pozitivnu definitnost, tačnije, ako je H_k pozitivno definitna i $y_k^T s_k > 0$, tada je i H_{k+1} takođe pozitivno definitna.

2.1. Spektralni gradijentni metod

Sada ćemo predstaviti modifikovanu verziju QN metoda - *Spektralni gradijentni (SG) metod*. Ovaj metod je poznat po svojoj efikasnosti i jednostavnosti u rešavanju optimizacionih problema [1, 4, 5, 6, 8]. Originalno je predložen od strane Barzilaija i Borweina [2], pa se često naziva BB metodom. Strategija odabira dužine koraka u SG metodu je ključna za bržu konvergenciju u poređenju sa klasičnim gradijentnim metodama, jer inkorporira informacije drugog reda vezane za spektar matrice Hesijana. Tačnije, oslanja se na jednostavnu aproksimaciju drugih izvoda, koja ima oblik dijagonalne matrice pomnožene sa takozvanim spektralnim koeficijentom. Grubo govoreći, ovaj koeficijent aproksimira karakteristični koren matrice Hesijana i pruža barem neku vrstu informacije drugog reda, što je ključno za brzu konvergenciju. Iako teorijski rezultati nisu tako snažni kao kod Njutnovih metoda koji se oslanjaju na prave druge izvode ili bolje aproksimacije matrice Hesijana, spektralni gradijentni metodi su jeftini, laki za implementaciju i pružaju vrlo dobre numeričke rezultate, što ih čini popularnim u praksi.

Spektralni koeficijent je konstruisan tako da najbolje odgovara jednačini sečice, gde je jedan od ključnih delova razlika između dva uzastopna gradijenta označena sa y_k. Stoga, cilj je pronaći dijagonalnu matricu D_k posebnog oblika

$$D_k = \gamma_k I, \quad \gamma_k \in \mathbb{R},$$

koja najbolje odgovara jednačini $H_{k+1}y_k = s_k$, pri čemu matrica $D_k \approx H_{k+1}$ predstavlja aproksimaciju inverzne matrice Hesijana $(\nabla^2 f(x_k))^{-1}$. Dakle, pravac pretrage je paralelan pravcu negativnog gradijenta, tj. važi

$$p_k = -\gamma_k I \nabla f(x_k) = -\gamma_k \nabla f(x_k).$$

Spektralni koeficijent γ_k je definisan uslovom iz jednačine sečice na sledeći način

$$\tilde{\gamma}_k = \arg\min_{\gamma \in \mathbb{R}} ||y_{k-1} - \gamma s_{k-1}||^2 \tag{2.4}$$

ili

$$\gamma_k = \arg\min_{\gamma \in \mathbb{R}} ||\gamma y_{k-1} - s_{k-1}||^2, \tag{2.5}$$

gde je $\gamma_k = \tilde{\gamma}_k^{-1}$. Iz (2.4) i (2.5) dobijamo spektralne koeficijente oblika

$$\gamma_k^{BB1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \quad \text{i} \quad \gamma_k^{BB2} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \tag{2.6}$$

Pored ova dva pravila za izračunavanje γ_k, u literaturi su predloženi brojni spektralni gradijentni metodi koji uopštavaju BB metode, kao na primer Adaptivni Barzilai–Borwein (ABB) [9] i njegova modifikacija ABBmin [3], koji se oslanjaju na adaptivne kriterijume korišćene za prelazak između γ_k^{BB1} i γ_k^{BB2}. Dužine koraka se definišu sledećim pravilima

$$\gamma_k^{ABB} := \begin{cases} \gamma_k^{BB2}, & \frac{\gamma_k^{BB2}}{\gamma_k^{BB1}} < \tau, \\ \gamma_k^{BB1}, & \text{inače}, \end{cases}$$

i

$$\gamma_k^{ABBmin} := \begin{cases} \min\{\gamma_j^{BB2} : j = \max\{1, k-m_a\}, ..., k\}, & \frac{\gamma_k^{BB2}}{\gamma_k^{BB1}} < \tau, \\ \gamma_k^{BB1}, & \text{inače}, \end{cases}$$

gde je $m_a \geq 0$ i $\tau \in (0, 1)$.

U slučaju kada uslov krivine $s_{k-1}^T y_{k-1} > 0$ nije zadovoljen, γ_k može biti negativno, tako da pravac pretraživanja nije opadajući pravac. Ova mana može biti prevaziđena korišćenjem safeguard-a [10] na sledeći način

$$\bar{\gamma}_k = \min\{\gamma_{max}, \max\{\gamma_k, \gamma_{min}\}\},$$

gde $0 < \gamma_{min} << 1 << \gamma_{max} < \infty$. Stavljajući da je $p_k = -\bar{\gamma}_k \nabla f(x_k)$ obezbeđeno je da je pravac opadajući i numerička stabilnost se može kontrolisati.

Literatura

[1] S. Bellavia, N. Krklec Jerinkić and G. Malaspina, "Subsampled nonmonotone spectral gradient methods", *Communications in Applied and Industrial Mathematics*, 11(1), pp. 19-34, 2020.

[2] J. Barzilai and J. M. Borwein, "Two-point step size gradient method", *IMA J. Numerical Analysis*, 8(1), pp. 141–148, 1988.

[3] G. Frassoldati , L. Zanni , G. Zanghirati, "New adaptive stepsize selections in gradient methods", *J. Ind. Manag. Optim.* 4 (2), pp. 299–312, 2008.

[4] N. Krejić and N. Krklec Jerinkić, "Spectral projected gradient method for stochastic optimization", *Journal of Global Optimization*, 73, pp. 59–81, 2018.

[5] N. Krejić and N. Krklec Jerinkić, T. Ostojić, "Spectral projected subgradient method for nonsmooth convex optimization problems", *Numerical Algorithms*, pp. 1-19, 2022.

[6] N. Krklec Jerinkić and T. Ostojić, "AN-SPS: Adaptive Sample Size Nonmonotone Line Search Spectral Projected Subgradient Method for Convex Constrained Optimization Problems", *Optimization Methods and Software*, 2024, DOI - 10.1080/10556788.2024.2324920.

[7] J. Nocedal and S. J. Wright, *Numerical Optimization, 2nd ed.*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.

[8] C. Tan, S. Ma, Y. H. Dai and Y. Qian, "Barzilai-borwein step size for stochastic gradient descent", *Advances in neural information processing systems*, 29, 2016.

[9] B. Zhou, L. Gao and Y.H. Dai, "Gradient methods with adaptive stepsizes", *Comput. Optim. Appl.* 35 (1), pp.69–86, 2006.

[10] R. Tavakoli and H. Zhang, "A nonmonotone spectral projected gradient method for large-scale topology optimization problems", *Numerical Algebra, Control and Optimization* vol. 2, No. 2, pp. 395-412, 2012.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

DEKOMPOZICIONE METODE U DISTRIBUIRANOJ OPTIMIZACIJI BEZ OGRANIČENJA[1]

Tijana Ostojić [2] **i Manojlo Vuković** [3]

https://doi.org/10.24867/META.2024.17 Review article

Sažetak. Zbog sve većeg obima podataka u mašinskom učenju, kao i dubokom učenju, potreba za rešavanjem optimizacionih problema u distribuiranom okruženju kontinuirano raste, jer ovi podaci rezultiraju problemima sa mnogo promenljivih. Jedna od metoda koja se može koristiti i u distribuiranoj optimizaciji je dekompozicija promenljivih, koja je inicijalno korišćena u ranim 1960-im godinama. Ključna ideja za rešavanje navedenih velikih optimizacionih problema leži u razdvajanju skupa promenljivih na manje podskupove. Ovi podskupovi se zatim obrađuju nezavisno ili u koordinaciji, što olakšava rešavanje složenih problema. Stoga, metoda je posebno korisna u distribuiranim sistemima, gde se problem može podeliti na različite čvorove ili procesore. U okviru ovog preglednog rada biće dat prikaz osnovnih ideja primene dekompozicije u optimizaciji.

AMS klasifikacija (2020): 49M27, 49M29

Ključne reči: optimizacija, distribuirana optimizacije, dekompozicija

1. Uvod

Usled rastućih dimenzija i složenosti modernih skupova podataka, potreba za proučavanjem metoda za rešavanje optimizacionih problema sa velikim brojem varijabli postaje sve veća. Jedan od načina da se poboljša efikasnost rešavanja takvih problema je da se oni podele na više manjih problema, koji se mogu rešavati paralelno i nezavisno jedan od drugog smanjujući ukupno vreme potrebno za optimizaciju. Dakle, ovi problemi se mogu predstaviti kao optimizacija funkcija koje su predstavljene kao suma velikog broja pojedinačnih funkcija, gde svaka od njih može biti povezana sa određenim delom podataka ili zadatkom. Tako, rešavanje originalnog problema, koji bi se rešavao na jednom računaru, može da se distribuira na više računara koji zajedno dolaze do rešenja posmatranog optimizacionog problema. Mnogi aktuelni problemi u mašinskom učenju (kao npr. problemi konačnih suma $f(x) = \sum_{i=1}^{n} f_i(x)$), statistici, nauci

[1]Ovo istraživanje je podržano od strane Fakulteta tehničkih nauka Univerziteta u Novom Sadu kroz projekat "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama".

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: tijana.ostojic@uns.ac.rs

[3]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: manojlo.vukovic@uns.ac.rs

o podacima se mogu efikasnije rešiti korišćenjem algoritmima distribuirane optimizacije [1]. Dakle, potreba za distribuiranim okruženjem se ogleda u tome što su radni podaci isuviše veliki da bi se obrađivali i skladištili na jednom računaru, kao i činjenica da lokalna optimizacija često zahteva manje resursa i vremena nego globalna optimizacija. [11].

U okviru ovog preglednog rada posmatra se problem optimizacije bez ograničenja sledećeg oblika

$$\min_x f(x), \tag{1.1}$$

pri čemu će akcenat biti stavljen na dekompoziciju promenljivih kao jednog od pogodnih načina za rešavanje tog problema [3]. Problem (1.1) ima jedinstveno rešenje ako važi sledeća pretpostavka.

Pretpostavka 1.1. Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ je konveksna i dva puta neprekidno diferencijabilna funkcija za koju postoje konstante $0 < \mu \leq L < \infty$ takve da za svako $x \in \mathbb{R}^n$ važi $\mu \mathbf{I} \preceq \nabla^2 f(x) \preceq L\mathbf{I}$.

U zavisnosti od osobina koje poseduje funkcija $f(x)$, u literaturi su predloženi razni metodi za rešavanje problema (1.1) [2, 4, 8, 9]. U ovom radu ćemo se koncentrisati na iterativni postupak linijskog pretraživanja koji je našao široku primenu u praksi [6, 8, 10]. Suština ovog postupka je da se dobije manja vrednost funkcije cilja u narednoj tački u odnosu na trenutnu. Iteracija se definiše na sledeći način

$$x_{k+1} = x_k + \alpha_k g_k. \tag{1.2}$$

Preciznije, x_{k+1} se računa tako da se prvo odredi opadajući pravac pretraživanja g_k[4], a potom dužina koraka α_k tako da se umanji vrednost funkcije cilja.

Dakle, aproksimacija rešenja originalnog problema se može dobiti korišćenjem sledećeg algoritma.

ALGORITAM 1.

za početnu tačku $x \in \mathbf{dom} f$

ponovi

S1 Nađi pravac pretraživanja g.

S2 Nađi dužinu koraka α,

S3 Ažuriraj $x = x + \alpha g$.

sve dok nije zadovoljen izlazni kriterijum.

Sledeća teorema daje uslove za konvergenciju niza $\{x_k\}$ koji je generisan sa (1.2).

[4]U literaturi su prisutni i razne metode koje ne pretpostavljaju da je pravac pretraživanja opadajući. U tom slučaju predložene su različite varijante nemonotonog linijskog pretraživanja (pogledati npr. [7].

Teorema 1.2. *[4] Neka je funkcija f zadovoljava pretpostavku 1.1. Neka je x_0 početna tačka i x_k dato sa* (1.2)*, gde je g_k vektor dimenzije n i $\alpha_k \geq 0$ skalar. Neka je:*

1. *skup $S = \{x : f(x) \leq f(x_0)\}$ je ograničen,*
2. *vektori g_k zadovoljavaju dovoljan uslov pada*

$$-\frac{g_k^\top \nabla f(x_k)}{\|g_k\|\|\nabla f(x_k\|} \geq \epsilon > 0,$$

3. *za vektore g_k važi*

$$\|g_k\| \geq m\|\nabla f(x_k)\| \text{ za sve } k \ (m > 0),$$

i

$$\|g_k\| \leq M \text{ za sve } k,$$

4. *skalar α_k je izabran kao prvi element niza $1, \frac{1}{2}, \frac{1}{4}, \ldots$ koji zadovoljava dovoljan uslov pada*

$$f(x_k + \alpha g_k) \leq f(x_k) + \mu\alpha_k g_k^\top \nabla f(x_k),$$

gde je $0 < \mu < 1$.

Tada je

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

S obzirom na to da funkcija f zadovoljava pretpostavku 1.1, Teorema 1.2 garantuje konvergenciju niza $\{x_n\}$ ka rešenju problema (1.1). U ovom radu će se razmatrati primena Algoritma 1 u dekompoziciji.

Dekompozicija u optimizaciji predstavlja relativno staru ideju koja datira još u 1960-im godinama [5]. Motivacija za uvođenje ovih metoda leži u rešavanju problema veoma velikih dimenzija koji se u to vreme nisu mogli rešiti standarnim poznatim alatima. Danas se ona može koristiti u distribuiranoj optimizaciji. Pod uslovom da se promenljive u funkciji cilja $f(x)$ mogu podeliti, tj. $x = (x_1, x_2)$, problem (1.1) se može rešiti distribuirano. U sledećoj definiciji se uvodi pojam separabilnog problema optimizacije.

Definicija 1.3. Problem optimizacije je separabilan ako se može napisati na sledeći način

$$\min_{x=(x_1,x_2)} \ f_1(x_1) + f_2(x_2). \tag{1.3}$$

Primetimo da funkcija f_1 zavisi samo od x_1, a funkcija f_2 zavisi samo od x_2. Dakle, kako x_1 nije ni u kakvoj vezi sa x_2, problem (1.3) se može rešiti tako što se podeli na dva odvojena potproblema optimizacije:

$$\underbrace{\min f_1(x_1)}_{\text{Potproblem 1}} \quad \text{i} \quad \underbrace{\min f_2(x_2).}_{\text{Potroblem 2}}$$

Neka funkcije f_1 i f_2 zadovoljavaju Pretpostavku 1.1. Potproblem 1 i Potroblem 2 se mogu rešiti korišćenjem Algoritma 1 na dva različita računara (procesora) i samim tim skratiti vreme potrebno za rešavanje originalnog problema (1.3). Međutim, većina realnih problema u praksi nisu separabilni, ali se u nekim slučajevima mogu, korišćenjem dekompozicije, napraviti da to budu. U nastavku sledi definicija skoro separabilnog problema optimizacije, koji se, iako to nije, ipak može rešiti distribuirano korišćenjem dekompozicije.

Definicija 1.4. Problem optimizacije je skoro separabilan ako se može napisati na sledeći način

$$\begin{aligned} \min\ f(x) &= f_1(x_1, y) + f_2(x_2, y), \\ x &= (x_1, x_2, y). \end{aligned} \tag{1.4}$$

Primetimo da za fiksirano y, problem (1.4) postaje separabilan po x_1 i x_2, pa se pomenuta dva potproblema mogu rešiti totalno nezavisno. Kako y komplikuje problem, jer se javlja kao promenljiva u obe funkcije, ona se naziva *komplikovana promenljiva*. S druge strane, x_1 i x_2 se nazivaju *privatnim* ili *lokalnim promenljivima* koje se odnose na funkcije f_1 i f_2, respektivno, dok je y *javna promenljiva* koja vezuje ta dva potproblema.

2. Primarna dekompozicija

U nastavku će biti razmotreno rešavanje skoro separabilnog problema (1.4) korišćenjem primarne dekompozicije. Ideja u primarnoj dekompoziciji je da se za fiksirano y posmatraju sledeća dva potproblema

$$\underbrace{\min_{x_1} f_1(x_1, y)}_{\text{Potproblem 1'}} \quad \text{i} \quad \underbrace{\min_{x_2} f_2(x_2)}_{\text{Potroblem 2'}}.$$

Označimo sa $\Phi_1(y)$ i $\Phi_2(y)$ optimalne vrednosti Potproblema 1' i 2', respektivno[5]. Ono što se sada može primetiti jeste da se Potproblemi 1' i 2' mogu rešiti nezavisno (paralelno), korišćenjem Algoritma 1, a čak mogu biti i fizički odvojeni, tj. da se rešavaju na različitim računarima.

Lako je primetiti da je problem (1.4) ekvivalentan problemu

$$\min_{y} \Phi_1(y) + \Phi_2(y). \tag{2.1}$$

Odnosno, promenljive u (2.1) su komplikovane promenljive y originalnog problema, dok je funkcija cilja master problema zapravo suma optimalnih vrednosti potproblema. Ovaj deo rada zaključujemo sa algoritmom za rešavnje problema (1.4) pomoću primarne dekompozicije:

ALGORITAM 2.

ponovi

[5]Pretpostavljamo da je za funkcije f_1 i f_2 Pretpostavka 1.1 zadovoljena.

S1 Reši potprobleme (nezavisno).
Nađi x_1 koji minimizira $f_1(x_1, y)$ i gradijent $g_1 = \nabla\Phi_1(y)$.
Nađi x_2 koji minimizira $f_2(x_2, y)$ i gradijent $g_2 = \nabla\Phi_2(y)$.

S2 Ažuriraj komplikovanu promenljivu.
$y := y - \alpha_k(g_1 + g_2)$.

Primetimo da se u Algoritmu 2 koristi Algoritam 1 i za rešavanje potproblema i za rešavanje problema (2.1) jer se u koraku S1 računaju: 1) x_1 i x_2 koji minimiziraju potprobleme, 2) gradijenti $\nabla\Phi_1(y)$ i $\nabla\Phi_2(y)$ čiji se zbir dalje koristi kao pravac u koraku S2 kada se radi linijsko pretraživanje problema (2.1).

3. Dualna dekompozicija

Originalni skoro separabilni problem (1.4) se može zapisati i kao

$$\begin{aligned} &\min f(x) = f_1(x_1, y_1) + f_2(x_2, y_2) \\ &\text{pod uslovom } y_1 = y_2, \end{aligned} \tag{3.1}$$

gde su y_1, y_2 lokalne verzije komplikovane promenljive y. Primetimo da se Lagranžova funkcija dualnog problema (3.1)

$$L(x_1, y_1, x_2, y_2) = f_1(x_1, y_1) + f_2(x_2, y_2) + \nu^T(y_1 - y_2)$$

može zapisati na sledeći način

$$L(x_1, y_1, x_2, y_2) = \underbrace{f_1(x_1, y_1) + \nu^T y_1}_{\text{zavisi samo od } (x_1, y_1)} + \underbrace{f_2(x_2, y_2) - \nu^T y_2}_{\text{zavisi samo od } (x_2, y_2)}.$$

Dualni problem će biti separabilan po (x_1, y_1) i (x_2, y_2), tj. rešavanje dualnog problema može se svesti na rešavanje dva odvojena potproblema

$$\underbrace{g_1(\nu) = \inf_{x_1, y_1} f_1(x_1, y_1) + \nu^T y_1}_{\text{Potproblem 1"}} \quad \text{i} \quad \underbrace{g_2(\nu) = \inf_{x_2, y_2} f_2(x_2, y_2) - \nu^T y_2,}_{\text{Potproblem 2"}} \tag{3.2}$$

gde je dualni problem dat sa

$$\max g(\nu) = g_1(\nu) + g_2(\nu).$$

Analogno, kao i primarnoj dekompoziciji, rešavanje dualnih potproblema se može uraditi u potpunosti nezavisno. Prilagođavanjem Algoritma 1 može dobiti rešenje Potproblema 1" i 2", kao i rešenje problema (3.2). Algoritam za rešavanje problema (1.4) dualnom dekompozicijom je dat sa:

ALGORITAM 3.

ponovi

S1 Reši dualne potprobleme (nezavisno).
Nađi x_1, y_1 koji minimizira $f_1(x_1, y_1) + \nu^T y_1$.
Nađi x_2, y_2 koji minimizira $f_2(x_2, y_2) - \nu^T y_2$.

S2 Ažuriraj dualne promenljive.
$\nu := \nu - \alpha_k(y_2 - y_1)$.

Dakle, u koraku S1 se rešavaju dualni potproblemi i u koraku S2 zatim ažuriramo dualnu promenljivu koristeći gradijent funkcija g_1 i g_2.

4. Zaključak

U okviru ovog preglednog rada su izložene dve mogućnosti za rešavanje skoro separabilnih optimizacionih problema korišćenjem primarne i dualne dekompozicije. Mogućnost primene primarne i dualne dekompozicije na distribuiranu optimizaciju se ne zaustavlja samo na optimizaciji bez ograničenja, već se slične ideje mogu iskoristiti i na optimizaciju sa ograničenjima [3]. Ipak, autori su odlučili da se u ovom radu fokusiraju isključivo na primenu dekompozicionih metoda u optimizaciji bez ograničenja kako bi jasno i sažeto objasnili osnovne principe primarne i dualne dekompozicije. Takođe, bitno je napomenuti da uz odgovarajuće modifikacije navedenih algoritama i izbora pravca pretraživanja, dekompozicija se može uspešno primeniti na klasu semi-glathih funkcija. Korišćenje subgradijentnih metoda omogućava efikasno rešavanje optimizacionih problema i u slučajevima gde funkcije nisu glatke, tako da ovo daje nove mogućnosti za primenu dekompozicionih tehnika u širem spektru praktičnih problema.

Literatura

[1] S.P. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, vol. 3 of *Foundations and Trends in Machine Learning*, 2011.

[2] S.P. Boyd and L. Vandenberghe, *Convex optimization, Cambridge university press,* 2004.

[3] S.P. Boyd, L. Xiao, A. Mutapcic and J. Mattingley, *Notes for EE364B, Stanford University*, 2007.

[4] I. Griva, S. G. Nash and A. Sofer, *Linear and Nonlinear Optimization 2nd Edition. Society for Industrial and Applied Mathematics,* 2008.

[5] G. B. Dantzig and P. Wolfe, *Decomposition principle for linear programs*, vol. 8 of *Operation research*, 1960.

[6] N. Krejić, and N. Krklec, *Line search methods with variable sample size for unconstrained optimization,* Journal of Computational and Applied Mathematics, 245, 213-231, 2013.

[7] N. Krejić, and N. Krklec, *Nonmonotone line search methods with variable sample size,* Numer Algor 68, 711–739, 2015.

[8] J. Nocedal and S. J. Wright, *Numerical optimization. New York, NY: Springer New York.* 1999.

[9] T. Ostojić, *Modifications of Newton-type methods for solving semi-smooth stochastic optimization problems.* PhD dissertation, Faculty of Sciences, Novi Sad, 2023.

[10] Z. J. Shi, *Convergence of line search methods for unconstrained optimization,* Applied Mathematics and Computation, 157(2), 393-405, 2004.

[11] Y. Zheng and Q. Liu, *A review of distributed optimization: Problems, models and algorithms* vol. 483 of *Neurocomputing*, 2022.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

TOMOGRAFSKA REKONSTRUKCIJA BAZIRANA NA PRETRAŽIVANJU NULA PROSTORA

Tamara Palalić [1] **i Tibor Lukić** [2]

https://doi.org/10.24867/META.2024.18 Review article

Sažetak. U ovom radu je dat pregled postupka tomografske rekonstrukcije slika, predloženog u radu Lukić i Kopanja [1], koji se bazira na pretraživanju nula prostora projektivne matrice. Do optimalnog rešenja se dolazi pretraživanjem kroz linearne kombinacije bazičnih vektora nula prostora projektivne matrice. Glavna prednost ovog algoritma jeste što u toku rekonstrukcije projektivna greška ostaje približno jednaka nuli. Pored teorijske analize, rad donosi i prikaz eksperimentalnih rezultata i poređenja sa drugim postojećim algoritmima.

AMS klasifikacija (2020): 65D18, 68U10, 94A08

Ključne reči: tomografija, nula prostor, projektivna greška

1. Uvod

Tomografija je područje obrade slike koje se bavi rekonstrukcijom nepoznate slike iz dostupnih projektivnih podataka [2]. Matematički, slika se može modelirati funkcijom čiji skup slika može biti neprekidan ili diskretni skup. U kompjuterizovanoj tomografiji funkcija slike je neprekidna. Diskretna tomografija (DT) [3, 4] je deo tomografije, gde je raspon funkcije slike konačan i diskretan skup. Specijalno, u binarnoj tomografiji (BT), nepoznata slika sadrži samo dve različite vrednosti intenziteta, obično nula i jedan. Spektar primene tomografskih metoda vrlo je širok, naširoko se koriste u različitim industrijskim istraživanjima problema, često u obliku nerazornog ispitivanja materijala.

Već postoje uspešne metode, predložene u literaturi za rešavanje problema tomografske rekonstrukcije, na primer [5, 6, 9]. Zajednička osobina ovih postupaka jeste što se u svakoj iteraciji menja vrednost piksela u cilju smanjivanja projektivne greške. Međutim, vrednost projektivne greške nije jednaka nuli. S obzirom da su projektivni podaci jedino što u startu znamo o nepoznatoj slici, cilj je da rekonstruišemo sliku koja što bliže odgovara datim podacima. U [1] predložili smo novi metod za tomografsku rekonstrukciju u kome projektivna greška teorijski uvek ostaje jednaka nuli - u praksi je približno jednaka nuli.

U poglavlju 2 dat je matematički model tomografske rekonstrukcije, dok je u poglavlju 3 dat pregled predloženog modela tomografske rekonstrukcije slika.

[1]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: palalict@uns.ac.rs

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: tibor@uns.ac.rs

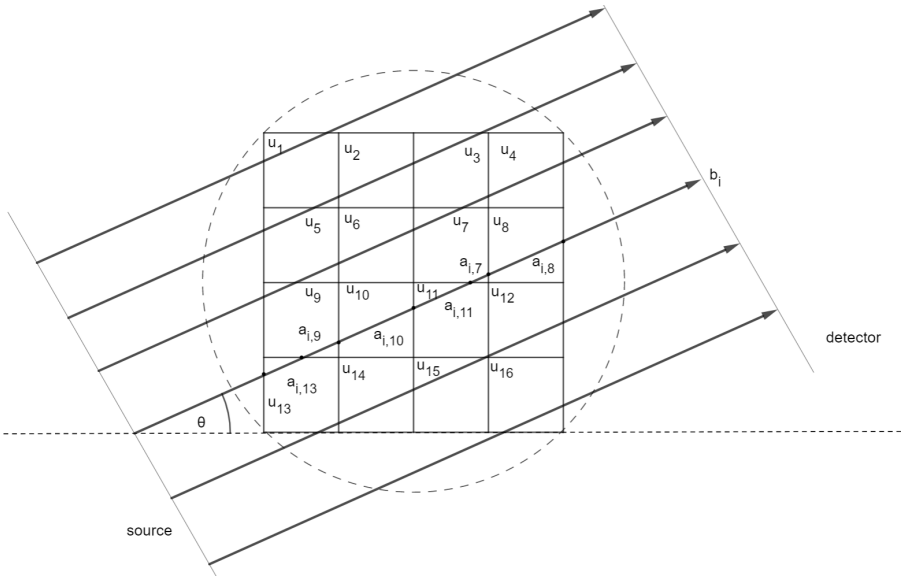
Poglavlje 4 je posvećeno eksperimentalnim rezultatima. Kratki zaključak dat je u poglavlju 5.

2. Definicije

Posmatrali smo model transmisione tomografske rekonstrukcije slika. Glavna karakteristika transmisione tomografije je da su i izvor i detektor postavljeni izvan posmatranog objekta. Matematički može se formulisati sledećim linearnim sistemom jednačina:

$$A\,u = b, \quad A \in \mathbb{R}^{m\times n}, \;\; u \in \mathbb{R}^{n}, \;\; b \in \mathbb{R}^{m}, \tag{2.1}$$

gde u predstavlja nepoznatu sliku koju treba rekonstruisati. U slučaju binarne tomografije, komponente vektora u imaju samo dve različite vrednosti, obično 0 i 1. Matrica A je tzv. projektivna matrica i njene vrste nose informaciju o dužini projektivnog zraka kroz piksele. Pretpostavka je da je svaki piksel predstavljen kao kvadrat s jediničnom dužinom stranice. Na slici 1 je predstavljena izračunata vrednost projekcije (elementa projektivne matrice A) za datu sliku iz jednog projektivnog pravca označenog uglom θ.



Slika 1: Paralelna transmisiona tomografija, i-ta projektivna vrednost je dobijena kao $b_i = a_{i,13}u_{13} + a_{i,9}u_9 + a_{i,10}u_{10} + a_{i,11}u_{11} + a_{i,7}u_7 + a_{i,8}u_8$.

Drugi projektivni pravac je dobijen rotacijom sistema izvor-detektor oko centra kruga. Svaki projektivni pravac doprinosi novom paralelnom skupu projektivnih zraka. Detektovane projektivne vrednosti stavljaju se u vektor projekcije b. Nije teško videti da je matrica projekcije A retka, tj. većina elemenata $a_{i,j}$ jednaki su nuli. U problemu tomografske rekonstrukcije projektivna matrica A i projektivni podaci b su dati, a potrebno je odrediti sliku u.

3. Metod tomografske rekonstrukcije baziran na pretraživanju nula prostora

U ovom poglavlju je dat pregled predložene metode za sivu i binarnu tomografiju. Osnovna ideja jeste da rešenje sistema jednačina (2.1)

$$A\,u = b,$$

može predstaviti kao zbir nekog partikularnog rešenja u_p, $(A\,u_p = b)$ i odgovarajućeg vektora koji pripada nula prostoru matrice A, definisanog sa

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid A\,x = 0\}.$$

Dakle, skup svih rešenja možemo predstaviti kao

$$u_p + z, \quad \text{where} \quad z \in \mathcal{N}(A).$$

Označimo bazu vektorskog prostora $N(A)$ skupom vektora $b_1, b_2, ...b_k$. Svako rešenje sistema 2.1 može se prikazati u sledećem obliku

$$\begin{gathered} w(\alpha) = u_p + \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 + \ldots + \alpha_k \mathbf{b}_k, \\ \text{where } \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathbb{R}^k \text{ and } \mathbf{b}_i \in \mathbb{R}^n \text{ for all } i = 1, \ldots, k. \end{gathered} \tag{3.1}$$

Odaberemo vrednosti koeficijenata $\alpha_1, \alpha_2, \ldots \alpha_k$ u α na način da dobijeno rešenje $w(\alpha) \in \mathbb{R}^n$ leži u određenom intervalu, na primer hiperkocki $[0,1]^n$. Da bismo to uspeli, posmatrali smo problem optimizacije

$$\arg\min_{\alpha} \sum_{i=1}^{n} W(w(\alpha)_i), \tag{3.2}$$

gde W predstavlja specijalno formiranu potencijalnu funkciju, definisanu sa

$$W(x) = \begin{cases} x^2, & x \leq 0 \\ (x-1)^2, & x \geq 1 \\ 0, & 0 < x < 1 \end{cases}. \tag{3.3}$$

W je neprekidno diferencijabilna funkcija koja se sastoji od dve kvadratne parabole i horizontalne linije. Njeno važno svojstvo je da je uvek veća od nule osim za vrednosti između 0 i 1, kada je nula, Slika 2 levo. Stoga, rešenje $\alpha^\star$ problema 3.2 daje kombinaciju koeficijenata baze vektora za koje odgovarajuće rešenje slike $W(\alpha^\star)$ pripada skupu $[0,1]^n$.

Za rešavanje problema 3.2 koristili smo Spectral Conjugate Gradient algoritam [7], koji je pokazao dobre rezultate u našem eksperimentalnom radu.

Pre svega, treba da odredimo jedno partikularno rešenje u_p. Predložili smo rešenje sa najmanjom normom u_{LN}, što se vrlo brzo pronađe korišćenjem algoritma konjugovanih gradijenata [8].

Napominjemo još jednom da u toku rekonstrukcije projektivna greška ostaje vrlo niska, praktično jednaka nuli. To je vrlo važno jer je to jedini podatak od koga polazimo, te želimo da rekonstruisana slika što verodostojnije odgovara projektivnim podacima. U toku procesa se menjaju koeficijenti uz bazične

vektore nula prostora čime se menja i rešenje, odnosno vrednosti piksela slike, ali projektivni podaci ostaju zadovoljeni. Ta činjenica je osnovna prednost predloženog metoda, nazvanog *Null Space Search based Tomography* (NSST).

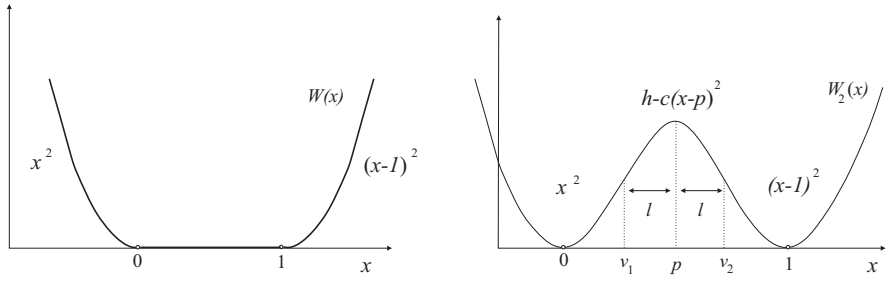
Dati algoritam se lako modifikuje za potrebe traženja binarnog rešenja, tj. binarne tomografije menjanjem potencijalne funkcije W.

$$W_2(x) = \begin{cases} x^2, & x \leq v_1 \\ (x-1)^2, & x \geq v_2 \\ h - c(x-p)^2, & v_1 < x < v_2 \end{cases}, \tag{3.4}$$

gde smo za parametar p podesili vrednost:

$$p = \frac{1}{2},\ v_1 = p - l,\ v_2 = p + l,\ c = \frac{1}{2l} - 1 \text{ i}$$

$$h = \frac{1}{4}(1 - 2l).$$



Slika 2: Potencijalna funkcija W definisana za rekonstrukciju sivih rešenja (levi grafik) i potencijalna funkcija W_2 definisana za rekonstrukciju binarnih rešenja (desni grafik)

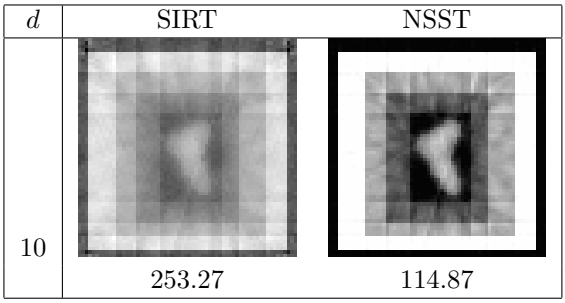
4. Rezultati

U ovom poglavlju dali smo kratki pregled eksperimentalnih rezultata u kome su koriš ćene dve binarne slike (IM1 i IM2) i jedna siva slika (IM3). Sve slike imaju piksele čije su vrednosti iz intervala $[0, 1]$.

Slika 3: Originalne test slike korisćene u eksperimentu

Tabela 1: Eksperimentalni rezultati za IM1 i IM2 koristeći 3 različita metoda rekonstrukcije. Parametar d označava broj projektivnih pravaca

		IM1				IM2			
d		2	3	4	6	2	3	4	6
NSST	E_P	2.22e-12	8.61e-08	6.44e-08	8.26e-08	1.73e-12	7.24e-08	8.63e-08	8.52e-08
	E_R	313.51	18.5770	5.30e-07	1.19e-06	1296	573.89	6.16	8.51e-07
	rE_R	23.93%	1.42%	$\approx 0\%$	$\approx 0\%$	78.83%	34.91%	0.37%	$\approx 0\%$
SPG	E_P	2.82	0	0	0	4.89	5.39	0	0
	E_R	21	0	0	0	1184	595	0	0
	rE_R	1.60%	0%	0%	0%	72.02%	36.19%	0%	0%
DC	E_P	23.49	6.91	6.69	7.87	23.49	17.96	16.50	10.65
	E_R	1325	27	12	10	1325	1007	236	30
	rE_R	1.01.15%	2.06%	0.92%	0.76%	80.60%	61.25%	14.36%	1.82%



Slika 4: Rekonstrukcija test slike IM3. Predstavljene vrednosti ispod rekonstrukcije pokazuju odgovarajuće greške rekonstrukcije E_R.

Upoređen je metod tomografske rekonstrukcije NSST sa tri poznata rekonstrukcijska postupka: sa SPG-om [6]i DC [9] algoritmi za slučaj binarne tomografije, te sa SIRT [5] algoritam za rekonstrukciju sive slike. Svi razmatrani algoritmi su implementirani u Matlab-u. O kvalitetu dobijenih rekonstrukcija govore sledeće tri funkcije merenja greške:

$$E_P(u^r) = \|Au^r - b\|, \qquad E_R(u^r) = \sum_{i=1}^{n} |u_i^r - u_i^*|, \qquad rE_R(u^r) = \frac{E_R(u^r)}{n_O} \cdot 100\,\%$$

gde je u^r rekonstruisana slika, $u^\star$ originalna slika i n_O ukupan broj piksela slike. Funkcija E_P je projektivna greška koja meri saglasnost rekonstrukcije sa projektivnim podacima. Greška rekonstrukcije E_R izražava udaljenost rekonstrukcije od originalne slike. Slika. 4 prikazuje rekonstrukcije sive slike. NSST pruža i vizuelno bolje rezultate u oba primera.

5. Zaključak

U ovom radu je dat pregled metoda tomografske rekonstrukcije baziranog na pretraživanju nula prostora projektivne matrice, predloženog u radu [1]. U eksperimentalnim rezultatima pokazao je dobre preformanse u odnosu na

nekoliko ranije predloženih algoritama za tomografsku rekonstrukciju. Teorijski, osnovna prednost ovog metoda jeste održavanje niske projektivne greške, gotovo jednake nuli. Za optimizaciju koristili smo SPG algoritam. Metod dozvoljava korišćenje dodatnih regularizacija i tu je mogućnost unapređivanja postupka u cilju bolje rekonstrukcije i smanjivanja broja projektivnih podataka.

Zahvalnica

Ovo istraživanje je podržano od strane Ministarstva nauke, tehnološkog razvoja i inovacija (br. ugovora: 451-03-65/2024-03/200156) i Fakulteta tehničkih nauka kroz projekat "Naučno istraživački i umetničko istraživački rad istraživača u nastavnim i saradničkim zvanjima Fakulteta tehničkih nauka u Novom Sadu"(br. 01-3394/1).

Takođe, autori su podržani od strane projekta Departmana za opšte discipline u tehnici, Fakulteta tehničkih nauka, Univerziteta u Novom Sadu "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama".

Literatura

[1] T. Lukić, T. Kopanja "Tomography Reconstruction Based on Null Space Search." u *Proceedings of the twenty-first International Workshop on Combinatorial Image Analysis (IWCIA 2022)*, vol. 13348, pp. 247–259, Messina, Italy, Lecture Notes in Computer Science, Springer-Verlag, 2023.

[2] G. T. Herman, *Image reconstruction from Projections.* Springer-Verlag, 1980.

[3] G. T. Herman, A. Kuba, *Advances in Discrete Tomography and Its Applications*, Birkhauser, 2007.

[4] G. T. Herman, A. Kuba *Discrete Tomography: Foundations, Algorithms and Applications*, Birkhauser, 1999.

[5] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections", u *Journal of Theoretical Biology 36(1)*, pp. 105-117, 1972.

[6] T. Lukić, A. Lukity "A Spectral Projected Gradient Optimization for Binary Tomography.", Computational Intelligence in Engineering, SCI, vol. 313, pp. 263–272. Springer-Verlag, 2010

[7] E. Birgin, J. Martinez, "Spectral Conjugate Gradient Method for Unconstrained Optimization", *Appl. Math. Optimization* 43, pp. 117-128, 2001.

[8] J. Nocedal, S. J. Wright, *Numerical Optimization.* Springer, New York, NY, USA, 2nd edn. 2006.

[9] T. Schule, C. Schnorr, S. Weber, J. Hornegger, "Discrete Tomography by Convexconcave Regularization and D.C. Programming." Discrete Appl. Math. 151, 229– 243, 2005

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

BULATOVLJEVA TEORIJA GRAFOVA U TEJLOR-MINIMALNIM ALGEBRAMA

Aleksandar Prokić [1] **i Simona Prokić** [2]

https://doi.org/10.24867/META.2024.19 Review article

Sažetak. Tejlorove algebre su usko vezane za problem zadovoljenja uslova (eng. Constraint Satisfaction Problem). Dihotomija problema zadovoljenja uslova pokazana je 2017. godine u nezavisnim radovima Bulatova i Žuka. U ovom radu ćemo proučavati određeni redukt Tejlorovih algebri, tzv. Tejlor-minimalne algebre. Pokazaće se da spomenute algebre imaju neke lepe osobine, koje Tejlorove algebre ne moraju da zadovoljavaju u opštem slučaju. To nas dovodi do pretpostavke da bi mogao da postoji jednostavniji dokaz dihotomije za problem zadovoljenja uslova.

AMS klasifikacija (2020): 08A30, 08A70

Ključne reči: Tejlorova algebra, univerzalna algebra, Problem zadovoljenja uslova, CSP, META

1. Uvod

Problem zadovoljenja uslova (skraćeno CSP) je privukao naučnike iz raznih oblasti matematike i teorijskog računarstva. Primeri problema zadovoljenja uslova su problem zadovoljenja logičkih formula, problem obojivosti grafova, rešavanje sistema jednačina nad konačnom algebarskom strukturom, kao i mnogi drugi primeri.

Jedan od prvih značajnijih rezultata vezanih za klasifikaciju CSP-a je dao Tomas Šefer 1978. godine u [10], gde je pokazao da vremenska složenost rešavanja CSP-a nad dvoelemntnim domenom, tzv. Bulovom strukturom, je ili polinomna ili NP-kompletna. Pretpostavku o dihotomiji za CSP su postavili Feder i Vardi i ona glasi da se pokazana teorema Šefera može uopštiti na domen proizvoljne konačne veličine. Što su Bulatov [3] i Žuk [11, 12] i pokazali u nezavisnim radovima 2017. godine.

U Šeferovom radu je prepoznato da vremenska složenost CSP-a zavisi od relacijskog klona koji generišu relacije iz posmatranog CSP-a. Galoove veze nam daju jedinstvenu korespodenciju između relacijskih klonova i klonova (skupa term operacija neke algebre). To dovodi do zaključka da je zapravo jedino potrebno posmatrati operacije kompatibilne sa relacijama CSP-a, što nas dovodi do toga da možemo koristiti i neke tehnike iz univerzalne algebre.

[1]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: aprokic@uns.ac.rs

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: simona.k@uns.ac.rs

Bulatov, Jeavons i Krohkin su 2005. godine u [5] pokazali da je svaki CSP ekvivalentan CSP-u nad idempotentnim domenom, odnosno gde za sve terme t važi $t(x, \dots, x) = x$. Drugi važan doprinos tog rada je da je pokazano da ako ne postoji Tejlorov term kompatibilan sa relacijama posmatranog CSP-a, onda je vremenska složenost tog CSP-a NP-kompletna. Tako je od 2005. godine postojala hipoteza da ako postoji Tejlorov term kompatibilan sa relacijama, u tom slučaju se CSP rešava u polinomnom vremenu. Iz navedenih razloga se podrazumeva da su sve algebre u ovom radu idempotentne i konačne.

I Bulatov i Žuk u svojim radovima nisu radili sa generalno Tejlorovim algebrama, nego su koristili redukt Tejlorove algebre (izbacili su neke operacije koje im nisu bile potrebne) tako da i dalje imaju Tejlorov term ali dobijaju i neke nove lepe osobine algebri. To nas dovodi do toga da posmatramo algebre koje su Tejlorove ali da nijedan njihov pravi redukt nije Tejlorov, odnosno do *Tejlor-minimalnih algebri*. Tejlor-minimalne algebre su prvo definisane u [1] i ovo je pregledni rad tog rada.

2. Definicije

Definicija 2.1. (Idempotentna, konačna) algebra $\mathbf{A}$ je Tejlorova ako za svako $n \in \mathbb{N}$ ne postoji količnička algebra podalgebre od $\mathbf{A}^n$ takva da se dobije dvoelementna algebra čije su sve operacije projekcije.

Ova definicija se može pojednostaviti.

Teorema 2.2 ([6])**.** *Algebra je Tejlorova ako i samo ako ne postoji količnička algebra podalgebre od* $\mathbf{A}$ *takva da se dobije dvoelementna algebra čije su sve operacije projekcije.*

Za osnovne definicije i teoreme iz univerzalne algebre pogledati knjigu [7]. Za nas najkorisnija karakterizacija Tejlorovih algebri je sledeća.

Teorema 2.3 ([2])**.** *Algebra* $\mathbf{A}$ *je Tejlorova ako za svaki prost broj* $p > |A|$, $\mathbf{A}$ *ima term operaciju arnosti* p *koja je ciklična, odnosno zadovoljava identitet*

$$t(x_1, x_2, \dots, x_p) = t(x_2, \dots, x_p, x_1).$$

Tri tipa operacija koje su Tejlorove i sa kojima ćemo raditi su:

- *polumrežna operacija* je binarna operacija $\vee$ koja je komutativna, idempotentna i asocijativna. Algebra $(A; \vee)$, gde je $\vee$ polumrežna operacija, se naziva polumreža.

- *majdžoriti operacija* (eng. majority) je ternarna operacija m koja zadovoljava $m(x, x, y) = m(x, y, x) = m(y, x, x) = x$.

- *Maljcevljeva operacija* je ternarna operacija p koja zadovoljava $p(y, x, x) = p(x, x, y) = y$. Jedan primer takve operacije je $d(x, y, z) = x - y + z$ na A, gde su $+$ i $-$ operacije abelove grupe na skupu A. U tom slučaju algebra $(A; x - y + z)$ se naziva *afina Maljcevljeva algebra* date abelove grupe.

Definicija 2.4. Algebra $\mathbf{A}$ je abelova ako je dijagonala $\Delta_A = \{(a, a) : a \in A\}$ jednaka klasi kongruencije za neku kongruenciju algebre $\mathbf{A}^2$.

Teorema 2.5. *Tejlorova algebra je abelova ako i samo ako je afini modul.*

Lako se može proveriti da na dvoelementnom skupu postoje samo četiri Tejlorove algebre: dve polumreže, majdžoriti i Maljcevljeva. Ove algebre su bile i inspiracija Bulatovu za definisanje njegovih ivica na algebri. U radu [1] koji ovde predstavljamo definisane su samo debele ivice, dok je pored njih Bulatov definisao i tanke ivice u [4]. Elementi algebre predstavljaju čvorove, dok ivice (grane) grafa definišemo na sledeći način.

Definicija 2.6. Neka je $\mathbf{A}$ algebra. Uređeni par $(a, b) \in A^2$ je ivica ako postoji kongruencija θ na $\mathbf{E} = \mathrm{Sg}_{\mathbf{A}}(a, b)$ različita od E^2, takva da je zadovoljen jedan od uslova:

- (polumrežna ivica) Postoji term operacija $f \in \mathrm{Clo}_2(\mathbf{A})$ koja deluje kao $\vee$-polumrežna operacija na $\{a/\theta, b/\theta\}$ sa maksimalnim elementom b/θ, tj. $f(a/\theta, b/\theta), f(b/\theta, a/\theta) \subseteq b/\theta$.
- (majdžoriti ivica) Postoji term operacija $m \in \mathrm{Clo}_3(\mathbf{A})$ koja deluje kao majdžoriti operacija na $\{a/\theta, b/\theta\}$.
- (abelova ivica) Algebra $\mathrm{Sg}_{\mathbf{A}}(a, b)/\theta$ je abelova.

Ivica (a, b) se naziva minimalna ako postoji maksimalna kongruencija θ koja svedoči da je (a, b) ivica i za sve $a', b' \in A$ takve da $(a, a'), (b, b') \in \theta$, imamo da $\mathrm{Sg}_{\mathbf{A}}(a', b') = \mathrm{Sg}_{\mathbf{A}}(a, b)$.

Sada ćemo dati definiciju apsorbujućih skupova neke algebre. Ovaj termin potiče iz teorije koja je razvijana za CSP i takođe je Žuk koristi u svom radu. Kasnije ćemo videti i neke veze između Bulatovljevog pristupa i Žukovog pristupa.

Definicija 2.7. Neka je $\mathbf{A}$ algebra i $B \subseteq A$. Kažemo da je B n-apsorbujući skup od $\mathbf{A}$ ako postoji term operacija $t \in \mathrm{Clo}_n(\mathbf{A})$ takva da $t(\mathbf{a}) \in B$ kada $\mathbf{a} \in A^n$ i $|\{i : a_i \in B\}| \geq n - 1$.

Pored apsorbujućih skupova važnu ulogu u Žukovoj teoriji ima i pojam centra. Levi centar relacije $R \subseteq A \times B$ je skup $\{a \in A : (\forall b \in B)(a, b) \in R\}$. Ako R ima neprazan levi centar, kažemo da je levo centralna relacija. Desni centar i desna centralna relacija se definišu analogno. Relacija je centralna ako je i levo i desno centralna.

Definicija 2.8. Podskup $B \subseteq A$ je centar algebre $\mathbf{A}$ ako postoji algebra $\mathbf{C}$ bez netrivijalnih 2-apsorbujućih poduniverzuma i $R \leq_{sd} \mathbf{A} \times \mathbf{C}$ takva da je B levi centar od R. Ako $\mathbf{C}$ može biti izabrano da je Tejlorova algebra, onda ćemo B nazivati Tejlorovim centrom algebre $\mathbf{A}$.

Kažemo da je algebra $\mathbf{A}'$ redukt algebre $\mathbf{A}$ ako je $\mathrm{Clo}(\mathbf{A}') \subseteq \mathrm{Clo}(\mathbf{A})$.

Definicija 2.9. Algebra $\mathbf{A}$ je Tejlor-minimalna ako je Tejlorova i ne postoji pravi redukt od $\mathbf{A}$ koji je Tejlorov.

3. Rezultati

Počnimo sa nekim lakšim, ali i veoma važnim tvrđenjima.

Lema 3.1. *Za svaku Tejlorovu algebru postoji redukt koji je Tejlor-minimalan.*

Ova lema nam omogućava da kod CSP-a sa obične Tejlorove algebre pređemo na Tejlor-minimalne. To možemo uraditi zato što ako smanjimo broj operacija kompatibilnih sa relacijama CSP-a dobićemo samo teži problem, ali kako i dalje imamo Tejlorov term, na osnovu hipoteze o dihotomiji koja je sada teorema o dihotomiji, problem je i dalje rešiv u polinomnom vremenu. Na dvoelementnom skupu, recimo $\{0, 1\}$, postoje samo tri Tejlor-minimalne algebre do na izomorfizam: polumreža, majdžoriti algebra i afina algebra sa termom $d(x, y, z) = x -_2 y +_2 z$ na $\mathbb{Z}_2$. Da su Tejlor-minimalne sledi iz Postove karakterizacije klonova na dvoelementnom skupu [9]. Na troelementnom skupu Brejdi je pronašao svih 24 Tejlor-minimalnih algebri. One se mogu pronaći i u master radu [8]. Sledeća lema važi za Tejlor-minimalne algebre, ali ne i za Tejlorove u opštem slučaju.

Lema 3.2. *Neka je* $\mathbf{A}$ *Tejlor-minimalna algebra i* $B \subseteq A$ *je zatvoren za operaciju* $f \in Clo(\mathbf{A})$ *takvu da* B *zajedno sa restrikcijom od* f *na* B *formira Tejlorovu algebru. Tada je* B *poduniverzum od* $\mathbf{A}$.

Lema 3.3. *Svaka podalgebra, konačan proizvod ili količnička algebra Tejlor-minimalne algebre je takođe Tejlor-minimalna.*

Teorema 3.4. *Neka je* $\mathbf{A}$ *Tejlor-minimalna algebra i* B *apsorbujući skup od* $\mathbf{A}$ *u odnosu na neki term. Tada je* B *poduniverzum od* $\mathbf{A}$.

Term $t(x_1, \dots, x_n)$ zavisi od i-te koordinate, $i \leq n$, ako postoje $a_1, \dots, a_{i-1}$, $a_{i+1}, \dots, a_n \in A$ takvi da $t(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n)$ nije konstantna funkcija. Kažemo i da je i-ta koordinata esencijalna za term t.

Definicija 3.5. Neka je $\mathbf{A}$ algebra i $B \subseteq A$. Skup B je jako projektivan poduniverzum od $\mathbf{A}$ ako za svako $f \in \mathrm{Clo}(\mathbf{A})$ i svaku esencijalnu koordinatu i za term f važi $f(\mathbf{a}) \in B$ gde je $\mathbf{a} \in A^n$ takav da $a_i \in B$.

Teorema 3.6. *Sledeća tvrđenja su ekvivalentna za sve Tejlor-minimalne algebre* $\mathbf{A}$ *i sve* $B \subseteq A$*:*

- B *je* 2*-apsorbujući skup od* $\mathbf{A}$.
- $R(x, y, z) = B(x) \vee B(y) \vee B(z)$ *je poduniverzum od* $\mathbf{A}^3$.
- B *je jako projektivan poduniverzum od* $\mathbf{A}$.

Teorema 3.7. *Sledeća tvrđenja su ekvivalentna za sve Tejlor-minimalne algebre* $\mathbf{A}$ *i sve* $B \subseteq A$*:*

- B *je* 3*-apsorbujući skup od* $\mathbf{A}$.
- $R(x, y) = B(x) \vee B(y)$ *je poduniverzum od* $\mathbf{A}^2$.
- B *je Tejlorov centar od* $\mathbf{A}$.

Lema 3.8. *Usmeren graf formiran pomoću minimalnih ivica neke proizvoljne algebre je (slabo) povezan.*

Teorema 3.9. *Neka je (a, b) ivica (polumrežna, majdžoriti ili abelova) u Tejlor-minimalnoj algebri $\mathbf{A}$ i neka je θ kongruencija algebre $\mathbf{E} = Sg(a, b)$ koja svedoči da je (a, b) ivica.*

a) Ako je (a, b) polumrežna ivica, onda je $\mathbf{E}/\theta$ term ekvivalentno dvoelementnoj polumreži sa apsorbujućim elementom b/θ.

b) Ako je (a, b) majdžoriti ivica, onda je $\mathbf{E}/\theta$ term ekvivalentno dvoelementnoj majdžoriti algebri.

c) Ako je (a, b) abelova ivica, onda je $\mathbf{E}/\theta$ term ekvivalentno afinoj Maljcevljevoj algebri, gde je Maljcevljeva operacija $x - y + z$ dobijena kao term abelove grupe koja je izomorfna grupi $\mathbb{Z}_m$, za neki prirodan broj m.

Takođe, kongruencija koja svedoči da je ivica (a, b) polumrežna je jedinstvena i ta kongruencija je maksimalna. Isto važi i za majdžoriti ivice.

Lema 3.10. *Neka je (a, b) minimalna polumrežna ivica u Tejlor-minimalnoj algebri. Tada je $\{a, b\}$ poduniverzum od $\mathbf{A}$, iz čega sledi da je $Sg_{\mathbf{A}}(a, b) = \{a, b\}$ i kongruencija koja svedoči da je (a, b) ivica je relacija $=$.*

Definicija 3.11. Neka je $\mathbf{A}$ algebra, $B \subseteq A$ i (b, a) ivica. Kažemo da je B stabilan za ivicu (b, a) ako za svaku kongruenciju θ algebre $\mathrm{Sg}_{\mathbf{A}}(b, a)$, koja svedoči da je (b, a) ivica, takvu da b/θ seče skup B važi da svaka θ-klasa seče skup B.

Teorema 3.12. *Sledeća tvrđenja su ekvivalentna za sve Tejlor-minimalne algebre $\mathbf{A}$ i sve $B \subseteq A$:*

- *B 2-apsorbuje algebru $\mathbf{A}$.*
- *B je stabilan za sve ivice.*

Teorema 3.13. *Svaki apsorbujući skup (u odnosu na neki term t) Tejlor-minimalne algebre $\mathbf{A}$ je stabilan u odnosu na polumrežne i abelove ivice. Takođe, za svako $b \in A$ sledeća tvrđenja su ekvivalentna:*

- *$\{b\}$ apsorbuje $\mathbf{A}$.*
- *$\{b\}$ je stabilan u odnosu na polumrežne i abelove ivice.*

4. Zaključak

U ovom radu smo prezentovali najvažnije rezultate iz rada [1]. Definisane su tri vrste ivica, a zatim su pokazane i veze između ivica i apsorbujućih skupova u Tejlor-minimalnim algebrama. Pokazana je i teorema povezanosti u grafu formiranom pomoću ivica koja važi za bilo koje algebre. Razlika u ivicama između Tejlorove algebre i obične algebre je u abelovim ivicama: kod Tejlorovih algebri abelova ivica (a, b) je zapravo *afina*, tj. $\mathrm{Sg}(a, b)/\theta$ je afina algebra, dok u običnim algebrama abelova ivica može biti ili afina ili *skupovna*, tj. $\mathrm{Sg}(a, b)/\theta$ ima samo projekcije kao operacije što kod Tejlorovih algebri nije moguće.

Zahvalnica

Ovo istraživanje je podržano od strane Departmana za opšte discipline, Fakulteta tehničkih nauka, Univerziteta u Novom Sadu, kroz projekat "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama".

Literatura

[1] L. Barto, Z. Brady, A. Bulatov, M. Kozik, D. Zhuk, "Unifying the Three Algebraic Approaches to the CSP via Minimal Taylor Algebras", *arXiv*:2104.11808v5, 2024.

[2] L. Barto, M. Kozik, "Absorbing Subalgebras, Cyclic Terms, and the Constraint Satisfaction Problem", *Logical Methods in Computer Science*, pages 1-27, 2012.

[3] A. Bulatov, "A dichotomy theorem for nonuniform CSPs", *In 58th IEEE Annual Symposium on Foundations of Computer Science*, FOCS 2017, Berkeley, CA, USA, pages 319–330, 2017.

[4] A. Bulatov, "Local structure of idempotent algebras I", *arXiv*, abs/2006.09599, 2020.

[5] A. Bulatov, P. Jeavons, A. Krokhin, "Classifying the complexity of constraints using finite algebras", *SIAM J. Comput.*, pages 720–742, 2005.

[6] A. Bulatov, P. Jeavons, "Algebraic structures in combinatorial problems", Technical Report MATH-AL-4-2001, Technische universität Dresden, Dresden, Germany, 2001.

[7] S. Burris, H. P. Sankappanavar, *A course in universal algebra*, Springer-Verlag, New York, 1981.

[8] F. Jankovec, *Minimal Taylor Clones on Three Elements*, MsC thesis, Faculty of Mathematics and Physics, Charles University, Prague, 2022.

[9] E. Post, *The Two-Valued Iterative Systems of Mathematical Logic*, (AM-5), Princeton University Press, 1941.

[10] T. Schaefer, "The complexity of satisfiability problems", *In Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, STOC '78, New York, NY, USA, pages 216–226, 1978.

[11] D. Zhuk, "A proof of CSP dichotomy conjecture", *In 58th IEEE Annual Symposium on Foundations of Computer Science*, FOCS 2017, Berkeley, CA, USA, pages 331–342, 2017.

[12] D Zhuk, "A proof of the CSP dichotomy conjecture", *J. ACM*, pages 1-78, 2020.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

A COMPOSITIONAL ENCODING OF π-CALCULUS INTO C_π-CALCULUS

Ivan Prokić [1]

https://doi.org/10.24867/META.2024.20 Original scientific paper

Abstract. Recently, a fragment of the π-calculus, named C_π-calculus, has been introduced as a model for confidential name passing in concurrent systems. The key feature of the C_π-calculus is disabling forwarding in name passing. It has been shown that even being a fragment of the π-calculus, it is possible to recover the full π-expressivity in this fragment. This was done via an encoding of π-processes into the C_π-processes and showing that such an encoding satisfies some desired properties. However, one property this encoding does not satisfy is compositionality. In this paper, we propose a new encoding that satisfies this property. We evaluate the encoding with an example and leave the proof of operational correspondence for future work.

AMS Mathematics Subject Classification (2020): 68Q85

Key words and phrases: process calculi, π-calculus, expressiveness, encodings

1. Introduction

The π-calculus is a well-known and well-developed model for expressing mobile and concurrent systems [1]. In the process algebra community, it is frequently considered the algebra for concurrent computation in the same way that λ-calculus is viewed as the model for sequential computation. In the past decades, great work has been conducted in proposing variants of the π-calculus, mostly by extending its syntax and semantics with many new constructs that allow for a direct expression of several distinguished features of concurrent systems. On the other hand, a few works spotted that to get some features of concurrent systems, instead of extending the π-calculus, one can consider only its fragment. The most important work in this direction is the modeling of asynchronous communications [2], but also the modeling of internal mobility [3], and locality [4].

Recently, another fragment of the π-calculus, called C_π-calculus, appeared [5] that allows for the modeling of confidential name passing in concurrent systems. This model makes a simple syntactic restriction that disallows one of the key features of the π-calculus - the forwarding, in such a way making

[1]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: prokic@ftn.uns.ac.rs

π	::=		*Prefixes*
		$a!k$	output
	|	$a?x$	input
	|	$[a = b]\pi$	matching
P	::=		*Process terms*
		0	termination
	|	$\pi.P$	prefix
	|	$P \mid P$	parallel composition
	|	$(\nu k)P$	name restriction
	|	$!P$	replication

Table 1: Syntax of C_π

it possible to model confidential communications directly. However, the same paper showed that the full expressivity of the π-calculus is not compromised in this fragment, since it is possible to encode it in the C_π. The encoding presented in [5] is shown correct through operational correspondence. However, this encoding comes with one limitation - it relies on the specific "handler" processes that make the encoding non-compositional. In this paper, we give another encoding that is compositional by the definition. Here, we only define the encoding and provide an example. Proof of the operational correspondence for this encoding will be our future work.

2. The syntax and semantics

The syntax of the C_π-calculus is given in Table 1. It consists of three prefixes: output, input, and matching. The prefixes are then used to build prefixed processes, which can also be terminated, a parallel composition, name restriction, and replication. All of the process constructs are explained in more detail later in the operational semantics. The syntax of processes relies on the notion of free and bound names. All names in processes are said to be free except for names in name restriction construct (k) and placeholder names in input (x) - these are called bound. The syntax of the π-calculus is the same as in Table 1, with only one distinction. In C_π we have a restriction that the name to be sent in output prefix (k) cannot be bound as a placeholder in the input prefix latter - thus preventing forwarding. For instance, $a?x.a!x.0$ is not a C_π process, but it is a valid π process.

The semantics of the C_π-calculus is given in terms of reduction relation in Table 3, which relies on the structural congruence relation given in Table 2. The first three rules of the structural congruence relation define the set of processes with respect to the parallel operator being a commutative monoid. The following three rules deal with name restriction, they allow for garbage collection, scope extrusion, and name swapping. The last two rules allow for removing matching if the matched names coincide, and the final allows for replication. The first reduction rule defines how two processes working in par-

$$P \mid 0 \equiv P \qquad P \mid Q \equiv Q \mid P \qquad (P \mid Q) \mid R \equiv P \mid (Q \mid R) \qquad (\nu k)0 \equiv 0$$

$$P \mid (\nu k)Q \equiv (\nu k)(P \mid Q) \text{ if } k \notin \mathsf{fn}(P) \qquad (\nu k)(\nu l)P \equiv (\nu l)(\nu k)P$$

$$[a = a]\pi.P \equiv \pi.P \qquad !P \equiv P \mid !P$$

Table 2: Structural congruence.

$$\text{(R-COMM)}\ k!l.P \mid k?x.Q \to P \mid Q\{l/x\} \qquad \text{(R-PAR)}\ \frac{P \to Q}{P \mid R \to Q \mid R}$$

$$\text{(R-RES)}\ \frac{P \to Q}{(\nu k)P \to (\nu k)Q} \qquad \text{(R-STRU)}\ \frac{P \equiv P' \to Q' \equiv Q}{P \to Q}$$

Table 3: Reduction relation.

allel can establish communication. The left process is output prefixed, it is ready to send l over name k. The right process is input prefixed, it is ready to receive a name over the same name k. Then, the synchronization of the output and the input happens and we obtain the processes where the prefixes have been consumed and the name l, that has been received in the inputting process is substituted for the placeholder x. The following three rules close the reduction relation under parallel composition, name restriction, and structural congruence relation. We remark that the semantics for the π-calculus is the same.

3. The encoding

Our new encoding from π-calulus terms into the C_π-calculus terms is given in Table 4. The encoding is presented as a function $[\![\cdot]\!]$. The encoding function is parameterized with a set of pairs of names ρ, whose purpose is to collect the names x bound in inputs (i.e. the placeholders) and pair them with m_x that are used as placeholders for name handlers for the received names (as explained later). Also, our target terms for our encoding are the polyadic C_π-processes. This is only for the simpler definition of the encoding, the polyadic C_π can be directly encoded into the one presented in the previous section in the standard way.

Our encoding distinguishes between the two types of output prefixes: ones that send the name directly (as in $a!k$) and those that forward the received name (as in $a?x.a!x$) The first send prefix is encoded in the scope of three fresh names m_k, e_1, and e_2. Name m_k is the name of the *handler process*, and the other two names are used only for blocking and synchronization of the sending and receiving processes. The matching construct does not change, but in the continuation, we have sending on name a fresh names e_1 and e_2 - first is going

$$\begin{aligned}
[\![[\tilde{c}=\tilde{d}]a!k.P]\!]_\rho &= (\nu m_k, e_1, e_2)([\tilde{c}=\tilde{d}]a!(e_1,e_2).m_k!(e_1).e_2!().[\![P]\!]_\rho \\
&\quad \mid !m_k?(x).x!(k,m_k).0) \\
[\![[\tilde{c}=\tilde{d}]a!x.P]\!]_{\rho,(x,m_x)} &= (\nu e_1, e_2)[\tilde{c}=\tilde{d}]a!(e_1,e_2).m_x!(e_1).e_2!().[\![P]\!]_{\rho,(x,m_x)} \\
[\![[\tilde{c}=\tilde{d}]a?x.P]\!]_\rho &= [\tilde{c}=\tilde{d}]a?(y,z).y?(x,m_x).z?().[\![P]\!]_{\rho,(x,m_x)} \\
[\![(\nu k)P]\!]_\rho &= (\nu k)[\![P]\!]_\rho \\
[\![P_1 \mid P_2]\!]_\rho &= [\![P_1]\!]_\rho \mid [\![P_2]\!]_\rho \\
[\![!P]\!]_\rho &= ![\![P]\!]_\rho \\
[\![0]\!]_\rho &= 0
\end{aligned}$$

Table 4: Encoding of π processes into C_π processes.

to be used in the receiving process to communicate with the handler process, and the second to synchronize unlocking the encoding of the continuation of sending and receiving processes. Then, we have sending e_1 on m_k - that is to the handler process, and sending on e_2 - to synchronize unlocking directly with the receiving process. Finally, the encoding of the continuation process P follows. In parallel we have the handler process: it is ready to receive on m_k repeatedly and reply along the received name with k and m_k.

Encoding of the other sending prefix (the one that represents forwarding) goes as follows. First in ρ we find pair (x, m_x). Then the rest is the same as for the other send prefix, except that now we do not have the handler process (since x is only a placeholder), and that instead of sending to the name handler (in the previous case m_k) here we have only its placeholder m_x.

The encoding of the receiving process goes as follows. The matching construct remains the same. Receiving happened along a with two names. The first (see e_1 in the encoding of output) is then used to receive from the handler, and, finally, the second name (see e_2 in the encoding of output) is used to synchronize with the sending process and to unlock the continuation. The encoding is homomorphism for the name restriction, parallel composition, replication, and inactive process.

Directly from the definition we have the compositionally result.

Theorem 3.1. *For all π-calculus processes P_1 and P_2 it holds that*

$$[\![P_1 \mid P_2]\!]_\rho = [\![P_1]\!]_\rho \mid [\![P_2]\!]_\rho$$

As noted earlier, the proof of the operational correspondence is left for future work. Here we give an example of how our encoding works.

Example 3.2. Consider the π-calculus process

$$P = a!k.0 \mid a?x_1.b!x_1.0 \mid b?x_2.0$$

where in parallel we have the left process sending k along a, the middle process receiving the name and forwarding it to the right process along name b. Hence, we have the following reductions

$$P \to b!k.0 \mid b?x_2.0 \to 0$$

Following the encoding introduced in this paper we have

$$\begin{aligned}[\![P]\!]_\emptyset = (\nu m_k, e_1, e_2)(a!(e_1, e_2).m_k!(e_1).e_2!().0 \mid !m_k?(x).x!(k, m_k).0) \\ \mid a?(y, z).y?(x_1, m_{x_1}).z?().(\nu e_1', e_2')b!(e_1', e_2').m_{x_1}!(e_1').e_2'!().0 \\ \mid b?(y, z).y?(x_2, m_{x_2}).z?().0\end{aligned}$$

Now we can see the encoding at work by showing the reductions step by step. The first synchronization is on name a, and the resulting process is

$$\begin{aligned}(\nu m_k, e_1, e_2)(m_k!(e_1).e_2!().0 \mid !m_k?(x).x!(k, m_k).0 \\ \mid e_1?(x_1, m_{x_1}).e_2?().(\nu e_1', e_2')b!(e_1', e_2').m_{x_1}!(e_1').e_2'!().0) \\ \mid b?(y, z).y?(x_2, m_{x_2}).z?().0\end{aligned}$$

Here, the scope of names m_k, e_1, and e_2 is now extruded to the receiving process, which is now ready to receive along e_1. But, before that, the sending process has to activate the handler process on m_k, which leads to

$$\begin{aligned}(\nu m_k, e_1, e_2)(e_2!().0 \mid e_1!(k, m_k).0 \mid !m_k?(x).x!(k, m_k).0 \\ \mid e_1?(x_1, m_{x_1}).e_2?().(\nu e_1', e_2')b!(e_1', e_2').m_{x_1}!(e_1').e_2'!().0) \\ \mid b?(y, z).y?(x_2, m_{x_2}).z?().0\end{aligned}$$

The receiving process (in the middle line) can receive the name k, along with the name of the handler m_k, reducing to

$$\begin{aligned}(\nu m_k, e_1, e_2)(e_2!().0 \mid !m_k?(x).x!(k, m_k).0 \\ \mid e_2?().(\nu e_1', e_2')b!(e_1', e_2').m_k!(e_1').e_2'!().0) \\ \mid b?(y, z).y?(x_2, m_{x_2}).z?().0\end{aligned}$$

The unlocking along e_2 happens, and we end up with

$$\begin{aligned}(\nu m_k, e_1, e_2)(!m_k?(x).x!(k, m_k).0 \\ \mid (\nu e_1', e_2')b!(e_1', e_2').m_k!(e_1').e_2'!().0) \\ \mid b?(y, z).y?(x_2, m_{x_2}).z?().0\end{aligned}$$

Now the communication along name b takes place

$$\begin{aligned}(\nu m_k, e_1, e_2)(!m_k?(x).x!(k, m_k).0 \\ \mid (\nu e_1', e_2')(m_k!(e_1').e_2'!().0 \\ \mid e_1'?(x_2, m_{x_2}).e_1'?().0))\end{aligned}$$

Because of our parameterized encoding, name m_k is now in place in the process in the middle line, hence it can connect with the handler process, which in turn sends k and m_k to the process in the last line. Hence, in two reduction steps, we end up with

$$\begin{aligned}(\nu m_k, e_1, e_2)(!m_k?(x).x!(k, m_k).0 \\ \mid (\nu e_1', e_2')(e_2'!().0 \\ \mid e_1'?().0))\end{aligned}$$

Now the unlocking happens and we finally obtain (using the structural congruence) only the restricted handler process

$$(\nu m_k)!m_k?(x).x!(k, m_k).0$$

Since the handler process starts with sending along a restricted name it is behaviorally equivalent to the terminated process 0.

When compared with the encoding given in [5], our encoding has the following differences:

- it does not rely on the global handler processes and does not use outer and inner encodings, instead
- it introduces handler processes for each sending prefix (that is not forwarding).

For these reasons, we believe encoding from this paper is more natural - confirmed by the compositionality result obtained here. However, considering the proof of operational correspondence, we predict certain complications, since in the approach presented here the handler processes can appear under replication. In [5] this was not the case, since there the handler processes were introduced at the top level.

Acknowledgement

The author acknowledges the financial support of the Department of Fundamentals Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Improving the teaching process in the English language in fundamental disciplines".

References

[1] D. Sangiorgi and D. Walker, *The Pi-Calculus - a theory of mobile processes*, Cambridge University Press, 2001.

[2] K. Honda and M. Tokoro, "An Object Calculus for Asynchronous Communication", *ECOOP'91 European Conference on Object-Oriented Programming*, Geneva, Switzerland, July 15-19, 1991.

[3] D. Sangiorgi, "pi-Calculus, Internal Mobility, and Agent-Passing Calculi", *Theoretical Computer Science*, vol. 167, pp. 235–274, 1996.

[4] M. Merro and D. Sangiorgi, "On asynchrony in name-passing calculi", *Mathematical Structures in Computer Science*, vol. 14, pp. 715–767, 2004.

[5] I. Prokić and H. T. Vieira, "The C_π-calculus: A model for confidential name passing", *Journal of Logical and Algebraic Methods in Programming*, vol. 119, 2021.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

PROBABILISTIC REASONING ABOUT TYPED PROGRAMS: TOWARDS COMPACTNESS

Simona Prokić [1]

https://doi.org/10.24867/META.2024.21 Original scientific paper

Abstract. Reasoning with uncertainty has gained an important role in various fields of computer science, artificial intelligence and cognitive science, while it is underdeveloped in typed calculi. For this reason, we have investigated different approaches used to introduce probability into typed calculi. In this paper, we develop PCL logic, which is a formal model for probabilistic reasoning about typed programs. The semantics of PCL is based on the possible world approach. Allowing the range of probability functions to be infinite results in non-compactness of the logic. As a consequence, a finite axiomatization of the logic cannot be sound and strongly complete. We propose the simplest method for resolving the non-compactness phenomenon of the logic PCL.

AMS Mathematics Subject Classification (2020): 03B38, 03B40, 03B48, 03F03

Key words and phrases: probabilistic reasoning, combinatory logic, simple types, possible world semantics, compactness

1. Introduction

The motivation for developing a new formal model for reasoning about typed terms is the fact that reasoning with uncertainty has gained an important role in various fields of computer science, artificial intelligence and cognitive science, while it is underdeveloped in typed calculi.

Introducing non-determinism and probabilities into the typed calculi has been topic of several papers, e.g. [1, 2, 3, 4, 8]. Nevertheless, the goal of these papers was to formalize computation in the presence of uncertainty and not to provide a framework that enables probabilistic reasoning about typed terms. Our goal is to introduce the logic in which we can express the following sentence:

> The probability that a term M inhabits a type σ is at least s.

We follow the method used for the logic LPP_2 [9], which is a probabilistic extension of the classical propositional logic. The language of the logic LPP_2

[1]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: simona.k@uns.ac.rs

is obtained by extending the language of the classical propositional logic with probability operators $P_{\geq s}$ whose meaning is "the probability is at least s".

The idea of formalization of probabilistic reasoning about typed terms has been introduced in [5, 6], where we have tried to developed the formal model upon the well-known models of λ-calculus and combinatory logic. However, these models are not suitable for propositional reasoning about typed terms. For this reason, we have first developed the logic LCL, a classical propositional extension of the simply typed combinatory logic, in [7]. Then, we have introduced the probability extension of the logic LCL in [10], and we present this extension in the next section.

2. The logic PCL

In this section, we introduce the syntax and semantics of the logic PCL. The logic PCL is a probabilistic extension of the logic LCL introduced in [7]. Hence, the language of PCL is layered into two sets of formulas: basic formulas (LCL-formulas) and probabilistic formulas.

The LCL-formulas are generated by the following grammar

$$\alpha := M : \sigma \mid \neg\alpha \mid \alpha \Rightarrow \alpha$$

where M is a term and σ a type such that there exists a basis Γ in which M inhabits σ.

Probabilistic formulas are obtained by applying the probability operator to LCL-formulas. For $s \in [0,1] \cap \mathbb{Q}$ and an LCL-formula α, the formula $P_{\geq s}\alpha$ is called *a basic probabilistic formula*. The set of all probabilistic formulas is generated by the following grammar

$$\varphi := P_{\geq s}\alpha \mid \neg\varphi \mid \varphi \wedge \varphi$$

The language of PCL is the union of basic and probabilistic formulas. The set of all PCL-formulas is denoted by For_{PCL} and is ranged over by $A, B, C, \ldots$.

Although we have used only negation and implication in the definition of LCL-formulas, and negation and conjunction in the definition of probabilistic formulas, other classical propositional connectives are defined as usual. Please notice that mixing of basic formulas and probabilistic formulas, and nested probability operators are not allowed

Following the approach used for the logic LPP_2 and other probability logics [9], the semantics of the logic PCL is based on the possible world approach, where the set of possible worlds is equipped with a finitely additive probability measure.

Definition 2.1. A PCL-model is a structure

$$\mathcal{M} = (W, \{D_w\}, \{A^{\sigma}_{w}\}, \{\cdot_w\}, \{\mathbf{s}_w\}, \{\mathbf{k}_w\}, \{\mathbf{i}_w\}, H, \mu, \rho)$$

where:

- W is a non-empty set of objects, called *possible worlds*,

- $\{D_w\} = \{D_w\}_{w \in W}$ is a family of sets indexed by worlds, where the set D_w is referred to as *the domain* of the world w,
- $\{A^\sigma_w\} = \{A^\sigma_w\}_{w \in W, \sigma \in \mathsf{Types}_\to}$ is a family of sets indexed by types σ and worlds w such that $A^\sigma_w \subseteq D_w$ for all $w \in W$ and $\sigma \in \mathsf{Types}_\to$.
- $\{\cdot_w\} = \{\cdot_w\}_{w \in W}$ is a family of binary operations indexed by worlds such that the following hold:
  - $\cdot_w$ is a binary operation on D_w, i.e. $\cdot_w : D_w \times D_w \to D_w$,
  - $\cdot_w$ is extensional, that is for every $w \in W$ and every $d_1, d_2 \in D_w$, if $(\forall e \in D_w)(d_1 \cdot_w e = d_2 \cdot_w e)$, then $d_1 = d_2$,
  - for every $\sigma, \tau \in \mathsf{Types}_\to$, it holds that the codomain of the restriction of function $\cdot_w$ to the set $A^{\sigma\to\tau}_w \times A^\sigma_w$ is A^τ_w,
- $\{\mathbf{s}_w\} = \{\mathbf{s}_w\}_{w \in W}$ is a family of elements indexed by worlds such that for every $w \in W$ the following hold:
  - $\mathbf{s}_w \in D_w$,
  - for every $\sigma, \tau, \rho \in \mathsf{Types}_\to$, $\mathbf{s}_w \in A^{(\sigma\to(\tau\to\rho))\to((\sigma\to\tau)\to(\sigma\to\rho))}_w$
  - and for every $d, e, f \in D_w$, $((\mathbf{s}_w \cdot_w d) \cdot_w e) \cdot_w f = (d \cdot_w f) \cdot (e \cdot_w f)$
- $\{\mathbf{k}_w\} = \{\mathbf{k}_w\}_{w \in W}$ is a family of elements indexed by worlds such that for every $w \in W$ the following hold:
  - $\mathbf{k}_w \in D_w$,
  - for every $\sigma, \tau \in \mathsf{Types}_\to$, $\mathbf{k}_w \in A^{\sigma\to(\tau\to\sigma)}_w$
  - and for every $d, e \in D_w$, $(\mathbf{k}_w \cdot_w d) \cdot_w e = d$
- $\{\mathbf{i}_w\} = \{\mathbf{i}_w\}_{w \in W}$ is a family of elements indexed by worlds such that for every $w \in W$ the following hold:
  - $\mathbf{i}_w \in D_w$,
  - for every $\sigma \in \mathsf{Types}_\to$, $\mathbf{i}_w \in A^{\sigma\to\sigma}_w$
  - and for every $d \in D_w$, $\mathbf{i}_w \cdot_w d = d$
- H is an algebra of subsets of W.
- μ is a finitely additive probability measure, $\mu : H \to [0, 1]$.
- $\rho : W \times V \to \bigcup_{w \in W} D_w$ provides for each world a valuation of term variables such that for every $w \in W$, $\rho(w, \cdot)$ is a map from the set of term variables to the domain D_w, i.e. $\rho(w, \cdot) : V \to D_w$.

The semantics of PCL is defined in such way that each world of a model represents one LCL-model. We formally state this in the next proposition.

Proposition 2.2. *Let*

$$\mathcal{M} = (W, \{D_w\}, \{A^\sigma_w\}, \{\cdot_w\}, \{\mathbf{s}_w\}, \{\mathbf{k}_w\}, \{\mathbf{i}_w\}, H, \mu, \rho)$$

be a PCL-model. For each $w \in W$, the structure

$$\mathcal{M}_w = \langle D_w, \{A^\sigma_w\}_\sigma, \cdot_w, \mathbf{s}_w, \mathbf{k}_w, \mathbf{i}_w \rangle$$

is an applicative structure for LCL and $\mathcal{M}_{\rho_w} = \langle \mathcal{M}_w, \rho(w, \cdot) \rangle$ is an LCL-model.

The satisfiability of a formula in a model is defined inductively. First, we introduce the notion of satisfiability of a basic formula α in a possible world w of a model $\mathcal{M}$.

Definition 2.3. Let $\mathcal{M} = (W, \{D_w\}, \{A^\sigma_w\}, \{\cdot_w\}, \{\mathbf{s}_w\}, \{\mathbf{k}_w\}, \{\mathbf{i}_w\}, H, \mu, \rho)$ be a PCL-moodel, w' a possible world in $\mathcal{M}$ and α a basic formula. The formula α is satisfied in a world w', denoted by $w' \models \alpha$, if and only if α is satisfied by the LCL-model $\mathcal{M}_{\rho_{w'}} = \langle \mathcal{M}_{w'}, \rho_{w'} \rangle$ where $\mathcal{M}_{w'} = \langle D_{w'}, \{A^\sigma_{w'}\}_\sigma, \cdot_{w'}, \mathbf{s}_{w'}, \mathbf{k}_{w'}, \mathbf{i}_{w'} \rangle$ and $\rho_{w'}(x) = \rho(w', x)$.

Now, the satisfiability of a probabilistic formula is defined only for the class of measurable PCL models as follows.

Definition 2.4. A PCL-model

$$\mathcal{M} = (W, \{D_w\}, \{A^\sigma_w\}, \{\cdot_w\}, \{\mathbf{s}_w\}, \{\mathbf{k}_w\}, \{\mathbf{i}_w\}, H, \mu, \rho)$$

is *measurable* if $[\alpha]_\mathcal{M} \in H$ for every formula $\alpha \in \mathsf{For_B}$, where $[\alpha]_\mathcal{M} = \{w \in W \mid w \models \alpha\}$. The class of all measurable PCL-models is denoted by PCL_{Meas}.

Definition 2.5. The satisfiability relation $\models \subseteq PCL_{\mathsf{Meas}} \times \mathsf{For}_{PCL}$ is defined in the following way:

- $\mathcal{M} \models \alpha$ if and only if for every $w \in W$, $w \models \alpha$.
- $\mathcal{M} \models P_{\geq s}\alpha$ if and only if $\mu([\alpha]) \geq s$.
- $\mathcal{M} \models \neg\phi$ if and only if it is not the case that $\mathcal{M} \models \phi$.
- $\mathcal{M} \models \phi \wedge \psi$ if and only if $\mathcal{M} \models \phi$ and $\mathcal{M} \models \psi$.

3. Towards compactness

In this section, we deal with non-compactness of the logic PCL. We say that a logic L satisfies the compactness theorem if the following holds:

> The set X of formulas of the logic L is satisfiable if and only if every finite subset of X is satisfiable.

The logic PCL does not satisfy the compactness theorem. Let us consider the following set

$$X = \{\neg P_{=0}(x : \sigma)\} \cup \{P_{<\frac{1}{n}}(x : \sigma) \mid n \in \mathbb{N}\}.$$

The set X is not satisfiable, and every finite subset of X is satisfiable. For more details about proving satisfiability of finite subsets of X and non-satisfiability of the set X we refer the reader to [10].

As the consequence of non-compactness, we have that a finite axiomatization can not be sound and strongly complete. More precisely, the inconsistency of the set X can not be proved using finite proof. Let us assume that there is a finite axiomatization of PCL, which is sound and strongly complete and let X be an infinite set of formulas such that every subset of X is satisfiable and X itself is not. From the strong completeness of the axiomatization, it follows that the set X is inconsistent, since it is unsatisfiable. So, it holds that $X \vdash \bot$. Since the axiomatization is finite, the proof of $X \vdash \bot$ has to be a finite sequence of formulas. Thus, there exists a finite subset $X' \subseteq X$ such that $X' \vdash \bot$. Then X' is also inconsistent. Furthermore, we conclude X' is unsatisfiable by the soundness of the axiomatization. This contradicts the assumption that every finite subset of X is satisfiable. We see that if we take a finite strongly complete axiomatization, there will be unsatisfiable sets, that are consistent, which results in unsoundness of the axiomatization. In [10], we have introduced the infinite axiomatization for PCL, that has one infinitary rule (the rule with countably many premises). This rule corresponds to the Archimedean axiom for real numbers and it guarantees that the set X is inconsistent ($X \vdash \bot$).

Now, we focus on the method for resolving non-compactness. The simplest method for this is to allow only probability measures with fixed finite ranges in models. We follow the approach used in [9] to resolve the non-compactness phenomenon of the logic LPP_2, and we set the range of probability measure to be the set $Fr(n) = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$. The obtained logic is denoted by $PCL^{Fr(n)}$. So, there is countably many more logics (for each positive integer n, one logic) which are similar to the logic PCL.

With this change in the semantics, the set X introduced above is still unsatisfiable, but now it is possible to give a finitary strongly complete axiomatization such that X is inconsistent.

The next step is to give sound and strongly complete axiomatizations for the logics $PCL^{Fr(n)}$. Finally, after we obtain the soundness and completeness results for the logics $PCL^{Fr(n)}$ we plan to prove Compactness theorem for these logics.

Acknowledgement

This work has been supported by Department of Fundamentals Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Improving the teaching process in the English language in fundamental disciplines".

References

[1] A. Bizjak and L. Birkedal, "Step-indexed logical relations for probability", In *Foundations of Software Science and Computation Structures - 18th International Conference, FoSSaCS 2015,* volume 9034 of Lecture Notes in Computer Science, pages 279–294, Springer, London, UK, April 11-18, 2015.

[2] R. Crubillé and U. Dal Lago, "On probabilistic applicative bisimulation and call-by-value λ-calculi", in *ESOP*, volume 8410 of Lecture Notes in Computer Science, pages 209–228, Springer, 2014.

[3] R. Crubillé, U. Dal Lago, D. Sangiorgi, and V. Vignudelli, "On applicative similarity, sequentiality, and full abstraction", in *Correct System Design*, volume 9360 of Lecture Notes in Computer Science, pages 65–82, Springer, 2015.

[4] U. Dal Lago and A. Rioli, "Applicative bisimulation and quantum λ-calculi" In *Fundamentals of Software Engineering - 6th International Conference, FSEN 2015*, volume 9392 of Lecture Notes in Computer Science, pages 54–68, Springer, Tehran, Iran, April 22-24, 2015.

[5] S. Ghilezan, J. Ivetić, S. Kašterović, Z. Ognjanović, and N. Savić, "Probabilistic reasoning about simply typed lambda terms", In *Logical Foundations of Computer Science - International Symposium - LFCS 2018*, volume 10703 of Lecture Notes in Computer Science, pages 170–189, Springer, Deerfield Beach, FL, USA, January 8-11, 2018.

[6] S. Ghilezan, J. Ivetić, S. Kašterović, Z. Ognjanović, and N. Savić, "Towards probabilistic reasoning in type theory - the intersection type case", In *Foundations of Information and Knowledge Systems - 11th International Symposium, FoIKS 2020*, volume 12012 of Lecture Notes in Computer Science, pages 122–139. Springer, Dortmund, Germany, February 17-21, 2020.

[7] S. Kašterović and S. Ghilezan, "Logic of Combinatory Logic", CoRR, `arXiv:2212.06675`.

[8] S. Kašterović and M. Pagani, "The discriminating power of the let-in operator in the lazy call-by-name probabilistic lambda-calculus", in *4th International Conference on Formal Structures for Computation and Deduction, FSCD 2019*, volume 131 of LIPIcs, pages 26:1–26:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dortmund, Germany, June 24-30, 2019.

[9] Z. Ognjanović, M. Rašković, and Z. Marković, *Probability Logics: Probability-Based Formalization of Uncertain Reasoning*, Springer Publishing Company, Incorporated, 1st edition, 2016. ISBN 3319470116.

[10] S. Prokić, *Probabilistic reasoning about computation and simple type theory*, PhD dissertation, Fakultet tehničkih nauka, Novi Sad, 2024. `http://www.ftn.uns.ac.rs/17547101/doktorska-disertacija`

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

METOD ZA UVEĆANJE DIGITALNE SLIKE BAZIRAN NA INTERVALNIM FAZI SKUPOVIMA[1]

Nebojša Ralević [2], Julija Šćekić [3] i Nataša Milosavljević [4]

https://doi.org/10.24867/META.2024.22 Original scientific paper

Sažetak. U ovom istraživanju predstavljamo inovativni pristup konstrukcije intervalnih fazi skupova (IVFS) primenom fazi skupova za obradu slika. Primenjujemo novi algoritam za smanjenje (uvećanje) rezolucije slike korišćenjem IVFS-a. Sam algoritam se temelji na principu blokovske ekspanzije, naglašavajući svoju jednostavnost i efikasnost kao ključne karakteristike. Implementacija ovog algoritma u Python-u omogućava jednostavnu primenu u praksi, pružajući korisnicima alat za poboljšanje kvaliteta slika uz minimalni gubitak informacija. Očekujemo da će naša metodologija imati značajan uticaj na razvoj tehnika obrade slika, posebno u domenima gde je potrebno zumiranje rezolucije bez značajnog uvođenja kompleksnosti.

AMS klasifikacija (2020): 68U10, 94A08

Ključne reči: intervalni fazi skup, obrada slike, uvećanje digitalne slike

1. Uvod

U članku je prikazana primena novog algoritma u obradi slika koji se oslanja na koncept intervalno vredonosnih fazi skupova (IVFS). Ovaj pristup omogućava detaljnije analize lokalnog okruženja svakog piksela na slici, čime se otvaraju nove mogućnosti u rešavanju problema uvećanja i smanjenja slika. Ključna karakteristika pomenutog algoritma je sposobnost da efikasno radi kako u procesu uvećanja, tako i u procesu smanjenja slike, koristeći istu osnovnu metodologiju. Ova fleksibilnost je od suštinskog značaja u situacijama gde je potrebno prilagoditi obradu slike različitim zahtevima i scenarijima. U ovom istraživanju, fokusiramo se na razvoj algoritma za smanjenje slike, ali istovremeno uzimamo u obzir i proces uvećanja slike kao bitan aspekt predstavljene metodologije. Svaki blok u novoj slici dobija se kroz ponderisanu agregaciju intenziteta piksela i njihovih suseda u originalnoj slici. Važno je napomenuti da

[1]Ovo istraživanje je delimično podržao Fond za nauku Republike Srbije, #GRANT br. 7632, projekat "Mathematical Methods in Image Processing under Uncertainty" MaMIPU. Prvi autor se zahvaljuje i Pokrajinskom sekretarijatu za visoko obrazovanje i naučnoistraživački rad (AP Vojvodina, Republika Srbija) na delimičnoj podršci kroz projekat broj 142-451-3188/2023-01: "Primena agregacionih operatora u prepoznavanju lica".

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: nralevic@uns.ac.rs

[3]Poljoprivredni fakultet, Univerzitet u Beogradu, e-mail: julijascekic997@gmail.com

[4]Poljoprivredni fakultet, Univerzitet u Beogradu, e-mail: natasaglisovic@gmail.com

se ovaj pristup implementira u programskom jeziku Python, koristeći odgovarajuće biblioteke, što omogućava jednostavnu primenu u praktičnim scenarijima. Ovaj rad je strukturiran u nekoliko delova, uključujući pregled preliminarnih definicija, metodologiju izgradnje IVFS-a, detaljan opis algoritma za uvećanje slike, eksperimentalne rezultate i zaključke. Detaljnije ćemo opisati razmatrani pristup, prikazati eksperimentalne rezultate i diskutovati o mogućnostima primene u različitim scenarijima.

2. Fazi intervalno vredonosni skupovi

Označimo sa $L([0,1]) = \{\mathbf{x} | \ [\underline{x}, \overline{x}],\ \underline{x}, \overline{x} \in [0,1],\ \underline{x} \leq \overline{x}\}$ skup svih zatvorenih podintervala u $[0,1]$. Relacija $\leq_L$ definisana sa: $\underline{x} \leq \underline{y} \wedge \overline{x} \leq \overline{y}$, je tranzitivna i antisimetrična, i izražava činjenicu da je $\mathbf{x}$ jako povezan sa $\mathbf{y}$.

Lema 2.1. *Algebarska struktura $L([0,1]), \leq_L)$ je kompletna mreža.*

Najmanji element je $0_L = [0,0]$, a najveći je $1_L = [1,1]$.

Definicija 2.2. Intervalni fazi skup A na univerzalnom skupu $U \neq \emptyset$ je preslikavanje $A : U \to L([0,1])$. Sa $IVFS_s(U)$ označavamo skup svih IVFS na U. Slično, $FS_s(U)$ je skup svih fazi skupova na U. Od sada ćemo sa $W([\underline{x}, \overline{x}])$ označavati dužinu intervala $[\underline{x}, \overline{x}]$, odnosno $W([\underline{x}, \overline{x}]) = \overline{x} - \underline{x}$.

Definicija 2.3. Neka je $\alpha \in [0,1]$. Operator $K_\alpha : L([0,1]) \to [0,1]$ se definiše kao konveksna kombinacija njegovih granice argumenta, tj.

$$K_\alpha(\mathbf{x}) = \underline{x} + \alpha(\overline{x} - \underline{x}), \quad \mathbf{x} \in L([0,1]).$$

Lema 2.4. *Za sve $\mathbf{x} \in L([0,1]))$, važi:*
1) $K_0(\mathbf{x}) = \underline{x}$,
2) $K_1(\mathbf{x}) = \overline{x}$,
3) $K_\alpha(\mathbf{x}) = K_\alpha([K_0(\mathbf{x}), K_1(\mathbf{x})]) = K_0(\mathbf{x}) + \alpha(K_1(\mathbf{x}) - K_0(\mathbf{x}))$.

Neka je $A \in IVFS_s(U)$ i $\alpha \in [0,1]$. Označimo sa $K_\alpha(A)$ fazi skup

$$K_\alpha(A) = \{u_i, K_\alpha(A(u_i)) \mid u_i \in U\}.$$

Tvrđenje 2.5. *Za sve $\alpha, \beta \in [0,1]$ i $A, B \in IVFS_s(U)$, važi:*
(a) $\alpha \leq \beta \ \Rightarrow \ K_\alpha(A) \leq K_\beta(A)$; *(b)* $A \leq_L B \ \Rightarrow \ K_\alpha(A) \leq K_\beta(B)$.

Nadalje će biti predstavljen metod konstrukcije elemenata u $L([0,1])$ tako da je njihova dužina unapred fiksirana što je veoma važno jer je poželjno predstaviti varijaciju intenziteta oko svakog piksela po dužini intervala. Pored ovoga varijacija se podešava za faktor skaliranja.

Tvrđenje 2.6. *Za preslikavanje $F : [0,1]^2 \times [0,1] \to L([0,1])$ dato sa*

$$F(x, y, \delta) = [\underline{F}(x, y, \delta), \overline{F}(x, y, \delta)]$$

gde $\underline{F}(x, y, \delta) = x(1 - \delta x)$, $\overline{F}(x, y, \delta) = x(1 - \delta y) + \delta y$, važi:
1) $\underline{F}(x, y, \delta) \leq x \leq \overline{F}(x, y, \delta)$ *za sve* $x \in [0,1]$;

2) $F(x, 0, \delta) = [x, x]$;
3) $F(0, y, \delta) = [0, \delta y]$;
4) $F(x, y, 0) = [x, x]$;
5) $W(F(x, y, \delta)) = \delta y$.
6) Ako je $y_1 \leq y_2$*, onda je* $W(F(x, y_1, \delta)) \leq W(F(x, y_2, \delta))$ *za sve* $x, \delta \in [0, 1]$.

Teorema 2.7. *Neka je* $AF \in FS_s(U)$ *i neka je* $\omega, \delta : U \to [0, 1]$ *biti dva preslikavanja. Onda*

$$A = \{(u_i, A(u_i)) = F(\mu_{A_F}(u_i), \omega(u_i), \delta(u_i)))|u_i \in U\}$$

je intervalno vredonosni fazi skup.

Ako se uzme da je $\delta(u_i) = 1$, sledi $\omega(u_i) = W(F(\mu_{A_F}(u_i), \omega(u_i), 1))$. No, tada je $A = \{(u_i, \mu_{A_F}(u_i)(1 - \omega(u_i)), \mu_{A_F}(u_i)(-\omega(u_i)) + \omega(u_i))|u_i \in U\}$.

Tvrđenje 2.8. *Za sve* $x, y, \delta \in [0, 1]$ *važi* $K_x(F(x, y, \delta)) = x$.

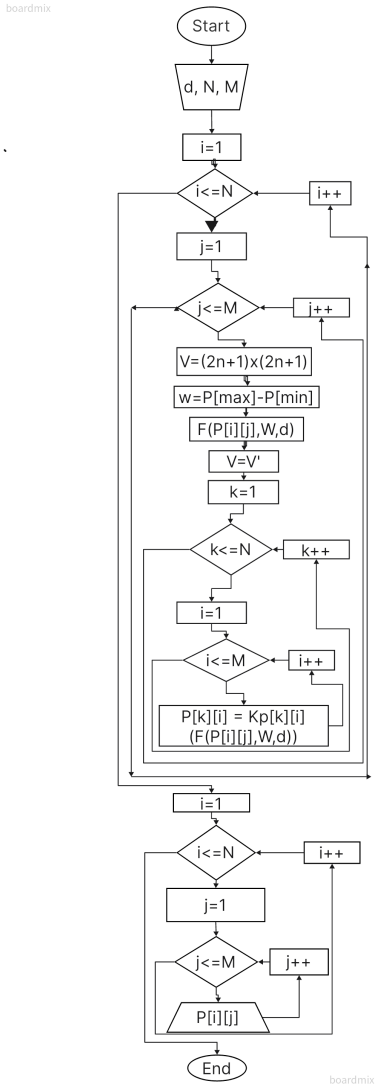
3. Uvećanje slike

U ovom poglavlju je izložen korigovani algoritam za uvećanje slike u sivoj skali koji koristi IVFS i K_α operatore, predložen od strane autora rada [1].

Razmatra se slika od $N \times M$ piksela kao matrica formata $N \times M$, čiji elementi q_{ij} označavaju intenzitete piksela na poziciji (i, j), $i \in \{1, ..., N\}$, $j \in \{1, ..., M\}$. Radi se sa sivim slikama čiji intenziteti uzimaju vrednosti iz intervala $[0, 255]$, odnosno posle normalizovanja imaju vrednosti iz $[0, 1]$. Cilj je da se slika dimenzije $N \times M$ uveća $2n + 1$ puta, odnosno gradi nova slika dimenzije $N' \times M'$ sa $N' = (2n + 1) \cdot N$, $M' = (2n + 1) \cdot M$, za $n \in \mathbb{N} \setminus \{0\}$ sa $2n + 1 \leq N$ i $2n + 1 \leq M$.

- Korak 1. Uzmemo $\delta \in [0, 1]$. Na primer, uzmemo $\delta = 0.5$.
- Korak 2. Za svaki piksel na slici, postavljamo rešetku V sa dimenzijama $(2n + 1) \times (2n + 1)$, tako da je centrirana na tom pikselu. Ova rešetka predstavlja okolinu koju koristimo kada povećavamo veličinu tog piksela sa slike. Intenziteti piksela unutar ove rešetke pružaju informacije koje su nam potrebne za izračunavanje dužine intervala pripadnosti koji se formira kroz funkciju F. Na primer, za piksel sa koordinatama $(4, 4)$, koji je označen tamnijom narandžastom bojom na slici 3, fiksiramo rešetku dimenzija 3×3 oko tog piksela, koja je označena svetlo narandžastom bojom.
- Korak 3. Izračunavamo W kao razliku između najvećeg i najmanjeg intenziteta piksela u V Za piksel (4, 4) izračunavamo W kao:
  $W = \max(0.196, 0.188, 0.184, 0.188, 0.176, 0.176, 0.160, 0.156, 0.149)$
  $- \min(0.196, 0.188, 0.184, 0.188, 0.176, 0.176, 0.160, 0.156, 0.149) = 0.196 - 0.149 = 0.047$
- Korak 4 Konstruišemo interval $F(q_{ij}, W, \delta)$. Svakom pikselu dodeljujemo interval dužine $\delta \cdot W$ koristeći metod objašnjen u prethodnoj sekciji.

$$F(p_{ij}, W, \delta) = [p_{ij}(1 - \delta \cdot W), p_{ij}(1 - \delta \cdot W) + \delta \cdot W].$$



Slika 1: Grafički prizak algoritma IVFS metode

Na primeru, interval koji je dodeljen pikselu (4, 4) je dat sa:
$F(0.176, 0.047, 0.5) = [0.176(1 - 0.0235), 0.176(1 - 0.0235) + 0.0235] = [0.17186, 0.19536]$.

- Korak 5. Konstruišemo blok V' koji je jednak bloku V.
- Korak 6. Izračunavamo $K_{p_{kl}}(F(p_{ij}, W, \delta))$ za svaki piksel. Proširićemo svaki piksel (i, j) sa slike Q preko novog bloka V'. Na primeru, blok V' koji je povezan sa pikselom $(4, 4)$ je proširen.

Da bismo zadržali vrednost originalnog piksela u centru novog bloka, Tvrđenje 2.8 navodi da bi α trebalo da bude jednako intenzitetu tog piksela. U slučaju piksela $(4, 4)$ imamo $0.176 = p'_{22} = K_{p_{44}}([0.17186, 0.19536]) = 0.17186 + p_{44} \cdot 0.047 = 0.180$. Iz 2.8 dobijamo :
$\alpha = p_{33} \Rightarrow p'_{11} = 0.17186 + p_{33} \cdot 0.047 = 0.181072$
$\alpha = p_{34} \Rightarrow p'_{12} = 0.17186 + p_{34} \cdot 0.047 = 0.180696$
$\alpha = p_{55} \Rightarrow p'_{33} = 0.17186 + p_{55} \cdot 0.047 = 0.178863$

Slika 2: Slika na kojoj je vršeno istraživanje

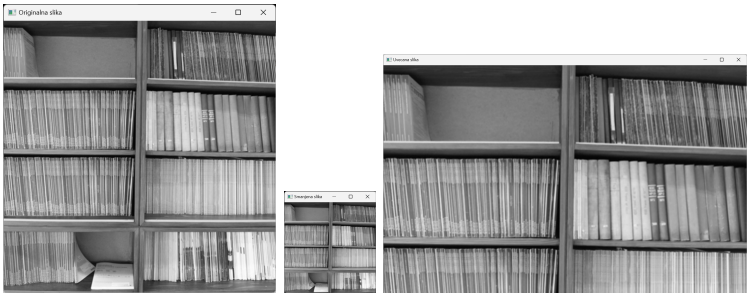
0.180	0.188	0.192	0.188	0.188
0.184	0.188	0.188	0.184	0.192
0.192	0.188	0.196	0.188	0.184
0.184	0.180	0.188	0.176	0.176
0.160	0.160	0.160	0.156	0.149

0.180	0.188	0.192	0.188	0.188
0.184	0.188	0.188	0.184	0.192
0.192	0.188	0.196	0.188	0.184
0.184	0.180	0.188	0.176	0.176
0.160	0.160	0.160	0.156	0.149

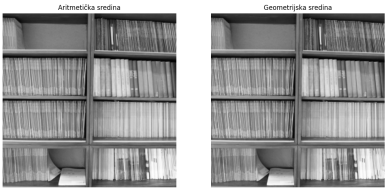
Slika 3: Tabelerni prikaz slike/Prozor na koji se primenjuje metoda

p_{33}	p_{34}			p_{35}
p_{43}	p'_{11}	p'_{12}	p'_{13}	p_{45}
	p'_{21}	$p'_{22}=p_{44}$	p'_{23}	
	p'_{31}	p'_{32}	p'_{33}	
p_{53}	p_{54}			p_{55}

Slika 4: Matrica piksela nakon uvećanja putem metode IVFS



Slika 5: Originalna slika/Smanjena slika/Uvećana slika



Slika 6: Smanjenje slike pomoću aritmetičke i geometrijske sredine

4. Zaključak

U ovom radu je prikazan algoritam putem grafičkog prikaza, a njegov alternativni prikaz možete pronaći u [1] Istraživanje je usredsređeno na istu metodu kao i u prethodnom radu, ali s drugim indeksiranim pikselom i sa drugačijom fiksnom delta vrednošću. Sve to smo realizovali u jednom od popularnih jezika, Pythonu i ilustrovali na slici 2 koja je javno dostupna. Koristeći sliku 1 dat je slikovni prikaz algoritma za uvećanje i smanjenje slike primenom ove metode. Ključna tema istraživanja je smanjenje slike korišćenjem različitih pristupa, a predstavljena je IVFS metoda. Algoritam je funkcionisao na principu izbora prozora sa neparnim brojem piksela, na čijim se vrednostima osvetljenosti primenjuje agregaciona funkcija. Naš fokus je bio na korišćenju nekih poznatih agregacionih funkcija prvenstveno aritmetičke i geometrijske sredine u algoritmu smanjenja slike. Dobijene rezultate smo uporedili i utvrdili da je aritmetička sredina pružila bolji kvalitet smanjenja slike.

Literatura

[1] A. Jurio, D. Paternain, C. Lopez-Molina, H. Bustince, R. Mesiar and G. Beliakov, "A construction method of interval-valued Fuzzy Sets for image processing", 2011 IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems (T2FUZZ), 2011.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

FILTER ZA UKLANJANJE ŠUMA NA DIGITALNOJ SLICI ZASNOVAN NA AGREGACIONIM FUNKCIJAMA[1]

Nebojša Ralević [2] **i Julijana Kapor** [3]

https://doi.org/10.24867/META.2024.23 Original scientific paper

Sažetak. U obradi slika neizbežan je problem pojave šuma u većini digitalnih slika. Najčešći uzroci za to su izvor snimanja, prenos slike ili nešto drugo. Rešenje ovog problema može biti implementacija algoritama koji su otporni na šum. Drugo rešenje je u predobradi slike korišćenjem filtera za uklanjanje šuma, dobijajući sliku bez šuma. Ovi filteri moraju tražiti ravnotežu između uklanjanja zašumljenih piksela i npr. očuvanja ivica i tekstura koji postoje na slici. Tako su poslednjih godina predloženi neki filteri za uklanjanje so i biber šuma kao npr. filter baziran na ponderisanoj aritmetičkoj odnosno na ponderisanoj stepeno-korenskoj sredini kao i fazi matematičkoj morfologiji. U ovom istraživanju predložen je jedan filter sličnog tipa. Takođe se vrši provera performansi predloženog filtera i on se poredi sa drugim odgovarajućim filtrima za različite nivoe šuma.

AMS klasifikacija (2020): 68U10, 94A08

Ključne reči: fazi matematička morfologija, obrada slike, impulsni šum, ponderisana aritmetička sredina, ponderisana stepeno-korenska sredina, uklanjanje šuma

1. Uvod

Čest problem u radu sa digitalnim slikama jeste pojava šuma. U načinu na koji su pikseli slike zašumljeni razlikujemo više tipova šuma. Pikseli mogu biti pogođeni jednakom količinom šuma (aditivni šum) ili u zavisnosti od njihovog intenziteta (multiplikativni šum). Kod svih ovih tipova šuma, svi pikseli na slici zahvaćeni su šumom. Za razliku od njih, kod impulsnog šuma nasumično su pogođeni samo neki pikseli i to određenom vrednosti. U ovom radu razmatraće se so i biber šum, gde su pikseli pogođeni ekstremnim vrednostima koje se pojavljuju kao crne i bele tačkice.

[1]Ovo istraživanje je delimično podržao Fond za nauku Republike Srbije, #GRANT br. 7632, projekat "Mathematical Methods in Image Processing under Uncertainty" MaMIPU i Pokrajinski sekretarijat za visoko obrazovanje i naučnoistraživački rad (AP Vojvodina, Republika Srbija) kroz projekat broj 142-451-3188/2023-01: "Primena agregacionih operatora u prepoznavanju lica".

[2]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: nralevic@uns.ac.rs

[3]Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: julijanakapor142@gmail.com

Osnova za novi algoritam koji će uklanjati ovaj tip šuma jeste fazi matematička morfologija i ponderisana aritmetička sredina. Najpre se detektuju pikseli zahvaćeni šumom a potom se prelazi na filtriranje. U fazi filtriranja rekonstruišu se pogođeni pikseli na osnovu njegovih suseda koji nisu zašumljeni. Svakom susedu dodeljuje se određena vrednost težine koja zavisi od njegove udaljenosti od posmatranog zašumljenig piksela.

Ovaj rad je baziran na radu [1] u kojem je predstavljen AWAM filter. Novina je da se u tom filtru umesto ponderisane aritmetičke sredine koristi ponderisana stepeno-korenska sredina (vidi npr. [2]). Takođe su umesto euklidske metrike u $\mathbb{Z}^2$ razmatrane i druge metrike, između ostalih i fazi metrike (vidi [3]).

2. Morfološke operacije

Morfološke operacije na slici predstavljaju rad sa objektima koji predstavljaju skup tačaka P slike koje imaju neku zajedničku osobinu.

Sem standardnih operacija sa pikselima kao što su: unija, presek, komplement (pozadina objekta P) razlika, translacija ($(P)_q = P+q = \{p+q|p \in P\}$), refleksija ($\hat{P} = \{-p|p \in P\}$), odnosno Minkowski sabiranje ($P \oplus Q = \bigcup_{q\in Q}(P+q)$) i oduzimanje ($P \ominus Q = \bigcap_{q\in Q}(P-q)$), postoje i druge. Navedimo neke.

Definicija 2.1. Dilatacija skupa P skupom Q je skup

$$D(P,Q) = P \oplus Q = \bigcup_{q\in Q}(P+q). \tag{2.1}$$

Skup P predstavlja sliku nad kojom se vrši dilatacija, dok od izgleda skupa Q zavisi priroda specifične dilatacije te se stoga naziva strukturni element.

Operacija dilatacije poseduje osobine: komutativnost ($D(P,Q) = D(Q,P)$), asocijativnost ($D(P,D(Q,R)) = D(D(P,Q),R)$), invarijantnost na translaciju ($D(P,Q+r) = P \oplus (Q+r) = (P \oplus Q) + r = D(P,Q) + r$).

Binarna erozija smanjuje sliku uklanjanjem ivičnih piksela objekta.

Definicija 2.2. Erozija slike P strukturnim elementom skupom Q je skup

$$E(P,Q) = \{r|(Q)_r \subseteq P\}. \tag{2.2}$$

Dakle, erozija slike P strukturnim elementom Q je skup tačaka takvih da se sve tačke strukturnog elementa poklope sa tačkama objekta (svetlim tačkama slike), kad se strukturni element translira za r.

Erozija i dilatacija nisu inverzne operacije (sem u specijalnim sličajevima): $D(E(P,Q),Q) \neq P \neq E(D(P,Q),Q)$. One su na neki način dualne operacije: $D^C(P,Q) = E(P^C,Q)$ i $E^C(P,Q) = D(P^C,Q)$.

Operacija erozije poseduje osobine: antikomutativnost ($E(P,Q) \neq E(Q,P)$), invarijantnost na translaciju ($E(P,Q+r) = P \ominus (Q+r) = (P \ominus Q) + r = E(P,Q) + r$).

Definicija 2.3. Otvaranje je operacija u kojoj eroziju neposredno sledi dilatacija istim strukturnim elementom:

$$O(P,Q) = D(E(P,Q),Q). \tag{2.3}$$

Operacija nastoji da proširi male otvore ili ukloni slabo povezane piksele.

Definicija 2.4. Zatvaranje je operacija u kojoj dilataciju neposredno sledi erozija istim strukturnim elementom:

$$C(P,Q) = E(D(P,Q),Q). \tag{2.4}$$

Zatvaranje popunjava male procepe i izolovane piksele pozadine. Ono ukida i izolovane piksele pozadine.

Definicija 2.5. Fazi ("Gray-scale") dilatacija slike P strukturnim elementom Q je skup

$$(P \oplus Q)(x,y) = \max\{P(x-x', y-y') + Q(x',y') | (x',y') \in D_Q\}). \tag{2.5}$$

gde je D_Q oblast definisanosti strukturnog elementa Q, a podrazumevana vrednost $P(x,y)$ izvan okvira slike iznosi $-\infty$.

Domen D_Q je matrica čiji su elementi 0 i 1, a koja definiše koja od lokacija iz okruženja se upotrebljava pri izračunavanju definisanog maksimuma što je slučaj samo ako se na odgovarajućem mestu nađe vrednost 1. Ovaj postupak se primenjuje na svaki par tačaka domena.

Definicija 2.6. Fazi ("Gray-scale") erozija slike P strukturnim elementom Q je skup

$$(P \ominus Q)(x,y) = \min\{P(x+x', y+y') + Q(x',y') | (x',y') \in D_Q\}), \tag{2.6}$$

gde je D_Q oblast definisanosti strukturnog elementa Q, a podrazumevana vrednost $P(x,y)$ izvan okvira slike iznosi $+\infty$.

Definicija 2.7. Fazi otvaranje je operacija:

$$P \circ Q = (P \ominus Q) \oplus Q = D(E(P,Q),Q). \tag{2.7}$$

Definicija 2.8. Fazi ("Gray-scale") zatvaranje je operacija:

$$P \bullet Q = (P \oplus Q) \ominus Q = E(D(P,Q),Q). \tag{2.8}$$

3. Filtriranje slike - algoritam

3.1. Detekcija zašumljenih piksela

Umesto standardnog načina detekcije korišćenjem "gray-scale" morfoloških operatora, predložena funkcija detekcije šuma poboljšana je uvođenjem fazi matematičkih morfoloških operatora. Cilj je pronaći odgovarajuću kombinaciju fazi otvaranja i zatvaranja za postizanje željenih rezultata.

Algoritam 1 - Faza detekcije

1. Odrediti minimalnu (S_{min}) i maksimalnu (S_{max}) vrednost kvadratnog prozora veličine N sa centrom (i,j).

2. Izračunati

$$d(i,j) = |\frac{C_{T,I_T}(O_{T,I_T}(A,B),B)(i,j) + O_{T,I_T}(C_{T,I_T}(A,B),B)(i,j)}{2} - A(i,j)|,$$

gde je T t-norma, a I_T njena R implikacija.

3. Izračunati

$$b(i,j) = \begin{cases} 255, & A(i,j) \in \{S_{min}, S_{max}\} \text{ i } d(i,j) \geq t \\ 0, & \text{inače} \end{cases},$$

gde je t prag koji mora biti unapred definisan.

4. $A(i,j)$ će se smatrati pikselom sa šumom kada je $b(i,j) = 255$.

3.2. Faza filtriranja bazirana na agregacionim funkcijama

Kao rezultat prethodne faze, detektovani su oštećeni pikseli, odnosno pikseli koji sadrže određenu količinu šuma. Sledi faza filtriranja, čiji je cilj da koristeći informacije iz preostalih piksela, rekonstruiše prethodno detektovane piksele. U fazi filtriranja, dok se krećemo kroz sliku, posmatraćemo susedne piksele. Osnovna zamisao za rekonstrukciju odgovarajućeg oštećenog piksela jeste kombinovanje vrednosti zdravih susednih piksela. Ova zamisao, takođe, glavno je načelo filtera zasnovanih na srednjoj vrednosti (median, mean-based), međutim, u ovom slučaju biće implementirane određene izmene u cilju što boljih rezultata. Za rekonstrukciju piksela potrebno je da u njegovom susedstvu postoji određeni broj zdravih piksela kako bi njegova vrednost mogla da se utvrdi. Što slika sadrži više šuma, utoliko će se smanjiti broj zdravih piksela u susedstvu piksela pogođenih šumom. Iz ovoga proizilazi zaključak da je potrebno da filter uzima u obzir promenu veličine prozora koji obuhvata susedne piksele. U ovom slučaju, razmatraćemo prozore veličine 3×3, 5×5 i 7×7. Veličina ovog prozora može varirati od piksela do piksela na istoj slici, kako za neke piksele će već u prvom (najmanjem) prozoru biti dovoljno susednih piksela za rekonstrukciju, a za neke će biti potrebno povećanje prozora kako bismo imali dovoljan broj susednih piksela. Međutim, ovde je potrebno razmotriti i udaljenost susednih piksela u odnosu na dati piksel koji želimo da rekonstruišemo. Što je veća udaljenost suseda od centralnog piksela, manja je verovatnoća da su oni slični. Naravno, postoje izuzeci za ovu ideju. Pikseli koji su u blizini ivice, a sa njenih različitih strana, imaju veću razliku u vrednostima iako se nalaze blizu jedan drugom. Prethodno navedena ideja o razmatranju udaljenosti susednih piksela može se implementirati korišćenjem agregacione funkcije koja će za svaki piksel iz susedstva uključiti njegovu težinu, koja mu je dodeljena u zavisnosti od udaljenosti tog piksela u odnosu na posmatrani.

Obeležimo sa $p_0 = A(i_0, j_0)$ piksel koji će biti obrađen, a sa $p_k = A(i_k, j_k)$, $k = 1, ..., m_0$ neoštećenih m_0 piksela koji će biti agregirani da bi se rekonstruisao p_0. Može se primetiti da m_0 zavisi od piksela p_0, i traži se da njegova vrednost mora biti veća od n, što je najmanji broj nezašumljenih piksela koji će se posmatrati za odšumljavanje p_0. Koristimo specijalnu agregacionu funkciju, ponderisanu stepeno-korensku sredinu:

$$WAM(p_1, ..., p_{m_0}) = \Big(\sum_{k=1}^{m_0} w_k (p_k)^\kappa \Big)^{1/\kappa}, \tag{3.1}$$

gde su težine date sa

$$w_k = \frac{\alpha_k}{\sum_{r=1}^{m_0} \alpha_r}, \quad k = 1, ..., m_0, \qquad \alpha_r = \frac{1}{\beta^{d((i_r,j_r),(i_0,j_0))}}, \quad r = 1, ..., m_0, \tag{3.2}$$

za odabranu vrednost $\beta > 1$ i rastojanje d definisano na $\mathbb{Z}^2$. Jasno je da što se rastojanje između dostupnog nezašumljenog piksela i piksela koji se trenutno obrađuje poveća, težina neoštećenog piksela se eksponencijalno smanjuje. Dakle, na performanse filtera u velikoj meri utiče n, eksponencijalna baza težina β i razmatrano rastojanje d.

Ukoliko u prvom prolazu imamo dovoljan/minimalan broj zdravih piksela, piksel će biti rekonstruisan računanjem 3.1. U suprotnom, vrednost distance se povećava za 1 i ovaj korak se završava sa dovoljno pronađenih zdravih piksela.

Algoritam 2 - Faza filtriranja

1. Neka je $C_D = \{p_k = A(i_k, j_k) | d((i_k, j_k), (i_0, j_0)) \leq D, b(i_k, j_k) = 0\}$ skup piksela koji su bez šuma u prvoj fazi, a na rastojanju do p_0 ne većem od D.

2. Uzmimo da je $D = 1$.

3. Ako je $card(C_D) < n$, onda je $D = D + 1$ i ovaj korak se ponavlja. U suprotnom, preći na sledeći korak.

4. Izlazna vrednost obrađenog nivoa sivila procesuiranog piksela p_0 je data sa

$$p_0 = WAM(p_1, ..., p_{m_0}),$$

gde je $p_i \in C_D$ za sve $1 \leq i \leq m_0 = card(C_D)$ kroz jednačinu (3.1) sa osnovom β i rastojanjem d.

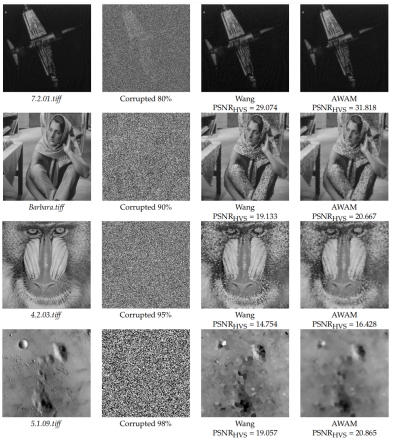
4. Rezultati i zaključci

U cilju komparativne analize, izvršeno je računanje vrednosti mera sličnosti za sve razmatrane filtere na određenom setu slika [5], [6]. Slike su zašumljene so biber šumom, u rasponu vrednosti $5\% - 95\%$ sa korakom 5, kao i vrednošću od 98%. AWAM filter pokazao se kao znatno bolje rešenje i proizveo je bolje rezultate u većini slučajeva. Samo u nekoliko slučajeva (kod šumova vrednosti $5\%, 10\%, 40\%, 50\%, 55\%$) neki od drugih uporednih filtera dali su bolje rezultate. Zaključak je da AWAM filter time značajno pospešuje tretiranje so i biber šuma u odnosu na druge posmatrane filtere.

Osim statističkih rezultata dobijenih merenjem, potrebno je razmotriti i vizuelnu komparaciju rezultata. U slučajevima manje i srednje vrednosti šuma, gde je AWAM filter dao nešto lošije merne rezultate, vizuelni prikaz filtriranja pokazuje veoma slične rezultate kao i kod filtera koji se za te vrednosti šuma pokazao boljim. Vizuelna komparacija rađena je i za veće procente šuma $(80\% - 98\%)$. U tim slučajevima, merenja pokazuju nešto bolje rezultate kod AWAM filtera. Međutim, vizuelnim posmatranjem može se uočiti veći stepen zamagljenja na slikama sa primenjenim AWAM filterom, ali su nivoi intenziteta približniji originalnoj slici te je zbog toga on bolje rešenje.

Na slici 1. prikazani su rezultati primene Wang [4] i AWAM filtera kod većih nivoa šuma, $80\% - 98\%$. AWAM filter daje bolje rešenje i u mernoj i u vizuelnoj komparaciji. Na slikama su jasno vidljivi oštriji i nagli prelazi kod Wang filtera. Kod AWAM filtera se za rekonstrukciju zašumljenog piskela koristi veći broj okolnih piksela u odnosu na Wang filter što je dalo realističnije rezultate. Može se zaključiti da iz vizuelne tačke gledišta, AWAM filter je superiorniji u poređenju sa ostalim filterima.

Navedeni rezultati koji su dobijeni od strane autora rada [1] (slučaj $\kappa = 1$) su i provereni. Testiran je i i naš uopšteni AWAM filter i za druge vrednosti parametra κ, kao i za različite vrednosti metrika. U nekim slučajevima su dobijeni bolji, a u nekim lošiji rezultati.



Slika 1: Slika 1. Rezultati primene Wang i AWAM filtera kod nivoa šuma $80\% - 98\%$. Originalne slike prikazane su u prvoj koloni

Literatura

[1] M. González-Hidalgo, S. Massanet, A. Mir and D. Ruiz-Aguilera "Impulsive Noise Removal with an Adaptive Weighted Arithmetic Mean Operator for Any Noise Density", *Appl. Sci.* Vol. 11, 560, 2021.

[2] Klement, E.; Mesiar, R.; Pap, E. "Triangular Norms", *Kluwer Academic Publishers* London, UK, 2000.

[3] N. M. Ralević, D. Karaklić, N. Pištinjat, "Fuzzy metric and its applications in removing the image noise", *Soft Computing*, Vol. 23, Issue 22, pp. 12049-12061, 2019.

[4] Wang, Y.; Wang, J.; Song, X.; Han, L. An Efficient Adaptive Fuzzy Switching Weighted Mean Filter for Salt-and-Pepper Noise Removal", *IEEE Signal Process. Lett.* Vol. 23, 1582-1586, 2016.

[5] http://sipi.usc.edu/database/misc.tar.gz.

[6] https://www.researchgate.net/figure/Barbara-original-image_fig1_220501736

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

A SHORT NOTE ON THE EXTENDED GEVREY REGULARITY

Filip Tomić [1]

https://doi.org/10.24867/META.2024.24 Review article

Abstract. In this paper we give an overview of the results concerning extended Gevrey classes. They are studied in detail by the author and his collaborators in the series of papers. Here we present an equivalent definition of such classes, recall some of their properties and briefly discuss some applications.

AMS Mathematics Subject Classification (2020): 46E10; 46F05

Key words and phrases: Gevery classes, weight sequences, associated function, ultradifferentable functions

1. Introduction

Classes of ultradifferentiable functions are determined by the sequences of positive numbers M_p, $p \in \mathbb{N}$. They are often called *weight sequences*, and they control the derivatives of the function. By imposing some technical conditions on M_p, one can prove different properties of the corresponding classes such as closedness under pointwise multiplication, closednes under composition or closedness under action of differential operators of infinite order. For instance, *moderate growth* condition

$$M_{p+q} \leq CM_pM_q, \quad C > 0, \quad p, q \in \mathbb{N}, \tag{1.1}$$

is important in contruction of differential oparators of infinite order (ultradifferentiable operators) which acts continuously on such spaces (see [5]). If $M_p = p^{tp}$, $t > 1$, we obtain well known Gevrey spaces (see [10]). For $t = 1$ we have spaces of locally analytic functions.

Extended Gevrey classes $\mathcal{E}_{\tau,\sigma}(U)$ are introduced in [7]. They contain functions whose derivativers are controlled by $M_p^{\tau,\sigma} = p^{\tau p^\sigma}$, $\tau > 0$, $\sigma > 1$. Such sequences do not satisfy the condition (1.1) for any choice of τ and σ, and therefore it is not possible to use arguments from the standard theory to analyze their properties. However, it turns out that the choice of two parameters τ and σ gives an advantage in a topological sense. More precisely, it is possible to define productive and inductive limit topologies on $\bigcap_{\tau>0} \mathcal{E}_{\tau,\sigma}(U)$ and $\bigcup_{\tau>0} \mathcal{E}_{\tau,\sigma}(U)$ prove that such classes have "nice" properties. For detail exposure on extended Gevrey classes we refer to [7, 8, 9, 11, 12].

[1]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: filip.tomic@uns.ac.rs

The aim of this paper is to recall some of the basic properties of $\mathcal{E}_{\tau,\sigma}(U)$. We will also give an equivalent definition of such classes by using products of sequences that increases to infinity. At the end we will discuss some recent applications.

Let us start with preliminary notation.

1.1. Notation

We use the standard notation: $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{C}$, denote sets of positive integers, nonnegative integers, integers, real numbers, positive real numbers and complex numbers, respectively. The length of a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ is denoted by $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$ and $\alpha! := \alpha_1! \cdots \alpha_d!$. For $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ we denote: $|x| := \left(x_1^2 + \ldots + x_d^2\right)^{1/2}$, $x^\alpha := \prod_{j=1}^d x_j^{\alpha_j}$, and $\partial^\alpha = \partial_x^\alpha := \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$, $j = 1, \dots, d$. We write $\lfloor x \rfloor := \max\{m \in \mathbb{N} : m \leq x\}$.

2. Extended Gevrey classes

We start with the properties of weight sequences $M_p^{\tau,\sigma} = p^{\tau p^\sigma}$, $\tau > 0$, $\sigma > 1$, $p \in \mathbb{N}$, which are given in the following Lemma.

Lemma 2.1. *Let $\tau > 0$, $\sigma > 1$, $M_0^{\tau,\sigma} = 1$, and $M_p^{\tau,\sigma} = p^{\tau p^\sigma}$, $p \in \mathbb{N}$. Then the following properties hold:*

$(M.1)$ $(M_p^{\tau,\sigma})^2 \leq M_{p-1}^{\tau,\sigma} M_{p+1}^{\tau,\sigma}, \quad p \in \mathbb{N},$

$\widetilde{(M.2)}$ $M_{p+q}^{\tau,\sigma} \leq C^{p^\sigma + q^\sigma} M_p^{\tau 2^{\sigma-1},\sigma} M_q^{\tau 2^{\sigma-1},\sigma}$, $p, q \in \mathbb{N}_0$, *for some constant $C \geq 1$,*

$\widetilde{(M.2)}'$ $M_{p+1}^{\tau,\sigma} \leq C^{p^\sigma} M_p^{\tau,\sigma}, \quad p \in \mathbb{N}_0,$ *for some constant $C \geq 1$,*

$(M.3)'$ $\displaystyle\sum_{p=1}^{\infty} \frac{M_{p-1}^{\tau,\sigma}}{M_p^{\tau,\sigma}} < \infty.$

Remark 2.1. Note that $M_p^{\tau,1} = p^{\tau p}$ are Gevrey sequences. They satisfy condition $(M.1)$ for any $\tau > 0$, and $(M.3)'$ only for $\tau > 1$. Moreover, conditions $\widetilde{(M.2)}$ and $\widetilde{(M.2)}'$ are classical Komatsu's conditions $(M.2)$ and $(M.2)'$, where $(M.2)$, given by (1.1), is stronger (see [5]).

When $\sigma > 1$ note that in $\widetilde{(M.2)}$ parameter τ increases to $\tau 2^{\sigma-1}$, $\tau > 0$. Therefore, $\widetilde{(M.2)}$ and $\widetilde{(M.2)}'$ are not comparable.

Now we can define extended Gevrey classes.

Definition 2.1. Let $\tau > 0$ and $\sigma > 1$. The extended Gevrey class $\mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ is the set of all $\phi \in C^\infty(\mathbb{R}^d)$ such that for every compact set $K \subset\subset \mathbb{R}^d$ there exists constant $h > 0$ such that

$$|\partial^\alpha \phi(x)| \leq h^{|\alpha|^\sigma + 1} M_{|\alpha|}^{\tau,\sigma}, \quad x \in K, \quad \alpha \in \mathbb{N}_0^d. \tag{2.1}$$

Remark 2.2. Note that $\mathcal{E}_{t,1}(\mathbb{R}^d) := \mathcal{G}(\mathbb{R}^d)$, $t > 1$, and $\mathcal{E}_{1,1}(\mathbb{R}^d) := \mathcal{A}(\mathbb{R}^d)$ are Gevery class of order t and class of locally analytic functions on $\mathbb{R}^d$, respectively. Moreover, $\mathcal{E}_{\tau,1}(\mathbb{R}^d)$, $0 < \tau < 1$ are classes of quasi-analytic functions.

It is also possible to distinguish extended Gevrey classes of *Roumieu* and *Beurling* type. Here we will discuss only Rouimeu case. For details see [7].

Basic properties of classes $\mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ are given in the following proposition. The proof is based on the properties of $M_p^{\tau,\sigma}$ given in Lemma 2.1.

Proposition 2.1. *([7, 8, 9])*

i) *For $\sigma_2 > \sigma_1 > 1$ we have*

$$\mathcal{A}(\mathbb{R}^d) \subset \bigcup_{t>1} \mathcal{G}_t(\mathbb{R}^d) \subset \bigcup_{\tau>0} \mathcal{E}_{\tau,\sigma_1}(\mathbb{R}^d) \subset \bigcap_{\tau>0} \mathcal{E}_{\tau,\sigma_2}(\mathbb{R}^d) \subset C^\infty(\mathbb{R}^d),$$

with strict inclusions.

ii) *$\mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ is closed under the pointwise multiplication.*

iii) *$\mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ is closed under finite order derivation.*

iv) *$\mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ is closed under superposition. More precisely, if F is an entire function on $\mathbb{R}$ and $f(x) \in \mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ then $F(f(x)) \in \mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$.*

v) *Classes $\bigcap_{\tau>0} \mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ and $\bigcup_{\tau>0} \mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ are ultradifferentiable, i.e., for each class there exist a differential operator of infinite order which is continuous.*

In the classical theory (see [5]) expression h^{p^σ} appearing in (2.1) is usually written as h^p. This modification in our case is due to properties $\widetilde{(M.2)}$ and $\widetilde{(M.2)}'$ of $M_p^{\tau,\sigma}$. In the sequel we prove that h^{p^σ} in the definition of $\mathcal{E}_{\tau,\sigma}$ can be replaced by the product of sequences that increases to infinity.

We start with the following lemma.

Lemma 2.2. *Let $\sigma = 2, 3, 4...$ and $a_p > 0$ for $p \in \mathbb{N}$.*

i) *There exists a constant $h > 0$ such that*

$$\sup\left\{\frac{a_p}{h^{p^\sigma}} : p \in \mathbb{N}_0\right\} < \infty \tag{2.2}$$

if and only if

$$\sup\left\{\frac{a_p}{R_{p,\sigma}} : p \in \mathbb{N}_0\right\} < \infty \tag{2.3}$$

for arbitrary positive sequence (r_j) that increase to infinity, where

$$R_{0,\sigma} = 1, \quad R_{p,\sigma} := \prod_{j=1}^{p^\sigma} r_j \quad p \in \mathbb{N}. \tag{2.4}$$

ii) *There exist positive sequence (r_j) that increase to infinity such that*

$$\sup\{R_{p,\sigma} a_p : \quad p \in \mathbb{N}_0\} < \infty, \tag{2.5}$$

where R_p is given by (2.4)*, if and only if*

$$\sup\left\{h^{p^\sigma} a_p : p \in \mathbb{N}_0\right\} < \infty \tag{2.6}$$

for every $h > 0$.

Proof. i) Let $a_p \le Ch^{p^\sigma}$, $p \in \mathbb{N}_0$, for some $C, h > 0$, and let (r_j) be arbitrary sequence of positive numbers that increase to infinity. Choose j_0 such that $\dfrac{h}{r_j} \le 1$, $j \ge j_0$. Then

$$a_p \le Ch^{p^\sigma} = C \prod_{j=1}^{j_0} r_j \frac{h}{r_j} \prod_{j=j_0+1}^{p^\sigma} r_j \frac{h}{r_j} \le CC_1^{j_0} \prod_{j=1}^{p^\sigma} r_j,$$

for large p and suitable $C_1 > 0$. This proves (2.3).

The opposite part we prove by contradiction. Assume that (2.3) holds for arbitrary (r_j), and that (2.2) is violated, i.e., $\sup\left\{\dfrac{a_p}{h^{p^\sigma}} : p \in \mathbb{N}_0\right\} = \infty$ for every $h > 0$. Arguing in a similar way as in [6] (see also [1]), we can choose a subsequence $(a_{p_m})_{m\in\mathbb{N}}$ of $(a_p)_{p\in\mathbb{N}}$ and strictly increasing sequence $(h_m)_{m\in\mathbb{N}}$ such that

$$p_{m+1} > p_m, \quad a_{p_m} > m\, h_m^{p_m^\sigma} \quad m \in \mathbb{N}_0.$$

Note that $p_{m+1}^\sigma > p_m^\sigma$ for all $\sigma = 2, 3, \ldots$.

Let us define a step sequence

$$r_j := h_1,\ 1 \le j \le p_1^\sigma, \quad r_j := \left(\frac{h_m^{p_m^\sigma}}{h_{m-1}^{p_{m-1}^\sigma}}\right)^{\frac{1}{p_m^\sigma - p_{m-1}^\sigma}},\ p_{m-1}^\sigma < j \le p_m^\sigma;$$

where $m = 2, 3, \ldots$. Note that (r_j) increase to ∞, and

$$R_{p_m,\sigma} = \prod_{j=1}^{p_m^\sigma} r_j = \prod_{j=1}^{p_1^\sigma} h_1^{p_1^\sigma} \cdots \prod_{j=p_{m-1}^\sigma+1}^{p_m^\sigma} \left(\frac{h_m^{p_m^\sigma}}{h_{m-1}^{p_{m-1}^\sigma}}\right)^{\frac{1}{p_m^\sigma - p_{m-1}^\sigma}} = h_m^{p^\sigma}, \quad m \in \mathbb{N}_0.$$

Hence we obtain, $\dfrac{a_{p_m}}{R_{p_m,\sigma}} > m$ and therefore (2.3) is not satisfied.

ii) The if part follows simlarly as in i).

Let us assume now that condition (2.6) is satisfied for every $h > 0$. Put $C_h := \sup\left\{h^{p^\sigma} a_p : p \in N_0\right\}$ for $h \ge 1$. We define

$$H_j := \sup\left\{\frac{h^j}{C_h} : h \ge 1\right\}, \quad j \in \mathbb{N}_0.$$

Note that

$$H_{p^\sigma} a_p = \sup\left\{\frac{h^{p^\sigma} a_p}{C_h} : h \ge 1\right\} \le 1.$$

From the arguments of [6, Lemma 3.4] it follows that H_j satisfies $(M.1)$ and H_j/h^j tends to infinity for all $h \ge 1$. Therefore we may choose $r_j = \dfrac{H_j}{H_{j-1}}$, $j \in \mathbb{N}$, to obtain

$$R_{p,\sigma}a_p = \Big(\prod_{j=1}^{p^\sigma} r_j\Big)a_p = H_{p^\sigma}a_p \leq 1,$$

and (2.5) follows.

□

As an easy cosequence of the previous Lemma we obtain the following Theorem.

Theorem 2.1. *Let $\tau > 0$ and $\sigma > 1$. A function $\phi \in C^\infty(\mathbb{R}^d)$ belongs to $\mathcal{E}_{\tau,\sigma}(\mathbb{R}^d)$ if and only if for any compact set K and any sequence (r_j) that increases to infinity there exists $C > 0$ such that*

$$|\partial^\alpha \phi(x)| \leq C R_{p,\sigma} M_{|\alpha|}^{\tau,\sigma}, \quad x \in K,\ \alpha \in \mathbb{N}_0,$$

where $R_{p,\sigma} = \prod_{j=1}^{\lfloor p^\sigma \rfloor} r_j$.

Proof. Without loss of generality we may assume that $h \geq 1$. From the simple inequality $\lfloor p^\sigma \rfloor \leq p^\sigma \leq 2\lfloor p^\sigma \rfloor$, $p \in \mathbb{N}$, it follows that $h^{\lfloor p^\sigma \rfloor} \leq h^{p^\sigma} \leq h^{2\lfloor p^\sigma \rfloor}$. Therefore (2.1) holds if and only if for every compact set $K \subset\subset \mathbb{R}^d$ there exists a constant $h_1 > 0$ such that

$$|\partial^\alpha \phi(x)| \leq h_1^{\lfloor |\alpha|^\sigma \rfloor + 1} M_{|\alpha|}^{\tau,\sigma}, \quad x \in K, \quad \alpha \in \mathbb{N}_0^d.$$

Now the statements follows from part $i)$ of the Lemma (2.2), putting $R_{p,\sigma} = \prod_{j=1}^{\lfloor p^\sigma \rfloor} r_j$ for $\sigma > 1$ instead of $R_{p,\sigma} = \prod_{j=1}^{p^\sigma} r_j$ for $\sigma = 2, 3, 4 \ldots$. □

Remark 2.3. Second part of the Lemma (2.2) is used to prove similar characterization in the Beurling case. We refer to [13] for details.

3. Applications

In [2] spaces $\mathcal{E}_{1,2}$ are used as an appropriate solution spaces for strictly hyperbolic equations with low regularity with respect to time variable. Recently it has been proved that extended Gevrey classes fits in the theory of weight matrix spaces. Namely, it is possible to define classes $\mathcal{E}_{\tau,\sigma}$ using familes of sequences, called *weight matrices*, of the form $\{\tau^{p^\sigma} p^{\tau p^\sigma}\}_{\tau>0}$ and to characterize them by the functions of the Braun-Meise-Taylor type. For the details we refer to see [4, 12]. Finally, in [3] properties of the sequences $M_p^{\tau,\sigma}$ are used to prove the surjectivity of Borel mappings.

Acknowledgement

The author acknowledge financial support of Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, in the frame of Project "Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama", "Improving the teaching process in the English language in fundamental disciplines."

References

[1] R. Carmichael, A. Kaminski, S. Pilipović, *Notes on Boundary Values in Ultradistribution Spaces*, Lecture Notes Series of Seul University, 49, 1999.

[2] M. Cicognani, D. Lorenz, "Strictly hyperbolic equations with coefficients low-regular win time and smooth in space", *J. Pseudo-Differ. Oper. Appl.*, 9, pp. 643–675, 2018.

[3] J. Jiménez-Garrido, A. Lastra, J. Sanz, "Extension Operators for Some Ultraholomorphic Classes Defined by Sequences of Rapid Growth", *Constructive Approximation*, 2023. https://doi.org/10.1007/s00365-023-09663-z

[4] J. Jiménez-Garrido, D.N. Nenning, G. Schindl, "On generalized definitions of ultradifferentiable classes", *J. Math. Anal. Appl.* 526, 127260, 2023.

[5] H. Komatsu, "Ultradistributions, I: Structure theorems and a characterization", *J. Fac. Sci. Univ. Tokyo*, Sect. IA Math., 20 (1), pp. 25–105, 1973.

[6] H. Komatsu, "Ultradistributions, III: Structure theorems and a characterization", *J. Fac. Sci. Univ. Tokyo*, Sect. IA Math., 20 (1), 1973.

[7] S. Pilipović, N.Teofanov, F. Tomić, "On a class of ultradifferentiable functions", *Novi Sad Journal of Mathematics*, 45 (1), pp. 125–142, 2015.

[8] S. Pilipović, N.Teofanov, F. Tomić, "Beyond Gevrey regularity: Superposition and propagation of singularities", *Filomat*, 32 (8), pp. 2763–2782, 2018.

[9] S. Pilipović, N.Teofanov, F. Tomić, "A Paley–Wiener theorem in extended Gevrey regularity", *J. Pseudo-Differ. Oper. Appl.*, 11, pp. 593–612, 2020.

[10] L. Rodino, *Linear Partial Differential Operators in Gevrey Spaces*. World Scientific, 1993.

[11] N.Teofanov, F. Tomić, "Inverse closedness and localization in extended Gevrey regularity", *J. Pseudo-Differ. Oper. Appl.*, 8, pp. 411-421, 2017.

[12] N.Teofanov, F. Tomić, "Extended Gevrey regularity via weight matrices", *Axioms* 2022, 11, 576, 6pp, 2022.

[13] N.Teofanov, F. Tomić, M. Žigić, "An introduction to Extended Gevrey regularity", arXiv:2402.18341, 2024.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

ORTHOGONAL DECOMPOSITION OF STOCHASTIC TRANSPORT EQUATION

Stefan Tošić [1]

https://doi.org/10.24867/META.2024.25 Original scientific paper

Abstract. Uncertainty quantification seeks to provide a quantitative means to understand complex systems that are impacted by parametric uncertainty. The generalized polynomial chaos (gPC), see[3] and [2], as a tool of stochastic Galerkin method, is used for numerical solving of those complex systems which are described by stochastic ordinary differential equations (SODE) and stochastic partial differential equations (SPDE). We take transport equation as an example, and show that the Galerkin procedure results in a system of deterministic partial differential equations whose solving require an additional effort. Moreover, taking as a model various examples of SPDEs, numerical results with appropriate errors and deviations from exact solutions are presented. To show these results in a clearly visible manner, we used a number of figures, obtained using programming language Matlab.

AMS Mathematics Subject Classification (2020): 65C30, 60H10

Key words and phrases: Polynomial chaos expansion, Galerkin method, Stochastic partial differential equations, transport equation

1. General approach

Let us consider general form of prominent transport stochastic partial differential equation, given by

$$\frac{du(x,t,Z)}{dt} = C(Z)\frac{du(x,t,Z)}{dx} + f(x,t,Z), \quad x \in (-1,1), \quad t > 0, \tag{1.1}$$

with initial condition,

$$u(x,0,Z) = u_0(x,Z). \tag{1.2}$$

and boundary conditions

$$\begin{aligned} u(1,t,Z) =& u_+(t,Z), \qquad C(Z) > 0, \\ u(-1,t,Z) =& u_-(t,Z), \qquad C(Z) < 0. \end{aligned} \tag{1.3}$$

It is important for further work to say that x is spatial variable, t is time, and $C(Z) \in L^2(\Omega)$ is an arbitrary function in terms of a random variable Z.

[1]Faculty of Technical Sciences, University of Novi Sad, e-mail: stefan.tosic@uns.ac.rs

By theory of generalized polynomial chaos(gPC) and Galerkin method(see [1] and [4]), approximation of solution of the governing equation has following form,

$$u(x,t,Z) \approx v_N(x,t,Z) = \sum_{k=0}^{P} v_k(x,t)\Phi_k(Z), \qquad P = \binom{N+d}{N}, \tag{1.4}$$

where N is order of gPC expansion, and $\{\Phi_k\}$ is d-variate gPC basis of orthogonal polynomials.
By substituting (1.4) into the governing equation, and taking projection of

$$\frac{dv_N(x,t,Z)}{dt} = C(Z)\frac{dv_N(x,t,Z)}{dx} + f(x,t,Z), \tag{1.5}$$

onto space spanned by $\{\Phi_k(Z)\}$, we get

$$\frac{dv_k(x,t)}{dt} = \sum_{i=0}^{P} a_{ik}\frac{dv_i(x,t)}{dx} + f_k(x,t), \quad k = 0,1,\ldots,P, \tag{1.6}$$

where

$$a_{ik} = \mathbb{E}[c(Z)\Phi_i(Z)\Phi_k(Z)],\ f_k(x,t) = E[f(x,t,Z)\Phi_k(Z)] \quad 0 \le i,k \le P. \tag{1.7}$$

We have obtained a coupled system of $(P+1)\times(P+1)$ deterministic transport equations, whose notation can be simplified into a matrix form, i.e.

$$\frac{d\mathbf{v}(x,t)}{dt} = \mathbf{A}\frac{d\mathbf{v}(x,t)}{dx} + \mathbf{f}, \tag{1.8}$$

where $\mathbf{v} = (v_0, v_1, \ldots, v_P)^T$, $\mathbf{f} = (f_0, f_1, \ldots, f_P)$ and $\mathbf{A}$ is $(P+1)\times(P+1)$ matrix whose entries are a_{ik}. Note, by definition $a_{ik} = a_{ki}$, so that $\mathbf{A}$ is a symmetric matrix, i.e. $\mathbf{A}^T = \mathbf{A}$.

Theorem 1.1. *Consider the deterministic system* (1.8) *where the coefficients are defined in* (1.7). *Then if* $C(z) \ge 0$ *(resp.* $C(z) \le 0$*) for all* $z \in \mathbb{R}$*, then the eigenvalues of* $\boldsymbol{A}$ *are all non-negative (resp. non-positive); if* $C(z)$ *changes sign, then* $\boldsymbol{A}$ *has both positive and negative eigenvalues for sufficiently large* P.

Therefore, the obtained system (1.8) is symmetric, so there exists an orthogonal matrix $\mathbf{Q}$ ($\mathbf{Q}^T = \mathbf{Q}^{-1}$), such that $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \Lambda$ or equivalently $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T$, where Λ is diagonal matrix with eigenvalues of $\mathbf{A}$, i.e.,

$$\Lambda = diag(\lambda_0, \ldots, \lambda_{j_+}, \ldots, \lambda_{j_-}, \ldots, \lambda_N).$$

Entries $\lambda_0, \ldots \lambda_{j_-}$ are booked for negative eigenvalues and $\lambda_{j_+}, \ldots, \lambda_N$ are for positive, the rest, if they exist, are for zeros. If we multiply (1.8) with $\mathbf{Q}^T$ by left side, and denote $\mathbf{r} = (r_0, r_1, \ldots, r_P) = \mathbf{Q}^T\mathbf{v}$, $\hat{\mathbf{f}} = (\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_P) = \mathbf{Q}^T\mathbf{f}$, we obtain following diagonal system

$$\frac{d\mathbf{r}(x,t)}{dt} = \Lambda\frac{d\mathbf{r}(x,t)}{dx} + \hat{\mathbf{f}}, \tag{1.9}$$

with initial condition,

$$\mathbf{r}(x,0) = \mathbf{Q}^T v(x,0). \tag{1.10}$$

In order to impose boundary conditions to diagonal deterministic system (1.9)-(1.10) we ought to use sign of eigenvalues, i.e.,

$$\begin{aligned} r_j(1,t) &= \sum_{k=0}^{P} q_{kj} v_k(1,t), \qquad j = 0, \dots, j_-, \\ r_j(-1,t) &= \sum_{k=0}^{P} q_{kj} v_k(-1,t), \qquad j = j_+, \dots, P, \end{aligned} \tag{1.11}$$

where q_{jk} are entries of matrix $\mathbf{Q}$. Coefficients $v_k(1,t)$ are obtained using gPC expansion of functions $u_+(t,Z)$, and $u_-(t,Z)$,

$$\begin{aligned} u_+(t,Z) &= \sum_{k=0}^{P} v_k(1,t)\Phi_k(Z), \\ u_-(t,Z) &= \sum_{k=0}^{P} v_k(-1,t)\Phi_k(Z). \end{aligned} \tag{1.12}$$

2. Numerical simulations

Example 2.1. Now, we present some numerical examples to support the elaborated theory. We are going to use different random variables, as well as initial and boundary conditions. First, we consider

$$\begin{aligned} \frac{du(x,t,Z)}{dt} &= Z\frac{du(x,t,Z)}{dx}, \qquad -1 < x < 1,\ t \geq 0, \\ u(x,0,Z) &= \cos x, \qquad -1 < x < 1, \end{aligned} \tag{2.1}$$

where $Z \sim \mathcal{U}(-1,1)$ is uniformly distributed random variable, with PDF $\rho(z) = 1/2$ and suitable Legendre polynomials. The exact solution is

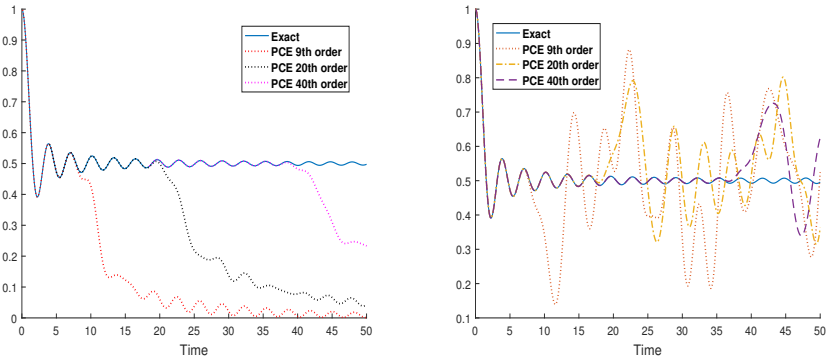
$$u_{exact} = \cos(x - Zt).$$

We suppose that solution of (2.1) has form

$$u_{approx}(x,t,Z) = \sum_{k=0}^{N} u_k(x,t)\Phi_k(Z).$$

There are two ways to obtain coefficients $u_k(x,t)$, namely, first one is to use exact solution u_{exact} where in every step we have to find expectation $u_k(x,t) = E[u_{exact}\Phi_k]$. Second way is to use described gPC Galerkin method and obtain system of deterministic equations, whose solutions are $u_k(x,t)$.
Figure 1, shows these mentioned methods. Namely, on both graphs we see that for fixed $x = x_0$, obtained approximations are accurate up to some point in time. Obviously, it depends on height of gPC order which is employed. As the order of polynomial chaos increases, accuracy in time is longer and better. Nevertheless, graphs are not the same, on the left one we see that after losing accuracy solutions converge to zero, whereas, on the right one, solutions continue to diverge. Obviously, the first method is inapplicable when the exact solution is unknown.
Accuracy will be measured by the mean-square error

$$e_m(N,t) = \max_x (E[(u_{exact} - u_{approx})^2])^{1/2},$$



(a) Direct method (b) Galerkin method

Figure 1: Solution of PDE (2.1) for $x = 0$, with $Z \sim \mathcal{U}(-1, 1)$

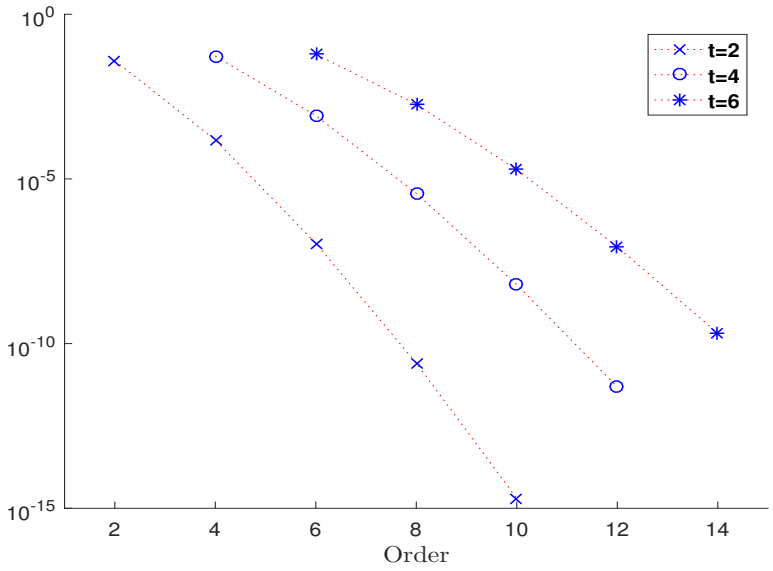
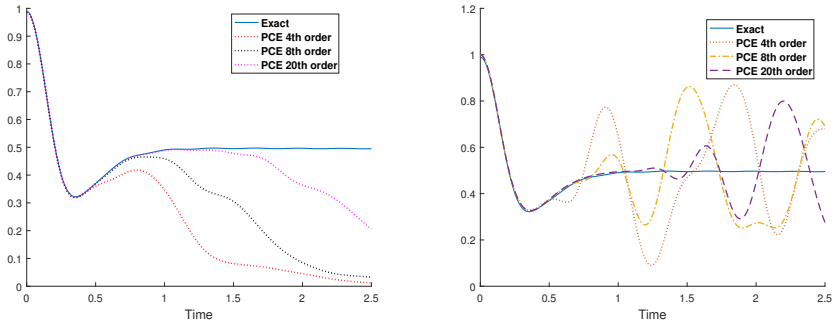


Figure 2: Mean-square error for $Z \sim \mathcal{U}(-1, 1)$ and different t

and results which are presented in Figure 2, shows exponential convergence of error in each time moment.

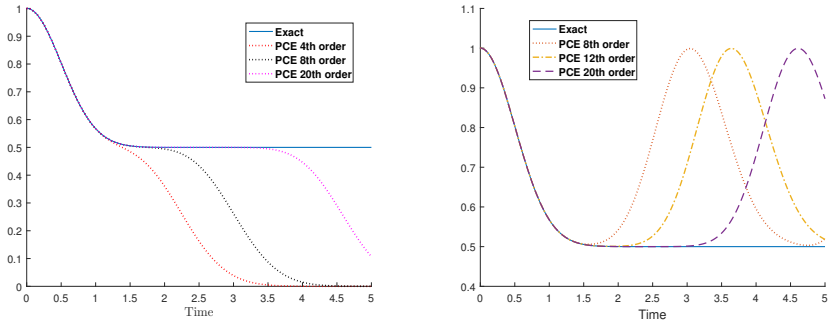
Also, we consider equation (2.1) for gamma and normal distributions. The obtained numerical results are presented in Figures 3 and 4. Conclusions for these two distributions are same as in previous case, but we notice that higher orders are required for the same length in time as were needed for uniform distribution.

Example 2.2. Now, let us consider following example with boundary conditions,



(a) Direct method (b) Galerkin method

Figure 3: Solution of PDE (2.1), for $x = 0$, with $Z \sim \Gamma(3, 1)$



(a) Direct method (b) Galerkin method

Figure 4: Solution of PDE (2.1), for $x = 0$, with $Z \sim \mathcal{N}(0, 1)$

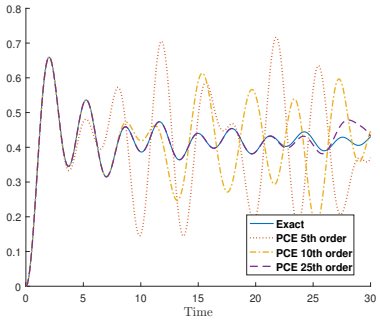
and discontinuity in random space.

$$
\begin{aligned}
u_t(x,t,Z) &= C(Z)u_x(x,t,Z), && -1 \le x \le 1,\ t > 0 \\
u(x,0,Z) &= kZ\sin(kx), && -1 \le x \le 1,\ Z > 0 \\
u(x,0,Z) &= 2kZ\sin(2kx), && -1 \le x \le 1,\ Z < 0.
\end{aligned} \tag{2.2}
$$

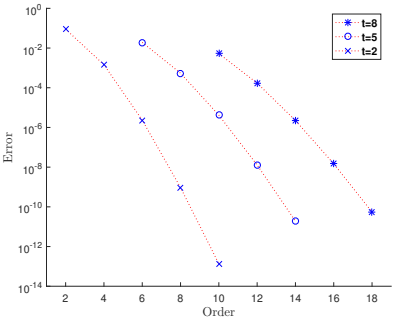
Here, $C(Z) = \sigma Z$, where $0 < \sigma < 1$, controlling the variability of random input and $k > 0$ is a real constant. We give boundary conditions as

$$
\begin{aligned}
u(1,t,Z) &= kZ\sin(k(1 + C(Z)t)), && Z > 0, \\
u(-1,t,Z) &= 2kZ\sin(2k(-1 + C(Z)t)), && Z < 0.
\end{aligned} \tag{2.3}
$$

Exact solution of (2.2)-(2.3) is $u_e = \sin(k(x + C(Z)t))$ for $Z > 0$, and $u_e = \sin(2k(x + C(Z)t))$ for $Z < 0$. The numerical solutions are solved with $\sigma = 0.5$ and $k = 1$. The boundary conditions are implemented via the eigenvalue analysis explained in theoretical part of section. We can see exponential convergence of the mean-square error from 10^0 to 10^{-12} in only few steps, in Figure 5b. Obviously, discontinuity in random space is not a problem for Galerkin method, because we obtain continuous functions in spatial and time domain.



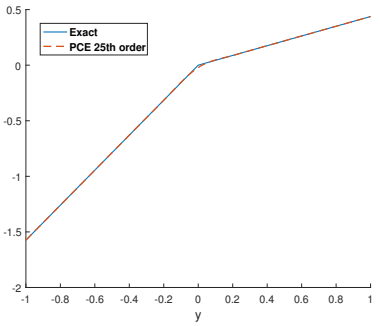
(a) Solutions for different orders of PCE



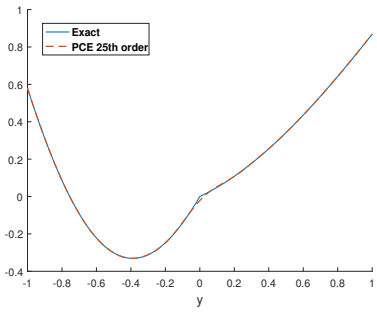
(b) Error for different values of t

Figure 5: Numerical solution of (2.2)-(2.3), for $x = 0$, with $Z \sim \mathcal{U}(-1, 1)$

Nevertheless, we notice in Figure 6 that the exact solution has some kind of discontinuity around point $y = 0$, where approximation has oscillations which are consequence of Gibbs phenomena. Problem (2.2)-(2.3) can be solved by taking other



(a)



(b)

Figure 6: Approximation of $u(x, t, Z)$ for $x = 0.453$ and (a) $t = 0$ and (b) $t = 1.2$

types of distribution for random input Z. Such an extension is more-less straightforward, because all we need is to use suitable type of gPC basis of polynomials, and follow the described steps.

3. Acknowledgments

The author gratefully acknowledge the financial support of the Department of fundamental sciences, Faculty of Technical Sciences, University of Novi Sad through project "Improving the teaching process in the English language in fundamental disciplines".

References

[1] R. G. Ghanem and P. D. Spanos, *Stochastic finite elements: a spectral approach,* Courier Corporation, 2003.

[2] G.E. Karniadakis and D. Xiu, "The Wiener-Askey polynomial chaos for stochastic differential equations", *SIAM J. Sci. Comput.*, 24(2) pp. 619-644, 2002.

[3] N. Wiener, "The homogeneous chaos", *Amer. J. Math.*, 60, pp. 897-936, 1938.

[4] D. Xiu, *Numerical methods for stochastic computations,* Princeton University Press, 2010.

The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

BASIC COMPUTATIONAL GEOMETRY APPLICATIONS IN COMPUTER GRAPHICS [1]

Daria Varga [2], Lana Dujmović[2], Isidora Đurić[2], Lidija Krstanović[2], Danijela Milekić[2] and Ana Perišić[2]

https://doi.org/10.24867/META.2024.26 Professional paper

Abstract. This paper explores fundamental applications of computational geometry in computer graphics. It demonstrates the implementation and visualization of examples including the art gallery problem, utilizing Catalan numbers, as well as applications of 2D and 3D convex hulls. The demonstrations are carried out using the Python programming language in conjunction with Blender 3D modeling software. The paper illustrates the broad utility of computational geometry concepts in computer graphics, such as automatically extracting parts of 2D images or calculating convex hulls for various types of 3D objects.

AMS Mathematics Subject Classification (2020): 68U05

Key words and phrases: computational geometry, convex hull, The Art Gallery, Catalan number, Python

1. Introduction

Computer geometry is one of the branches of computer science, and as such it deals with research of algorithms for solving geometric problems. It is applied in many fields such as mathematics, robotics, statistics and computer graphics. The basic idea is to provide output based on input information such as a set of points, a set of lines or vertices of a polygon. Inputs defined in different ways can provide 2D or 3D output. The problems that computational geometry deals with are found in the answers to questions about objects: do they intersect with each other, how can their cross-section be calculated, is the

[1]This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451 − 03 − 65/2024 − 03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project Improving the teaching process in the English language in fundamental disciplines (No. 01-3394/1), by the Science Fund of the Republic of Serbia, #7449, Multimodal multilingual human-machine speech communication – AI-SPEAK, and by the Fund of Autonomous Province of Vojvodina for the project: Application of modern methods of computer graphics and simulation for the development and 3D printing of a tactile map of the Campus, University of Novi Sad(No.000661458/2024/09318/004/000/000/001/04/003)

[2]Department of Fundamental Sciences, Faculty of Technical Sciences, University of Novi Sad, e-mail: varga.daria01@gmail.com / lana.dujmovic.ns@gmail.com / isidoradjuric@uns.ac.rs / lidijakrstanovic@uns.ac.rs / milekicdanijela30@gmail.com / anaperisic@uns.ac.rs

cross-section convex. From such defined questions arise other problems such as defining new geometric objects based on some input data and similar [1,2]. This paper explores basic computational geometry applications in computer graphics, including the Catalan number and convex hull applications in both two-dimensional (2D) and three-dimensional (3D) space, along with a visualization of the art gallery problem. The implementation and visualization of the aforementioned examples are demonstrated using the Python programming language and 3D modeling software (Blender).

2. Basic computational geometry problems

For further understanding of this paper, it will be necessary to define the problems that will be analysed. A brief overview of all important concepts will be given in this chapter.

2.1. The art gallery problem

In problems handling polygons, oftentimes it is necessary to subdivide it in some way. Seeing as the polygons we will be focusing on in this paper can also be classified as straight line planar graphs (a graph that can be embedded in the plane without crossings), we can define triangulation as a planar subdivision if all its bounded regions are triangles [6]. Out of many algorithms devised for triangulation, the Delaunay triangulation was chosen in this paper due to the fact that it can be found in optimal time [6]. The art gallery problem is one that seeks to find the minimum number of guards necessary to protect a museum of a simple polygonal shape. First proposed in 1973 by mathematician Victor Klee to Václav Chvátal, this problem found its solution two years later, becoming known as Chvátal's Art Gallery Theorem [3]. Chvátal's proof, originally complex, was later simplified by Steve Fisk through a 3-coloring visual approach [3], which was used in this paper. It consists of three steps: triangulation, 3-colouring of the triangulation graph and counting the number of vertices of each colour and selecting the least frequently used. By placing a guard in the vertex of a polygon, he is able to cover all triangles that the vertex belongs to [7]. A visualization of these three steps will be displayed in chapter 3.1.

2.2. 2D and 3D Convex hulls

By definition, a convex hull is the smallest convex set containing a set of points. Many algorithms are used to create convex hulls in 2D, and some of them are: Wrapping or Jarvis's Algorithm, Graham Scan, Divide and Conquer [2]. In this paper, the QuickHull algorithm is used which is a representation of a Divide and Conquer algorithm. Divide and Conquer algorithm begins by sorting the points by their x-coordinate. Then it partitions the point set into two sets A and B, where A consists of half the points with the lowest x-coordinates and B consists of half of the points with the highest x-coordinates. With created sets the two separate convex hulls are created. After that, the two hulls are merged into a common convex hull, by computing upper and lower tangents for the hulls and discarding all the points lying between these

two tangents [5]. The convex hull of a set P of n points in 3D space is a convex polytope. Polytope is a geometric object with flat faces [1]. The algorithms used to create convex hulls in 3D are: Gift Wrapping in 3D, 3D Divide and Conquer Merge and Randomized Incremental. In fact, they represent a direct extension of the algorithms used in 2D. 3D Divide and Conquer Merge has the same idea as 2D algorithm. The only difference is the merge step. The algorithm finds the tangent of two already created convex hulls and the edge of one of them connected to the tangent. Based on these inputs it creates triangles and repeats this process until convex hulls are joint.

3. Applications

In this chapter we present the visualization of the aforementioned concepts. The implementation and visualization are demonstrated using the Python programming language NumPy library, as well as Blender modeling software. NumPy library is employed for working with arrays, linear algebra, and matrices. The OpenCV library was also used. Google Colab was used as a workspace, enabling the writing and executing python code within browsers, without the need for any computer configurations. It also offers free GPU access and easy sharing.

3.1. The art gallery problem and Catalan numbers visualization

The visualization of the art gallery problem and the occurrence of Catalan numbers in triangulation is realized by applying the three steps of Chvátal's proof for the art gallery theorem to a user-selected simple polygon and calculating all possible triangulations of the polygon using the definition of Catalan numbers. Chvátal's Art Gallery Theorem states that [n/3] guards with 360° vision, with n being the number of walls, are always sufficient and sometimes necessary to guard a simple polygonal art gallery.

Catalan numbers are a sequence of integers that have interesting appearances in many areas of mathematics. Relevant to the topic of this paper, Catalan numbers occur in Euler's triangulation problem. The problem entails finding the number of ways the interior of a convex n-gon can be divided into triangles by drawing nonintersecting diagonals, where $n \geq 3$ [4].

In order to acquire a polygon suitable for this task, PolygonSelector class was used, after which we ensured the resulting vertices create a closed, simple polygon. The vertices of a successfully selected polygon are then passed to the Polygon class which is then displayed, along with the vertices (Figure 1a). This allows for the first of the three steps of the art gallery proof to be performed, which is triangulation. Due to the fact that the user is allowed to select a polygon that is not convex, the chosen triangulation method was the constrained Delaunay triangulation. The following step included using the triangulation information to perform the 3-colouring of the vertices. This means that every vertex of a given triangle will have a different value (colour) assigned to it. The results of the 3-colouring implementation used were then displayed, along with the constrained Delaunay triangulation (Figure 1b). To complete this visualization, checkboxes for displaying different sets of 3-coloured vertices

by their colours were added, completing the third step of the art gallery proof, choosing the smallest set of vertices. The resulting number of sufficient guards for the given polygon is then displayed, along with the Catalan number for the same polygon, showing how many different triangulations of the polygon exist. This was calculated using the definition of Catalan numbers (Figure 1c).

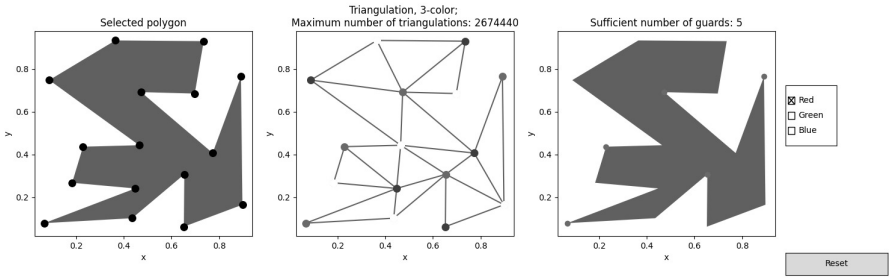


Figure 1: (a) The selected polygon; (b) triangulation and 3-colouring of the vertices, (c) display of the selected set of vertices

3.2. Convex hull in 2D

The practical application of the convex hulls in 2D is realized by finding convex hulls based on binary images. This technique was used on the example of the world map with highlighted areas with lower and higher population density. The first step is loading the images and converting them to gray color space. Images were downloaded from the internet [3] and modified within the Adobe Photoshop program, so that population densities are sufficiently highlighted (Figure 2a). The next step is the application of the blur function, in order to get rid of unnecessary information for the sake of easier calculation of the contours themselves. With the appropriate settings for the function blur, we get separated parts of higher and lower population density (Figure 2b).

Figure 2: (a) Loaded image with highlighted higher and lower population density converted to Gray color space; (b) applied blur function.

Then the contours of the obtained polygonal surfaces were extracted using the function findContours. Contours were accessed with a for loop and a convex hull was defined and assigned for each of the contours of the image using the

[3]Our World in Data, Population density (2024),https://ourworldindata.org/grapher/population-density

function ConvexHull. The procedure was repeated for each image separately. The resulting contours and convex hull are drawn on one of the initial maps with applied blur, using the drawContours function (Figure 3).

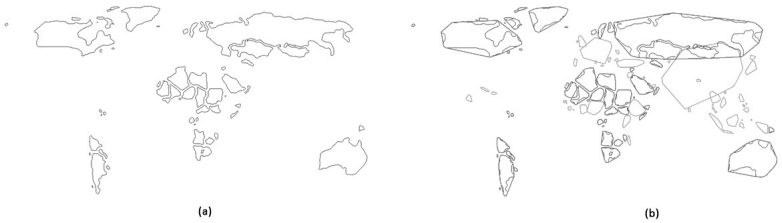


Figure 3: (a) Contours and (b) their convex hulls

3.3. Convex hull in 3D

The generation of the 3D convex hull using Python was demonstrated on an imported 3D mesh. The libraries needed for array operations, 3D plotting and reading mesh files were imported. The function ConvexHull from spatial module of scipy was used to compute the convex hull of a set of points. Before displaying it, the aspect ratio of the plot is set to be equal to ensure that the scale is uniform along each axis. Figure 4 shows the step by step visualization of the 3D convex hull constructed for the imported mesh.

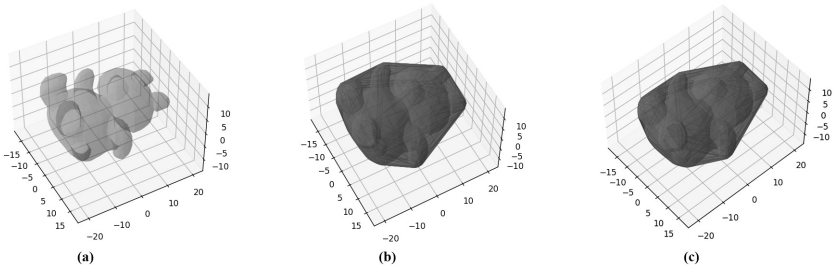


Figure 4: (a) Plotted mesh; (b) Created Convex hull, plotted simplexes and faces of the hull; (c) Equal aspect ratio of the plot

The method of finding convex hulls in 3D also finds its application in software for 3D modeling, such as Blender. Scripts in Blender are written in the Python programming language, so the example presented bellow is based on the already described processes. For an input set of points a predefined function determines a convex hull. The difference is that instead of the libraries loaded so far, the Blender modules Bpy and BMesh are used. Bpy package provides Blender as a Python module for use in studio pipelines, web services, scientific research, and more. BMesh module provides access to Blenders BMesh data structures, which contains information about UVs, vertex color, edge crease,

etc [4]. Two 3D models were created, one that accepts all the vertices of the loaded object as input points (Figure 5a), while the other allows users to select the vertices of a part of the object in Edit Mode and creates a convex hull for them only (Figure 5b).

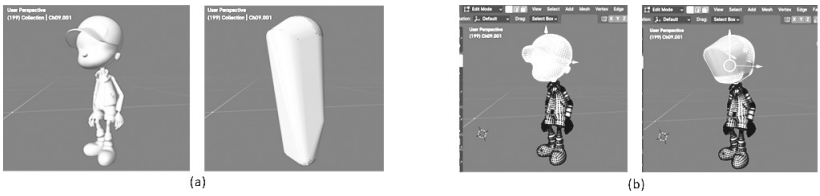


Figure 5: (a) Convex hull of the entire object and (b) of the selected part

4. Conclusion

In this paper, fundamental computational geometry problems are analyzed. The implementation and visualization of examples, such as the art gallery problem utilizing Catalan numbers, as well as applications of 2D and 3D convex hulls, are demonstrated using the Python programming language and Blender, 3D modeling software. The field of computational geometry has wide applications, and the examples created as part of this research can be used for various purposes. For instance, the method of extracting regions of the world with different population densities from images can be applied to other forms of input data. Additionally, convex hulls in Blender can be improved to facilitate calculations for intersections or collisions between objects in computer animation.

References

[1] M. Berg, O. Cheong, M. Kreveld, M. Overmars, Computational Geometry: Algorithms and Applications (2005)

[2] M. Errazki (Medium, 2022), Computing the Convex Hull in Python, accessed in 2024

[3] O. Joseph, Art gallery theorems and algorithms, Vol. 57. Oxford: Oxford University Press, (1987)

[4] T. Koshy, Catalan Numbers with Applications, Oxford University Press (2008)

[5] D. M. Mount, Department of Computer Science, University of Maryland, College Park, Computational geometry (2012)

[4]Blender Authors, Blender 4.1 Python API Documentation, https://docs.blender.org/api/current/index.html

[6] F. P. Preparata, M. I. Shamos, Computational geometry: an introduction, Springer Science & Business Media (2012)

[7] N. Petruzelli, How to Guard an Art Gallery: A Simple Mathematical Problem, The Review: A Journal of Undergraduate Student Research vol. 23 (2022)

CIP - Каталогизација у публикацији
Библиотеке Матице српске, Нови Сад

51:62(082)

CONFERENCE on Mathematics in Engineering: Theory and Applications (9 ; 2024 ; Novi Sad)
Conference proceedings [Elektronski izvor] / The 9th Conference on Mathematics in Engineering: Theory and Applications, META 2024, May 31- June 2, 2024, Novi Sad, Serbia. - Novi Sad : Faculty of Technical Sciences, 2024. - (Mathematics in Engineering: Theory and Applications (Online), ISSN 3042-0466)

Način pristupa (URL): https://imft.ftn.uns.ac.rs/meta-mathematics-in-engineering-theory-and-applications/. - Opis zasnovan na stanju na dan 12.6.2024. - Radovi na engl. i srp. jeziku. - Bibliografija uz svaki rad.

ISBN 978-86-6022-662-6

a) Математика -- Примена -- Инжењерство -- Зборници

COBISS.SR-ID 146984457