УНИВЕРЗИТЕТ У БЕОГРАДУ

МАТЕМАТИЧКИ ФАКУЛТЕТ

# ЗБОРНИК РАДОВА

VI СИМПОЗИЈУМ „МАТЕМАТИКА И ПРИМЕНЕ”

16. и 17. октобар 2015.

УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

ЗБОРНИК РАДОВА – VI СИМПОЗИЈУМ „МАТЕМАТИКА И ПРИМЕНЕ”

16. и 17. октобар 2015.

*Издавач:*
Универзитет у Београду
Математички факултет

*За издавача:*
проф. др Зоран Ракић, декан

*Главни и одговорни уредник:*
проф. др Миодраг Матељевић

*Уредник:*
доц. др Миљан Кнежевић

*Припрема за штампу:*
Александра Делић
Марек Светлик

*Илустација на корицама:*
Славиша Радовић

*Штампа:*
Донат Граф

*Тираж:*
100 примерака

CIP - Каталогизација у публикацији –
Народна библиотека Србије, Београд

51-7(082)
371.3::51(082)

СИМПОЗИЈУМ Математика и примене (6 ; 2015 ; Београд)
Зборник радова / VI симпозијум Математика и примене, 16. и 17. октобар 2015. ; [организатори] Универзитет у Београду, Математички факултет [и Српска академија наука и уметности] ; [уредник Миљан Кнежевић]. - Београд : Универзитет, Математички факултет, 2016 (Београд : Донат Граф). - 125 стр. : илустр. ; 25 cm

Радови на срп. и енгл. језику. - Текст лат. и ћир.
- Тираж 100. - Библиографија уз сваки рад.

ISBN 978-86-7589-112-3

1. Математички факултет (Београд)
a) Математика - Зборници b) Математика - Настава - Зборници

COBISS.SR-ID 226973452

# П Р Е Д Г О В О Р

Шести симпозијум „МАТЕМАТИКА И ПРИМЕНЕ”, национални скуп са међународним учешћем, одржан је 16. и 17. октобра 2015. године у организацији Математичког факултета Универзитета у Београду и Српске академије наука и уметности. Скуп је одржан уз подршку Министарства просвете, науке и технолошког развоја Републике Србије.

Програм Симпозијума се одвијао у три паралелне секције:

- Математика и примене - данас
- Математика и информатика у образовању
- Научноистраживачки и стручни рад студената

Отварање Симпозијума одржано је у петак, 16. октобра у свечаној сали САНУ. На отварању Симпозијума, учеснике и госте су поздравили академик Драгош Цветковић, секретар одељења САНУ за математику, физику и геонауке и Миодраг Матељевић, дописни члан САНУ и председник Програмског одбора Симпозијума. Након отварања, одржана су два пленарна предавања, а затим је рад на Симпозијуму настављен по секцијама.

Излагања у секцији „Математика и примене - данас” била су посвећена актуелним темама у применама математике у различитим областима, новим правцима у истраживањима и постигнутим резултатима. У оквиру секције „Математика и информатика у образовању”, предавачи су скренули пажњу на актуелне проблеме у настави математике и информатике и предложили неке идеје за решавање тих проблема. Традиционално, секција „Математика и информатика у образовању” је окупила многе наставнике математике и информатике из основних и средњих школа, који су активно учествовали у дискусији поводом различитих тема које се тичу процеса учења, наставе, мотивације ученика, популаризације математике, итд. Трећа секција била је посвећена научноистраживачком и стручном раду студената са свих нивоа студија. Студенти неколико факултета су у оквиру ове секције представили своје научне и стручне радове, као и резултате пројеката на којима учествују.

Шестом симпозијуму „Математика и примене” је присуствовало око 200 учесника и гостију из земље и иностранства. Кроз секције Симпозијума, своје резултате представило је око 95 истраживача из реномираних научноистраживачких институција из земље и иностранства. У секцији Математика и информатика у образовању” активно је учествовало око 100 професора математике и информатике из основних и средњих школа широм Србије. Са задовољством можемо констатовати да су испуњени главни циљеви скупа: сагледавање постојећих и отварање нових могућности примене математике у различитим областима, унапређивање наставе математике и рачунарства, активно учешће студената у научним и стручним активностима.

Захваљујемо свим учесницима на успешној реализацији скупа и постигнутим резултатима и унапред се радујемо VII Симпозијуму „Математика и примене”, који ће се одржати 4. и 5. новембра у организацији Математичког факултета Универзитета у Београду и Српске академије наука и уметности.

Програмски одбор VI Симпозијума:

- **проф. др Миодраг Матељевић**,
  дописни члан САНУ, редовни професор Математичког факултета Универзитета у Београду - председник одбора,
- **проф. др Зоран Ракић**,
  Универзитет у Београду, декан Математичког факултета,
- **проф. др Градимир Миловановић**,
  академик САНУ,
- **проф. др Синиша Врећица**,
  Универзитет у Београду, редовни професор Математичког факултета,
- **проф. др Зоран Петровић**,
  Универзитет у Београду, ванредни професор Математичког факултета,
- **проф. др Зорица Станимировић**,
  Универзитет у Београду, ванредни професор Математичког факултета,
- **проф. др Мирослав Марић**,
  Универзитет у Београду, ванредни професор Математичког факултета,
- **доц. др Драгана Илић**,
  Универзитет у Београду, доцент Математичког факултета

Организациони одбор VI Симпозијума:

- **доц. др Миљан Кнежевић**,
  Универзитет у Београду, доцент Математичког факултета - председник одбора,
- **проф. др Зорица Станимировић**,
  Универзитет у Београду, ванредни професор Математичког факултета,
- **Марек Светлик**,
  Универзитет у Београду, асистент Математичког факултета,
- **Ђорђе Стакић**,
  Универзитет у Београду, Рачунарска лабораторија Математичког факултета,
- **Сања Косановић**,
  Универзитет у Београду, менаџер за односе са јавношћу Математичког факултета,
- **Божидар Радивојевић**,
  Универзитет у Београду, студент мастер студија Математичког факултета,
- **доц. др Александра Делић**,
  Универзитет у Београду, доцент Математичког факултета,
- **Душко Вишић**,
  Универзитет у Београду, Рачунарска лабораторија Математичког факултета

У Београду, октобар 2016.

# САДРЖАЈ

Symposium MATHEMATICS AND APPLICATIONS, Faculty of Mathematics, University of Belgrade, 2015 ,Vol. VI(1)

# Hyperbolic geometry and Schwarz lemma

**Miodrag Mateljević**

*Faculty of Mathematics, University of Belgrade*
*e-mail: miodrag@matf.bg.ac.rs*

## 1. Hyperbolic geometry

The "flat" geometry of everyday intuition is called Euclidean geometry (or parabolic geometry), and the non-Euclidean geometries are called hyperbolic geometry (or Lobachevsky-Bolyai-Gauss geometry). The definition of parallel lines (in both Euclidean and hyperbolic geometry) is: Parallel lines are infinite lines in the same plane that do not intersect.

In Euclidean geometry, we can use this definition to prove the theorem that "parallel lines are equidistant along their length". When students are asked to prove this theorem, they often complain: "It is obvious, I can see that they are equidistant - what are you asking me to do?" In elementary and high school we accept some concepts (when we were very young) and the statements and later it is difficult to change it. The Eucldean geometry is an example of it. Since we use mental images of parallel lines, squares and circles as our definitions, it is difficult to accept non-Euclidean postulate. This, in mathematics and in particular in geometry, can be completely wrong.

In hyperbolic geometry the parallel postulate of Euclidean geometry is replaced with *non-Euclidean postulate*: For any given line $l$ and point $P$ not on $l$, in the plane containing both line $l$ and point $P$ there are at least two distinct lines through $P$ that do not intersect $l$ (compare this with Playfair's axiom, the modern version of Euclid's parallel postulate).

There are a lot of differences between Euclidean and non-Euclidean geometry. For example, in Euclidean geometry, the notion of area is based on the area of a rectangle. The area of parallelograms and triangles follows easily from this definition. The area of other regions is determined by a limiting process involving an approximation by rectangles. In hyperbolic geometry, we do not have figures analogous to rectangles. A hyperbolic quadrilateral has angle sum less than $\pi$ so cannot have four right angles. Instead, we use triangles as basic figures.

The strangeness and counter-intuitiveness of non-Euclidean geometry helps students to understand the differences between definitions and theorems, the concepts of axioms and proofs as they are used in geometry.

The development of non-Euclidean geometry caused a profound revolution, not just in mathematics, but in science and philosophy as well. Einstein and Minkowski found in non-Euclidean geometry a geometric basis for the understanding of physical time and space.

In this paper we give an introduction to the fascinating subject of planar Hyperbolic geometry. Our approach is related to Schwarz's lemma and methods of Complex Analysis.

Our discussion in Section 1 includes postulates which are equivalent to the parallel postulate, properties of the cross-ratio, model of non-Euclidean (Lobachevsky) geometry, non-Euclidean and psedo-hyperbolic distance and the hyperbolic law of cosines and sines.

In Section 2 we give approach to Hyperbolic geometry via Schwarz lemma. In particular, we derive the formula for the area of a hyperbolic triangle.

In section 3 we shortly discuss a few unpublished results of the author. Theorems 5-6 are versions of the Earle-Hamilton fixed point theorem (independently obtained by the author).

In Section 4 we shortly discuss the place hyperbolic geometry in mathematic and science.

### 1.1. Euclidean geometry

The Pythagorean theorem forms the basis of trigonometry and it has been applied to real-world problems since at least 1500 B.C. Mathematicians often include The Pythagorean theorem in top dozen candidates for favorite mathematical theorems ( that are relatively easy for non-mathematicians to understand). For example, David Joyce, Professor of Mathematics at Clark University Updated Jul 28, 2015, said there are so many important theorems, but two I would list in any listing are:
The Pythagorean theorem - anything to do with geometry depends on it.

The Fundamental Theorem of Calculus, in particular, the version that says

$$\int_a^b f'(x)dx = f(b) - f(a).$$

It is what makes analysis work.

There are large literature related to the Pythagorean theorem, see for example [12] and literature cited there. In mathematics, the Pythagorean theorem or Pythagoras' theorem is a relation in Euclidean geometry among the three sides of a right triangle. It states that the square of the hypotenuse (the side opposite the right angle) is equal to the sum of the squares of the other two sides. The theorem can be written as an equation relating the lengths of the sides $a$, $b$ and $c$, often called the Pythagorean equation:

$$a^2 + b^2 = c^2,$$

where $c$ represents the length of the hypotenuse, and $a$ and $b$ represent the lengths of the other two sides.

Let us take one practical example.

**Question**. If the pole height 8m breaks at a height of 3m (but not quite) as much as the top after falling to the ground away from the foot. Using an approximation of reality to the ideal world of mathematics we can find the distance $d$ by formula $d^2 + 3^2 = 5^2$. Hence, $d = 4m$.

The Pythagorean theorem is named after the Greek mathematician Pythagoras (ca. 570 BC - ca. 495 BC), who by tradition is credited with its proof, although it is often argued that knowledge of the theorem predates him. There is evidence that Babylonian mathematicians understood the formula, although there is little surviving evidence that they used it in a mathematical framework. Also, Mesopotamian, Indian and Chinese mathematicians have all been known for independently discovering the result, some even providing proofs of special cases. The converse of the theorem is also true:

For any three positive numbers $a$, $b$, and $c$ such that $a^2 + b^2 = c^2$, there exists a triangle with sides $a$, $b$ and $c$, and every such triangle has a right angle between the sides of lengths $a$ and $b$.

An alternative statement is: For any triangle with sides $a, b, c$, if $a^2 + b^2 = c^2$, then the angle between $a$ and $b$ measures $90°$.

This converse also appears in Euclid's Elements (Book I, Proposition 48):

"If in a triangle the square on one of the sides equals the sum of the squares on the remaining two sides of the triangle, then the angle contained by the remaining two sides of the triangle is right."

The Pythagorean theorem is a special case of the more general theorem relating the lengths of sides in any triangle, the law of cosines:

$$a^2 + b^2 - 2ab\cos\theta = c^2,$$

where $\theta$ is the angle between sides $a$ and $b$.

When $\theta$ is $90°$, then $\cos\theta = 0$, and the formula reduces to the usual Pythagorean theorem.

In a Euclidean space, the sum of measures of these three angles of any triangle is invariably equal to the straight angle, also expressed as $180°$, $\pi$ radians, two right angles, or a half-turn.

It was unknown for a long time whether other geometries exist, where this sum is different. The influence of this problem on mathematics was particularly strong during the 19th century. Ultimately, the answer was proven to be positive: in other spaces (geometries) this sum can be greater or lesser, but it then must depend on the triangle. Its difference from $180°$ is a case of angular defect and serves as an important distinction for geometric systems.

In Euclidean geometry, the triangle postulate states that the sum of the angles of a triangle is two right angles. This postulate is equivalent to the parallel postulate [10]. In the presence of the other axioms of Euclidean geometry, the following statements are equivalent:

*Triangle postulate*: The sum of the angles of a triangle is two right angles.

*Playfair's axiom*: Given a straight line $l$ and a point $M$ not on the line, exactly one straight line may be drawn through the point parallel to the given line in the plane defined by $l$ and $M$.

*Proclus' axiom*: If a line intersects one of two parallel lines, it must intersect the other also.

*Equidistance postulate*: Parallel lines are everywhere equidistant (i.e. the distance from each point on one line to the other line is always the same).

*Triangle area property*: The area of a triangle can be as large as we please.
*Three points property*: Three points either lie on a line or lie on a circle.
*Pythagoras' theorem*: In a right-angled triangle, the square of the hypotenuse equals the sum of the squares of the other two sides.

Euclidean disks in the hyperbolic model $\mathbb{U}$ are also hyperbolic disc and vice versa. So, at first glance one can think that *Three points property* holds in a hyperbolic model, but it is easy to construct counter example. Namely, in hyperbolic model $\mathbb{U} \subset \mathbb{C}$ take a point $z_0$ on the unit circle and a circle $K$ with center at $z_0$ and let $K'$ be the intersection of $K$ and $\mathbb{U}$. If we take three different points $z_1, z_2$ and $z_3$ on $K'$, then Three points property is not true for those points.

### 1.2. The cross-ratio

The cross-ratio of a 4-tuple of distinct points on the real line with coordinates $z_1, z_2, z_3, z_4$ is given by

$$(z_1, z_2; z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_2 - z_3)(z_1 - z_4)}.$$

We also use notation $(\,)_{cr}$ or $[\,] = [\,]_{cr}$ for cross-ratio if there is no possibility of confusion. It can also be written as a "double ratio" of two division ratios of triples of points:

$$(z_1, z_2; z_3, z_4) = \frac{z_1 - z_3}{z_2 - z_3} : \frac{z_1 - z_4}{z_2 - z_4}.$$

The same formulas can be applied to four different complex numbers and can also be extended to the case when one of them is the symbol $\infty$, by removing the corresponding two differences from the formula. The formula shows that cross-ratio is a function of four points, generally four numbers $z_1,\ z_2,\ z_3,\ z_4$ taken from $\overline{\mathbb{C}}$.

From the definition of the cross-ratio it follows $(z_1, z_2; z_3, z_4)$ tends to $\frac{z_1 - z_3}{z_2 - z_3}$, when $z_4$ tends to $\infty$, and we define $(z_1, z_2; z_3, \infty) = \frac{z_1 - z_3}{z_2 - z_3}$. Hence $(z_1, z_2; 0, \infty) = \frac{z_1}{z_2}$ and, in particular, $(z, 1; 0, \infty) = z$.

There are 24 possible permutations of the four coordinates; it is clear from definition of the cross-ratio that some permutations leave the cross-ratio unaltered. In fact, exchanging any two pairs of coordinates preserves the cross-ratio:

$$(z_1, z_2; z_3, z_4) = (z_2, z_1; z_4, z_3) = (z_3, z_4; z_1, z_2) = (z_4, z_3; z_2, z_1).$$

Using these symmetries, there can then be 6 possible values of the cross-ratio, depending on the order in which the points are given. These are:

$$(z_1, z_2; z_3, z_4) = \lambda,\ (z_1, z_2; z_4, z_3) = \frac{1}{\lambda}, (z_1, z_3; z_4, z_2) = \frac{1}{1 - \lambda}, \tag{1}$$

$$(z_1, z_3; z_2, z_4) = 1 - \lambda\,, (z_1, z_4; z_3, z_2) = \frac{\lambda}{\lambda - 1}, (z_1, z_4; z_2, z_3) = \frac{\lambda - 1}{\lambda}. \tag{2}$$

We give two interesting application of the cross-ratio technique.

**Proposition 1** (Ptolemy theorem)**.** *Let $z_1, z_2, z_3, z_4$ be cyclic quadrilateral, and $a, b, a', b'$ the lengths of sides $z_1z_2, z_2z_2, z_2z_4, z_4z_1$ respectively and $d_1, d_2$ the lengths of diagonals $z_1z_3, z_4z_2$. Then $d_1d_2 = aa' + bb'$.*

Outline of the proof: We have $\lambda = [z_1, z_2; z_4, z_3] = \frac{bb'}{d_1d_2}$, $\mu = [z_2, z_3; z_1, z_4] = \frac{aa'}{d_1d_2}$ and $\lambda + \mu = 1$.

The projection from a point onto a plane or central projection: If $C$ is a point, called the center of projection, then the projection of a point $P$ different from $C$ onto a plane that does not contain $C$ is the intersection of the line $CP$ with the plane. The points P such that the line $CP$ is parallel to the plane do not have any image by the projection, but one often says that they project to a point at infinity of the plane (see projective geometry for a formalization of this terminology).

**Definition 1.** Let $l_1$ and $l_2$ be two lines in a plane and let $S$ be a point in the same plane which does belong to $l_1 \cup l_2$. The mapping $f : l_1 \to l_2$ which sends a point $M \in l_1$ to the point $f(M) = SM \cap l_2$ we call central projection of the line $l_1$ on the line $l_2$ from the center $S$.

**Proposition 2.** *The cross ratio is invariant with respect central projection.*

*Proof.* Suppose that $L_i$: $y = k_i x$ and $L$: $y = kx + m$ are lines and let the points $z_i$ be the intersection of $L_i$ with $L$. Then $x_i = \frac{m}{k_i - k}$, $z_i = x_i(1 + ik_i) = (1 + ik_i)\frac{m}{k_i - k} = mz_i'$ and $z_i' - z_j' = [\frac{1}{k_i - k} - \frac{1}{k_j - k}](1 + ik)$. If we set $a_k = \frac{1}{k_i - k}$, then $[z_1, z_2; z_3, z_4] = [z_1', z_2'; z_3', z_4'] = [a_1, a_2; a_3, a_4] = [k_1, k_2; k_3, k_4]$. □

### 1.3. Hyperbolic geometry

An angle equal to $1/2$ turn ($180°$ or $\pi$ radians) is called a straight angle.

Angles larger than a right angle and smaller than a straight angle (between $90°$ and $180°$) are called obtuse angles.

Absolute geometry is another term for "neutral geometry" and refers to what can be deduced from using all of Euclid's axioms, except for any axiom equivalent to the the parallel postulate, also called Euclid's fifth postulate, because it is the fifth postulate in Euclid's Elements.

*Parallel postulate*: If a line segment intersects two straight lines forming two interior angles on the same side that sum to less than two right angles, then the two lines, if extended indefinitely, meet on that side on which the angles sum to less than two right angles.

Euclidean geometry is the study of geometry that satisfies all of Euclid's axioms, including the parallel postulate. A geometry where the parallel postulate does not hold is known as a non-Euclidean geometry. Geometry that is independent of Euclid's fifth postulate (i.e. only assumes the modern equivalent of the first four postulates) is known as absolute geometry (or, in other places, known as neutral geometry).

One can also prove in absolute geometry the exterior angle theorem (an exterior angle of a triangle is larger than either of the remote angles), as well as the Saccheri-Legendre theorem, which states that the sum of the measures of the angles in a triangle has at most $180°$.

The alternate interior angle theorem states that if lines $a$ and $b$ are cut by a transversal $t$ such that there is a pair of congruent alternate interior angles, then $a$ and $b$ are parallel. However, the converse is equivalent to the parallel postulate and is not true in the Absolute geometry.

Proposition 31, in Euclid's Elements, is the construction of a parallel line to a given line through a point not on the given line. As the proof only requires the use of Proposition 27 (the Alternate Interior Angle Theorem), it is a valid construction in absolute geometry. More precisely, given any line $l$ and any point P not on $l$, there is at least one line through $P$ which is parallel to $l$. This can be proved using a familiar construction: given a line $l$ and a point P not on $l$, drop the perpendicular m from P to $l$, then erect a perpendicular $n$ to $m$ through $P$. By the alternate interior angle theorem, $l$ is parallel to $n$.

Absolute geometry is an incomplete axiomatic system, in the sense that one can add extra independent axioms without making the axiom system inconsistent. One can extend absolute geometry by adding different axioms about parallel lines and get incompatible but consistent axiom systems, giving rise to Euclidean or hyperbolic geometry. Thus every theorem of absolute geometry is a theorem of hyperbolic geometry and Euclidean geometry. However the converse is not true.

The sum of the angles of a hyperbolic triangle is less than $180°$. The relation between angular defect and the triangle's area was first proven by Johann Heinrich Lambert.

One can easily see how hyperbolic geometry breaks Playfair's axiom, Proclus' axiom (the parallelism, defined as non-intersection, is intransitive in an hyperbolic plane), the equidistance postulate (the points on one side of, and equidistant from, a given line do not form a line), and Pythagoras' theorem. A circle cannot have arbitrarily small curvature, so the three points property also fails.

The sum of the angles can be arbitrarily small (but positive). For an ideal triangle, a generalization of hyperbolic triangles, this sum is equal to zero.

The Poincaré disk model also called the conformal disk model, is a model of $2-$dimensional hyperbolic geometry in which the points of the geometry are inside the unit disk, and the straight lines consist of all segments of circles contained within that disk that are orthogonal to the boundary of the disk, plus all diameters of the disk. Hyperbolic Straight lines consist of all arcs of Euclidean circles contained within the disk that are orthogonal to the boundary of the disk, plus all diameters of the disk.

### 1.4. Model of non-Euclidean (Lobachevsky) geometry

Recall if we suppose that axioms of absolute geometry are true then Euclid's fifth postulate is equivalent to the following.

*Playfair axiom:* In a plane, given a line and a point not on it, at most one line parallel to the given line can be drawn through the point.

The famous Pythagorean Theorem is also Equivalent to the Parallel Postulate.

From now on, we assume the existence of a model for Euclidean Geometry. Within the euclidian model we will construct a hyperbolic model.

By $\mathbb{U}$ we denote the unit circle in Euclidean plane; sometimes the notation $\mathbb{D}$ is used. Points of $\mathbb{U}$ are $\mathbb{U}$-hyp points, $\mathbb{U}$-hyp lines are intersection of circles orthogonal on $\mathbb{T}$ and lines through coordinate origin $O$ with $\mathbb{U}$.

Points in $\mathbb{H}$ are $\mathbb{H}$-hyp points and $\mathbb{H}$-hyp lines are intersections of circles and Euclidean lines orthogonal on $\mathbb{R}$ with $\mathbb{H}$.

Points of $\mathbb{U}$ are $\mathbb{U}$-klein points. $\mathbb{U}$-klein lines are intersection of Euclidean lines with $\mathbb{U}$.

If $K$ is $\mathbb{H}$-hyp line defined by Euclidean circle $\hat{K}$, the points $a, b$ which belongs to the intersection of $\hat{K}$ and $\mathbb{R}$ are ideal point of $K$.

Recall *Parallel postulate*: If $l$ is a line in a plane $P$, then $l$ divides the plane on two half-plane say $P^1$ and $P^2$. Let $p$ and $q$ be two line at $P$ which intersect $l$ respectively at point $P$ and $Q$ and let $P'$ and $Q'$ be points in say $P^1$ respectively on $p$ and $q$. If the sum of measure of angles $QPP'$ and $PQQ'$ is less then two right angles, then there is a common point of $p$ and $q$ in $P^1$.

An important equivalent of Euclid's parallel postulate is the following postulate.

*Parallel-Transversal postulate*: If the two lines are parallel and $l$ a transversal which intersects them, then consecutive interior angles are supplementary (they add up to two right angles), corresponding angles are equal, and alternate angles are equal.

In geometry, a transversal is a line that passes through two lines in the same plane at two distinct points. Transversals play a role in establishing whether two other lines in the Euclidean plane are parallel. The intersections of a transversal with two lines create various types of pairs of angles: consecutive interior angles, corresponding angles, and alternate angles. By an equivalent of Euclid's parallel postulate, if the two lines are parallel, consecutive interior angles are supplementary (they add up to two right angles), corresponding angles are equal, and alternate angles are equal.

Recall that:

L1. In Euclidean geometry corresponding and alternate angles on transversal which intersects parallel lines are equal.

L2. In Euclidean geometry, the triangle postulate states that the sum of the angles of a triangle is two right angles ($\pi$). This postulate is equivalent to the parallel postulate.

**Proposition 3.** *In Euclidean geometry, the triangle postulate is equivalent to the parallel postulate.*

*Proof.* We will only outline a proof. Let ABC be a triangle with angles $\alpha$, $\beta$ and $\gamma$ at corners $A$, $B$ and $C$ respectively. By Playfair's axiom there is exactly one straight line $c'$ through the point parallel to the given line $c$ defined by the points $A$ and $B$. Let $b$ and $a$ respectively be lines defined by the points $A$ and $C$, and $B$ and $C$ respectively. The lines $b$ and $a$ respectively intersects parallel line $c$ and $c'$ at points $A$ (respectively $B$) and $C$ and by Euclid's parallel postulate alternate angle $\alpha'$ at corner $C$ and angle $\alpha$ are equal. In a similar way alternate angle $\beta'$ at corner $C$ and angle$\alpha$ are equal. The union of of these three angles $\alpha'$, $\gamma$ and $\beta'$ is a half-plane and the sum of measures of these three angles equal to the straight angle, also expressed as $180°$, $\pi$ radians, two right angles, or a half-turn. □

Motions in Euclidean geometry are isometry. It is easy to check that Translations $T_a(z) = z+a$ and rotations $R_\alpha = e^{i\alpha}z$ are Euclidean isometry. It is well known that :

(T1) Every isometry is composition translation and rotation.

(T2) Every isometry is composition at most 3 reflections.

On the hand, the isometry in geometry of Lobachevsky are Möbius transformation.

It is convenient to consider parallel model on $\mathbb{U}$ and $\mathbb{H}$.

### 1.5. Hyperbolic distance

In this subsection we derive the formulas for for non-Euclidean distance.

Let $I = [iy_1, iy_2]$, $y_1 < y_2$. Then $d(iy_1, iy_2) = \sum \int_I \frac{dy}{y} = \ln \frac{y_2}{y_1}$.

Let $z_1$ and $z_2$ points in $\mathbb{H}$. Then there exists a unique $\mathbb{H}$-hyp line $l$ throughout those points and let $a$ and $b$ be ideal points of line $l$.

Suppose that $a < b$ and that the point $z_1$ is between $a$ and $z_2$. Define $\{z_1, z_2\} = [z_2, z_1; a, b]$. Möbius transformation $L(z) = \frac{z-a}{b-z}$ maps $l$ on $Y^+ = \{iy : y > 0\}$. $L$ maps points $z_2, z_1; a, b$, respectively, to the points $iy_2, iy_1, 0, \infty$. Hence, $\{z_1, z_2\} = \{iy_1, iy_2\} = \frac{y_2}{y_1}$ and $d = \ln\{z_1, z_2\}$.

Model on the unit disk: Using conformal mappings $A(z) = \frac{z-i}{z+i}$ we can define hyperbolic distance na $\mathbb{U}$. Let $K = K(z_1, z_2)$ circle orthogonal on $\mathbb{T}$ throughout points $z_1$ i $z_2$ and denote by $a$ and $b$ the intersection points $K$ and $\mathbb{T}$. Usually we denote the intersection points such that $z_1$ is between $a$ and $z_2$.

Recall that $[z_1, z_2; a, b] = \frac{z_1-a}{z_1-b} : \frac{z_2-a}{z_2-b}$. For example, if $a = -1$ and $b = 1$, then, according to the notation, we have $z_1 = 0$ and $z_2 = r$, $0 < r < 1$. Since $[0, r; -1, 1] = \frac{1-r}{1+r}$, it is $0 < [0, r; -1, 1] < 1$. Therefore, it is convenient to define $\{z_1, z_2\} = [z_1, z_2; a, b]_2 = \frac{z_2-a}{z_2-b} : \frac{z_1-a}{z_1-b}$. It follows that, according to our convention on notation, $\{z_1, z_2\} > 1$ and $\{z_1, z_2\} \cdot \{z_2, z_3\} = \{z_1, z_3\}$. Also, if we define $d_{hyp}(z_1, z_2) = \ln\{z_1, z_2\}$, then $\{0, r\} = \frac{1+r}{1-r}$.

Define $T_{z_1}(z) = \frac{z-z_1}{1-\overline{z_1}z}$, $\varphi_{z_1} = -T_{z_1}$ and $\sigma(z_1, z_2) = |T_{z_1}(z_2)| = |\frac{z-z_1}{1-\overline{z_1}z}|$.

Schwarz's lemma yields motivation to introduce hyperbolic distance: If $f \in \mathrm{Hol}(\mathbb{U}, \mathbb{U})$, then

$$\sigma(fz_1, fz_2) \leqslant \sigma(z_1, z_2).$$

Consider $F = \varphi_{w_1} \circ f \circ \varphi_{z_1}$, $w_k = f(z_k)$. Then $F(0) = 0$ and $|\varphi_{w_1}(w_2)| \leqslant |\varphi_{z_1}(z_2)|$.

Hence,

$$|f'(z)| \leqslant \frac{1 - |fz|^2}{1 - |z|^2}.$$

By notation $w = f(z)$ and $dw = f'(z)dz$, so

$$\frac{|dw|}{1 - |w|^2} \leqslant \frac{|dz|}{1 - |z|^2}.$$

Define the density $\rho(z) = \frac{1}{1-|z|^2}$.

For a vector $\mathbf{v} \in T_z\mathbb{C}$ we define $|\mathbf{v}|_\rho = \rho(z)|\mathbf{v}|$ and we set $\mathbf{v}^* = df_z(\mathbf{v})$. Then $|\mathbf{v}^*|_\rho \leqslant |\mathbf{v}|_\rho$.

If $\gamma$ piecewise smooth then define $|\gamma|_\rho = \int_\gamma \rho(z)|dz|$ and $d(z_1, z_2) = \inf |\gamma|_\rho$, where the infimum is taken over all paths $\gamma$ in $\mathbb{U}$ joining the points $z_1$ and $z_2$.

Let $G$ be a simply connected domain different from $\mathbb{C}$ and let $\phi : G \to \mathbb{U}$ be a conformal isomorphism. Define the pseudo hyperbolic distance on $G$ by $\varphi_a^G(z) = \varphi_b(\phi)$, where $b = \phi(a)$, and $\delta_G(a, z) = |\varphi_a^G(z)|$. Verify that the pseudo hyperbolic distance on $G$ is independent of conformal mapping $\phi$. In particular, using conformal isomorphism $A(w) = A_{w_0}(w) = \frac{w-w_0}{w-\overline{w_0}}$ of $\mathbb{H}$ onto $\mathbb{U}$, we find $\varphi_{H,w_0}(w) = A(w)$ and therefore $\delta_H(w, w_0) = |A(w)|$.

**Proposition 4.** *If $G$ and $D$ are conformally isomorphic to $\mathbb{U}$ and $f \in \mathrm{Hol}(G, D)$, then*

$$\delta_D(fz, fz') \leqslant \delta_G(z, z'),$$

*for all $z, z' \in G$.*

Whether $\delta_H$ is related to a density function? Note that, for $z = x + iy \in \mathbb{H}$, that

(*) $\sigma_{\mathbb{H}}(z, z+h)/|h| = \frac{2}{z+h-\overline{z}}$ tends to $y^{-1}$ if $h$ tends to 0. Thus $y^{-1}$ is hyperbolic density and it defines $d_{hyp}$ hyperbolic metric in a standard way.

Check that

$$d_{hyp}(z_1, z_2) = \ln \frac{1 + \sigma(z_1, z_2)}{1 - \sigma(z_1, z_2)} = \ln \frac{|1 - \overline{z_1}z_2| + |z_1 - z_2|}{|1 - \overline{z_1}z_2| - |z_1 - z_2|},$$

where $d = d^{hyp} = d_U^{hyp}$.

By $\mathrm{arcosh}$ and $\mathrm{arsinh}$ we denote inverses of hyperbolic functions:

$\mathrm{arsinh}\, x = \ln\left(x + \sqrt{x^2 + 1}\right)$ , $\mathrm{arcosh}\, x = \ln\left(x + \sqrt{x^2 - 1}\right); x \geqslant 1.$

**Proposition 5.** *a) If $d = d_H^{\mathrm{hyp}} = \mathrm{Hyp}_H$, then*

$$\operatorname{dist}(\langle x_1, y_1\rangle, \langle x_2, y_2\rangle) = \operatorname{arcosh}\left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2y_1 y_2}\right).$$

*b) If $d = d_{\mathbb{U}}^{\mathrm{hyp}}$, then*

$$\cosh d(z, w) = \frac{(1 + |z|^2)(1 + |w|^2) - 4|z||w|\cos\alpha}{(1 - |z|^2)(1 - |w|^2)},$$

*for all $z, w \in \mathbb{U}$.*

*Proof.* We have $\cosh d = \frac{1}{2}(e^d + e^{-d}) = \frac{1}{2}(\frac{1+\sigma}{1-\sigma} + \frac{1-\sigma}{1+\sigma}) = \frac{1+\sigma^2}{1-\sigma^2} = 1 + \frac{2\sigma^2}{1-\sigma^2}$ on $\mathbb{H}$. Hence, $\cosh d = 1 + 2\frac{|z_1 - z_2|^2}{|z_1 - \overline{z_2}|^2 - |z_1 - z_2|^2}$, and since $|z_1 - \overline{z_2}|^2 - |z_1 - z_2|^2 = 4y_1 y_2$, we find

$$\cosh d = 1 + \frac{|z_1 - z_2|^2}{2y_1 y_2}.$$

In general, the distance between two points measured in this metric along such a geodesic is:

$$\operatorname{dist}(\langle x_1, y_1\rangle, \langle x_2, y_2\rangle) = \operatorname{arcosh}\left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2y_1 y_2}\right).$$

b) Let $\alpha$ be the measure of angle for between $z$ and $w$. Then $I_1 = |z - w|^2 = |z|^2 + |w|^2 - 2|z||w|\cos\alpha$ and $I_2 = |1 - z\overline{w}|^2 = 1 + |zw|^2 - 2|z||w|\cos\alpha$ and therefore $I = I_1 + I_2 = (1 + |z|^2)(1 + |w|^2) - 4|z||w|\cos\alpha$ and $I = I_2 - I_1 = (1 - |z|^2)(1 - |w|^2)$. Since $\cosh d = \frac{1+\sigma^2}{1-\sigma^2} = I/J$, then b) follows. □

**Example 1.** *(a) Let $Q = \{z = x+iy : x, y > 0\}$ and $Z(z) = \frac{1}{2}(z+z^{-1})$. Then $Z(Q) = \{w : Rew > 0\}\setminus[1, \infty)$ and $Z(\mathbb{U}) = \overline{C} \setminus [-1, 1]$.*

*(b) Let $G = \{z = x + iy : |z| < 1, x, y > 0\}$, $l_1 = [0, 1]$, $l_2 = \{e^{it} : 0 < t < \pi/2\}$ and $l_3 = [i, 0]$. Then $Z$ maps respectively $l_1, l_2$ and $l_3$ on $l_1' = [+\infty, 1]$, $l_2' = (1, 0)$ and $l_3' = [0, -i\infty]$. Hence $Z$ maps $G$ onto IV quadrant $Q_4 = \{w = u + iv : u > 0, v < 0\}$.*

*(c) If $\underline{P} = \{w : Rew > 0\}$, then $\delta_{\underline{P}}(w', w) = |\frac{w'-w}{w'+\overline{w}}|$.*

*(d) Note that $Z(Q) \cap Z(\mathbb{U}) = \{w : Rew > 0\} \setminus (0, \infty)$. If $A, B$ and $Y$ are sets and $f$ injective mapping on $A \cup B$ into $Y$, then (1) $f(A \cap B) = f(A) \cap f(B)$. Explain way we can apply (1) in the item (b). Also, for $f$ given by $f(z) = z^2$, that maps $Q_4$ onto $\mathbb{H}^- = \{w : \mathrm{Im}w < 0\}$. Hence, $\delta_{\mathbb{H}^-}(w', w) = |\frac{w'-w}{w'-\overline{w}}|$ and $\delta_G(z', z) = |\frac{w'^2-w^2}{w'^2-\overline{w}^2}|$, where $w = Z(z)$ and $w' = Z(z')$.*

### 1.6. The hyperbolic law of cosines and sines

A good reference for this subsection is Ahlfors book [2]. For a right triangle in hyperbolic geometry with sides $a, b, c$ and with side $c$ opposite a right angle, the relation between the sides takes the form:

$$\cosh c = \cosh a \ \cosh b,$$

where cosh is the hyperbolic cosine. This formula is a special form of the hyperbolic law of cosines that applies to all hyperbolic triangles:

**Proposition 6** (I. The hyperbolic law of cosines)**.**

$$\cosh c = \cosh a \ \cosh b - \sinh a \ \sinh b \ \cos\gamma$$

*with $\gamma$ the angle at the vertex opposite the side c.*

By using the Maclaurin series for the hyperbolic cosine, $coshx = 1 + x^2/2 + o(x^2)$, it can be shown that as a hyperbolic triangle becomes very small (that is, as $a$, $b$, and $c$ all approach zero), the hyperbolic relation for a right triangle approaches the form of Pythagoras' theorem.

In hyperbolic geometry when the curvature is $-1$, the law of sines becomes:

**Proposition 7** (II. The hyperbolic law of sines)**.** *For a hyperbolic triangle $ABC$,* $\dfrac{\sin A}{\sinh a} = \dfrac{\sin B}{\sinh b} = \dfrac{\sin C}{\sinh c}$.

In the special case when $B$ is a right angle, one gets $\sin C = \dfrac{\sinh c}{\sinh b}$, which is the analog of the formula in Euclidean geometry expressing the sine of an angle as the opposite side divided by the hypotenuse.

Proof of I: By an abuse of notation, we use the same symbols for vertices and the measures of corresponding angles. Without loss of generality we can suppose $C = 0$, $0 < B < 1$ and $A = e^{iC}s$, where $0 < s < 1$. Then $B = \tanh(a/2)$, $A = \tanh(b/2)e^{iC}$ and $\sigma(A, B) = \tanh(c/2)$. As in Euclidean trigonometry all trigonometric function we can express by $\tanh$. For hyperbolic $\cos$ and $\sin$ (see also Proposition 5),

$$\cosh x = \frac{1+\tanh^2(x/2)}{1-\tanh^2(x/2)}, \quad \sinh x = \frac{2\tanh(x/2)}{1-\tanh^2(x/2)}\,.$$

Also,

$$\cosh c = \tag{3}$$

$$\frac{(1+\tanh^2(a/2))(1+\tanh^2(b/2)) - 4\tanh(a/2)\tanh(b/2)\cos C}{(1-\tanh^2(a/2))(1-\tanh^2(b/2))} \tag{4}$$

$$= \cosh a\ \cosh b - \sinh a\ \sinh b\ \cos\gamma. \tag{5}$$

Proof of II: By the hyperbolic law of cosines,

$$\cos C = \frac{\mathrm{ch}a\,\mathrm{ch}b - \mathrm{ch}c}{\mathrm{sh}a\,\mathrm{sh}b}\,,$$

$$\sin^2 C = \frac{(\mathrm{ch}^2a-1)(\mathrm{ch}^2b-1) - (\mathrm{ch}a\,\mathrm{ch}b - \mathrm{ch}c)^2}{\mathrm{sh}^2a\mathrm{sh}^2b},$$

$$\frac{\sin^2 C}{\mathrm{sh}^2 c} = \frac{1-\mathrm{ch}^2a-\mathrm{ch}^2b-\mathrm{ch}^2c+2\mathrm{ch}a\,\mathrm{ch}b\,\mathrm{ch}c}{\mathrm{sh}^2a\mathrm{sh}^2b\mathrm{sh}^2c}\,.$$

Since the formula is symmetric, II follows.

## 2. Schwarz lemma and Hyperbolic geometry

In this section we give more details related to connections between Schwarz lema and Hyperbolic geometry. In particular, we derive the formula for the area of a hyperbolic triangle,the gauss-bonnet formula: If the hyperbolic triangle $ABC$ has angles $\alpha, \beta, \gamma$, then its area is $\pi - (\alpha + \beta + \gamma)$.

If $D$ and $G$ are two domains in complex plane $\mathbb{C}$ by $\mathrm{Hol}(G, D)$ we denote the family of all holomorphic mappings $f : G \to D$.

For a $G$ domain in $\mathbb{C}$, by $\mathrm{Hyp}_G$ we denote the hyperbolic distance on $G$. Note that

$$\delta_G = \tanh(\mathrm{Hyp}_G/2). \tag{6}$$

The considerations in Section 1 lead to the following result:

**Proposition 8.** *If $G$ and $D$ are conformally isomorphic to $\mathbb{U}$ and $f \in \mathrm{Hol}(G, D)$, then*
*(a) $\delta_D(fz, fz') \leqslant \delta_G(z, z')$, $z, z' \in G$ and*
*(b) $\mathrm{Hyp}_D(fz, fz') \leqslant \mathrm{Hyp}_G(z, z')$, $z, z' \in G$.*
*If the equality holds for $z \neq z' \in G$ holds in (a) or (b), then $f$ is conformal isomorphism of $G$ onto $D$.*

This result can be considered as a version of Schwarz lemma and gives an close connection between holomorphic functions and hyperbolic distances.

### 2.1. The Schwarz lemma 1

If $|a| < 1$ define the Möbius transformation

$$\varphi_a(\zeta) = \frac{\zeta - a}{1 - \overline{a}\,\zeta}\,. \tag{7}$$

**Example 2.** *Fix $a \in \mathbb{D}$. Then $\varphi_a(0) = -a$, $\varphi_a(a) = 0$, $\varphi_a$ is a one-to-one mapping which carries $\mathbb{T}$ onto $\mathbb{T}$, $\mathbb{D}$ onto $\mathbb{D}$. The inverse of $\varphi_a$ is $\varphi_{-a}$.*

*Check that $\varphi_a'(z) = (1-|a|^2)(1-\bar{a}z)^{-2}$ and in particular $\varphi_a'(0) = (1-|a|^2), \varphi_a'(a) = (1-|a|^2)^{-1}$.*

Suppose that $f : \mathbb{D} \to \mathbb{D}$ is an analytic map and $f(0) = 0$. The classic Schwarz lemma states : $|f(z)| \leqslant |z|$ and $|f'(0)| \leqslant 1$.

A standard proof is based on an application of the Maximum Modulus Theorem to the function $g$ defined by $g(z) = \frac{f(z)}{z}$ for $z \neq 0$ and $g(0) = f'(0)$.

Now we shall drop the assumption $f(0) = 0$. Let $f : \mathbb{D} \to \mathbb{D}$ is an arbitrary analytic map. Fix an arbitrary point $z \in \mathbb{D}$ and consider the mapping $F = \varphi_w \circ f \circ \varphi_{-z}$, where $w = f(z)$. Since $\varphi_{-z}(0) = z$, $F(0) = 0$. By an application of Schwarz lemma,

$$|F(\zeta)| = |\varphi_w \circ f \circ \varphi_{-z}(\zeta)| \leqslant |\zeta|, \quad \zeta \in \mathbb{D}, \tag{8}$$

and $|F'(0)| \leqslant 1$. Hence, since $F'(0) = \varphi_w'(w) f'(z) \varphi_{-z}'(0)$, we find

$$\frac{|f'(z)|}{1-|f(z)|^2} \leqslant \frac{1}{1-|z|^2}, \quad z \in \mathbb{D}\,, \tag{9}$$

with equality only if $F = e^{i\alpha} Id$, that is $f = \varphi_{-w} \circ (e^{i\alpha}\varphi_z)$.

Hence, equality holds in (9) if and only if $f$ is a Möbius transformation of $\mathbb{D}$ onto itself.

Let $\omega$ be an arbitrary point in $\mathbb{D}$ and $\zeta = \varphi_z(\omega)$, $\zeta' = \varphi_w(f(\omega))$. Then $\varphi_{-z}(\zeta) = \omega$, $F(\zeta) = \zeta'$ and by (8), we find $|\varphi_w(f(\omega))| \leqslant |\varphi_z(\omega)|$.

It is convenient to introduce a pseudo-distance

$$\delta(z, \omega) = |\varphi_z(\omega)| = \left| \frac{z-\omega}{1-\overline{\omega}\, z} \right|, \tag{10}$$

which is a *conformal invariant*.

Thus

$$\delta(f(z), f(\omega)) \leqslant \delta(z, \omega), \tag{11}$$

with equality only if $f$ is a Möbius transformation of $\mathbb{D}$ onto itself.

This shows that the Riemannian metric, whose element of length is

$$\mathrm{d}s = \lambda(z)|\mathrm{d}z| = \frac{2\,|\mathrm{d}z|}{1-|z|^2}, \tag{12}$$

is invariant under conformal self-mappings of the disk.
In this metric every rectifiable arc $\gamma$ has length

$$|\gamma|_{\mathrm{hyp}} = \int\limits_{\gamma} \frac{2\,|\mathrm{d}z|}{1-|z|^2}$$

and $|f \circ \gamma|_{\mathrm{hyp}} = |\gamma|_{\mathrm{hyp}}$, if $f$ is a Möbius transformation of $\mathbb{D}$ onto itself.

We call the distance determined by this metric the non-Euclidean distance (hyperbolic) and denote by $\lambda$; we also use notation $\lambda(z) = \frac{2}{1-|z|^2}$ for metric density and $||h||_\lambda = \lambda(z)|h|$ for $h \in T_z$.

The fact that the hyperbolic distance is invariant under self-mapping of the disk we can state in the form: If $h \in T_z$, $A \in \mathrm{Aut}(\mathbb{D})$ and $h_* = A'(z)h$, then $||h_*||_\lambda = ||h||_\lambda$ for every $z \in \mathbb{D}$ and every $h \in T_z$.

If $\gamma$ is a piecewise continuously differentiable path which joins 0 and $r$, $0 \leqslant r < 1$, and $I_0 = [0, r]$ using obvious geometric interpretation and the circular projection $p(z) = |z|$, we find $|\gamma|_{\mathrm{hyp}} \geqslant |I_0|_{\mathrm{hyp}}$ and hence

$$\lambda_{\mathbb{H}}(0, r) = |I_0|_{\mathrm{hyp}} = \ln \frac{1+r}{1-r}.$$

Thus, the shortest arc from 0 to any other point is along a radius. So, the geodesics are circles orthogonal to $\mathbb{T} = \{|z| = 1\}$. The non-Euclidean distance from 0 to $r$ is

$$\lambda(0, r) = \int_0^r \frac{2\,\mathrm{d}t}{1 - t^2} = \ln \frac{1 + r}{1 - r}. \tag{13}$$

Since $\delta(0, r) = r$, it follows that non-Euclidean distance $\lambda$ is connected with $\delta$ through $\delta = \tanh \frac{\lambda}{2}$.

Hence, the hyperbolic distance on the unit disk $\mathbb{D}$ is

$$\lambda(z, \omega) = \ln \frac{1 + \left|\frac{z-\omega}{1-z\bar{\omega}}\right|}{1 - \left|\frac{z-\omega}{1-z\bar{\omega}}\right|}. \tag{14}$$

If $f : \mathbb{D} \to \mathbb{D}$ is an arbitrary analytic map, then

$$\lambda(fz, f\omega) \leqslant \lambda(z, \omega).$$

**Exercise 1.** *Check the formula (13).*

**Solution**. $f(t) = \frac{2}{1-t^2}$, $f(t) = \frac{1}{1-t} + \frac{1}{1+t}$, $F = \int f(t) = -\ln(1-t) + \ln(1+t)$. *Hence* $\lambda(0, r) = \int_0^r f(t) = -\ln(1-t)|_0^r + \ln(1+t)_0^r = \ln(1+r) - \ln(1-r) = \ln \frac{1+r}{1-r}$.

**Exercise 2.** *If $\gamma$ is a piecewise continuously differentiable path in $\mathbb{D}$, whether $|\gamma|_{\mathrm{hyp}} = |\gamma|_\delta$?*

### 2.2. The upper half plane

A region $G$ is conformally equivalent to a region $D$ if there is an analytic bijective function $f$ mapping $G$ to $D$; we call $f$ conformal isomorphism. Conformal equivalence is an equivalence relation. Conformal isomorphism of a domain onto itself is called conformal automorphism. Conformal automorphisms of a domain $D$ form a group which we denote by $\mathrm{Aut}D$.

If $f_0 : G \to D$ is a fixed conformal isomorphism, then every conformal isomorphism $f : G \to D$ can be represented in the form

$$f = \phi \circ f_0, \quad \phi \in \mathrm{Aut\,D}. \tag{15}$$

**Example 3.** *Describe* $\mathrm{Aut}(\mathbb{H})$.

*If $A \in \mathrm{Aut}(\mathbb{H})$, there is a point $x_0 \in \mathbb{R}$ such that $A(x_0) = \infty$. We consider two cases.*
*Case (i) $x_0 = \infty$. Then $A = L$, where $L(z) = \lambda z + s$, $\lambda > 0$ and $s \in \mathbb{R}$.*
*Case (ii) $x_0 \in \mathbb{R}$. Define $w = T(z) = -\frac{1}{z} + x_0$. Then $T^{-1}(w) = \frac{1}{x_0 - w}$ and $A \circ T$ maps $\infty$ to $\infty$. Hence $A \circ T = L$, for some $\lambda > 0$ and $s \in \mathbb{R}$, and therefore $f = L \circ T^{-1}$. That is $A(w) = \lambda T^{-1}(w) + s = \lambda \frac{1}{x_0 - w} + s = \frac{a_1 z + b_1}{x_0 - w}$, where $a_1 = -s$ and $b_1 = \lambda + s x_0$. Therefore, every $A \in \mathrm{Aut}(\mathbb{H})$ can represented in the form*

$$f(z) = \frac{az + b}{cz + d}, \tag{16}$$

*where a, b, c, $d \in \mathbb{R}$ and $\mathrm{D} = \mathrm{D}(f) = ad - bc = 1$. If A is represented by (16), then*

$$Az - \overline{Az} = \frac{z - \bar{z}}{|cz + d|^2}. \tag{17}$$

*Hence, it is clear that $A \in \mathrm{Aut}(\mathbb{H})$.*

*There is another way to describe $\mathrm{Aut}(\mathbb{H})$ using $(w, 1; 0, \infty) = w$. Namely, if A carries points $x_2, x_3, x_4$ $(x_2 > x_3 > x_4)$ into $1, 0, \infty$, then $w = (z, x_2, x_3, x_4)$.*

If $L \in \mathrm{Aut}(\mathbb{H})$, then $L$ is Möbius transformation and maps $\mathbb{R}$ onto itself and symmetric points with respect to $\mathbb{R}$ onto symmetric points with respect to $\mathbb{R}$. Hence, if $z_1, z_2 \in \mathbb{H}$ and $w_1 = Lz_1$ and $w_2 = Lz_2$, then $\overline{w_1} = L\overline{z_1}$ and $\overline{w_2} = L\overline{z_2}$. Since the cross-ratio is invariant under Möbius transformation, we get

$$(z_1, \overline{z_1}; z_2, \overline{z_2}) = (Lz_1, L\overline{z_1}; Lz_2, L\overline{z_2}) = (w_1, \overline{w_1}; w_2, \overline{w_2}). \tag{18}$$

Set $Tz = \dfrac{z-z_2}{z-\overline{z_2}}$. Then $(z_1, \overline{z_1}; z_2, \overline{z_2}) = T(z_1)/T(\overline{z_1})$. $T$ maps $\mathbb{H}$ onto $\mathbb{D}$ and symmetric points, with respect to $\mathbb{R}$, $z_1$ and $\overline{z_1}$ onto points $T(z_1)$ and $T(\overline{z_1})$, that are symmetric with respect to $\mathbb{T}$, respectively. Hence, $T(\overline{z_1})\overline{T(z_1)} = 1$ and therefore $(z_1, \overline{z_1}; z_2, \overline{z_2}) = |T(z_1)|^2$. The pseudo-hyperbolic distance on $\mathbb{H}$ can be defined by

$$\delta_H(z_1, z_2) = \left|\frac{z_1 - z_2}{z_1 - \overline{z_2}}\right|.$$

It is invariant with the group $\mathrm{Aut}(\mathbb{H})$ because of (18) and $\delta_H^2(z_1, z_2) = (z_1, \overline{z_1}, z_2, \overline{z_2})$. We will give another proof of this fact in subsection on Schwarz lemma (below).

Often, in the literature, a Riemannian metric on a domain $D$ is given by $ds = \rho|dz|$, where $\rho > 0$, or in its fundamental form

$$ds^2 = \rho^2(dx^2 + dy^2).$$

In some situations it is convenient to to call $\rho$ shortly metric density. If $\mathbf{v} \in T_z$, we define $\rho$-norm of $\mathbf{v}$ by $|\mathbf{v}|_\rho = \rho(z)|\mathbf{v}|_e$, where by the subscript $e$ we denote Euclidean norm.

For a piecewise continuously differentiable path $\gamma$, we define $|\gamma|_\rho = \int\limits_\gamma |\mathrm{d}z|/y$. We use this infinitesimal form to obtain $\rho$- distance between two points $p$ and $q$ in $D$ by putting

$$d_\rho(p, q) = \inf |\gamma|_{\mathrm{hyp}} = \inf \int\limits_\gamma |\mathrm{d}z|/y\,,$$

where the infimum is taken over all piecewise continuously differentiable paths $\gamma$ joining $p$ to $q$ in $D$. For a fixed $z \in \mathbb{H}$, moving on to the limit value of $\delta_H(z, w)/e(z, w)$, where $e$ is Euclidean distance, when $w \to z$ we get an infinitesimal invariant $\mathrm{d}s = |\mathrm{d}z|/y$ (we drop multiple 2), where $y = \mathrm{Im}z$. For a piecewise continuously differentiable path $\gamma(t) = (x(t), y(t))$, $0 \leqslant t \leqslant 1$, in $\mathbb{H}$, we define $|\gamma|_{\mathrm{hyp}} = \int\limits_\gamma |\mathrm{d}z|/y = \int\limits_0^1 \frac{|\gamma'(t)|}{y(t)}\, dt$. We use this infinitesimal form to obtain Poincaré distance between two points $p$ and $q$ in $\mathbb{H}$ by putting

$$d_{\mathrm{hyp}}(p, q) = \inf |\gamma|_{\mathrm{hyp}} = \inf \int\limits_\gamma |\mathrm{d}z|/y,$$

where the infimum is taken over all paths $\gamma$ joining $p$ to $q$. The curve for which infimum is attained we call geodesic. We also use shorter notation $\lambda$ ($\lambda_{\mathbb{H}}(p, q)$ ) instead of $d_{\mathrm{hyp}} = d_{\mathrm{hyp},\mathbb{H}}$ if it is clear that our considerations is related to $\mathbb{H}$.

To find geodesic which joins $p$ and $q$ we use $A \in \mathrm{Aut}(\mathbb{H})$ which maps $z_1$ and $z_2$ to $iy_1$ and $iy_2$. It is easy to conclude that a minimum is attained along the vertical segment $\mathrm{I}_0$ that connects $iy_1$ and $iy_2$. If $\gamma$ is a piecewise continuously differentiable path which joins $iy_1$ and $iy_2$, using obvious geometric interpretation, we find $|\gamma|_{\mathrm{hyp}} \geqslant |I_0|_{\mathrm{hyp}}$ and hence

$$\lambda_{\mathbb{H}}(iy_1, iy_2) = |I_0|_{\mathrm{hyp}} = |\ln(y_2/y_1)|.$$

Hence it follows that geodesics are the arcs of circles orthogonal to the real axis.

There is circular arc $K$ perpendicular to the real axis that contains $z_1$ and $z_2$ and connects real points $a_1$ and $a_2$. We can compute $\omega = (p, q, a_1, a_2)$. Suppose that $a_1 > a_2$ and define $A(z) = \dfrac{z-a_1}{z-a_2}$, then $\det(1, -a_1; 1, -a_2) = a_1 - a_2 > 0$ and therefore $A \in \mathrm{Aut}(\mathbb{H})$. Hence it maps $K$ on one half of the imaginary axis. If $A(p) = iy_1$ and $A(q) = iy_2$, the cross ratio $\omega$ equals

$$(iy_2, iy_1, 0, \infty) = y_2/y_1.$$

Hence $\lambda_H(p, q) = |\ln(y_2/y_1)| = \ln(p, q, a_1, a_2)$. Since, for $y_2 \geqslant y_1$, we get

$$\delta(iy_1, iy_2) = \frac{y_2 - y_1}{y_2 + y_1} = \frac{e^\lambda - 1}{e^\lambda + 1}$$

and hence

$$\delta = \tanh(\lambda/2). \tag{19}$$

In a similar way one can prove that this formula is valid if $y_2 < y_1$. We consider the canonical Möbius transformation $T$ of $\mathbb{H}$ onto $\mathbb{D}$ that maps the points $0, i, \infty$ onto the points $-1, 0, 1$, respectively, and let $S$ denote the inverse of $T$. Then we find

$$w = Tz = \frac{z-i}{z+i},\ z = Sw = i\frac{1+w}{1-w}.$$

Note that if $z, a \in \mathbb{H}$, $b = Ta$, then $(z, \bar{z}, a, \bar{a}) = (w, w^*, b, b^*) = |\varphi_b(w)|^2$.

It is convenient to introduce the mapping $\phi_a = T^{-1} \circ \varphi_b \circ T$ and the pseudo-distance

$$\delta(z, \omega) = |\varphi_z(\omega)| = \left|\frac{z-\omega}{1-\overline{\omega}\, z}\right|, \tag{20}$$

which is a *conformal invariant*. It is easy to check that $\delta_H(a, z) := |\phi_a(z)| = \delta_U(T(a), T(z))$.

Moving on to the limit value, when $\omega \to z$, we get infinitesimal invariant $\mathrm{d}s = \lambda(z)|\mathrm{d}z|$, where $\lambda(z) = 2(1-|z|^2)^{-1}$ is the hyperbolic density (we add multiple 2 so that the Gaussian curvature of the hyperbolic density is $-1$ see below).

The shortest arc from 0 to any other point is along a radius. Hence the geodesics are circles orthogonal to $\mathbb{T}$.

Since $\delta(0, r) = r$ it follows that non-Euclidean distance $\lambda$ is connected with $\delta$ through $\delta = \tanh\dfrac{\lambda}{2}$.

There is another way of calculating that exhibits additivity.

Let $\gamma$ be a circular arc (geodesic), orthogonal to $T$ at the points $w_1$ and $w_2$, that contains the points $z_1$ and $z_2$ of the unit disk (suppose that the points $w_1, z_1, z_2, w_2$ occur in this order). Since $(r, 0, -1, 1) = (1+r)/(1-r)$, we find

$$\lambda(z_1, z_2) = \ln(z_2, z_1, w_1, w_2).$$

We leave to the interested reader to check that $\{z_1, z_2\} = (z_2, z_1, w_1, w_2) > 0$, if the points are in the order indicated above.

In this form we can consider $\lambda$ as the oriented distance which changes the sign of the permutation $z_1$ and $z_2$. Additivity of the distance on geodesics follow from $(z_2, z_1, w_1, w_2) = (z_2, z_3, w_1, w_2)(z_3, z_1, w_1, w_2)$.

We summarize those arguments in the following:

**Theorem 1.**

$$\lambda_{\mathbb{U}} = \ln\frac{1+\delta_{\mathbb{U}}}{1-\delta_{\mathbb{U}}},\quad \lambda_{\mathbb{H}} = \ln\frac{1+\delta_{\mathbb{H}}}{1-\delta_{\mathbb{H}}}\,. \tag{21}$$

**Example 4.**

1. *Let $f$ be a Möbius transformation, $z_1$, $z_2$, $z_3$, $z_4$ four numbers from $\overline{\mathbb{C}}$, $w_k = f(z_k)$, $A = (z_2, z_3, z_4)$ and $B = (w_2, w_3, w_4)$. Then $S_B \circ f \circ S_A^{-1} = \mathrm{Id}$ and therefore $f = S_B^{-1} \circ S_A$ and $S_B \circ f = S_A$. In particular, $S_B(w_1) = S_A(z_1)$ and it says that the cross-ratio is invariant under Möbius transformation.*

2. *If $z_1, z_2, z_3, z_4$ are four distinct points, then $(z_1, z_2; z_3, z_4)$ is a real number if and only if all four points belong to a circle. We say that a function $f : U \to U$ is conformally conjugate to a function $g : V \to V$, if there is a conformal map $\phi : U \to V$ such that $g = \phi \circ f \circ \phi^{-1}$.*

3. *Let $K = \mathbb{T}(a, R)$ and $w = Az = a + Rz$. Then $A$ maps $\mathbb{T}$ onto $K$. Define $S = A \circ \overline{J} \circ A^{-1}$. Since $z = \dfrac{w-a}{R}$, we find $\overline{J} \circ A^{-1}(w) = \dfrac{R}{\bar{w}-\bar{a}}$ and hence $Sw = a + \dfrac{R^2}{\bar{w}-\bar{a}}$.*
*We say that $w$ and $Sw$ are symmetric with respect to $K$.*

4. *Let $K$ be a circle through the points $z_2, z_3, z_4$. Show that $z$ and $z^*$ are symmetric with respect to $K$ if $(z^*, z_1, z_2, z_3) = \overline{(z, z_1, z_2, z_3)}$.*

5. *For $a \in \mathbb{D}$ ($a \neq 0$), define $R = R(a) = (|a^*|^2 - 1)^{1/2} = \sqrt{1-|a|^2}/|a|$. The circle $K = \mathbb{T}(a^*, R(a))$ is orthogonal to the unit circle $\mathbb{T}$. The reflection (inversion) with respect to this circle is given by*

$$\sigma_a z = a^* + R(a)^2(z-a^*)^* = -\frac{a}{\bar{a}}\frac{\bar{z}-\bar{a}}{1-a\bar{z}}. \tag{22}$$

*If $a = |a|e^{i\alpha}$, the mapping $r$, defined by $r(w) = -e^{i2\alpha}\bar{w}$, is a reflection with respect to the line orthogonal to $a$ which contains the origin 0. Define $\varphi_a(z) = \dfrac{z-a}{1-\bar{a}z}$ and $\varphi_a = r \circ \sigma_a$. Check that $\varphi_a$ maps $\mathbb{D}$ onto itself.*

6. *Let $f : \mathbb{D} \to \mathbb{D}$ is an arbitrary analytic map. Fix an arbitrary point $z \in \mathbb{D}$ and consider the mapping $F = \varphi_w \circ f \circ \varphi_{-z}$, where $w = f(z)$. Since $\varphi_{-z}(0) = z$, $F(0) = 0$. By an application of Schwarz lemma, $|F(\zeta)|| \leqslant |\zeta|, \quad \zeta \in \mathbb{D}$. Hence, if $\omega$ is an arbitrary point in $\mathbb{D}$ and $\zeta = \varphi_z(\omega)$, then $\varphi_{-z}(\zeta) = \omega$ and we find $|\varphi_w(f(\omega))| \leqslant |\varphi_z(\omega)|$. Thus,*

$$\delta(f(z), f(\omega)) \leqslant \delta(z, \omega), \tag{23}$$

*with equality only if $f$ is a Möbius transformation of $\mathbb{D}$ onto itself. Note, in particular, if $f \in \mathrm{Aut}\mathbb{D}$, then $F \in \mathrm{Aut}\mathbb{D}$ and $F(0) = 0$. Hence $F = e^{i\alpha}\mathrm{Id}$ and therefore $\varphi_w(f(\omega)) = e^{i\alpha}\varphi_z(\omega)$, which shows that $|\varphi_w(f(\omega))| = |\varphi_z(\omega)|$. This gives another motivation to define pseudo hyperbolic distance $\delta$ by $\delta(z, \omega) = |\varphi_z(\omega)|$, which is invariant under automorphisms of $\mathbb{D}$. For this exercise see the subsection Schwarz lemma below.*

7. *Let $0 < r < \rho < 1$, $K_\rho = \{|z| = \rho\}$, $\varphi_r(z) = \frac{z-r}{1-rz}$, $a = \varphi_r(\rho)$ and $b = -\varphi_r(-\rho) = \frac{\rho+r}{1+r\rho}$. Show that $a < b$ and that $\varphi_r$ maps the circle $K_\rho$ onto the circle $K(\rho, r) = \{|z - \frac{a-b}{2}| = \frac{a+b}{2}\}$. Hence, $\delta(r, z) \leqslant b$, for $z \in K_\rho$ and therefore*
*(i1) $\delta(z, w) \leqslant |z| + |w|$.*
*Let $a, z, w \in \mathbb{D}$. Using automorphism $\varphi_a$, (i1) yields*
*(i2) $\delta(z, w) \leqslant \delta(z, a) + \delta(w, a)$.*
*Show that $[-r, r]$ is geodesic with respect the metric $\delta = \delta_{\mathbb{D}}$. Since $\delta(-r, r) = \frac{2r}{1+r^2}$, $\delta(-r, 0) = \delta(0, r) = r$, we have $\delta(-r, r) < \delta(-r, 0) + \delta(0, r)$ and hence $\delta$ is not additive on geodesics. Therefore we call $\delta$ pseudo hyperbolic distance.*

8. *If $\gamma$ piecewise continuously differentiable path in $\mathbb{D}$, show that $|\gamma|_{\mathrm{hyp}} = |\gamma|_\delta$, where $|\gamma|_\delta$ denotes pseudo-hyperbolic length of the curve $\gamma$.*

9. *Let $0 < y_1 < y_2$ and $\mathrm{I}_0$ the vertical segment that connects $iy_1$ and $iy_2$. Show*
*(i3) $|\mathrm{I}_0|_{\mathrm{hyp}} = \int\limits_{y_1}^{y_2} \frac{1}{y}\, dy = \ln(y_2/y_1)$.*
*If $\gamma$ is a path which joins $iy_1$ and $iy_2$, using obvious geometric consideration, show that*
*(i4) $|\gamma|_{\mathrm{hyp}} \geqslant |\mathrm{I}_0|_{\mathrm{hyp}}$*
*and therefore $\lambda_{\mathbb{H}}(iy_1, iy_2) = \ln(y_2/y_1)$.*
*Hint. Take a partition $P_n : u_0 < u_1 < \cdots < u_n$, $u_0 = y_1, u_n = y_2$ and set $\delta_n = \max\{u_k - u_{k-1} : 1 \leqslant k \leqslant n\}$ and $l_k = \{x + iu_k : x \in \mathbb{R}\}$, $1 \leqslant k \leqslant n$. If $\gamma_k$ is a subarc of the path $\gamma$ which joins $l_{k-1}$ and $l_k$, $1 \leqslant k \leqslant n$, then $|\gamma_k|_e \geqslant u_k - u_{k-1}$ and therefore*

$$\sum_{k=1}^{n} \frac{u_k - u_{k-1}}{u_k} \leqslant \sum_{k=1}^{n} \frac{|\gamma_k|_e}{u_k}. \tag{24}$$

*Hence, by letting $n$ to $\infty$, we find (i4).*

### 2.3. The area of a hyperbolic triangle

For this subsection see also [2]. For $z, z + h \in \mathbb{D}$, set $e = e(h) = \sigma_{\mathbb{D}}(z, z + h)$. Since $e(h) = 2(1 - |z|^2 - h\overline{z})^{-1}$ and $d_{hyp,\mathbb{D}}(z, z + h) = \ln \frac{1+e(h)}{1-e(h)} = 2e(h) + o(e(h))$, when $e(h)$ tends to 0, we find
(1) $\quad d_{hyp,\mathbb{D}}(z, z + h)/|h|$ tends to $\frac{2}{1-|z|^2}$, if $h$ tends to 0.
In a similar way, we prove, for $z = x + iy \in \mathbb{H}$, that
(1') $\quad d_{hyp,\mathbb{D}}(z, z + h)/|h|$ tends to $y^{-1}$, if $h$ tends to 0.
Note that, for $z = x + iy \in \mathbb{H}$, that
(*) $\quad \sigma_{hyp,\mathbb{D}}(z, z + h)/|h|$ tends to $y^{-1}$, if $h$ tends to 0. Thus $y^{-1}$ is hyperbolic density and it defines $d_{hyp}$ hyperbolic metric in a standard way. Since $\sigma_{hyp}$ is invariant for $Aut(\mathbb{H})$, also $d_{hyp,\mathbb{H}}$ is invariant. $L_0 = \{iy : y > 0\}$ is a right line (geodesic) and $T(L_0)$ is a right line (geodesic), for every $T \in Aut(\mathbb{H})$.

Now, consider model on $\mathbb{H}$. Let $z = x + iy \in \mathbb{H}$, $T_z$ denote tangent space at $z$; vector $h$ belongs $T_z$ if begins at $z$. By (1') define for $v \in T_z$ non Euclidean norm (L -norm ) by $|v|_L = |v|/y^2$. If $v_1, v_2 \in T_z$, then Euclidean area of parallelogram $R$ defined by these vectors is
(2) $\quad P = P_e = P_{euc} = |v_1||v_2| \sin\theta$, where $\theta$ is measure number of convex angle between these vectors.

We define non Euclidean area of parallelogram $R$ replacing Euclidean norms in (2) with non Euclidean norms:

$$P_L = |v_1|_L |v_2|_L \sin\theta. \tag{3}$$

We will use also notation $A_{hyp}$ for non Euclidean area. Note that $A_{hyp}(R) = \frac{1}{y^2} P_{euc}(R)$.

Let $W$ be domain in $\mathbb{H}$ and let $Q_n$ be square net with edges of length $1/n$, $n \geqslant 1$, and denote by $A_n$ the sum of non Euclidean area of squares which belong to $W$; it follows that $A_n$ tends to

$$A = |W|_L = \int_W \frac{dxdy}{y^2} \tag{4}$$

if $n$ tends to $\infty$. We call $|W|_L$ non Euclidean area of domain $W$.

We can count area of suitable domains.

Let $y = u(x), x_1 \leqslant x \leqslant x_2$ be an arc and $W$ elementary domain defined by

$$W = \{(x, y) : y > u(x), x_1 \leqslant x \leqslant x_2\}.$$

By Fubini's theorem we find

$$A = \int_W \frac{dxdy}{y^2} = \int_{x_1}^{x_2} dx \int_{u(x)}^{\infty} \frac{dy}{y^2}$$

and hence

$$|W|_L = \int_{x_1}^{x_2} \frac{1}{u(x)} dx. \tag{5}$$

If we denote by $c$ arc defined by $c(x) = x + iu(x), x_1 \leqslant x \leqslant x_2$, we can rewrite formula as

$$|W|_L = \int_c \frac{dx}{y} = \int_{x_1}^{x_2} \frac{1}{u(x)} dx\,. \tag{6}$$

In the special case when $c$ is a circular arc we find interesting formula.

Let $K$ be the circle $(x-a)^2 + y^2 = r^2$ and $A, C \in K$ points in $\mathbb{H}$ and denote by $l$ the arc on of the circle which join $A = a + re^{i\theta_1}$ and $C = a + re^{i\theta_2}$, $\theta_1 < \theta_2 < \theta_1 + \pi$ in positive direction . Since $yy' = -(x-a)$ we find $\sqrt{1 + (y')^2} = r/|y|$, Euclidean length of arc $l$ is $|l| = r \int_l dx/|y| = r|l|_L$.

Using polar coordinates $x - a = rcos\theta$, $y = rsin\theta$, we find

(7) $\quad |l| = r \int_l d\theta = r(\theta_2 - \theta_1)$.

Note that the formula (7) follows from measure of angle (see [7, 8]). Denote by $T = l'(\theta)$ tangent vector i defini.imo $argT = \theta + \pi/2$; if $r = 1$, Euclidean length of $l$ equals to change of tangent angle. In special case, if $a \in R$, arc $l$ is a non Euclidean line. Non Euclidean triangle $\Delta(A, \infty, C)$, whose two vertex $A$ and $C$ belong to the unit circle, and third vertex $B = \infty$, we call special triangle; it is interesting that, by (6) and (7), it follows that non Euclidean area of special triangle equals Euclidean length of the arc (the edge of the triangle), which belongs to the unit circle. Hence non Euclidean area of special triangle $\Delta(A, \infty, C)$ equals change of tangent angle along arc $AC$ of the triangle. For triangle $\Delta(A, \infty, C)$ let $r$ be radius of the circle which contains Euclidean arc $AC$. The altitude change of angle that the tangent of arc $AC$ makes with the positive $x$- axis along arc $AC$ is $(\pi - \gamma) - \alpha)$; hence, if $r = 1$, the triangle is special and non Euclidean area $\Delta(A, \infty, C)$ equals Euclidean length of the arc AC and hence $(\pi - \gamma) - \alpha = \pi - (\gamma + \alpha)$.

Note that we first consider the case $r = 1$ only from pedological reasons. If $r \neq 1$, using a homothety we can map $K$ on unit circle. Note that non Euclidean length and area are homothety invariant. Let $\Delta(A, B, C)$ non Euclidean triangle; without loss of generalization we can assume that vertices $B$ and $C$ on $y$ - axis as on the picture. Denote by $A_L, A, A$ respectively non Euclidean area of triangle $\Delta(A, B, C)$, $\Delta(A, \infty, C)$ and $\Delta(A, \infty, C)$. Since, by (13) , $A_1 = \pi - (\alpha_1 + \gamma_1)$, $A_2 = \pi - (\alpha_1 + \beta_1)$, and $A_1 = A_L + A_1$, it follows (14) $A_L = \pi - (\gamma + \alpha + \beta)$.

Thus we have proved:

**Theorem 2.** *In a hyperbolic triangle the sum of the angles A, B, C (respectively opposite to the side with the corresponding letter) is strictly less than a straight angle. The difference between the measure of a straight angle*

*and the sum of the measures of a triangle's angles is called the defect of the triangle. The area of a hyperbolic triangle is equal to its defect* $(\pi - A - B - C)$.

By Stokes' formula area $A$ of non Euclidean polygon is

$$A = \int_P \frac{dxdy}{y^2} = -\int_P d(\frac{1}{y})\, dx = \int_{\partial P} \frac{dx}{y}\,.$$

Using polar coordinates $z - a = re^{i\theta}$, $dx/y = -d\theta$, if we denote by $\alpha_k$ and $\beta_k$ measures of interior and exterior angles of polygon, then

$$A = \sum \beta_k - 2\pi = (n-2)\pi - \sum \alpha_k\,.$$

Now, we will compute area $A$ of non Euclidean $n$- polygon. We will not use partition of non Euclidean $n$-polygon on non Euclidean triangles. We need:

**Theorem 3** (Theorem of Turning Tangents)**.** *Let $c$ be a positively oriented parametrization of $C$ and let $\varphi(s)$ be the angle from $e_1$ to $c'$ at the point $c(s)$. Then $\int \varphi'(s)ds = 2\pi$.*

1. First, we outline an argument.
By Stokes' formula area $A$ of non Euclidean polygon is

$$A = \int_P \frac{dxdy}{y^2} = -\int_P d(\frac{1}{y})\, dx = \int_{\partial P} \frac{dx}{y}\,.$$

Using polar coordinates $z - a = re^{i\theta}$, $dx/y = -d\theta$, if we denote by $\alpha_k$ and $\beta_k$ measures of interior and exterior angles of polygon, then

$$A = \int_{\partial P} d\theta = \sum \beta_k - 2\pi = (n-2)\pi - \sum \alpha_k\,.$$

Thus, if $S$ is sum of measures of interior angles of non Euclidean $n$- polygon, then non Euclidean area $A = (n-2)\pi - S$.

2. Now we give additional details.
Let $c$ be circle arc of circle $(x-a)^2 + y^2 = r^2$, $a \in \mathbb{R}$. Using polar coordinates $z - a = re^{i\theta}$, $T = c'(\theta) = re^{i(\theta+\pi/2)}$ and therefore $argT = \theta + \pi/2$ and $\int_c d\theta = \Delta_c Arg\, T$.

Here we used Theorem of Turning Tangents: Let $\partial P$ consist of circle arcs $c_k$. Set $a_k = \Delta_{c_k} Arg\, T$ and $S = \sum \alpha_k$. Then $A = \int_{\partial P} \theta = -\sum \Delta_{c_k} Arg\, T = -\sum a_k$. By Theorem 3, $\sum(a_k + \beta_k) = 2\pi$ and therefore $A = \sum \beta_k - 2\pi$. Since $\beta_k = \pi - \alpha_k$, we find $A = (n-2)\pi - S$.

In terms of the Poincaré half-plane model absolute length corresponds to the infinitesimal metric $ds = \dfrac{|dz|}{\mathrm{Im}(z)}$ and in the Poincaré disk model to $ds = \dfrac{2|dz|}{1-|z|^2}$.

In terms of the (constant and negative) Gaussian curvature $K$ of a hyperbolic plane, a unit of absolute length correspond to a length of $R = \dfrac{1}{\sqrt{-K}}$.

In a hyperbolic triangle the sum of the angles $A, B, C$ (respectively opposite to the side with the corresponding letter) is strictly less than a straight angle. The difference between the measure of a straight angle and the sum of the measures of a triangle's angles is called the defect of the triangle. The area of a hyperbolic triangle is equal to its defect multiplied by the square of $R$:

$$(\pi - A - B - C)R^2.$$

In all the formulas stated below the sides $a, b$, and $c$ must be measured in absolute length, a unit so that the Gaussian curvature K of the plane is $-1$. In other words, the quantity R in the paragraph above is supposed to be equal to 1.

## 3. Further results

There is interesting connection of hyperbolic geometry with complex geometry. The author published a paper [6] about holomorphic fixed point theorem on Riemann surfaces.

Let $M$ and $N$ be hyperbolic Riemann surfaces and $f : M \to N$ an analytic function. If $p$ is fixed point, then $|f'(p)| \leqslant 1$.

**Theorem 4.** *(i) Let $M$ be hyperbolic Riemann surface and $f : M \to M$ analytic function and $F$ compact subset of $M$. If $f$ is not isometry, then $f$ is contraction on $F$. In addition, if $f(F) \subset F$, then there is a unique fixed point $p_0 = f(p_0) \in F$.*

Let $G$ be bounded connected open subset of complex Banach space, $p \in G$ and $v \in T_pG$. We define $k_G(p, v) = \inf\{|h|\}$, where infimum is taking over all $h \in T_0\mathbb{C}$ for which there exists a holomorphic function such that $\phi : \mathbb{U} \to G$ such that $\phi(0) = p$ and $d\phi(h) = v$.

One can prove the following theorems.

**Theorem 5.** *Suppose that $G$ and $G_1$ are bounded connected open subset of complex Banach space and $f : G \to G_1$ is holomorphic. Then $k_{G_1}(fz, fz_1) \leqslant k_G(z, z_1)$, for all $z, z_1 \in G$.*

**Theorem 6.** *Suppose that $G$ is bounded connected open subset of complex Banach space and $f : G \to G_*$ is holomorphic, $G_* \subset G$, $s_0 = dist(G_*, G^c)$, $d_0 = diam(G)$ and $q_0 = \frac{d_0}{d_0+s_0}$. Then $k_{G_*}(fz, fz_1) \leqslant q_0 k_{G_*}(z, z_1)$, for $z, z_1 \in G_*$.*

We worked on the subject from time to time between 1980 -1990 and in that time we proved Theorems 6-5([1]). But we realized these days that it is a version of the Earle-Hamilton (1968) fixed point theorem, which may be viewed as a holomorphic formulation of Banach's contraction mapping theorem. A version of this result was proved in 1968 (when I enroled Math Faculty) by Clifford Earle and Richard Hamilton by showing that, with respect to the Carathéodory metric on the domain, the holomorphic mapping becomes a contraction mapping to which the Banach fixed-point theorem can be applied. Perhaps there are applications of this result in the Teichmüller theory.

## 4. Concluding comments

In this section we only touch some questions related to place of Euclid's theory, hyperbolic geometry in science and real words.

Euclid's Elements is a mathematical and geometric treatise consisting of 13 books attributed to the ancient Greek mathematician Euclid in Alexandria, Ptolemaic Egypt circa 300 BC. It is a collection of definitions, postulates (axioms), propositions (theorems and constructions), and mathematical proofs of the propositions.

The Elements is still considered a masterpiece in the application of logic to mathematics. In historical context, it has proven enormously influential in many areas of science.

The reason that Euclid was so influential is that his work is more than just an explanation of geometry or even of mathematics. He was first to the develop concepts as axioms and proofs. The way in which he used logic and demanded proof for every theorem shaped the ideas of western philosophers right up until the present day. Great philosopher mathematicians such as Descartes and Newton presented their philosophical works using Euclid's structure and format, moving from simple first principles to complicated concepts.

The Pythagorean theorem is false if the parallel postulate does not hold.

Unless you have studied mathematics based on an axiomatic theory (like Euclid's, or number theory based on the Dedekind-Peano axioms, or set theory based on Zermelo-Fraenkel axioms) you have not seen real mathematics. All that arithmetic, algebra, trigonometry, and calculus you have studied for years and years is not real mathematics until every statement is proved, and that requires axioms, definitions, and theorems with proof.

As early as 1899, Hilbert proposed a whole new formal set of geometrical axioms, known as Hilbert's axioms, to substitute the traditional axioms of Euclid. Hilbert's approach signaled the shift to the modern axiomatic method. In this, Hilbert was anticipated by Moritz Pasch's work from 1882. Axioms are not taken as self-evident truths. Geometry may treat things, about which we have powerful intuitions, but it is not necessary to assign any explicit meaning to the undefined concepts. The elements, such as point, line, plane, and others, could be

[1] we found a my hand written manuscript 1990 and did not pay much attention to it at that time

substituted, as Hilbert is reported to have said to Schoenflies and Kötter, by tables, chairs, glasses of beer and other such objects.It is their defined relationships that are discussed.

Hilbert first enumerates the undefined concepts: point, line, plane, lying on (a relation between points and lines, points and planes, and lines and planes), betweenness, congruence of pairs of points (line segments), and congruence of angles. The axioms unify both the plane geometry and solid geometry of Euclid in a single system. If we study mathematics we ask some question. Whether there is a rectangular triangle in the real world, whether the version of Pythagorean theorem for Hyperbolic geometry or Euclidean geometry is valid? What is a straight line in the real world? Some physicists believe that the a straight line path of light and brightness beam bend when passing by the country. Such questions do not study the mathematics. The mathematics is defined Riemann manifold and geodesic (but such objects do not exist in the real world). Mathematics is applied to the models and approximations. All science together contribute to the progress of knowledge about the real world. I personally believe that the human mind will never understand the real world completely. "Because of experimental error, a physical experiment can never prove conclusively that space is Euclidean - it can prove only that space is non-Euclidean"(Greenberg, 1993, p. 291). "Hyperbolic Geometry is a "curved" space, and plays an important role in EinsteinŠs General theory of Relativity" (Castellanos,2002).

Here we follow shortly discussion in [14]. In mathematics we assuming some things are true (axioms), and then we ask what other things would be true as well. We do not discuss whether the things we are assuming to be true really are true, but if we run across a world where they are really true, then we may be sure that anything else we deduce logically from them will also be true in that world. So there are statements we take for granted, called axioms or postulates or assumptions, and then there are statements called theorems that we deduce or ”prove” from our assumptions. So we do not know that our theorems are really true, but in any world where the assumptions are true, then the theorems are also true. In Euclidean geometry we describe a special worlds, a Euclidean plane, line and point. Those object do not really exist in the real world we live in, but we pretend it does, and we try to learn more about that perfect world. So when we "prove" a statement in Euclidean geometry, the statement is only proved to be true in a perfect or "ideal" Euclidean plane, but not on the paper we are drawing on, or the world we are living in.

Edward Frenkel, a mathematician at Berkeley, said: I argue, as others have done before me, that mathematical concepts and ideas exist objectively, outside of the physical world and outside of the world of consciousness. We mathematicians discover them and are able to connect to this hidden reality through our consciousness. If Leo Tolstoy had not lived we would never have known Anna Karenina. There is no reason to believe that another author would have written that same novel. However, if Pythagoras had not lived, someone else would have discovered exactly the same Pythagoras theorem. Moreover, that theorem means the same to us today as it meant to Pythagoras 2,500 years ago.

## References


[1] **L.V. Ahlfors.** Conformal invariants. *McGraw-Hill Book Company*, 1973.

[2] **L.V. Ahlfors.** Möbius transformations in several dimensions. *Lecture Notes, University of Minnesota*, 1981.

[3] **J.W. Anderson.** Hyperbolic Geometry. *Springer, London, UK*, 1999.

[4] **J.W. Cannon, W.J. Floyd, R. Kenyon and W.R. Parry.** Hyperbolic Geometry, *Flavors of Geometry MSRI Publications* Volume 31, 1997.
http://library.msri.org/books/Book31/files/cannon.pdf

[5] **K.J. Devlin**. The Language of Mathematics: Making the Invisible Visible. *Macmillan*, 2000, p. 161. ISBN 0-8050-7254-3

[6] **M. Mateljević.** Holomorphic fixed point theorem on Riemann surfaces. *Math. Balkanica* **12** (1-2) (1998), 1-4.

[7] **M. Mateljević.** Kompleksne Funkcije 1 & 2. *Društvo matematičara Srbije*, 2006.

[8] **M. Mateljević.** Topics in Conformal, Quasiconformal and Harmonic maps. *Zavod za udžbenike*, Beograd 2012.

[9] **S. Vukmirović.** *Modeli geometrije Lobačevskog.* http://alas.matf.bg.ac.rs/ vsrdjan/files/geomlob.pdf.

[10] **E.W. Weisstein.** "The parallel postulate is equivalent to the Equidistance postulate, Playfair axiom, Proclus axiom, the Triangle postulate and the Pythagorean theorem." *CRC concise encyclopedia of mathematics (2nd ed.)*, 2003, p. 2147. ISBN 1-58488-347-2.

[11] https://en.wikipedia.org/wiki/Parallel_postulate

[12] https://en.wikipedia.org/wiki/Pythagorean_theorem

[13] https://en.wikipedia.org/wiki/Absolute_geometry

[14] http://alpha.math.uga.edu/ roy/camp2011/10.pdf

Симпозијум МАТЕМАТИКА И ПРИМЕНЕ, Математички факултет, Универзитет у Београду, 2015 ,Vol. VI(1)

# Гребнерове базе – од топологије до алгебарске комбинаторике

**Зоран З. Петровић**

*Математички факултет, Београд*
*e-mail: zoranp@matf.bg.ac.rs*

**Апстракт.** У раду је дат кратак приказ неких проблема који се могу решавати коришћењем Гребнерових база. У ту сврху, разјашњене су основни појмови у вези Гребнерових база. Показано је како се оне могу користити за рачунање у кохомологији Грасманових многострукости. Дате су и примене добијених резултата у диференцијалној топологији и алгебарској комбинаторици.

**Кључне речи:** Гребнерове базе; Грасманијани; имерзије; Косткини бројеви.

## 1. Увод

Појам Гребнерове базе уведен је у докторату Бруна Бухбергера из 1965. године. Гребнерове базе имају велику примену у разним областима. Ми ћемо се овде позабавити неким новим применама до којих је дошла група београдских математичара. Резултати ових радова су приказани у десетак радова објављених у математичким часописима, као и у две докторске дисертације (видети [16], [18]).

Овај рад је прерађена варијанта предавања са VI Симпозијума Математика и примене. Како је ово било пленарно предавање, коме су присуствовали сви учесници симпозијума, то је и овај рад таквог карактера да од њега могу имати користи и они који се не баве темама о којима је реч, а надамо се да може бити од користи и наставницима математике, који редовно узимају учешће на овим симпозијумима. Наведимо укратко о чему ће бити речи у одељцима који следе.

Одељак 2, посвећен је кратком прегледу основних појмова који се тичу многострукости. Учињен је покушај да се ови појмови учине доступним свима, уз подсећање на појмове са којима смо се сретали током школовања. Наведени су и неки занимљиви примери, који нису баш директно везани за Гребнерове базе. Искоришћена је прилика да се укратко укаже на основне методе које алгебарска топологија користи за доказивање неких геометријских резултата.

Одељак 3, посвећен је увођењу појма Гребнерових база и ту је дато и упутство како се може доказати постојање Гребнерове базе за произвољни идеал. Доста пажње је посвећено и једноставном случају прстена полинома са једном неодређеном у покушају да се питање постојања Гребнерове базе мотивише и на тај начин.

Одељак 4 је посвећен основним појмовима теорије векторских раслојења и карактеристичних класа, као и вези проблема имерзије са теоријом хомотопије.

У одељку 5 приказани су нови резултати који су добијени у гореспоменутим радовима. Наравно, они су само делимично могли бити приказани, а идеја је да се можда читаоци привуку да консултују оригиналне референце.

## 2. Многострукости

### 2.1. Основни појмови

У школи смо учили о кривама, било да се ради о графицима функција, било да се ради о кривама задатим једноставним (линеарним, или квадратним) једначинама. Ако посматрамо криву у равни задату неком једначином, на пример, кружницу, знамо да не можемо у потпуности једну од координата изразити преко друге координате, али то можемо да урадимо локално. На пример, све тачке кружнице задате једначином $x^2 + y^2 = 1$ за које је $x < 0$ имају својство да је вредност координате $x$ потпуно одређена вредношћу координате $y$; ту је, наиме, $x = -\sqrt{1 - y^2}$. Заправо, постоји бијекција између тог дела кружнице и дела $y$-осе задатог са $-1 < y < 1$. На аналогни начин имамо и бијекцију између дела кружнице задате

са $x > 0$ и горенаведеног дела $y$-осе, као и бијекције за случајеве $y < 0$ и $y > 0$ и одговарајућег дела $x$-осе. Овакве бијекције називамо и *карте*. Дакле, свака тачка на кружници је *локално* одређена једном координатом (док је број потребних карата 4). Стога је кружница један *једнодимензиони* објекат. Слично се може закључити и за за хиперболу задату једначином $x^2 - y^2 = 1$, при чему је хипербола, за разлику од кружнице неповезана, али то није чињеница од централног значаја. Минималан број карата за хиперболу је 2.

На факултету смо учили и о површима. Било да су то површи задате као графици функција две промењљиве, било да су задате једначинама (попут елипсоида, хиперболоида, дакле конусних пресека), или су задате на неки други начин – на пример торус је таква површ. И, слична ствар се дешава и у овом случају. Без обзира на различитост задавања ових површи, увек се *локално* оне могу задати помоћу, овај пут две, координате. Број карата ће зависити од конкретне површи са којом радимо.

Но, чак и ако желимо да изучавамо неке особине тродимензионог простора, не можемо се ограничити само на три димензије. На пример, ако желимо да посматрамо простор који чине сви могући положаји 6 различитих тачака у тродимензионом простору, биће нам потребно 18 координата (по три за сваку тачку) од којих неке морају бити различите (дакле, тај простор није $\mathbb{R}^{18}$). Још је занимљивије ако изучавамо све могуће положаје 6 различитих тачака на торусу. Евидентно се ради о занимљивом објекту, који је доста правилан, али свакако се не може видети као подскуп тродимензионог простора.

Ови, али и многи други примери из разних области математике и њених примена су условили потребу за увођењем појма *многострукости*. Наведимо дефиницију тог појма.

**Дефиниција 1.** Многострукост димензије $n$, или $n$-димензиона многострукост је тополошки простор који има следећа својства:
(1) $M$ је Хауздорфов.
(2) $M$ је локално еуклидски димензије $n$.
(3) $M$ има пребројиву базу отворених скупова.

Прокоментаришимо мало ову дефиницију. Својство (1) има за циљ да избегне неке необичне примере многострукости. Замислимо да имамо два примерка реалне праве, рецимо $\mathbb{R} \times \{-1\}$ и $\mathbb{R} \times \{1\}$ и да их *слепимо* свуда сем у координатном почетку (другим речима да идентификујемо тачке $(x, -1)$ и $(x, 1)$ за све $x \in \mathbb{R} \setminus \{0\}$. Тако бисмо добили помало егзотичну праву са два координатна почетка. То није објекат који нас занима. Својство (3) је технички услов који неки стављају, а неки и не, али се нећемо задржавати даље на њему. Оно што је централно је својство (2). Оно означава да свака тачка $p$ многострукости $M$, има отворену околину $U$ у $M$ и да постоји хомеоморфизам $\varphi$ између те околине и неке отворене кугле у $\mathbb{R}^n$ (или и целог $\mathbb{R}^n$ – то је еквивалентан услов). Хомеоморфизам је непрекидна бијекција, чији је инверз такође непрекидан. То одговара горенаведеној бијекцији између дела кружнице и дела једне од координатних оса. Са $(U, \varphi)$ означаваћемо једну такву координатну карту. Дакле, $M$ је унија тих околина $M = \cup_{i \in I} U_i$, при чему постоји хомеоморфизам $\varphi_i : U_i \longrightarrow \mathbb{R}^n$. Заправо, с обзиром на својство (3) у дефиницији многострукости, може се претпоставити да је скуп $I$ пребројив; у даљем ћемо се искључиво бавити *компактним* многострукостима код којих се додатно може претпоставити да је тај скуп $I$ коначан.

Овако смо задали појам *тополошке* многострукости. Нас заправо занима појам *диференцијабилне* (*глатке*) многострукости, што је, кратко речено, многострукост на којој се може изводити диференцијални рачун. Она се добија малом модификацијом претходне дефиниције. Наиме, претпоставља се и да су *функције промене координата* $\varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \longrightarrow \varphi_j(U_i \cap U_j)$ глатке функције (ово су функције чији су и домен и кодомен отворени подскупови у $\mathbb{R}^n$ и зато има потпуно смисла говорити о њиховој глаткости, тј. о постојању и непрекидности извода свих редова).

На овај начин можемо да разматрамо и глатка пресликавања међу многострукостима. Наиме, ако су $M$ и $N$ глатке многострукости (које не морају имати исту димензију) и $f \colon M \longrightarrow N$ непрекидно пресликавање, онда кажемо да је оно глатко уколико су све функције $\psi_j \circ f \circ \varphi_i^{-1}$ глатке као функције међу отвореним подскуповима одговарајућих еуклидских простора (овде функције $\psi_j$ и $\varphi_i$ задају одговарајуће карте на датим многострукостима).

Нас ће у даљем посебно занимати следећи примери многострукости.

1) Реални пројективни простор димензије $n$, у ознаци $\mathbb{R}P^n$ је скуп свих правих у $\mathbb{R}^{n+1}$ које садрже координатни почетак. Нећемо сада улазити у то како се задаје структура глатке многострукости (за овај и наредни пример, препоручујемо читаоцу да погледа, на пример, књигу [7], која је корисна и за многе друге појмове који ће се у даљем разматрати). Истакнимо само да се природно на овом скупу може задати метрика тако што се за растојање између две праве узме угао између њих. Тада је неједнакост троугла заправо један познати резултат из осмог разреда основне школе (који?).

2) Општије од претходног: можемо разматрати све $k$-димензионе векторске потпросторе векторског простора $\mathbb{R}^{n+k}$ (у претходном примеру су разматрани једнодимензиони потпростори простора $\mathbb{R}^{n+1}$). Многострукост која се овде добија назива се Грасманова многострукост, или, краће, Грасманијан. Ознака је $G_{k,n}(\mathbb{R})$. Димензија ове многострукости је $kn$ (кратко, али не и потпуно објашњење: простор $\mathbb{R}^{n+k}$ је директна сума потпростора $\mathbb{R}^k$ и $\mathbb{R}^n$ – тада се $k$-димензиони потпростор може видети као график линеарног пресликавања из $\mathbb{R}^k$ у $\mathbb{R}^n$, односно као баш то линеарно пресликавање, односно као матрица формата $n \times k$, а она има $kn$ компоненти).

3) Најсложенији пример добијамо када у $\mathbb{R}^n$ разматрамо све низове потпростора $(V_1, \ldots, V_r)$, који чине ортогоналну (у односу на стандардни скаларни производ) декомпозицију простора $\mathbb{R}^n$ и чије су димензије редом $n_1, \ldots, n_k$ (стога је $n_1 + \cdots + n_r = n$). Скуп свих ових потпростора означавамо са $F_{\mathbb{R}}(n_1, \ldots, n_r)$ и на њему постоји структура многострукости. Та многострукост се назива многострукост застава типа $(n_1, \ldots, n_r)$. Њена димензија је $\sum_{1 \le i < j \le r} n_i n_j$.

Наведимо на крају овог одељка само још то да се одговарајуће многострукости задају и над комплексним бројевима. Сви ови примери су примери компактних многострукости.

### 2.2. Утапања и имерзије

Као што смо видели у претходном пододељку, појам многострукости није везан за конкретно представљање у неком еуклидском простору као што смо на то навикли у случају површи које смо изучавали на студијама. Стога се природно може поставити питање да ли се уопште произвољна многострукост може реализовати у неком еуклидском простору. Највише ће нас занимати наши конкретни примери, а овде ћемо навести неке познате резултате који се односе на компактне многострукости (дакле на оне који су компактне као тополошки простори и које стога имају и коначне атласе).

Да бисмо објаснили појам утапања, пођи ћемо од појма имерзије. Посматрамо глатко пресликавање $f\colon M \longrightarrow N$. Нека је $p$ произвољна тачка многострукости $M$. Ако су $(U, \varphi)$ и $(V, \psi)$ одговарајуће карте око тачака $p$, односно $f(p)$, онда се локално функција $f$ види као функција $g$ из неког отвореног подскупа од $\mathbb{R}^m$ (где је $m$ димензија од $M$) у неки отворени подскуп од $\mathbb{R}^n$ ($n = \dim N$): $g(x_1, \ldots, x_m) = (h_1(x_1, \ldots, x_m), \ldots, g_n(x_1, \ldots, x_m))$.

У самој тачки $\varphi(p) \in \mathbb{R}^m$ можемо израчунати Јакобијеву матрицу ове функције $g$, односно матрицу свих парцијалних извода:

$$\mathcal{J}(g) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_m} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1} & \frac{\partial h_n}{\partial x_2} & \cdots & \frac{\partial h_n}{\partial x_m} \end{pmatrix}$$

Наравно, сви парцијални изводи су израчунати у тачки $\varphi(p)$. Лако се може показати да ранг ове матрице не зависи од избора координата (карата).

**Дефиниција 2.** Глатко пресликавање $f\colon M \longrightarrow N$ је *имерзија* уколико је ранг одговарајуће Јакобијеве матрице у свакој тачки многострукости $M$ једнак димензији те многострукости.

Јасно је да Јакобијеву матрицу рачунамо у тачки из $\mathbb{R}^m$, али смо овде због краткоће, написали да се ради о Јакобијевој матрици у тачки многострукости – стандардан пример

злоупотребе језика. Наравно, да би ранг уопште могао да буде једнак $m = \dim M$, мора важити $\dim M \leq \dim N$.

Зашто је овај појам имерзије значајан? Коришћењем теореме о имплицитној функцији из Анализе 2, лако се може установити да је имерзија *локално* '1–1', тј. да за сваку тачку $p \in M$ постоји нека њена околина $W$ за коју је рестрикција функције $f$ на ту околину '1–1'. Према томе, ова функција, локално 'верно представља' многострукост $M$ у многострукости $N$. Али, не и глобално – тј. $f$ уопште не мора бити '1–1'.

**Дефиниција 3.** Глатко пресликавање$f\colon M \longrightarrow N$ је *утапање* уколико је $f$ имерзија, '1–1' и успоставља хомеоморфизам између $M$ и слике $f(M)$.

Ако желимо да истакнемо да је $f$ имерзија, онда то означавамо са $f\colon M \looparrowright N$ (ознака која указује на постојање 'самопресека'), док се утапање означава са $f\colon M \hookrightarrow N$.

Сада се можемо вратити на питање о могућности 'представљања' дате многострукости у неком еуклидском простору. Добро је позната Витнијева (Whitney) теорема из четрдесетих година прошлог века.

**Теорема 1.** *Нека је $M$ компактна $n$-димензиона многострукост. Тада постоји утапање $f\colon M \hookrightarrow \mathbb{R}^{2n}$ и имерзија $g\colon M \looparrowright \mathbb{R}^{2n-1}$.*

Концентрисаћемо се на питање постојања имерзије. Наравно, природно се може поставити питање: може ли се добити и бољи резултат од овог? Заправо није тешко наћи пример да за неке $n$ постоји пример компактне $n$-димензионе многострукости за коју не постоји имерзија у $\mathbb{R}^{2n-2}$. Конкретно, за $n = 2^r$ не постоји имерзија $\mathbb{R}P^{2^r}$ у $\mathbb{R}^{2^{r+1}-2}$ (видети [7]).

Дакле, не може се на овај начин поправити тај резултат. Али, можда може тако да смањење димензије зависи од $n$? Другим речима, можда постоји функција $h(n)$, која није константно једнака 1 и за коју важи да за сваку компактну $n$-димензиону многострукост постоји имерзија у $\mathbb{R}^{2n-h(n)}$. Заправо, одговор је потврдан. Као кулминација дугогодишњег рада многих математичара, Ралф Коен (Ralph Cohen) је доказао следећу теорему (видети [4]).

**Теорема 2.** *Нека је $M$ компактна многострукост. димензије $n$ Тада постоји имерзија $f\colon M \looparrowright \mathbb{R}^{2n-\alpha(n)}$, где је $\alpha(n)$ број јединица у бинарном развоју броја $n$.*

Наравно, практично свако ко види овај резултат први пут, изненади се због појављивања необичне функције $\alpha(n)$. Но, заправо њено појављивање није неприродно. Претпоставимо да је $n = 2^{r_1}+2^{r_2}+\cdots 2^{r_k}$, где је $r_1 > r_2 > \cdots > r_k > 0$ приказ броја $n$ у облику суме степена двојки (дакле говоримо о бинарном запису). Тада је $\alpha(n) = k$. Навели смо да за многострукост $\mathbb{R}P^{2^r}$ не постоји имерзија у $\mathbb{R}^{2^{r+1}-2}$. Није много теже показати да за многострукост $\mathbb{R}P^{2^{r_1}} \times \mathbb{R}P^{2^{r_2}} \times \cdots \times \mathbb{R}P^{2^{r_k}}$ не постоји имерзија у $\mathbb{R}^{2n-k-1}$. Дакле, видимо да се овај Коенов општи резултат не може поправити – за свако $n$ постоји многострукост те димензије за коју не постоји имерзија у $\mathbb{R}^{2n-\alpha(n)-1}$. За више детаља о овој теми читалац може консултовати и рад [9].

Сасвим другачије стоје ствари када постављамо такво питање за неку конкретну многострукост, или неку ужу фамилију многострукости. Видели смо да су реални пројективни простори важни за ову тематику. Заправо, немали број врло озбиљних радова је исписан, а и данас се појављују нови, мада знатно, знатно ређе, који се баве питањем поправљања овог општег резултата у случају конкретних реалних пројективних простора. Занимљиво је да општи резултат за све пројективне просторе није много бољи од општег Коеновог резултата за све компактне многострукости: општи резултат се може поправити највише за 4 (видети [6]). Но, ти су резултати знатно бољи за конкретне многострукости, а има и низ резултата, који доказују непостојање имерзија. Ипак, разлика између димензија за које је доказано постојање имерзије и димензија за које је доказано непостојање, расте неограничено са растом димензије. Тако да ту има доста могућности за поправку. Питање је наравно, с обзиром на тежину проблема, колико је то реално. За детаљније информације о овом питању, упућујемо читаоца на страницу тополога Доналда Дејвиса (Donald Davis) где се могу наћи многи занимљиви радови о овој тематици, као и табеле имерзија за пројективне просторе.

Један од метода за истраживање постојања имерзија је базиран на теорији опструкција. Но, најпре морамо да се позабавимо појмом кохомологије многострукости што је тема следећег пододељка.

### 2.3. Кохомологија

Метод алгебарске топологије састоји се у следећој једноставној идеји. Геометријски (тополошки) објекти су превише 'сложени' и често је тешко директно одговорити на питање које се поставља за њих. Стога је добро превести то питање на неко алгебарско питање којим се можемо лакше позабавити, које је потенцијално 'лакше', мада не обавезно тривијално. У сваком случају је више 'коначне' природе него оригинално питање.

Како то извести? Поједностављено гледано, у топологији проучавамо тополошке просторе и непрекидна пресликавања међу њима, док у алгебри проучавамо алгебарске структуре и хомоморфизме међу тим структурама. Дакле, просторима треба придружити неке алгебарске структуре, а непрекидним пресликавањима хомоморфизме. Али не било како, него тако да се нека правилност ипак очувава. Да идентичка пресликавања прелазе у идентичка и да се комутативност дијаграма очувава. Схематски гледано, имамо придруживања $X \mapsto S(X)$, где је $X$ тополошки простор, а $S(X)$ нека алгебарска структура и, ако је $f\colon X \longrightarrow Y$ непрекидно пресликавање, онда је $S(f)\colon S(X) \longrightarrow S(Y)$ (или $S(f)\colon S(Y) \longrightarrow S(X)$) хомоморфизам, при чему важи да је $S(\mathrm{id}_X) = \mathrm{id}_{S(X)}$ и $S(g \circ f) = S(g) \circ S(f)$ (или $S(g \circ f) = S(f) \circ S(g)$) ако $X \xrightarrow{f} Y \xrightarrow{g} Z$. Заправо, овакво придруживање се назива коваријантни (односно контраваријантни) функтор. Овде није згорег напоменути да је функторијалност веома значајна особина не само у теоријској математици него и у применама.

Један од проблема који се може разматрати је проблем *проширења* пресликавања: претпоставимо да је $A$ потпростор простора $X$ и да постоји пресликавање $f\colon A \longrightarrow Y$; питање је да ли постоји пресликавање $g\colon X \longrightarrow Y$ такво да је рестрикција од $g$ на $A$ једнака $f$. Ако са $i$ означимо инклузију $A$ у $X$, онда се поставља питање да ли постоји $g$ такво да је $g \circ i = f$. Пређимо сада на алгебарско питање. Нека је $S$ неки коваријантни функтор. Тада $S(f)\colon S(A) \longrightarrow S(Y)$ и $S(i)\colon S(A) \longrightarrow S(X)$. Ако пресликавање $g$ постоји, онда је $S(g) \circ S(i) = S(f)$. Али, размотримо једноставније питање. Да ли уопште постоји *било какав* хомоморфизам $\phi : S(X) \longrightarrow S(Y)$ такав да је $\phi \circ S(i) = S(f)$. Ако не постоји, онда сигурно не постоји ни тражено $g$ (јер би хомоморфизам $S(g)$ 'одрадио посао'). На пример, нека је $S(Z)$ Абелова група за сваки тополошки простор $Z$ и нека је $A = Y$ и $f = \mathrm{id}_A$, при чему је $S(A) \neq \{0\}$, а $S(X) = \{0\}$. Тада бисмо се питали да ли постоји хомоморфизам $\phi$ из $S(X) = \{0\}$ у $S(Y) = S(A) \neq \{0\}$, такав да је $\mathrm{id}_{S(A)} = \phi \circ S(i)$. Но, ако је $p$ неки ненула елемент из $S(A)$ онда је $p = \mathrm{id}_{S(A)}(p) = \phi(S(i)(p)) = \phi(0) = 0$ (јер је домен од $\phi$ тривијална група) што даје контрадикцију. Овај пример није измишљен само за ову прилику, заправо на овај начин се показује да не постоји ретракција $n$-димензионог диска $D^n$ на његову рубну сферу $S^{n-1}$ (ретракција $r$ простора $X$ на потпростор $A$ је непрекидно пресликавање $r : X \longrightarrow A$ које не помера тачке из $A$, тј. за које је $r \circ i = \mathrm{id}_A$, где је $i\colon A \longrightarrow X$ инклузија). Последица ове чињенице је чувена Брауерова (Brower) теорема о фиксној тачки која установљава да свако непрекидно пресликавање $f$ диска $D^n$ у самог себе има фиксну тачку, тј. постоји $x \in D^n$ такво да је $f(x) = x$. Ова теорема има многе примене.

Нека је $X$ произвољна повезана, компактна $n$-димензиона многострукост. Тој многострукости могу се придружити такозване *кохомолошке групе* са $\mathbb{Z}_2$ коефицијентима:

$$H^0(X;\mathbb{Z}_2), H^1(X;\mathbb{Z}_2), \ldots, H^n(X;\mathbb{Z}_2),$$

При чему је (због повезаности многострукости $X$) $H^0(X;\mathbb{Z}_2) \cong \mathbb{Z}_2$. Нећемо се бавити (јер просто и не можемо) детаљима како се до ових група долази, али само прокоментаришимо да се природа $\mathbb{Z}_2$ коефицијената манифестује и у чињеници да је $z+z = 0$ за произвољан елемент неке од ових група. Заправо су све те групе векторски простори над пољем од два елемента: $\mathbb{Z}_2$. Ово 'ко' у називу означава контраваријантност, тј. да при овом придруживању долази до 'окретања' стрелица (у старијим изворима може се наћи термин контрахомологија): ако је $f\colon M \longrightarrow N$ непрекидно пресликавање, онда $H^k(f)\colon H^k(N) \longrightarrow H^k(M)$ и то пресликавање је

заправо линеарно пресликавање међу овим векторским просторима. У даљем ћемо уместо $H^k(f)$ писати $f^*$ без обзира на $k$. Напоменимо да се кохомолошке групе могу придружити широј класи тополошких простора, а не само многострукостима, као и да коефицијенти не морају бити у пољу $\mathbb{Z}_2$. Заправо, касније ћемо имати и случај где су коефицијенти цели бројеви.

Оно што је посебно занимљиво је да, не само да су ово векторски простори над пољем $\mathbb{Z}_2$, него је могуће дефинисати и операцију множења:

$$H^k(X;\mathbb{Z}_2) \times H^l(X;\mathbb{Z}_2) \xrightarrow{\cup} H^{k+l}(X;\mathbb{Z}_2),$$

при чему је елемент из $H^0(X;\mathbb{Z}_2)$ који одговара $1 \in \mathbb{Z}_2$ заправо неутрал за множење и означаваћемо га кратко са 1. Ово пресликавање има и одговарајуће својство асоцијативности и, пошто су коефицијенти у $\mathbb{Z}_2$, а ту је $z = -z$ за све $z$, и својством комутативности. Заправо, ако све ове групе 'скупимо заједно', односно ако посматрамо директну суму

$$H^*(X;\mathbb{Z}_2) = H^0(X;\mathbb{Z}_2) \oplus H^1(X;\mathbb{Z}_2) \oplus \cdots \oplus H^n(X;\mathbb{Z}_2),$$

онда је $H^*(X:\mathbb{Z}_2)$ једна *алгебра* над пољем $\mathbb{Z}_2$. Дакле, ради се о прилично богатој структури.

Наведимо један једноставан пример. У случају $n$-димензионе сфере $\mathbb{S}^n$ имамо да је $H^*(\mathbb{S}^n;\mathbb{Z}_2) = H^n(\mathbb{S}^n;\mathbb{Z}_2) \cong \mathbb{Z}_2$, док је $H^k(\mathbb{S}^n;\mathbb{Z}_2)$ тривијална група за све $k \neq 0, n$. Реални пројективни простор је сложенији пример. У овом случају је $H^k(\mathbb{R}P^n;\mathbb{Z}_2) \cong \mathbb{Z}_2$ за све $k = \overline{0,n}$ и тривијалне иначе. Заправо, имамо и изоморфизам алгебри:

$$\mathbb{H}^*(\mathbb{R}P^n;\mathbb{Z}_2) \cong \mathbb{Z}_2[w]/\langle w^{n+1}\rangle,$$

где је $w$ нетривијални елемент из групе $H^1(\mathbb{R}P^n;\mathbb{Z}_2)$, а са $\langle w^{n+1}\rangle$ је означен идеал у прстену полинома $\mathbb{Z}_2[w]$ генерисан елементом $w^{n+1}$. Ова алгебра се назива и *сасечена полиномска алгебра.*

Скицирајмо једну примену овог резултата. Добро је позната Борсук-Уламова теорема.

**Теорема 3.** *Нека је $f\colon \mathbb{S}^n \longrightarrow \mathbb{R}^n$ непрекидно пресликавање. Тада постоји $x \in \mathbb{S}^n$ такво да је $f(x) = f(-x)$.*

Разне су интерпретације резултата ове теореме. Попут оне да увек на Земљиној површини постоје две антиподалне тачке у којој су исте,рецимо, температура и притисак. Нећемо улазити у та популарна тумачења (зашто је то баш тако, да ли је Земљина површина баш сфера или не, да ли је притисак непрекидан и слично), али рецимо да ова теорема има заиста широке и врло занимљиве примене.

**Скица доказа.** Претпоставимо да ово није тачно, тј. да је $f(x) \neq f(-x)$ за све $x \in \mathbb{S}^n$. Тада се може дефинисати непрекидно пресликавање $g\colon \mathbb{S}^n \longrightarrow \mathbb{S}^{n-1}$ са:

$$g(x) = \frac{f(x) - f(-x)}{|f(x) - f(-x)|},$$

где је са $|f(x) - f(-x)|$ означена норма вектора $f(x) - f(-x) \in \mathbb{R}^n$. Тада је

$$g(-x) = \frac{f(-x) - f(x)}{|f(-x) - f(x)|} = -\frac{f(x) - f(-x)}{|f(x) - f(-x)|} = -g(x).$$

Напоменимо да се реални пројективни простор $\mathbb{R}P^n$ може добити и идентификацијом антиподалних тачака на сфери $\mathbb{S}^n$ Наиме, подсетимо се да је $\mathbb{R}P^n$ простор свих правих у $\mathbb{R}^{n+1}$ које пролазе кроз координатни почетак. Свака таква права сече сферу $\mathbb{S}^n$ у две антиподалне тачке. Тако да уместо да посматрамо све такве праве у $\mathbb{R}^{n+1}$ можемо да посматрамо парове антиподалних тачака на сфери, при чему сматрамо тај пар за једну тачку – идентификујемо антиподалне тачке.

Пажљива анализа, која излази изван оквира овог рада, може показати да ово пресликавање $g$ индукује непрекидно пресликавање $\tilde{g} : \mathbb{R}P^n \longrightarrow \mathbb{R}P^{n-1}$ за које је $\tilde{g}^*(u) = v$, где смо са $u \in H^1(\mathbb{R}P^{n-1}; \mathbb{Z}_2)$, односно $v \in H^1(\mathbb{R}P^n; \mathbb{Z}_2)$ означили одговарајуће генераторе алгебри $H^*(\mathbb{R}P^{n-1}; \mathbb{Z}_2)$, односно $H^*(\mathbb{R}P^n; \mathbb{Z}_2)$. Дакле, добијамо хомоморфизам алгебри

$$\tilde{g}^* \colon H^*(\mathbb{R}P^{n-1}; \mathbb{Z}_2) \longrightarrow H^*(\mathbb{R}P^n; \mathbb{Z}_2).$$

за који је $\tilde{g}^*(u) = v$. Но, тада је $v^n = (\tilde{g}^*(u))^n = \tilde{g}^*(u^n) = \tilde{g}^*(0) = 0$, али $v^n \neq 0$ у $H^n(\mathbb{R}P^n; \mathbb{Z}_2)$. Ова контрадикција завршава нашу скицу доказа теореме Борсук-Улама.

### 2.4. Грасманова многострукост; Борелов опис кохомологије

Позабавимо се сада кохомологијом Грасманијана. Ево Бореловог описа те кохомологије.

**Теорема 4.** *$H^*(G_{k,n}(\mathbb{R}); \mathbb{Z}_2) \cong \mathbb{Z}_2[w_1, \ldots, w_k]/I_{k,n}$ где је идеал $I_{k,n} = \langle \overline{w}_{n+1}, \ldots, \overline{w}_{n+k} \rangle$, док су такозване дуалне класе $\overline{w}_j$ одређене релацијом:*

$$(1 + w_1 + \cdots + w_k)(1 + \overline{w}_1 + \overline{w_2} + \cdots) = 1.$$

Овде је важно напоменути да $w_i, \overline{w}_i \in H^i(G_{k,n}(\mathbb{R}); \mathbb{Z}_2)$. Такође, овде нема никакве бесконачне суме. Наиме, с обзиром да је $\dim G_{k,n}(\mathbb{R}) = kn$ и да су кохомолошке групе многострукости тривијалне за експоненте веће од димензије те многострукости, овде имамо једну, додуше потенцијално „дугачку", али ипак коначну суму. Касније ћемо појаснити овај термин 'дуалне класе', покушајмо за сада да проверимо да ли можемо боље да видимо опис ове кохомологије у неком конкретном случају.

За почетак узмимо тривијалан случај: $G_{1,n}(\mathbb{R})$. Наиме, то су једнодимензиони потпростори у $\mathbb{R}^{n+1}$, тј. $G_{1,n}(\mathbb{R})$ није ништа друго до $\mathbb{R}P^n$, а ту кохомологију знамо. Да проверимо да ли се горенаведени опис слаже са описом кохомологије пројективног простора.

Дакле, $H^*(G_{1,n}(\mathbb{R}); \mathbb{Z}_2) \cong \mathbb{Z}_2[w_1]/I_{1,n}$, где је $I_{1,n} = \langle \overline{w}_{n+1} \rangle$, при чему је

$$(1 + w_1)(1 + \overline{w}_1 + \overline{w}_2 + \cdots) = 1.$$

Горња једнакост је заправо краћи запис за следеће једнакости (у $k$-том реду исписујемо једнакост која важи у $H^k(G_{1,n}(\mathbb{R}); \mathbb{Z}_2)$):

$$\begin{aligned} \overline{w}_1 + w_1 &= 0 \\ \overline{w}_2 + \overline{w}_1 w_1 &= 0 \\ \overline{w}_3 + \overline{w}_2 w_1 &= 0 \\ &\vdots \end{aligned}$$

Имајући у виду да је овде $z = -z$, добијамо:

$$\begin{aligned} \overline{w}_1 &= w_1 \\ \overline{w}_2 &= w_1^2 \\ \overline{w}_3 &= w_1^3 \\ &\vdots \end{aligned}$$

Дакле, $\overline{w}_{n+1} = w_1^{n+1}$ и добијамо да је $H^*(G_{1,n}(\mathbb{R}); \mathbb{Z}_2) \cong \mathbb{Z}_2[w_1]/\angle w_1^{n+1}\rangle$, као што је и требало.

Позабавимо се сада нешто сложенијим примером $G_{2,4}(\mathbb{R})$. У овом случају је

$$H^*(G_{2,4}(\mathbb{R}); \mathbb{Z}_2) \cong \mathbb{Z}_2[w_1, w_2]/I_{2,4},$$

где је $I_{2,n} = \langle \overline{w}_5, \overline{w}_6 \rangle$, при чему је

$$(1 + w_1 + w_2)(1 + \overline{w}_1 + \overline{w}_2 + \cdots) = 1.$$

Покушајмо да мало убрзамо рачунање. Видимо да је елемент $1+\overline{w}_1+\overline{w}_2+\cdots$ заправо инверз елемента $1+w_1+w_2$ у тој алгебри. Дакле

$$1+\overline{w}_1+\overline{w}_2+\cdots=\frac{1}{1+w_1+w_2}=1+\sum_{r\geq 1}(w_1+w_2)^r,$$

при чему смо користили познату суму за геометријску прогресију уз корисну чињеницу да је овде $1=-1$. Нас интересују елементи $\overline{w}_5$ и $\overline{w}_6$, који се налазе у димензији 5 и 6. Ово сада помало подсећа на неки задатак са пријемног испита на факултет: наћи у горњој суми све елементе у димензији 5, односно 6. Овде је $w_1$ у димензији 1, а $w_2$ у димензији 2, па је $w_1^k w_2^l$ у димензији $k+2l$. Добијамо да је

$$\overline{w}_5=w_1^5+\binom{4}{3}w_1^3w_2+\binom{3}{1}w_1w_2^2=w_1^5+w_1w_2^2;$$

$$\overline{w}_6=w_1^6+\binom{5}{4}w_1^4w_2+\binom{4}{2}w_1^2w_2^2+w_2^3=w_1^6+w_1^4w_2+w_2^3.$$

Наравно, овде смо користили да је $2=1+1=0$ у $\mathbb{Z}_2$, па је, на пример $\binom{4}{2}=6=0$.

Коначно, добијамо да је

$$H^*(G_{2,4}(\mathbb{R});\mathbb{Z}_2)\cong\mathbb{Z}_2[w_1,w_2]/\langle w_1^5+w_1w_2^2,w_1^6+w_1^4w_2+w_2^3\rangle.$$

Лепо. Ми сада знамо, на пример, да је у кохомологији овог Грасманијана $w_1^5=w_1w_2^2$, али колико ми заиста можемо успешно да рачунамо у овој кохомологији? На пример, да ли је елемент $w_2^4$ у овој кохомологији једнак нули, или није. Ово питање није произвољно. Елемент $w_k^n$ лежи у димензији $kn$, а то је заправо димензија Грасманијана. Из опште теорије о многострукостима познато да је $H^{kn}(G_{k,n}(\mathbb{R});\mathbb{Z}_2)\cong\mathbb{Z}_2$, то је овај елемент ту генератор или је једнак нули. Заправо, баш ово питање јесте повезано са једним занимљивим резултатом – видети рад [22].

Вратимо се на наш конкретан случај. У димензији 8, која је димензија овог Грасманијана, налазе се и други елементи, који могу бити генератори. На пример, елемент $w_1^8$. Или $w_1^4w_2^2$, или њихов збир. Ми не знамо који од тих кандидата јесте генератор, који су нула, а такође је могуће да се генератор може изразити као полином по $w_1$ и $w_2$ на више начина.

Питање да ли је $w_2^4=0$ у $H^*(G_{2,4}(\mathbb{R});\mathbb{Z}_2)$ еквивалентно је питању да ли $w_2^4$ припада идеалу генерисаном елементима $w_1^5+w_1w_2^2$ и $w_1^6+w_1^4w_2+w_2^3$ у прстену полинома $\mathbb{Z}_2[w_1,w_2]$. На таква питања уопште није лако одговорити. Како бисмо могли да покушамо да то урадимо? Покушали бисмо да одговоримо на питање да ли постоје полиноми $p(w_1,w_2)$ и $q(w_1,w_2)$ из $\mathbb{Z}_2[w_1,w_2]$ за које је

$$w_2^4=p(w_1,w_2)(w_1^5+w_1w_2^2)+q(w_1,w_2)(w_1^6+w_1^4w_2+w_2^3).$$

Идеја би могла да буде да 'поделимо' полином $w_2^4$ једним од ова два полинома, па да онда тај остатак поделимо преосталим полиномом, те ако добијемо 0 као тај последњи остатак, онда полином јесте у идеалу, а ако не добијемо 0, онда није. Наравно, ако мало погледате полиноме, видите да имате велики проблем да $w_2^4$ 'поделите' полиномом $w_1^5+w_1w_2^2$ – како да делите? Други полином више обећава, само га треба другачије написати – као $w_2^3+w_1^4w_2+w_1^6$. Онда се понашате као да се ради о полиномима по неодређеној $w_2$ са коефицијентима у $\mathbb{Z}_2[w_1]$ и, с обзиром да се ради о моничном полиному, заиста можете да извршите дељење у том прстену и добијате остатак $w_1^4w_2^2+w_1^6w_2$. Сада сте завршили са једним од ова два полинома из идеала. Или можда нисте? Сада је опет погодно да тај полином 'гледате' као полином по неодређеној $w_1$, са коефицијентима у $\mathbb{Z}_2[w_2]$... И опасно се ближимо зачараном кругу. Излаз из тога нам нуди теорија Гребнерових база која је тема следећег одељка.

## 3. Гребнерове базе

Видели смо да је кохомологија реалне Грасманове многострукости са $\mathbb{Z}_2$ коефицијентима изоморфна количничкој алгебри полиномске алгебре над $\mathbb{Z}_2$ по одређеном идеалу. Да би смо установили да ли два полинома задају исти елемент у кохомологији, морамо проверити да ли њихова разлика припада идеалу. Дакле, природно нам се појављује питање *припадности идеалу*.

Размотримо ово питање за општи случај алгебре $K[X_1, \ldots, X_n]/I$, где је $K$ неко поље, $I$ идеал генерисан полиномима $h_1, \ldots, h_k$: $I = \langle h_1, \ldots, h_k \rangle$. Како установити да ли је дати полином $f$ у идеалу $I$?

Погледајмо најпре најједноставнији случај: $n = 1, k = 1$. Дакле, питамо се да ли полином $f(X) \in K[X]$ припада идеалу генерисаном полиномом $h_1(X) \in K[X]$. Јасно је како то радимо. Једноставно поделимо полином $f(X)$ полиномом $h_1(X)$ и проверимо да ли је остатак једнак 0. Ако јесте, онда је $f(X) = q(X)h_1(X)$ за неки полином $q(X)$ и $f(X)$ јесте у том идеалу. У случају да постоји ненула остатак, полином није у идеалу. Дакле, ништа простије. Подсетимо се како се врши дељење. Нека је $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$ и $h_1(X) = b_mX^m + b_{m-1}X^{m-1} + \cdots + b_1X + b_0$. Уколико је $n < m$ ништа не можемо да урадимо, зато посматрајмо случај $n \geq m$. Шта радимо? Посматрамо *искључиво* мономе $a_nX^n$ и $b_mX^m$ и први моном поделимо другим. Добијамо да је резултат $\frac{a_n}{b_m}X^{n-m}$. То ће нам бити први моном у количнику. Потом помножимо тим мономом полином $h_1(X)$ и одузмемо резултат од полинома $f$. Добијамо нови полином $f_1(X) = f(X) - \frac{a_nX^n}{b_mX^m}h_1(X)$. Ово можемо записати и овако:

$$f \xrightarrow{h_1} f_1,$$

и то можемо читати: полином $f$ је сведен (редукован) на полином $f_1$ помоћу полинома $h_1$. Да бисмо поједноставили запис, полиноме смо писали без експлицитног навођења неодређене. То ћемо често радити и даље, посебно у случају више неодређених.

Поступак даље примењујемо на полином $f_1$. Ово се наставља све док не дођемо или до нуле или до полинома степена мањег од степена полинома $h_1$. Тај добијени полином, који можемо (не превише маштовито) да означимо са $r$, заправо је остатак при дељењу полинома $f$ полиномом $h_1$. То се у претходној симболици записује и овако:

$$f \xrightarrow{h_1} f_1 \xrightarrow{h_1} f_2 \cdots \xrightarrow{h_1} f_s = r.$$

Дакле, дељење једног полинома другим заправо се састоји из више корака које називамо редукције. Ово дељење нам једноставно омогућава да разрешимо питање припадности идеалу у случају полинома са једном неодређеном и идеала генерисаног једним елементом.

Шта се дешава у случају $n = 1, k = 2$? Тада имамо идеал $I = \langle h_1(X), h_2(X) \rangle$. Како установити да ли дати полином $f$ припада овом идеалу? Рекло би се, ништа посебно. Рецимо, поделимо полином $f$ полиномом $h_1$. Ако је остатак 0, онда јесте у идеалу, ако није, онда тај остатак поделимо полиномом $h_2$. Сад, ту постоји извесна произвољност – зашто прво $h_1$, па после $h_2$, али добро. Али, да ли ово ради?

Размотримо следећи пример: $f = X^3 + X$, $h_1 = X^2 - X$, $h_2 = X^2$. Вршимо дељење полиномом $h_1$:

$$X^3 + X \xrightarrow{h_1} X^2 + X \xrightarrow{h_1} 2X.$$

У првој редукцији смо од полинома $f$ одузели полином $h_1$ помножен мономом $X = \frac{X^3}{X^2}$, а потом смо од добијеног полинома одузели полином $h_1$ помножен мономом $1 = \frac{X^2}{X^2}$. Све по правилима којима вршимо дељење. Добили смо полином $2X$ и њега не можемо више да делимо полиномом $h_1$. Добро, пређимо на полином $h_2$.... Али, не можемо да га делимо ни полиномом $h_2$! Чини се да он није у идеалу. А шта би се десило да смо прво делили полиномом $h_2$?

$$X^3 + X \xrightarrow{h_2} X.$$

Опет не можемо да наставимо даље. Приметимо да нисмо добили исти резултат. Но, ми смо дељење раставили на појединачне кораке. Можда можемо да мало редукујемо помоћу $h_1$, а

мало помоћу $h_2$?

$$X^3 + X \xrightarrow{h_1} X^2 + X \xrightarrow{h_2} X.$$

Како год да радимо, не добијамо да полином припада идеалу. Али, јасно се види да полином јесте у идеалу: $X = h_2 - h_1$, па $X \in I$, а $f = (X^2 + 1) \cdot X$, па $f \in I$.

Дакле, овај наш поступак не ради баш добро. Требало би мало да размислимо. Очигледно је било погрешно користити само полиноме $h_1$ и $h_2$. Треба гледати и друге полиноме који су у идеалу. Наравно, лако ћемо се досетити како ово поправити. Не само за овај посебан случај, него уопште. Ми врло добро знамо да је прстен полинома $K[X]$ главноидеалски, тј. у њему је сваки идеал генерисан једним полиномом. Знамо и који је то полином. То је полином који је заправо највећи заједнички делилац тог коначног броја полинома који генеришу идеал. У случају два полинома, највећи заједнички делилац се добија Еуклидовим алгоритмом, а за више полинома се поступак итерира. У нашем случају се лако добија да је $\mathrm{nzd}(X^2 - X, X^2) = X$ и онда је све јасно (и лако).

Дакле, проблем припадности идеалу $I = \langle h_1, \ldots, h_k \rangle$ решили смо преласком на „бољи" генераторни скуп. У случају идеала генераторни скуп се назива и база, при чему, наравно, не мислимо на базу у смислу векторских простора. Дакле, од базе $\{h_1, \ldots, h_k\}$ прелазимо на једночлану базу $\{\mathrm{nzd}(h_1, \ldots, h_k)\}$ и тада се проблем припадности идеалу једноставно решава.

Позабавимо се сада случајем полинома са више неодређених.

Потребно је, пре свега, да разјаснимо дељење у прстену полинома са више неодређених. Као што смо видели, дељење је заправо низ редукција, те ћемо стога разјаснити појам редукције.

Када вршимо редукцију полинома са једном неодређеном, ми се концентришемо на два монома, који су заправо мономи највећег степена у два полинома које разматрамо. Дакле, они су нека врста *водећих монома*. Како то изгледа у случају полинома са више неодређених? На пример, који је то водећи моном у полиному $X^3Y + 4X^2Y^2 + 3X^2Y + Y^3 - X^2 + 2Y - 1$? Ако гледамо по тоталним степенима, онда су и $X^3Y$ и $4X^2Y^2$ степена 4. Наравно, можемо да се концентришемо на највиши степен од $X$. Али, шта рећи за овај полином: $X^2Y^2Z + X^2YZ^2 - Y^2 + 3Z - 5$? Овде можемо да гледамо већи степен од $Y$. Треба то пажљивије осмислити.

Ево мало пригодне терминологије. Ако имамо *моном* $cX_1^{\alpha_1}X_2^{\alpha_2}\cdots X_n^{\alpha_n}$, онда је ту $c$ *коефицијент*, а $X_1^{\alpha_1}X_2^{\alpha_2}\cdots X_n^{\alpha_n}$ је *производ степена неодређених*, кратко *производ*. Овај производ ћемо кратко означавати са $\boldsymbol{X^\alpha}$ (овде су $\boldsymbol{X}$ и $\boldsymbol{\alpha}$ одговарајуће $n$-торке). Желимо да уведемо неки поредак $\preceq$ на скупу свих производа. Најпре желимо да тај поредак проширује поредак у смислу дељивости, тј. да из $\boldsymbol{X^\alpha} \mid \boldsymbol{X^\beta}$ следи $\boldsymbol{X^\alpha} \preceq \boldsymbol{X^\beta}$. Такође желимо да тако добијемо добро уређени скуп свих производа, јер не желимо бесконачан опадајући низ: $\boldsymbol{X^{\alpha_1}} \succ \boldsymbol{X^{\alpha_2}} \succ \ldots$. Посебно, то значи да имамо једно линеарно уређење, тј. свака два производа су упоредива. Приметимо да је и $1 \preceq \boldsymbol{X^\alpha}$ за свако $\boldsymbol{\alpha}$ ($1 = \boldsymbol{X^0}$, $\boldsymbol{0} = (0, 0, \ldots, 0)$). Сваки такав поредак назива се *мономни* поредак (мада је заправо дефинисан на производима, а не мономима, али не постоји усаглашеност термина, те се нећемо даље бринути о томе).

Заправо, ево прецизне дефиниције која нам даје све тражене услове.

**Дефиниција 4.** *Линеарно уређење на скупу свих производа је мономни поредак уколико испуњава следећа два услова.*

1. *За све* $\boldsymbol{\alpha} \neq \boldsymbol{0}$ *је* $1 \prec \boldsymbol{X^\alpha}$

2. *За све* $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ *из* $\boldsymbol{X^\alpha} \prec \boldsymbol{X^\beta}$ *следи* $\boldsymbol{X^\alpha X^\gamma} \prec \boldsymbol{X^\beta X^\gamma}$.

Нећемо овде детаљно проверавати испуњеност горенаведених својстава.

Мономних поредака има уистину много. Наведимо два најједноставнија. Први је *лексикографски* поредак, ознака lex. У том случају је

$$X_1^{\alpha_1}X_2^{\alpha_2}\cdots X_n^{\alpha_n} \prec_{\mathrm{lex}} X_1^{\beta_1}X_2^{\beta_2}\cdots X_n^{\beta_n} \overset{\mathrm{def}}{\iff} (\exists i \in \{1, \ldots, n\})((\forall j < i)(\alpha_j = \beta_j) \wedge \alpha_i < \beta_i)$$

Кратко речено, производи се пореде као у речнику. Замислите да сте исписали производ као реч у речнику (при чему су неодређене слова и то тако да је $X_1$ прво слово итд.). Онда

је већа она која се прва појави. Рецимо $X_1^2X_2X_3^7 \prec_{\text{lex}} X_1^2X_2^2X_3$: $\alpha_1 = 2 = \beta_1$, а $\alpha_2 = 1 < 2 = \beta_2$. Заправо, прва одговара речи (речник је на ћирилици): аабвввввв, а друга ааббв. Сигурно се ни у једном речнику српског језика ове две речи не појављују, али знамо која би се прва појавила. Посебно, овде је $X_1 \succ_{\text{lex}} X_2 \succ_{\text{lex}} \cdots \succ_{\text{lex}} X_n$. Наиме, $X_2 = X_1^0X_2^1 \cdots X_n^0$, а $X_1 = X_1^1X_2^0 \cdots X_n^0$, па је $X_2 \prec_{\text{lex}} X_1$, што се види поређењем степена неодређене $X_1$. Уосталом, слово а се појављује пре слова б, зар не? Наравно, поређење са речником је натегнуто: наше неодређене комутирају, а слова у речима сигурно не, али ово објашњава терминологију.

Други једноставан поредак, који ћемо користити у даљем је такозвани *степенасти лексикографски* поредак, у ознаци grlex. У овом поретку, ако гледамо аналогију са речником, прво поредите дужину речи (дуже речи су веће од краћих), а речи исте дужине поредите лексикографски: $X_1^{\alpha_1}X_2^{\alpha_2} \cdots X_n^{\alpha_n} \prec_{\text{grlex}} X_1^{\beta_1}X_2^{\beta_2} \cdots X_n^{\beta_n}$ ако и само ако је $\sum_{i=1}^n \alpha_i < \sum_{i=1}^n \beta_i$ или је $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i$, а $X_1^{\alpha_1}X_2^{\alpha_2} \cdots X_n^{\alpha_n} \prec_{\text{lex}} X_1^{\beta_1}X_2^{\beta_2} \cdots X_n^{\beta_n}$. Јасно је да је и овде $X_i \succ_{\text{grlex}} X_j$ за $i < j$.

Заправо, ми смо и дефинисали оба ова поретка тако да неодређене овако буду уређене. Може се наравно, аналогно, задати лексикографски и степенасти лексикографски поредак да неодређене буду уређене на неки други начин.

Изаберимо неки мономни поредак $\preceq$. Сваки полином $f \in K[X_1, X_2, \ldots, X_n] \setminus \{0\}$ можемо записати на следећи начин:

$$f = a_1\boldsymbol{X}^{\boldsymbol{\alpha}_1} + a_2\boldsymbol{X}^{\boldsymbol{\alpha}_2} + \cdots + a_r\boldsymbol{X}^{\boldsymbol{\alpha}_r},$$

при чему је $\boldsymbol{X}^{\boldsymbol{\alpha}_1} \succ \boldsymbol{X}^{\boldsymbol{\alpha}_2} \succ \cdots \succ \boldsymbol{X}^{\boldsymbol{\alpha}_r}$. Природно је увести следећу терминогију (и ознаке). *Водећи производ* полинома $f$, у ознаци $LP(f)$, је $\boldsymbol{X}^{\boldsymbol{\alpha}_1}$, *водећи коефицијент* тог полинома, у ознаци $LC(f)$ је $a_1$, док је $a_1\boldsymbol{X}^{\boldsymbol{\alpha}_1}$ *водећи моном*: $LM(f) = \alpha_1\boldsymbol{X}^{\boldsymbol{\alpha}_1}$.

Нека је задат идеал $I = \langle h_1, h_2, \ldots, h_k\rangle$. Занима нас питање припадности полинома $f \neq 0$ овом идеалу. На аналоган начин можемо увести појам редукције. Уочимо $LP(f)$ и $LP(h_i)$. Ако $LP(h_i) \mid LP(f)$, онда вршимо редукцију:

$$f \xrightarrow{h_i} f_1,$$

где је

$$f_1 = f - \frac{LM(f)}{LM(h_i)}h_i.$$

Приметимо да је сада $LP(f_1) \prec LP(f)$, јер смо водећи моном 'скинули', а све се радило производима који су мањи од $LP(f)$ (користили смо водећи моном и у $h_i$). Дакле, можемо понављати ове редукције користећи све ове полиноме $h_1, \ldots, h_k$. На крају долазимо до полинома $r$ који даље не можемо да редукујемо. У случају полинома са једном неодређеном, то се дешава када степен добијеног полинома буде мањи од степена полинома $h_i$, а у случају полинома са више неодређених, то се дешава када ниједан од водећих производа $LP(h_i)$ не дели $LP(r)$. Ако је $r = 0$, онда знамо да је $f$ у идеалу, а ако није, онда...

Као што смо видели у случају полинома са једном неодређеном, овај поступак баш није идеалан. Пре свега тај редуковани полином ('остатак') зависи од редоследа којим радимо редукције. Посебно, то значи да се може добити да 'остатак' није 0, мада полином јесте у идеалу. Проблем је био у генераторном скупу (бази) датог идеала. У случају полинома са једном неодређеном, знали смо да постоји једночлана база и она је решила проблем, али у случају полинома са више неодређених, који није главноидеалски прстен, не можемо тако закључивати. Но, размотримо мало полиноме са једном неодређеном. Проблем је био у томе што су генератори били вишег степена од неких полинома који припадају идеалу, па се ти полиноми нису могли даље редуковати. Дакле, ту је пожељно наћи полином најмањег степена који је у идеалу. Ако се присетите доказа чињенице да је сваки идеал у прстену $K[X]$ главни, он се баш заснива на тој идеји – да се покаже да је ненула полином најмањег степена који је у идеалу заправо генератор! У овом случају је тај најмањи степен заправо такав да дели остале степене (полинома који су у идеалу). То долази од тога што је у

случају полинома са више неодређених релација дељивости међу производима $X^n$ заправо линеарно уређење. Тако нешто немамо у случају полинома са више неодређених, па самим тим не можемо имати ни једночлану базу за сваки идеал. Али, имамо нешто мало другачије, али ипак довољно добро. Можемо посматрати *минималне производе*. На пример, у скупу производа $\{X^3YZ, XY^2, Z^3, X^3YZ^5, Z^2\}$ производи $X^3YZ, Z^2, XY^2$ јесу минимални у смислу дељивости (нису дељиви другим производима из тог скупа), али нису најмањи – они су заправо међусобно неупоредиви у смислу дељивости.

Може се поставити питање: да ли у произвољном скупу производа постоји само коначно много минималних елемената (у смислу дељивости) Одговор је потврдан. Ово је заправо чисто комбинаторно питање и може се разматрати независно од ове приче. Амбициозном читаоцу препоручујемо да сам проба да ово докаже, а ако не успе, може да потражи доказ Диксонове леме у литератури.

Вратимо се на питање редукције. Уместо да радимо са датом базом $h_1, \ldots, h_k$, ми желимо да добијемо бољу базу, у којој ће редукција решити проблем припадности идеалу (а и у којој 'остатак' не зависи од редоследа редукција). Да ли можемо да нађемо такву базу? Посматрајмо скуп $\mathcal{LP}(I) = \{LP(g) : g \in I \setminus \{0\}\}$. То је скуп свих водећих производа полинома из $I$. Према горенаведеном резултату, постоји коначно много минималних елемената (у смислу дељивости) у овом скупу, тј. постоји коначно много полинома $g_1, \ldots, g_s \in I$ таквих да је скуп $\{LP(g_1), \ldots, LP(g_s)\}$ скуп свих минималних елемената скупа $\mathcal{LP}(I)$.

Заправо овако добијени елементи $g_1, \ldots, g_s$ и представљају тражену погодну базу – то је Гребнерова база идеала $I$! Покажимо да је то заиста база тог идеала, а и да је погодна у горенаведеном смислу.

Показаћемо да:

$$\text{за сваки } g \in I \setminus \{0\} \text{ постоји } i \in \{1, \ldots, s\} \text{ тако да } LP(g_i) \mid LP(g). \tag{1}$$

У супротном, нека је $g$ ненула полином из идеала $I$ чији водећи производ није дељив ниједним од водећих производа полинома $g_i$. Како су они минимални, то, за све $i$: $LP(g) \nmid LP(g_i)$. Постоји само коначно много производа који деле $LP(g)$. Самим тим постоји само коначно много производа из скупа $\mathcal{LP}(I)$ који деле $LP(g)$. Изаберимо неки минималан такав (можда је то баш $LP(g)$, нема везе). Тај производ ће онда бити и минималан у целом скупу $\mathcal{LP}(I)$, а то противречи чињеници да је $\{LP(g_1), \ldots, LP(g_s)\}$ скуп свих минималних елемената (тај производ није ни један од ових – они не деле $LP(g)$).

Дакле, сада имамо полиноме $g_1, \ldots, g_s$ из идеала $I$ са горенаведеним својством (1). Покажимо да они генеришу $I$. Заправо ћемо показати да је $f \in I$ ако и само ако се помоћу $g_i$ може редуковати до 0.

Наравно, један смер је лак. Ако се полином редукује до 0 коришћењем полинома $g_i$, онда је тај полином облика $\sum_i p_i g_i$ за неке полиноме $p_i$, па самим тим припада идеалу $I$ (при редукцији од датог полинома $f$ одузимамо неки $g_i$ помножен неким мономом; ако дођемо на крају од нуле, онда, радећи уназад, добијамо да је $f$ сума умножака $g_i$ неким мономима). Претпоставимо стога да $f \in I$. То значи да је $LP(g_i) \mid LP(f)$ за неко $i$, па можемо извршити редукцију:

$$f \xrightarrow{g_i} f_1,$$

при чему је $LP(f_1) \prec LP(f)$ и $f_1 \in I$. Уколико је $f_1 = 0$, добили смо тражену; у супротном настављамо поступак. Тако добијамо низ полинома: $f = f_0, f_1, \ldots$ за које је $LP(f_0) \succ LP(f_1) \succ \ldots$. На основу својстава мономног поретка, знамо да такав низ не може бити бесконачан, те се процес мора завршити и добијамо 0 после коначно много корака.

Сада можемо дати експлицитну дефиницију појма Гребнерове базе.

**Дефиниција 5.** *Полиноми $g_1, \ldots, g_s \in I$ чине Гребнерову базу за идеал $I$ ако испуњавају горенаведено својство (1).*

Видели смо да Гребнерова база увек постоји. Друго је питање како је наћи. Постоје алгоритми за налажење Гребнерове базе конкретног идеала, а они су и имплементирани у

разним пакетима за симболичка израчунавања. Како се ми нећемо тиме бавити овде, поново упућујемо читаоца на литературу. Само ћемо прокоментарисати да, наравно, Гребнерова база зависи од избора мономног поретка, али је занимљива и следећа једноставна чињеница: ако је $\{g_1, \ldots, g_s\}$ Гребнерова база за поредак $\preceq_1$ и посматрамо поредак $\preceq_2$ и скуп полинома $\{h_1, \ldots, h_s\}$ за које је $LP_{\preceq_1}(g_i) = LP_{\preceq_2}(h_i)$ за све $i$, онда је скуп $\{h_1, \ldots, h_s\}$ Гребнерова база за поредак $\preceq_2$. Надамо се да је читаоцу потпуно јасно зашто ово важи.

За крај овог одељка, наведимо да се појам Гребнерове базе може задати и за случај полинома над неким прстеном коефицијената. Наравно да је ту сложеније разматрање (на пример, одмах се може приметити да морамо разматрати да ли $LM(g) \mid LM(f)$, а не само да ли $LP(g) \mid LP(f)$), али за правилне прстене, попут главноидеалских домена, теорија се може фино развити. У одељку о кохомологији Грасманијана, имаћемо такав случај у коме је прстен коефицијената $\mathbb{Z}$.

## 4. Векторска раслојења

У овом одељку навешћемо основне резултате о векторским раслојењима, карактеристичним класама, као и начин свођења проблема о имерзијама на питање из теорије хомотопије.

Реално векторско раслојење $\xi$ са тоталним простором $E$ и базом $B$ је пресликавање

$$\begin{array}{c} E \\ \Big\downarrow{\scriptstyle p} \\ B \end{array}$$

При томе, за свако $b \in B$ слој (фибра) $p^{-1}(b)$ је векторски простор над пољем $\mathbb{R}$ (исте димензије). Осим овог услова захтева се још један технички, али важан услов, локалне тривијалности који нећемо формулисати – желимо само да скицирамо основне идеје.

Наведимо и пар примера.

0. Ако је $E = B \times \mathbb{R}^n$ и $p$ пројекција на прву координату, онда је у питању тривијално раслојење (ознака $\varepsilon^n$).

1. Тангентно раслојење на многострукост ($\tau$). Заправо нам је ово раслојење добро познато из ранијег школовања, мада га нисмо експлицитно наводили: свака глатка површ у свакој тачки има своју тангентну раван. Унија свих тих равни представља тангентно раслојење те површи.

2. Канонско раслојење над Грасманијаном: тотални простор овог раслојења је

$$E = \{(\pi, v) \in G_{k,n}(\mathbb{R}) \times \mathbb{R}^{n+k} : v \in \pi\}.$$

Дакле, над сваком тачком из $G_{k,n}(\mathbb{R})$, која није ништа друго до раван димензије $k$ у $\mathbb{R}^{n+k}$, лежи баш та раван.

Могуће је на природан начин дефинисати и појам изоморфизма раслојења (те ће бити хомеоморфизам тоталних простора, који 'поштује' структуру векторског простора у свакој фибри). Такође се на векторским раслојењима могу вршити и операције сабирања и множења:

$$(\xi, \eta) \mapsto \xi \oplus \eta, \xi \otimes \eta$$

Врло важан резултат је следећи: ако је базни простор $B$ компактан, онда за свако $\xi$ постоји $\theta$ тако да је $\xi \oplus \theta \cong \varepsilon^N$ за неко $N$. Дакле, сваком раслојењу над компактном базом, може се додати друго раслојење тако да је добијено раслојење изоморфно тривијалном. Ово нам даје и јасан појам нормалног раслојења за тангентно раслојење многострукости: то је раслојење $\nu$ такво да је $\tau \oplus \nu \cong \varepsilon^N$. Природно је видети $\nu$ као адитивни инверз за $\tau$. Није лоше имати на уму следећи једноставан пример: замислите дводимензиону сферу и тангентну раван у некој тачки. Нормално раслојење је овде димензије 1: за сваку тачку посматрамо праву кроз ту тачку која је нормална на тангентну раван. Све те праве заједно, чине нормално раслојење.

Реалном векторском раслојењу $\xi$, ради лакшег испитивања његових својстава методама алгебарске топологије, придружујемо Штифел-Витнијеве карактеристичне класе $w_i(\xi) \in H^i(B;\mathbb{Z}/2)$. Заправо, можемо посматрати и тоталну класу: $\xi \mapsto w(\xi) \in H^*(B;\mathbb{Z}/2)$:

$$w(\xi) = 1 + w_1(\xi) + w_2(\xi) + \cdots + w_m(\xi),$$

где је $m$ димензија слоја (векторског простора). При томе, за тривијално раслојење важи: $w(\varepsilon^N) = 1$, док је:

$$w(\xi \oplus \eta) = w(\xi) \cdot w(\eta).$$

Посебно: $w(\tau) \cdot w(\nu) = 1$.

Реално векторско раслојење $\xi$ над базом $B$ у коме је димензија слоја $n$ (реално $n$-раслојење над $B$) потпуно је одређено пресликавањем

$$f_\xi \colon B \longrightarrow BO(n) (= G_{n,\infty}(\mathbb{R})).$$

Овде је са $G_{n,\infty}$ означен бесконачни Грасманијан. Нећемо се њиме даље бавити, довољно је рећи да се може добити као унија коначних Грасманијана $n$-димензионих равни у просторима све веће димензије. Уколико се ради само о раслојењу $\xi$ чију димензију не знамо, онда га видимо као пресликавање

$$f_\xi \colon B \longrightarrow BO.$$

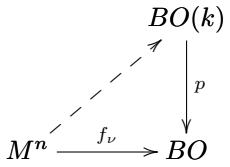
Овај простор $BO$ такође нећемо даље изучавати (он се пак може видети као унија претходних $BO(n)$ када $n$ неограничено расте). Примена теорије хомотопије на питање имерзије је садржана у следећем.

Нормално раслојење одређује димензију простора у који можемо „имерзовати" многострукост (подсетимо се да је $\tau \oplus \nu \cong \varepsilon^N$, за неко $N$). Дакле, нормално раслојење одговара пресликавању

$$f_\nu \colon B \longrightarrow BO(M),$$

за неко $M$, односно пресликавању у $BO$ (тада га зовемо стабилно нормално раслојење).

Наш задатак је да решимо „проблем подизања". Хиршова теорема каже да се компактна многострукост $M^n$ може имерзовати у $\mathbb{R}^{n+k}$ акко се може позитивно решити следећи проблем подизања:




Ово се решава факторисањем пресликавања $p$ у једноставнија, која се могу алгебарски описати. За испитивање постојања подизања за та једноставнија пресликавања (подизање се врши „спрат по спрат") користи се детаљно познавање кохомологије многострукости $M^n$, карактеристичних класа тангентног раслојења, као и познавање дејства Стинродове алгебре (о којој такође овде нисмо говорили) на овој многострукости.

## 5. Рачунање у кохомологији Грасманијана и примене

Видели смо да је кохомологија Грасманијана $G_{k,n}(\mathbb{R})$ са $\mathbb{Z}_2$ коефицијентима заправо полиномска алгебра посечена по идеалу $I_{k,n}$. У раду [8], Монкс је одредио Гребнерове базе за идеал $I_{2,n}$ за случај $n = 2^s - 3$ и $n = 2^s - 4$. Помоћу ових база добио је нове резултате о кохомологији одговарајућих Грасманијана, а такође, користећи методу модификованих кула Постникова (то су оне факторизације које су споменуте на крају претходног одељка), добио је резултате који се тичу имерзија Грасманијана $G_{2,2^s-3}$ у еуклидске просторе. Ове резултате о имерзијама је касније поправио Шимкус у раду [21].

У случају $k = 2$, који је разматран у овим радовима, ради се о количичкој алгебри $\mathbb{Z}_2[w_1, w_2]/\langle \overline{w}_{n+1}, \overline{w}_{n+2}\rangle$. У раду [10], одређене су Гребнерове базе за све идеале $I_{2,n}$ за степенасти лексикографски поредак за који је $w_1 \succ w_2$. Прокоментаришимо укратко, без доказа, ове резултате.

Гребнерову базу за идеал $I_{2,n}$ чине полиноми $g_m$, за $0 \leq m \leq n+1$ задати са:

$$g_m := \sum_{a+2b=n+1+m} \binom{a+b-m}{a} w_1^a w_2^b. \tag{2}$$

Наравно, овде се биномни коефицијенти посматрају по модулу 2, те су или 0 или 1. Раније смо разматрали случај $G_{2,4}(\mathbb{R})$, па их израчунајмо у овом случају:

$$g_0 = w_1^5 + w_1 w_2^2, \quad g_1 = w_1^4 w_2 + w_1^2 w_2^2 + w_2^3, \quad g_2 = w_1^3 w_2^2, \quad g_3 = w_1^2 w_2^3 + w_2^4, \quad g_4 = w_1 w_2^4, \quad g_5 = w_2^5.$$

Као што се види, $LP(g_i) = w_1^{5-i} w_2^i$ (у случају $\mathbb{Z}_2$ коефицијената сви водећи коефицијенти су наравно 1, па се водећи мономи и водећи производи подударају). Дакле, сви производи $w_1^a w_2^b$ за које је $a + b = 5(= 4 + 1)$ појављују се као водећи у Гребнеровој бази. То значи да сви полиноми из $\mathbb{Z}_2[w_1, w_2]$ могу бити редуковани до полинома у којима се појављују само производи $w_1^a w_2^b$ за које је $a+b \leq 4$. Стога заправо базу за саму кохомологију $H^*(G_{2,4}(\mathbb{R}); \mathbb{Z}_2)$ чине мономи (или производи, свеједно) $w_1^a w_2^b$ за које је $a+b \leq 4$ – било која њихова нетривијална линеарна комбинација над $\mathbb{Z}_2$ не може даље бити редукована, дакле не припада идеалу $I_{2,n}$, па није једнака 0 у кохомологији; дакле ти мономи не само што нису једнаки 0 у кохомологији, нису ни линеарно зависни као елементи векторског простора $H^*(G_{2,4}(\mathbb{R}); \mathbb{Z}_2)$. Како се сваки полином може редуковати до таквог у коме се појављују искључиво такви мономи, закључујемо да они чине базу кохомологије $H^*(G_{2,4}(\mathbb{R}); \mathbb{Z}_2)$ као векторског простора над $\mathbb{Z}_2$. Посебно, ту је и елемент $w_2^4$ за кога смо се питали да ли је једнак 0 у кохомологији. Дакле, одговор је да није.

Заправо, из (2) лако се може извести закључак у општем случају, да мономи $w_1^a w_2^b$ за $a + b \leq n$ чине базу векторског простора $H^*(G_{2,n}(\mathbb{R}); \mathbb{Z}_2)$.

Ови резултати су затим проширени за случај $k = 3$ у [11]. У том раду се, између осталог, добило да векторски простор $H^*(G_{3,n}(\mathbb{R}); \mathbb{Z}_2)$ има базу који чине мономи облика $w_1^a w_2^b w_3^c$ за $a + b + c \leq n$. Природно се поставило питање да ли је у општем случају тачно да векторски простор

$$H^*(G_{k,n}(\mathbb{R}); \mathbb{Z}_2) \text{ има базу коју чине мономи } w_1^{\alpha_1} w_2^{\alpha_2} \cdots w_k^{\alpha_k} \text{ где је } \sum_{i=1}^{k} \alpha_i \leq n? \tag{3}$$

Но, било је јасно да би поступак примењен у ова два наведена рада било тешко наставити за $k \geq 4$. Оно што је било могуће урадити је да се добије одговарајући резултат за комплексне Грасманијане за $\mathbb{Z}$-кохомологијом. Заправо, кохомологија ових многострукости задата је на скоро идентичан начин као и кохомологија реалних Грасманијана – само $\mathbb{Z}_2$ треба заменити са $\mathbb{Z}$ и $w_i$ са $c_i$ (то су Чернове класе, при чему $c_i \in H^{2i}(G_{k,n}(\mathbb{C}); \mathbb{Z})$). Одговарајуће Гребнерове базе за комплексне Грасманијане за $k = 2$ и $k = 3$ одређене су у раду [17].

Познавање Гребнерових база, уз примену модификованих кула Постникова довело је до нових резултата о имерзијама Грасманових многострукости, који су приказани у радовима [10] и [11]. На пример:

$$2 \nmid n \text{ и } n \geq 5 : G_{2,n}(\mathbb{R}) \looparrowright \mathbb{R}^{4n-5};$$

$$n \equiv 6 \pmod 8 \Longrightarrow G_{3,n}(\mathbb{R}) \looparrowright \mathbb{R}^{6n-5};$$

Ако је $n$ степен двојке, онда је $\operatorname{imm}(G_{3,n}(\mathbb{R})) = 6n - 3$.

Подсетимо се да је димензија Грасманове многострукости $G_{k,n}(\mathbb{R})$ је $kn$, док се са $\operatorname{imm}(M)$ означава минимална димензија еуклидског простора у који се $M$ може имерзовати. Читалац

може упоредити ове резултате са раније наведеним општим Коеновим резултатом. Још један нови резултат о имерзијама Грасманијана може се наћи у [12]

У време истраживања која су резултирала претходно наведеним радовима, нисмо знали да је одговор на претходно постављено питање (3) потврдан. Када смо то сазнали, био је могућ други приступ проблему налажења Гребнерових база за идеале $I_{k,n}$. Новодобијени резултати су приказани у раду [14]. Наведимо укратко неке од главних резултата тог рада. Радићемо са комплексним Грасманијанима и са $\mathbb{Z}$ коефицијентима.

Уведимо најпре неке ознаке. За $\alpha = (a_1, \ldots, a_k)$, $\mu = (m_2, \ldots, m_k)$:

$$|\alpha| := \sum_{j=1}^{k} a_j, \|\alpha\| := \sum_{j=1}^{k} j a_j, |\mu| := \sum_{j=2}^{k} m_j, \|\mu\| := \sum_{j=2}^{k} (j-1) m_j$$

$$[\alpha, \mu]_t := \binom{\sum_{j=t-1}^{k} a_j - \sum_{j=t}^{k} m_j}{a_{t-1}}$$

$$[\alpha, \mu] := \prod_{t=2}^{k} [\alpha, \mu]_t$$

Посматрамо елементе

$$g_\mu := \sum_{\|\alpha\| = n+1+\|\mu\|} (-1)^{n+1+|\alpha|} [\alpha, \mu]\, C^\alpha,$$

где је $C^\alpha := c_1^{a_1} \cdots c_k^{a_k}$ ($c_i$ су Чернове класе канонског раслојења над $G_{k,n}(\mathbb{C})$). Тада је

$$\{g_\mu : |\mu| \leq n+1\}$$

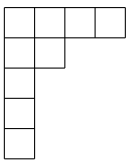
Гребнерова база одговарајућег идеала, док је

$$\begin{aligned} B_{k,n} &:= \{c_{m_1} c_{m_2} \cdots c_{m_s} : s \leq n, 1 \leq m_s \leq \cdots \leq m_1 \leq k\} \\ &= \{C^\alpha : |\alpha| \leq n\} \end{aligned}$$

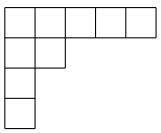
адитивна база за $H^*(G_{k,n}(\mathbb{C}); \mathbb{Z})$

Осим ових, у раду [14] добијени су и резултати који припадају области алгебарске комбинаторике. Да бисмо и њих приказали, морамо увести још неке појмове.

Партиција $\lambda := (l_1, \ldots, l_s)$ је нерастући низ целих бројева: $l_1 \geq l_2 \geq \cdots \geq l_s \geq 0$. Дужина партиције је $l(\lambda) := \max\{t : t \neq 0\}$, док је тежина $|\lambda| := l_1 + \cdots + l_s$. Партиције представљамо Јанговим дијаграмима:



$\lambda = (4, 2, 1, 1, 1)$



$\lambda^* = (5, 2, 1, 1)$

Нека су $k, n \in \mathbb{N}$, $V = \mathbb{C}^{n+k}$ и нека је дата комплетна застава

$$0 = V_0 \subset V_1 \subset \cdots \subset V_{n+k} = V.$$

Ова застава је заправо растући низ векторских потпростора $V_i$ таквих да је $\dim V_i = i$. Партиција $\lambda \subset k \times n$, $\lambda = (l_1, \ldots, l_n)$ одређује Шубертов варијетет:

$$X_\lambda := \{W \in G_{k,n}(\mathbb{C}) : \dim(W \cap V_{n+i-l_i}) \geq i, 1 \leq i \leq k\}.$$

Ако се са $\sigma_\lambda$ означи класа од $X_\lambda$ у $H^*(G_{k,n}(\mathbb{C});\mathbb{Z})$, тада је

$$\Sigma_{k,n} := \{\sigma_\lambda : \lambda \subset k \times n\}$$

адитивна база за $H^*(G_{k,n}(\mathbb{C});\mathbb{Z})$. Множење је одређено Пијеријевом формулом

$$\sigma_\lambda \cdot \sigma_{(m)} = \sum_\nu \sigma_\nu,$$

где је сума по свим партицијама $\nu$ које се могу добити додавањем $m$ квадратића Јанговом дијаграму за $\lambda$ при чему се никоја два не додају у исту колону.

Сада имамо две базе: $B_{k,n}$ добијене преко Бореловог приказа и $\Sigma_{k,n}$ помоћу Шубертових класа.

$$\sigma_{1^i} = (-1)^i c_i, \quad 1 \le i \le k$$
$$\sigma_{(i)} = \overline{c}_i, \quad 1 \le i \le n.$$

Добија се веза:

$$c_{m_1} c_{m_2} \cdots c_{m_s} = (-1)^{|\mu|} \sum_\lambda K_{\lambda\mu} \sigma_{\lambda^*},$$

где је $\mu = (m_1, \ldots, m_s)$ партиција за коју је $m_1 \le k$ и сума је по свим партицијама $\lambda \subset n \times k$ за које је $|\lambda| = |\mu|$. $K_{\lambda\mu}$ су Косткини бројеви.

Показаћемо како се, коришћењем добијених Гребнерових база, могу добити формуле Пијеријевог типа за базу $B_{k,n}$.

Уведимо нове ознаке. Нека је $\lambda = (l_1, \ldots, l_i, \ldots, l_k)$, $k$-торка целих бројева, за које је $l_1 + \cdots + l_k = n$. Тада је

$$\lambda^i := (l_1, \ldots, l_i + 1, \ldots, l_k),$$
$$\underline{\lambda} := (l_2, \ldots, l_k).$$

Коришћењем добијене Гребнерове базе добијамо формулу Пијеријевог типа:

$$c_i \cdot C^\lambda = \sum_{\|\alpha\| = n+1+\|\lambda^i\|, \alpha \neq \lambda^i} (-1)^{|\alpha|-|\lambda|} [\alpha, \lambda^i]\, C^\alpha.$$

Наиме, ако је $\lambda$ такво да је $|\lambda| < n$, онда је $c_i \cdot C^\lambda \in B_{k,n}$, док за $|\lambda| = n$, добијамо представљање производа $c_i \cdot C^\lambda$ преко елемената базе $B_{k,n}$.

Наведимо још пар занимљивих чињеница о Косткиним бројевима. Биће нам користан појам Јанговог таблоа. Јангов табло представља попуњавање Јанговог дијаграма природним бројевима, „без прескакања", тј. ако се појављује број $i+1$, мора се појављивати и $i$. Јангов табло је полустандардан уколико бројеви у врстама чине неопадајући низ, а по колонама растући низ.

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 2 | 3 | 4 |
| 2 | 3 | 3 | | |
| 3 | 5 | | | |
| 4 | | | | |
| 5 | | | | |

Облик таблоа је партиција одговарајућег Јанговог дијаграма, док је тежина таблоа $\mu = (m_i)$, где је $m_i$ број појављивања броја $i$ у овом таблоу. У наведеном примеру је облик $\lambda = (5, 3, 2, 1, 1)$, а тежина: $\mu = (1, 3, 4, 2, 2)$.

**Занимљивост:** Косткин број $K_{\lambda\mu}$ је број полустандардних Јангових таблоа облика $\lambda$ и тежине $\mu$.

Narayan је 2006. показао да је проблем израчунавања Косткиних бројева $K_{\lambda\mu}$, чак и када је $l(\lambda) = 2$, један $\sharp P$-комплетан проблем. Стога је занимљиво наћи неке рекурентне формуле помоћу којих би се могли рачунати ови бројеви. Покажимо како се то може остварити коришћењем претходних резултата.

Нека је $\alpha = (a_1, \ldots, a_k)$, $a_i \geq 0$. Са $\alpha_\rightarrow$ означимо партицију која садржи $a_i$ компоненти једнаких $i$. На пример, ако је $\alpha = (2, 0, 1, 2)$, онда је $\alpha_\rightarrow = (4, 4, 3, 1, 1)$. Приметимо да је $|\alpha| = l(\alpha_\rightarrow)$ и $\|\alpha\| = |\alpha_\rightarrow|$.

Коришћењем формуле

$$C^\alpha = (-1)^{\|\alpha\|} \sum_{|\lambda|=\|\alpha\|} K_{\lambda\alpha_\rightarrow}\sigma_{\lambda^*}$$

и добијене Гребнерове базе, добија се веза

$$\sum_{\|\alpha\|=|\lambda|} (-1)^\alpha [\alpha, \mu] K_{\lambda\alpha_\rightarrow} = 0.$$

Из ове везе, могу се добити рекурентне формуле за Косткине бројеве које их потпуно одређују.

У раду [15] ови резултати су поправљени, а добијени су и неки други резултати.

## 6. Закључак

Прича се овде не завршава. Осим резултата за Грасманијане, добијени су и резултати за просторе застава, чак и за случај ма које заставе. О томе се читаоци могу више информисати у радовима наведеним у литератури, односно на страницама аутора тих радова.

**Захвалница.** *Аутор овог рада је делимично подржан од стране пројекта Министарства просвете, науке и технолошког развоја Републике Србије* #174032.

## Библиографија

[1] **W. W. Adams, P. Loustaunau.** An Introduction to Gröbner Bases, Grad. Stud. Math., vol. 3, *American Mathematical Society, Providence*, 1994.

[2] **A. Borel.** Sur la cohomologie des espaces fibres principaux et des espaces homogenes de groupes de Lie compacts. *Ann. of Math.* **57** (1953) 115–207.

[3] **S. S. Chern.** On the multiplication in the characteristic ring of a sphere bundle. *Ann. of Math.* **49** (1948) 362–372

[4] **R. Cohen.** The immersion conjecture for differentiable manifolds, *Ann. of Math.* (2) 22 (1985), no. 2, 237-328.

[5] **D. R. Grayson, A. Seceleanu, M. E. Stillman.** Computations in intersection rings of flag bundles. *arXiv:1205.4190*

[6] **R. J. Milgram.** Immersing projective spaces. *Ann. of Math.* 85, no. 2, (1967), 473–482.

[7] **J. W. Milnor, J. D. Stasheff.** Characteristic Classes. Ann. of Math. Studies **76** *Princeton University Press, New Jersey* 1974.

[8] **K. G. Monks.** Groebner bases and the cohomology of Grassmann manifolds with application to immersion,*Bol. Soc. Mat. Mex (3)* **7** (2001), no. 1, 123–136.

[9] **Z. Z. Petrović.** On applied algebraic topology, *Proceedings of the symposium on contemporary mathematics: devoted to the 125th anniversary of the Faculty of Mathematics and to 190 years of teaching mathematics in Serbia, Belgrade, Serbia, December 18-20, 1998*; Belgrade: University of Belgrade, Faculty of Mathematics (2000) 29–37.

[10] **Z. Z. Petrović, B. I. Prvulović.** On Groebner bases and immersions of Grassmann manifolds $G_{2,n}$. *Homology Homotopy Appl.* **13**(2) (2011) 113–128.

[11] **Z. Z. Petrović, B. I. Prvulović.** Gröbner bases and some immersion theorems for Grassmann manifolds $G_{3,n}$, arXiv:1306.4814.

[12] **Z. Z. Petrović, B. I. Prvulović.** Note on immersion dimension of real Grassmannians, *Topol. Appl.*, **175** (2014) 38–42.

[13] **Z. Z. Petrović, B. I. Prvulović.** Groebner bases and non-embeddings of some flag manifolds. *J. Aust. Math. Soc.* **96**(3) (2014) 338–353.

[14] **Z. Z. Petrović, B. I. Prvulović, M. Radovanović.** Multiplication in the cohomology of Grassmannians via Gröbner bases. *J. Algebra* **438** (2015) 60–84.

[15] **Z. Z. Petrović, M. Radovanović.** Recurrence Formulas for Kostka and Inverse Kostka Numbers via Quantum Cohomology of Grassmannians. *Algebr. Represent. Theor.* Online First.

[16] **B. I. Prvulović.** Grebnerove baze i imerzije Grasmanovih mnogostrukosti, doktorska disertacija, Matematichki fakultet, Beograd 2012.
[17] **B. I. Prvulović.** Gröbner bases for complex Grassmann manifolds.*Publ. Inst. Math.* **90** (**104**) (2011) 23–46.
[18] **M. Radovanović.** Grebnerove baze za mnogostrukosti zastava i primene, doktorska disertacija, Matematichki fakultet, Beograd 2015.
[19] **M. Radovanović.** Gröbner bases for some flag manifolds and applications. *Mathematica Slovaca.* Прихваћен.
[20] **M. Radovanović.** On the $\mathbb{Z}_2$-cohomology cup-length of some real flag manifolds. *Filomat* **30** (6) (2016) 1577–1590.
[21] **T. A. Shimkus.** New immersions of Grassmann manifolds, *Bol. Soc. Mat. Mex (3)* **13** (2007), no. 2, 381–389.
[22] **S. T. Vrećica, R. T. Živaljević.** An extension of ham sandwich theorem. *Bulletin London Math. Soc.***20** (2) (1990) 183–186.

Simpozijum MATEMATIKA I PRIMENE, Matematički fakultet, Univerzitet u Beogradu, 2015 ,Vol. VI(1)

# Primena simetrične enkripcije za vertikalnu autorizaciju u bazama podataka

**Mladen Vidić**

*Matematički fakultet Beograd, Studentski trg 16, Beograd, Srbija*
*Saobraćajni fakultet Doboj, Vojvode Mišića 52, Doboj, RS, BiH*
*e-mail: mladen@matf.bg.ac.rs*

**Apstrakt.** Zaštita podataka u bazi podataka je važan aspekt funkcionisanja SUBP. Najvažnije funkcije zaštite podataka od neovlaštenog pristupa su provera identiteta korisnika, kontrola ovlaštenja korisnika za operacije u SUBP i nad objektima baze, autorizacija pristupa pojedinačnim podacima, zaštita od neovlaštenog čitanja podataka sa diska sistema. Relacione baze su najrasprostranjenije u primeni i zaštita u njima je najčešći izazov po pitanju autorizacije. Svi aspekti autorizacije se odnose na proveru i kontrolu dok je SUBP aktivan. Kada sistem nije aktivan neophodno je podatke u SUBP čuvati kriptovane za zaštitu od čitanja podataka sa diska. Operacije kriptovanja podataka u bazi omogućuju zaštitu od neovlaštenog uvida u podatke korisnicima koji ne poseduju ključ za dekripciju bez obzira da li je SUBP aktivan ili ne. Mehanizam kriptovanja podataka u bazi sa uvedenim sistemom upravljanja ključevima i njihovim obeležjima, sa definisanom relacijom poretka koja određuje ovlaštenja korisnika za definisanu vidljivost podataka, je dovoljan za vertikalnu (po kolonama, atributima) autorizaciju podataka u bazi. Složenost strukture obeležja utiče na izražajnost sistema vertikalne autorizacije. Višestruka primena istog ključa za enkripciju nad kolonama otvara potrebu za zaštitu od napada detekcije ključa. Poznate su neke tehnike zaštite od napada za detekciju ključa (ciphertext search attack). Kako na vertikalnu autorizaciju utiče opasnost od napada detekcijom ključa?

**Ključne reči:** Enkripcija; zaštita baza podataka; vertikalna autorizacija atributa; Encryption; database security; attribute based authorization; ciphertext search attack.

## 1. Uvod

U ovom radu bavićemo se primenom simetrične enkripcije za zaštitu podataka od neovlaštenog čitanja iz baze podataka u toku rada SUBP ili sa diska iz datoteka baze. Objasnićemo razliku između primene kriptografije za zaštitu podatka i primene kriptografije za autorizaciju pristupa podacima.

Za uspešno praćenje diskusije o autorizaciji podataka u bazi podataka, pretpostavlja se da je čitalac upoznat sa osnovnim pojmovima iz relacionog modela i relacionih baza podataka, odnosno razumevanje objektno-orjenitisanih baza podataka. O relacionom modelu baze podataka neophodno je poznavanje pojmova *relacija* i *atribut* relacije, *n-torka* relacije, *relaciona schema* baze podataka. Za relacione baze podataka neophodno je poznavati pojmove *tabela*, *kolona*, *red u tabeli* koji su ekvivalenti u relacionom modelu pojmovima relacija, atribut i n-torka relacije. Preporučujemo vezanu literaturu za bolje upoznavanje ovih pojmova u relacionim sistemima za upravljanje bazama podataka (RSUBP) [12]. Gde se ukaže potreba, dodefinisaćemo pojmove za potrebe praćenja izloženog teksta.

U slučaju objektno-orjenitisane baze podataka i SUBP, možemo diskutovati o analognim pojmovima. Osnovni su *klasa* objekata, *kolekcija* nad klasom objekata, *instanca objekta* date klase. Između pojmova *red tabele* u relacionoj bazi i *instanca objekta* klase u kolekciji objekata klase objektno-orjentisane baze postoji korespodencija u strukturi podataka. I jedan i drugi pojam asociraju na listu vezanih atributa u jednom složenom podatku, redu relacione tabele odnosno instanci objekta klase u objektnoj kolekciji podataka. U daljoj diskusiji zbog opštosti diskusije koristićemo pojam *pojavljivanje podataka* ili samo *pojavljivanje* kao sinonim za red relacione tabele odnosno instancu objekta klase u kolekciji objekata te klase.

Za pristup podacima u relacionoj bazi podataka koristimo upitni jezik SQL za postavljanje upita nad jednom ili više tabela baze podataka. Manipulacija podacima se postiže operacijama DML jezika (eng. DML - Data manipulation language). Za selektivni pristup podacima koriste se logički izrazi restrikcije koji sužavaju rezultat na skup pojavljivanja koja zadovoljavaju sve logičke izraze navedene u upitima ili operacijama ažuriranja. Logički kvantifikovani iskaz restrikcije koji sužava skup rezultata odnosno izmenjenih pojavljivanja tabele nazivamo *logičkim predikatom* ili samo *predikatom* koji po svojoj formi odgovaraju predikatima u predikatskom računu. Svi predikati se međusobno povezuju u listu logičkih iskaza operacijom konjunkcije (AND) i čine jedinstven logički iskaz koji filtrira u rezultatu samo ona pojavljivanja koja zadovoljavaju taj iskaz u celini. U upitima i operacijama

je moguće dopunjavati listu iskaza proizvoljnim predikatom (logičkim iskazom) ograničenim na domen podataka ekvivalentan tabelama podataka u upitima ili tabeli u operacijama nad bazom podataka.

## 2. Kriptografija u funkciji zaštite podataka

Daćemo pregled, za rad važnih, pojmova o kriptografiji. U radu se bavimo primenom kriptografije za autorizaciju podataka u bazama podataka. Nećemo se baviti detaljnim pitanjima kriptografije i kriptoanalize. Dotičemo se tipova algoritama prema načinu generisanja i razmene ključeva, ali ne i samim algoritmima, metodama generisanja ključeva, detaljnim metodama razmene ključeva za simetričnu enkripciju i ostalim pitanjima kvaliteta ključeva te optimitacijom samih algoritama. Bavićemo se primenom kriptografije da obezbedimo primarnu tajnost podataka kriptovanjem i sekundarnom kontrolom baze primenjenih ključeva. U slučaju primene kriptografije za autorizaciju u bazama podataka minimiziran je uticaj rizika razmene ključeva koji je vrlo značajan za stepen postignute zaštite, u poređenju sa slučajevima primene kriptografije u zaštiti mrežne komunikacije.

Kriptografija se primenjuje za zaštitu podataka od neovlaštenog uvida, ako za tim ima potrebe, maskiranjem zapisa originalne vrednosti podatka tekstualnim zapisom transformisane vrednosti u novu vrednost. Realizuje se transformacijom originalnog podatka u skriveni (maskirani) oblik, koji nazivamo *kriptat* (eng. *Cipher text or value*). Dakle, menjamo originalni podatak transformisanim podatkom (kriptatom). Transformacija u kriptat može biti rezultat primene funkcije koja transformaciju originalne vrednosti *A* vrši primenom neke funkcije *F* u *F(A)* ili funkcije koja pored algoritma *F* zavisi i od ključa *K* i transformiše u vrednost *F(A,K)*. Prvi tip transformacije nazivamo *Hash funkcije* koje tajnost i zaštitu podatka postižu dovoljno dobrom injektivnom (1-1) transformacijom u binarni oblik. Drugi tip funkcija su kriptografski algoritmi čija tajnost i kvalitet zaštite zavise od primenjenog ključa *K* i složenosti postupaka transformacije (obično se pretpostavlja da algoritam može biti poznat) [7,10,11].

Namena kriptogafije je u zaštiti originalnog podatka transformacijom u kriptovanu vrednost, iz koje može rekonstruisati inverznim postupkom originalnu vrednost samo korisnik ili proces koji zna inverzni algoritam Hash funkcije odnosno zna ključ za postupak demaskiranja. Demaskiranje inverznim algoritmom je rekonstrukcija originalnog podatka iz kriptata, a taj postupak nazivamo dekriptovanje.

**Definicija 1.** Postupak transformacije originalnog podatka u kriptat nazivamo *postupak kriptovanja* ili samo *kriptovanje* ili *enkripcija*. Inverzni postupak rekonstrukcije originalnog podatka iz kriptovane vrednosti nazivamo *postupak dekriptovanja* ili samo *dekriptovanje* ili *dekripcija*.

Ne umanjujući opštost, opredelićemo se u narednoj diskusiji za transformacije *F(A,K)* koje se zasnivaju na primeni algoritama kriptovanja pomoću tajnog ključa. Diskusija sa primenom Hash funkcija je jednostavniji slučaj pristupa diskusiji nego kriptovanje primenom ključa *K*. Za uspešnu zaštitu komunikacije ili skrivanje podataka, ne razmenjuje se direktno (ne prosleđuje) kroz kanal komunikacije originalni tekst *A* ili datoteka *D* nego kriptovani tekst *E=Enc(A,K)* ili datoteka *Enc(D,K)* algoritmom kriptovanja *Enc* primenjujući ključ *K* (simetrični) ili *PubK* (javni). Podacima (datoteci, tekstu, vrednosti prostog ili složenog tipa, binarnom sadržaju) može pristupiti korisnik ili proces koji znajući obrnuti algoritam *Dec* za dekriptovanje $(Dec = Enc^{-1})$ i ključ *K* ili odgovarajući privatni *PrivK* mogu dekriptovati kriptovani sadržaj i rekonstruisati originalnu vrednost *A=Dec(E,K)* ili *A=Dec(E,PrivK)*. Kao što je već naglašeno, postoje dve vrste algoritama za kriptovanje primenom ključa enkripcije, prema tome kako se konstruiše ključ koji se koristi za dekripciju. U algoritmima prve vrste koristi se isti ključ za enkripciju i dekripciju, a u drugoj klasi par ključeva *(PubK, PrivK)* koji su javni i tajni ključ za postupak enkripcije i dekripcije, redom. Algoritme *Enc* prve vrste nazivamo *simetrični*, a druge vrste *asimetrični* tip algoritama. Postupke enkripcije i dekripcije nazivamo zajedno kriptografskim postupcima, a mogu biti simetrični i asimetričini algoritmi prema načinu generisanja ključa dekripcije. Uz pretpostavku da je algoritam poznat i nepromenljiv, tajnost kriptovanog sadržaja *E* zavisi od kvaliteta ključa *K* i postignute zaštite ključa za dekripciju, *K* odnosno *PrivK*.

Hash funkcije nećemo u daljoj diskusiji razmatrati jer ne omogućuju pouzdanu rekonstrukciju originalnih podataka iz transformisane vrednosti. One služe da za identifikaciju objekata ne koristimo u radu programa originalnu vrednost identifikacionih podataka nego Hash funkcijom transformisanu, sakrivenu, nečitljivu, vrednost. Kada treba porediti zaštićene vrednosti moguće je to učiniti na osnovu hash vrednosti jer se hash vrednost ne menja, a daje dovoljnu zaštitu jer je transformacija samo u jednom smeru, od originalne ka hash vrednosti. Kada imamo potrebu da pouzdano rekonstruišemo original iz tajne vrednosti i da se originalna vrednost maskira transformisanom vrednošću primenjivaćemo algoritme kriptovanja.

Postupci koji imaju zadatak pronalaženja ključa za dekriptovanje (tzv. razbijanje šifre) nazivamo postupcima *kriptoanalize*. Kriptoanaliza nije poseban predmet ovog rada.

Obe vrste algoritama kriptovanja imaju svoju široku primenu koja zavisi od toga da li se primenjuju u sistemima koji imaju intenzivnu (frekventnu, u realnom vremenu) razmenu podataka i sadržaja binarne prirode (razmena datoteka, multimedijalni sadržaji zvuka, slike, videa) ili povremenu razmenu sa jednokratnim zahtevima za razmenom sadržaja sa uniformnom i frekventnom raspodelom pauza između dveju komunikacija. Intuitivno objašnjeno, u prvom slučaju često se razmenjuju podaci i ređe su pauze, a u drugom slučaju česte su pauze i retko se dešava komunikacija. Simetrični algoritmi su po svojoj konstrukciji brži za izvršenje i inverzna operacija dekripcije izvodi se iz algoritma enkripcije obrnutim smerom kompozicije operacija. Kod asimetričnih algoritama konstrukcija algoritma za enripciju i dekripciju i konstrukcija para *(PubK, PrivK)* zavisi od složenih matematičkih izračunavanja koja počivaju na složenim algebarskim principima. U primenama u kojima je akcenat na brzom izvršenju funkcija enkripcije i dekripcije i u kojima je moguće obezbediti pouzdano pohranjivanje i razmenu ključa preporučeno je koristiti simetrične kriptografske algoritme. Drugi odlučujući faktor za izbor vrste enkripcije, pored brzine algoritma, je zaštićeno čuvanje i razmena ključa, odnosno bezbednost razmene ključa. Kod asimetrične enkripcije razmenjuje se samo javni ključ *PubK* dok privatni *PrivK* za dekripciju nije potrebno razmenjivati. Privatni ključ se čuva samo kod vlasnika ključa i služi mu za dekripciju. Bezbedna je enkripcija i dekripcija srazmerno koliko je bezbedno pohranjen ključ kod njegovog vlasnika. Taj ključ se ne razmenjuje međusobnom komunikacijom dva sagovornika koji razmenjuju tekstualne poruke, vrednosti numeričkog ili drugog tipa ili binarne sadržaje datoteka i multimedije. Razmenjuje se samo javni ključ *PubK* koji možemo dostaviti svakom korisniku ili procesu koji imaju nameru slati poruke vlasniku ključa. Javni ključ ne može biti iskorišten za dekripciju poruka namenjenih vlasniku što obezbeđuju matematički algoritmi ugrađeni u generisanje para ključeva i algoritam dekripcije.

## 3. Čuvanje, razmena i primena ključa kao faktor pouzdanosti zaštite enkripcijom

Kod simetrične enkripcije postoji samo jedan ključ *K* i trebao bi biti razmenjen taj ključ *K*, koji bi generisala jedna strana i dostavila drugoj strani u komunikaciji.

Čuvanje i razmena ključa *K* je vrlo kritično pitanje i definiše nivo bezbednosti i upotrebljivost algoritma *Enc* i *Dec*. Ako se ključ *K* pošalje u otvorenoj neizmenjenoj (nezaštićenoj) formi *K* tada zaštita enkripcijom gubi svrhu jer svako ko bi u komunikaciji došao do sadržaja kriptovanih poruka mogao je doći prethodno i do ključa *K* kojim će dekriptovati razmenjene poruke. Druga ideja je da dva sagovornika u komunikaciji razmene neki specijalizovan ključ $K_s$ dok je bezbedna komunikacija (u bezbednoj zoni ili na početku primene softvera za komunkaciju) koji bi se koristio samo za razmenu ključa *K*. Dakle, razmenili bi $Enc(K, K_s)$ kriptovani ključ *K* ključem $K_s$. To je dobra opcija jer se ključ *K* povremeno generiše i ključ $K_s$ samo tada koristi za razmenu. Moguće je da se ključ *K* razmeni i pomoću nekog bezbednog kanala komunikacije kojim se štiti sva komunikacija sa drugim sagovornikom koja ne zahteva primenu simetrične enkripcije. Na primer, primenom asimetričnog algoritma se može razmeniti ključ *K* koji će biti korišten u daljoj komunikaciji za simetričnu enkripciju. Upotreba ključa $K_s$ ili asimetrične enkripcije za razmenu ključa *K* je ređa u frekvenciji i značajno je što usled retke primene ključa $K_s$ malo je prikupljenih poruka da bi se na osnovu njih mogao izvršiti napad na presretnute poruke i rekonstruisati ključ *K*, posebno zbog toga što se nasilnom primenom generisanih ključeva $K_{sFake}$ i pokušaja pogađanja ključa $K_s$ ne može zaključiti koja je validna vrednost ključa *K*, jer sam ključ *K* ne mora da predstavlja čitljiv i semantički jasan tekst nego binarni sadržaj predefinisane fiksne dužine. Dakle, zaštita zavisi od kvaliteta ključa *K* i koliko je on bezbedno prosleđen drugoj strani. Dodajmo prethodnom faktor bezbednog čuvanja ključa *K*. Ako napadač dođe do ključa *K* kod strana u komunikaciji, zaštita razmene ključa nema svrhu.

Kada se *K* ključ prosledi drugoj strani on se koristi sve dok se ne razmeni novi $K_{i+1}$. U nekim sistemima standard je da se za svaku novu komunikaciju $COM_{i+1}$ generiše novi ključ $K_{i+1}$ i razmenjuje. To već zahteva češću upotrebu ključa $K_s$ pa stoga i povremeno obnavljanje i razmenu tog ključa koji će biti korišten za razmenu regularnih ključeva $K_i$. Iako je to moguće raditi u bezbednim zonama, ispostavilo se da je tehnički teže primenjivo jer zavisi od tehničke zaštite komunikacionih kanala i od razmene ključa $K_s$ za razmenu ključeva $K_i$. Ključ $K_s$ obično razmenjujemo fizički u direktnoj komunikaciji bez komunikacionih kanala, a pomaže nam da razmenu ključa $K_i$ izvedemo elektronski bez fizičke razmene. Ako nije moguće fizički razmeniti $K_s$ ključ, možemo iskoristiti asimetičnu enkripciju za razmenu. U istraživanjima se došlo do zaključka da takve okolnosti mogu ipak doći raznim tehnikama napada u situaciju da se rekonstruiše ključ $K_i$ iako je bezbedno razmenjen pomoću ključa $K_s$. To je iz razloga što se $K_s$ ključ ili par ključeva *(PubK, PrivK)* koriste da se razmeni ključ $K_i$ i izbegnemo

njegovu fizičku razmenu što odgovara procesu automatizacije razmene samog ključa $K_i$, ali ključ $K_s$ ne učestvuje u zaštiti razmenjenih poruka koje se štite samo ključem $K_i$. Kod dužih i čestih komunikacija uz korišćenje ključa $K_i$ bez dovoljno valjane strategije osvežavanja ključa $K_i$ napadač može prikupiti dovoljno poruka za nasilni napad pogađanjem i rekonstrukciju ključa $K_i$ jer mu je poznat algoritam kriptovanja i dekriptovanja. Vreme nakon komunikacije je na raspolaganju napadaču da detektuje ključ $K_i$. Teorijski i praktično je to moguće usled sve većih tehničkih (procesorske, memorijske) mogućnosti i paralelizma savremenih računara. Iako ključ $K_i$ razmenjujemo kriptovan ključem $K_s$, on pre razmene i nakon razmene se fizički čuva kod dve strane u komunikaciji tako da se raznim tehnikama špijuniranja može doći do ključa $K_i$ ili čak do ključa $K_s$ koji se koristi za razmenu $K_{i+1}$ kod osvežavanja ključeva. Ako napadač otkrije ključ $K_i$ ili rekonstruiše, naredna komunikacija $COM_{i+1}$ će biti kompromitovana ako koristimo isti ključ $K_i$. Poželjno je koristiti strategiju generisanja novog ključa za svaku novu komunikaciju. Sledeći su ključni faktori pouzdanosti primene algoritama za simetričnu enkripciju:

- Izbor algoritma enkripcije/dekripcije,
- Generisanje ključeva i strategija osvežavanja ključa u novim razgovorima,
- Bezbednost lokacija čuvanja ključeva kod strana u komunikaciji,
- Mehanizam razmene ključeva sa drugom stranom u komunikaciji.

U svim primenama simetrične enkripcije najkritičniji faktori zaštite su lokacija čuvanja ključeva i razmena ključeva. Kvalitet mehanizma razmene ključeva utiče na izbor strategije i mehanizma osvežavanja ključeva. Izbor algoritma je nezavisan od ostala tri faktora jer oni utiču na sam proces generisanja pouzdanog ključa. Izbor algoritma je predmet naučnog istraživanja i standardizacije koji zavise od odluka unapred donesenih u okviru organizacije ili između dveju strana u komunikaciji. Naravno, nesporno je da kvalitet i pouzdanost algoritma znači za kvalitet zaštite kriptovanjem ali se u ovom radu polazi od pretpostavke da je korisnik u mogućnosti da izabere najbolje poznate algoritme i da ih koristi. Kada se već odluči za algoritam ono što je promenljivo i stalno se razmenjuje su ključevi koji su kritični za nevlašteno dekriptovanje sadržaja koji treba da se zaštićeno čuva u bazi podataka ili razmeni u komunikaciji. Zbog prethodnog se za razmenu ključa $K_i$ simetrične enkripcije retko koristi mehanizam "generiši samostalno ključ $K_i$ u celini i razmeni sa drugom stranom kriptovan unapred razmenjenim ključem $K_s$".

### 3.1. Napredne tehnike razmene ključa enkripcije

Ključ mogu generisati obe strane u dogovorenoj komunikaciji kada se postiže veći stepen pouzdanosti čuvanja ključa i minimizira mogućnost njegove kompromitacije. Sve napredne tehnike razmene ključa $K_i$ uzimaju u obzir faktore da i posedovanje ključa $K_s$ i razmena ključa $K_s$ su kritični i da ključ $K_s$ može biti kompromitovan. U tom slučaju razmena ključa $K_i$ nije pouzdana te i kasnija komunikacija kriptovana tim ključem $K_i$. Postoje dve osnovne tehnike (protokola): da se ključ $K_i$ generiše u celini pa se razmenjuje tehnikom trofazne razmene ključa $K_i$ ili tehnike pregovaranja između dva sagovornika kako bi oni umesto jednostrano kreiranog ključa $K_i$ i njegove razmene ispregovarali zajedničku izgradnju (generisanje ključa) $K_i$ bez razmene u celini. Prva tehnika je poznata kao *Three-pass protokol*[10,11] sa varijacijama za razmenu tajnog ključa. Za drugi tip tehnika sa pregovaranjem generisanog ključa najpoznatija je *Diffie-Hellman strategija*[10,11] razmene i generisanja ključa simetrične enkripcije gde se ključ koristi jednokratno nakon izgradnje i ne čuva se pohranjen kod strana u komunikaciji. To je protokol koji predviđa da korisnici za svaku novu komunikaciju generišu novi ključ $K_i$, na početku komunikacije. Pri tome se može koristiti deo prethodnog ključa $K_{i-1}$ koji je verifikovan i time inicijalozovan zajednički koren niza za generisanje novih delova ključa, koje razmenjuju i ponovo kombinuju tako da se na svakoj strani dobije generisan ključ $K_i$ za simetričnu enkripciju, a da nije razmenjen u celini. Jedna strana ne zna tajni deo druge strane celog ključa u komunikaciji. Tajni delovi se ne razmenjuju. Napadač koji presretne komunikaciju i deo ključa $K_i$ u razmeni ne zna ključ u celini jer ne zna tajni deo koji nedostaje, minimalno od jedne strane u komunikaciji, a koji se ne razmenjuju nego svaka strana čuva kod sebe. U svakoj narednoj komunikaciji deo ključa $K_i$ se može koristi za generisanje delova koji će biti razmenjeni za generisanje novog ključa $K_{i+1}$ u celini. To je procedura koja se ponavlja za generisanje novog ključa $K_{i+1}$ u svakoj novoj komunikaciji. Ključ $K_i$ se ne razmenjuje u celini između strana u komunikaciji, ne čuva u celini ni jedna strana u komunikaciji, jedna strana ne zna tajni deo u ključu druge strane. Ako bi napadač presreo trenutnu komunikaciju i uspeo rekonstruisati ključ $K_i$ , što je teorijski moguće ali ne i praktično ako je malobrojan skup razmenjenih poruka, isti ne može biti upotrebljen u narednim komunikacijama za koje je generisan novi ključ $K_{i+1}$ i koje se kriptuju pomoću njega. Detaljnije za izučavanje kriptografije i kriptoanalize se preporučuje upotreba literature [7,10,11].

## 4. Zaštita u bazama podataka i vertikalna autorizacija

U bazama podataka, koje su jedan podsistem za realizaciju informacionih sistema, zaštita podataka od neovlaštenog uvida, izmene, generisanja velike količine podataka, gubitka ili oštećenja podataka je važan aspekt zaštite u informacionim sistemima. Osnovne funkcije zaštite, koje samo navodimo, su sledeće:

**a)** Identifikacija korisnika - korisnik saopštava svoje identifikacione podatke;

**b)** Provera autentičnosti - sistem na osnovu dostavljenih informacija proverava da li je osoba koja se predsatvlja identifikacijom;

**c)** Autorizacija sistemskih operacija, upita i operacija nad objektima baze podataka (DKP autorizacija - Diskreciona kontrola pristupa) - kontrola dozvola za operacije u sistemu ili nad objektima;

**d)** Autorizacija podataka (pojavljivanja, kolona, ćelija) - kontrola pristupa pojedinačnim podacima u BP, u samim kolekcijama podataka;

**e)** Autorizacija poslovnih funkcija - kontrola izvršenja složenih transakscionih operacija i kreiranja izveštaja, pregleda i grafikona nad podacima u sistemu;

**f)** Praćenje rada sistema i ponašanja korisnika tog sistema;

**g)** Kreiranje zaštitnih kopija sistema i strategija i procedure oporavka sistema;

**h)** Kriptografija i primena za zaštitu lokalnih podataka, komunikacije i podataka na serveru paze podataka (operativni sistem ili u bazi podataka);

**i)** Mrežna zaštita komunikacije sa BP.

Za naše istraživanje, fokus je na autorizaciji podataka u bazi podataka primenom kriptografije. Za sveobuhvatniju i detaljniju diskusiju o zaštiti u informacionim sistemima i bazama podataka predlažemo [8].

Autorizacija podataka podrazumeva kontrolisanje pristupa podacima prema ovlaštenjima korisnika i deklarisanoj dostupnosti (*vidljivosti*) podataka. Definiše se pojam *vertikalna autorizacija podataka* kao *autorizacija pristupa vrednostima atributa* relacije u izrazima upita i operacija nad bazom podataka odnosno kolekcije pojavljivanja podataka.

**Definicija 2.** *Vertikalna autorizacija podataka* u bazi podataka je mehanizam definisanja selektivnih ovlašćenja korisnika i njihove kontrole kod pristupa korisnika u izrazima upita i operacija kolonama relacione tabele, odnosno kolekcije objekata.

U narednom tekstu korištene su oznake $B_1$,$B_2$,...,$B_p$,...,$B_n$ za kolone $B1$,$B2$,...,$Bp$,...,$Bn$, redom, u tabeli TP. $A_i$ je oznaka i-te izdvojene kolone $Ai$ u istoj tabeli, $LA_i$ je oznaka kolone $LAi$ obeležja pridruženih koloni $Ai$. Oznake u tekstu $A_i^j$ referišu vrednosti kolone $Ai$ u *j*-tom pojavljivanju. U opštem slučaju, redni broj pojavljivanja pozicioniramo iznad oznake vrednosti atributa jer su indeksi *i*, kao u prethodnom, uključeni u nazive atributa. Oznaka $V^j$ je referenca na vrednost promenljive ili atributa *V* u *j*-tom pojavljivanju, gde je *j* redni broj pojavljivanja. U bazama podataka imamo rešenu diskrecionu autorizaciju (*DKP* - diskreciona kontrola pristupa)

**Tabela 1.** Tabela TP. Vrednost $A_i^4$ atributa $A_i$ u 4. pojavljivanju potrebno je sakriti ako korisnik nema ovlaštenje da pristupi tom podatku.

| $Rbr:j$ | $B_1$ | $B_2$ | $B_3$ | ... | $A_i$ | $LA_i$ | $B_p$ | $B_{p+1}$ | ... | $B_n$ | $Rid$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | ... | ... | ... | ... | $A_i^1$ | $LA_i^1$ | ... | ... | ... | ... | $Rid^1$ |
| 2. | ... | ... | ... | ... | $A_i^2$ | $LA_i^2$ | ... | ... | ... | ... | $Rid^2$ |
| 3. | ... | ... | ... | ... | $A_i^3$ | $LA_i^3$ | ... | ... | ... | ... | $Rid^3$ |
| 4. | ... | ... | ... | ... | $A_i^4$ | $LA_i^4$ | ... | ... | ... | ... | $Rid^4$ |
| 5. | ... | ... | ... | ... | $A_i^5$ | $LA_i^5$ | ... | ... | ... | ... | $Rid^5$ |
| 6. | ... | ... | ... | ... | $A_i^6$ | $LA_i^6$ | ... | ... | ... | ... | $Rid^6$ |
| 7. | ... | ... | ... | ... | $A_i^7$ | $LA_i^7$ | ... | ... | ... | ... | $Rid^7$ |
| 8. | ... | ... | ... | ... | $A_i^8$ | $LA_i^8$ | ... | ... | ... | ... | $Rid^8$ |

nad kolonama tabela. Primer komande za autorizaciju nad tabelom TP za projekciju kolona je:

```
grant select on TabelaTP(B1,B2,B3,Ai,LAi,Bp,Bn) to korisnik1;
```

Pojam vertikalne autorizacije je proširenje pojma diskrecione kontrole pristupa nad kolonama (atributima). Potreban je složeniji mehanizam za kontrolu autorizacije podataka nego osnovni da/ne mehanizam za kontrolu operacija nad objektima baze podataka. Kada je potrebno kontrolisati pristup kolonama (atributima relacija) neophodno je skrivati sadržaj podatka od neovlaštenih korisnika.

Vertikalna autorizacija podrazumeva uslovno skrivanje vrednosti atributa u izrazima upita i operacija korisnika za pojavljivanja koja ulaze u rezultat upita i operacija nakon primene predikata i eventualnih mehanizama horizontalne autorizacije. Horiyontalna autorizacija je opcionalni mehanizam za finu granulaciju skupa pojavljivanja koja ulaze u rezultat na osnovu ovlaštenaj korisnika za ta pojavljivanja. Za korisnike taj mehanizam moč biti transparentan i ne utiče na dalju diskusiju o autorizaciji atributa (kolona). Uvodimo imenovane labele (oveležja) za identifikaciju ovlaštenja korisnicima nad podacima jednog atributa (projekcija podataka na jedan atribut). Imamo skup obeležja $L = \{L_1, L_2, L_3, \ldots, L_n\}$ nad kojim važi relacija poretka $L_i \geqslant L_j$, za $i, j = 1, \ldots, n$. Svaki korisnik dobija obeležje za čitanje i obeležje za upis/izmenu podataka nad atributom $A_i$ tabele *TP*.

Za vertikalnu autorizaciju neophodan je mahanizam poretka nad skupom obeležja. Definišemo relaciju poretka "$\geqslant$" nad skupom $L$ obeležja.

**Definicija 3.** "$L_1$ je jednako ili nadređeno $L_2$" znači da su podaci obeleženi sa $L_1$ višeg ranga zaštićenosti nego podaci obeleženi sa $L_2$. Korisnik sa obeležjem $L_1$ može videti podatke obeležene labelom $L_2$, odnosno sve podatke obeležene labelom $L_p$ za koje važi $L_{korisnika} \geqslant L_p$. Korisnik sa labelom $L_1$ vidi potencijalno više podataka nego korisnik sa $L_2$ obeležjem. Čitamo i "$L_1$ može da vidi minimalno $L_2$".

Autorizaciju atributa $A_i$ postižemo skrivanjem vrednosti atributa. Uvodi se dodatna kolona sa obeležjima. Ako je naziv atribut $A_i$, naziv nove kolone je $LA_i$ čiji domen vrednosti odgovara vrednostima obeležja. Ovde se autorizivani atribut štiti skrivanjem koristeći uslovni izraz. Nismo još došli do primene enkripcije. Poštuje se konvencija u svim upitima i izrazima:

$$\text{get A:= if} \quad L_{korisnika} \geqslant LA_i \quad \text{then} \quad A_i \quad \text{else} \quad NULL. \tag{1}$$

Još uvek možemo izostaviti proveru $L_{korisnika} \geqslant LA_i$ i narušiti autorizaciju. To je pokušaj postizanja autorizacije atributa $A_i$ preko obeležja $LA_i$ poštujući uslovni izraz. Korisnici se autoriziju nad atributom $A_i$ obeležjima $LA_i$. Obeležja se koriste za definisanje vidljivosti podataka. Ovaj pristup ima niz nedostataka. Dodaje se nova kolona $LA_i$ i povećava se fizički dužina zapisa reda tabele (pojavljivanja kolekcije). Svaki korisnik ili proces koji ima pristup atributu $LA_i$ "vidi" sva obeležja svih dostupnih pojavljivanja. I onih pojavljivanja čije vrednosti atributa $A_i$ "ne treba da vidi". Autorizacija atributa samo skrivanjem vrednosti uslovnim izrazom koji upoređuje vrednost obeležja podatka nije dobro rešenje. Zbog prethodnog, skrivanje $A_i$ nije pouzdano za autorizaciju i praktično upotrebljivo ali ilustruje namenu skrivanja.

## 5. Kriptografija u bazi podataka i vertikalna autorizacija atributa

Kako iskoristiti enkripciju za kontrolu pristupa ćelijama u tabelama? Za primenu kriptografije u bazama podataka, specijalno za realizaciju vertikalne autroizacije atributa primenom kriptografije, nije potrebno vršiti razmenu ključeva jer su svi podaci u bazi podataka i svi ključevi bezbedni u zaštićenoj bazi ključeva gde su sačuvani nakon generisanja. Iz te baze se koriste ključevi i za kriptovanje sadržaja podataka u kriptat i za dekriptovanje u originalnu vrednost. Presudni faktori za kvalitet kriptovanja u realizaciji autorizacije su izbor algoritma, generisanje i povremeno regenerisanje ključeva i zaštićenost lokacije u bazi podataka gde se čuva baza ključeva. Kvalitet vertikalne autorizacije postignute kripovanjem zavisi i od mehanizma upravljanja izborom ključeva koji su iskorišteni za kriptovanje sadržaja i izbor ključa koji treba primeniti za dekriptovanje podatka. Koristićemo oznake: *S* - sadržaj podatka, *ES=AE(S, K)* kriptovani sadržaj primenom algoritma *AE* i ključa *K*, *S=AD(ES, K)* je dekriptovani sadržaj primenom inverznog algoritma *AD* i ključa *K*.

Neka je $L = \{L_1, L_2, \ldots, L_n\}$ skup obeležja i njima pridruženi ključevi $KB = \{K_1, K_2, \ldots, K_n\}$ u bazi ključeva. Ako koloni $A_i$ pridružimo kolonu obeležja $LA_i$ , a ćelijama vrednosti $A_i^j$ kolone $A_i$ sa obeležjima $LA_i^j \in L$ kolone $LA_i$ koja imaju pridružene ključeve $KLA_i^j \in KB$ iz baze ključeva, možemo sve ćelije $A_i^j$ kriptovati pridruženim ključem u vrednost $E(A_i^j, KLA_i^j)$. Oznake: obeležjima $LA_i^j$ atributa $LA_i$ za ćeliju $A_i^j$ atributa $A_i$ pridružuje se ključ za enkripciju $KLA_i^j$.

Skup obeležja *L* je uređen relacijom poretka ako je realizovana relacija poređenja obeležja koja odgovara definiciji 3. Realizacija relacije poretka obeležja može biti izvedena iz numeričkog, alfanumeričkog ili implicitnog poretka na osnovu složene strukture i poretka obeležja. Najčešće je kompozitni poredak preko definisane relacije za upoređivanje dva obeležja $L_k \geqslant L_p$ koja ima sva svojstva relacije poretka na skupu obeležja *L*.

Primenićemo enkripciju vrednosti atributa $A_i$ i zameniti sa $E(A_i^j, KLA_i^j)$ da bismo postigli obevezu primene uslova poređenja obeležja predikatom:

$$L_{korisnika} \geqslant LA_i \tag{2}$$

u upitu ili operaciji, pri čitanju ili izmeni podataka u tabeli nad atributom $A_i$.

Upitom nad skupom obeležja pridruženih svim ključevima (podskup od *L*) iz baze ključeva *KB* i poređenjem sa obeležjem korisnika predikatom (2) izdvajamo podskup obeležja $L_k = \{L_k^1, L_k^2, \ldots, L_k^s\}$ "podređenih" obeležju korisnika. Možemo definisati, pomoću skupa $L_k$, da redom pridruženi ključevi $K_k = \{K_k^1, K_k^2, ..., K_k^s\}$ su "dostupni" korisniku sa obeležjem $L_{korisnika}$.

**Definicija 4.** Skup ključeva "dostupnih" korisniku sa obeležjem $L_{korisnika}$ je skup $K_k = \{K_k^1, K_k^2, ..., K_k^s\}$. Ključ iz *KB* je dostupan korisniku ako se nalazi u skupu $K_k$ za tog korisnika.

Dekriptovanje ključem $D(E(A_i^j, KLA_i^j), KLA_i^j) = A_i^j$ omogućava da podaci budu dostupni ako i samo ako korisnik ima ključ $KLA_i^j$ na raspolaganju u skupu $K_k$, inače se dobija markirana vrednost ili *NULL*. Možemo uvesti za atribut $A_i$ izabrano podrazumevano obeležje $LA_i^d \in L$ ako u nekom novom pojavljivanju vrednost $A_i$ nema pridruženo obeležje (nije dodeljeno korisniku). Za podrazumevano obeležje $LA_i^d$ pridružen je iz baze ključeva ključ za enkripciju $KLA_i^d \in KB$.

Autorizaciju pristupa podacima postižemo skrivanjem vrednosti atributa kriptovanjem ključem pridruženim obeležju $L_{korisnika}$ kreatora podatka ili podrazumevanom obeležju $LA_i^d$. Uvodi se dodatna kolona da bismo sačuvali vrednost sa obeležjima. Za atribut $A_i$ nova kolona je $LA_i$ koja sadrži obeležja ili njihove jedinstvene identifikatore ako normalizujemo model. Poštuje se konvencija u svim upitima i izrazima:

$$\text{get A:= if} \quad L_{korisnika} \geqslant LA_i \quad \text{then} \quad Decrypt(A_i, getkey(LA_i)) \quad \text{else} \quad NULL, \tag{3}$$

gde je $getkey(L)$ funkcija koja za rezultat vraća ključ $KL$ za kriptovanje pridružen obeležju $L$.

Ako se izbegne provera $L_{korisnika} \geqslant LA_i$ i naruši autorizacija dobiće se kriptovani tekst umesto originala, ako korisnik ne dobije odgovarajući ključ. To je pokušaj postizanja autorizacije atributa $A_i$ preko obeležja i njegovog ključa.

Moramo sadržaj $A_i$ kriptovati ključem $KLA_i$ koji je pridružen obeležju $LA_i$ tako da je korisniku dostupan sadržaj atributa $A_i$ ako i samo ako uspešno dekriptuje sadržaj. Ako mu je dostupan ključ $KLA_i$ (u skupu $K_k$) biće uspešno dekriptovana vrednost, inače neće dekriptovati uspešno. Uslov $L_{korisnika} \geqslant LA_i$ moramo preneti na izbor ključa za dekripciju umesto uslovnog vraćanja $A_i$ vrednosti kao rezultata. To obezbeđuje da aplikacija u upitima i operacijama nad bazom mora uvek da primenjuje taj uslov i nije moguće izostaviti uslov, a da se ne uoči greška, u funkciji $getkey(LA_i)$ za izbor ključeva. Ovim se kreira potreba da uz enkripciju različitim algoritmima uvedemo i mehanizam upravljanja hijerarhijom ključeva i njihovim skrivenim, bezbednim, čuvanjem u bazi ključeva *KB*.

## 6. Enkripcija vrednosti u ćeliji izabranim ključem

Naredna tabela ilustruje popunjavanje vrednosti u atributu tabele kriptovanim vrednostima gde je $E_i^j = E(A_i^j, KLA_i^j)$, $KLA_i^j = getkey(LA_i^j)$. Neka obeležja $LA_i^j$ se mogu ponavljati u koloni $LA_i$ zbog primene istih ključeva enkripcije zavisno od obeležja korisnika $L_{korisnika}$ vlasnika podatka ili primene podrazumevanog obeležja kolone $LA_i^d$. U (4) vidimo primer upotrebe funkcije dekripcije u SQL izrazima:

$$select \quad B1, B2, D(Ai, getKey(LAi)) \quad from \quad TabelaPodaci \quad where... \tag{4}$$

Kriptujemo i čuvamo u tabeli pod atributom $A_i$, umesto vrednosti atributa $A_i$, kriptovanu vrednost $E_i = E(A_i, KLA_i)$. U kompletnijoj realizaciji tabela baze ključeva iz primera u Tabela 3 je modelovana tako da se koristi posebna tabela za skup obeležja *L*, a druga tabela za bazu ključeva *KB*. U tabeli, obeležja je potrebno referencijalno povezati na identifikatore ključeva u bazi ključeva.

**Tabela 2.** Vrednosti pojavljivanja za atribut $Ai$ su zaštićene kriptovanim vrednostima

| $Rbr:j$ | $B_1$ | $B_2$ | $B_3$ | ... | $A_i$ | $LA_i$ | $B_p$ | $B_{p+1}$ | ... | $B_n$ | $Rid$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | ... | ... | ... | ... | $E(A_i^1, KLA_i^1)$ | $LA_i^1$ | ... | ... | ... | ... | $Rid^1$ |
| 2. | ... | ... | ... | ... | $E(A_i^2, KLA_i^2)$ | $LA_i^2$ | ... | ... | ... | ... | $Rid^2$ |
| 3. | ... | ... | ... | ... | $E(A_i^3, KLA_i^3)$ | $LA_i^3$ | ... | ... | ... | ... | $Rid^3$ |
| 4. | ... | ... | ... | ... | $E(A_i^4, KLA_i^4)$ | $LA_i^4$ | ... | ... | ... | ... | $Rid^4$ |
| 5. | ... | ... | ... | ... | $E(A_i^5, KLA_i^5)$ | $LA_i^5$ | ... | ... | ... | ... | $Rid^5$ |
| 6. | ... | ... | ... | ... | $E(A_i^6, KLA_i^6)$ | $LA_i^6$ | ... | ... | ... | ... | $Rid^6$ |
| 7. | ... | ... | ... | ... | $E(A_i^7, KLA_i^7)$ | $LA_i^7$ | ... | ... | ... | ... | $Rid^7$ |
| 8. | ... | ... | ... | ... | $E(A_i^8, KLA_i^8)$ | $LA_i^8$ | ... | ... | ... | ... | $Rid^8$ |

**Tabela 3.** Primer tabele obeležja i ključeva

| ObelezID | Naziv obelezja | Kljuc | KljucNaziv |
|---|---|---|---|
| Lid1 | Obelezje1 | ########### | Kljuc1 |
| Lid2 | Obelezje2 | ######## | Kljuc2 |
| Lid3 | Obelezje3 | ################ | Kljuc3 |
| Lid4 | Obelezje4 | ### | Kljuc4 |
| Lid5 | Obelezje5 | ############### | Kljuc5 |

U koloni $LA_i$ su obeležja koja se potencijalno ponavljaju, pa za pojavljivanje $P_j$ u kome je vrednost $E_i^j = E(A_i^j, KLA_i^j)$ nad atributom $A_i$, je i dalje vredost obeležja $LA_i^j$ kome odgovara ključ $KLA_i^j$. U izrazu upita ili operacije *DML* jezika, korisnik pristupa listi obeležja u tabeli pridruženih ključeva i može naći ključ $KLA_i^j$ izrazom $KLA_i = getKey(LA_i)$ i dekriptovati $E_i^j$ izrazom $D(A_i, KLA_i)$ jer je u $A_i^j$ upisana vrednost $E_i^j$.

Korisnik je pristupio obeležju i dobio ključ iz baze ključeva bez ograničenja za predikat (2) i bez provere da li je bio autorizovan za dobijanje ključa. Nije postignuta zaštita nad ćelijom jer ako korisnik ima odgovarajuću *DKP* dozvolu nad atributima $A_i$ i $LA_i$ omogućeno mu je da dobije proizvoljan ključ za dekripciju, što nije autorizacija.

Zbog ovog nedostatka pristup ključevima mora biti uslovljen obeležjem koje poseduje korisnik koji pristupa bazi, iz istog skupa obeležja kao i $LA_i$ vrednosti, i relacijom poretka nad tim skupom obeležja.

Korisnik ne sme pristupati svim obeležjima i pridruženim ključevima bez ogrančenja, nego samo onim ključevima $KLA_i$ takvim da je $L_{korisnika} \geqslant LA_i$. Potrebno je da korisniku dodelimo obeležje $L_{korisnika}$ koje "može da vidi" obeležje $LA_i$ da bi imao ovlaštenje preko obeležja $LA_i$ za pristup ključu $KLA_i$.

Ključevi ne smeju biti slobodno dostupni u bazi nego moraju biti u posebnom skladištu, bazi ključeva. Korisnici neće imati direktan pristup tabeli ključeva nego samo funkciji koja vraća ključeve za obeležja.

Kada korisnik pristupi bazi ključeva proverom se omogućava da pristupa samo ključevima nad čijim obeležjima ima ovlaštenje.

Imali smo u (1) get A:= if $L_{korisnika} \geqslant LA_i$ then $A_i$ else $NULL$; uz strogo poštovanje uslova provere (2) $L_{korisnika} \geqslant LA_i$. Ako se zaobiđe uslov (2), obeležja nemaju svrhu i vertikalna autorizacija nije postignuta. Teško je obavezati programere bez formalizma da tu proveru uvek poštuju, a moguće su i greške bez namere. Sada dobijamo novi izraz u (3) $getA := if L_{korisnika} \geqslant LA_i then D(E_i, KLA_i) else NULL$. Ubacivanje ovog kriterijuma (2) na sva mesta u upitima, pogledima i operacijama je logično i očekivano. Ako ovde zaobiđemo uslov $L_{korisnika} \geqslant LA_i$, rezultat izraza $D(E_i, KLA_i)$ zavisi od toga da li korisnik poseduje dozvolu za ključ $KLA_i$ u skupu $K_k$. Zaključujemo da funkcija za dekripciju treba da proverava i razrešava da li se za rezultat vraća dekriptovan sadržaj iz $E_i$ ili *NULL* vrednost.

## 7. Čuvanje ključeva u skladištu/bazi ključeva

Skladište ključeva je posebno izgrađen model u istoj bazi podataka koji nije dostupan drugim korisnicima i administratorima sem jednom korisniku, vlasniku scheme baze podataka u kojoj je baza ključeva.

Ako se čuvaju ključevi u zasebnoj tabeli sa identifikatorima i nazivima ključeva, a u tabeli obeležja se čuvaju samo uparena obeležja sa identifikatorima ključeva i nazivi obeležja, vraćanje rezultata $D(E_i, getKey(LA_i))$

gde je u tabeli obeležja uz $LA_i$ bio ključ $KLA_i$, svodi se na formulu:

$$D(E_i, getKeyFromKB(getKeyId(LA_i))) = \begin{cases} NULL & \text{ako} \big(E_i = NULL \vee LA_i = NULL\big), \\ ?^* & \text{ako} \big(returnedKey = NULL\big), \\ Decrypt(E_i, returnedKey) & \text{inače} \end{cases}$$

$$\text{gde je} \quad returnedKey = getKeyFromKB(getKeyId(LA_i)). \quad (5)$$

Ako je $returnedKey = NULL^*$ imamo dve mogućnosti za definisanje rezultata:

a) *NULL*, jer je vraćeni ključ bio nepoznat ili nedostupan;

b) $E_i$, jer ključ kojim je kriptovan $E_i$ je nedostupan tekućem korisniku.

U slučaju da imamo kriptovanu vrednost $E_i$, ključ je bio definisan i uneto je obeležje $LA_i$ sa pridruženim ključem. Ako je $L_{korisnika} \geqslant LA_i$ ključ treba da bude dostupan. Ako nije tačan uslov ključ nije dostupan i vraćeni ključ je *NULL* i prihvatljivo je da se kao rezultat prihvati zaštićena vrednost $E_i$, ili *NULL* kao u opciji pod a. Ovo je u slučaju da je uveden poredak obeležja i da je ključ nedostupan zbog neautorizovanog pristupa vrednosti $E_i$ sa obeležjem $LA_i$. Rezultat zavisi od toga da li je korisnik dobio ključ za obeležje $LA_i$ pridruženo kriptovanoj vredosti $E_i$ (kriptat za $A_i$) i da li je uvedena relacija poretka nad obeležjima?!

Ako je uvedena relacija poretka na obeležjima tada se može naći obeležje koje odgovara ključu i uporediti sa obeležjem korisnika. Ključ će biti iskorišćen za dekriptovanje ako je dostupan korisniku preko relacije obeležja. Primeniće se taj ključ samo ako je "aktivan" za tog korisnika.

Kada se prijavi na SUBP korisniku je poznato obeležje koje mu je pridruženo . Nisu svi ključevi enkripcije iz baze ključeva *KB* automatski primenjivi za dekripciju. Ako bi se u svakom upitu ponavljalo izvršenje funkcije za proveru aktuelnosti ključa performanse upita i operacija bi bile ugrožene zbog korelisanog ponavljanja izvršenja funkcije $getKey(LA_i)$ eksplicitno ili implicitno u izrazu poziva funkcije $D(E_i, getKey(LA_i))$ sa dva ili samo jednim argumentom $D(E_i)$.

Korisnik u toku izvršenja operacija i upita (*SUBP* online) vidi dekriptovane podatke tabele samo u ćelijama atributa $A_i$ koje se kriptovane ključevima koji su aktivni u tekućoj sesiji. Podaci su uvek čuvaju u bazi u kriptovanom formatu $E_i$ i čitanjem podataka iz datoteka baze, u slučaju pristupa, mogu se videti te kriptovane vrednosti. Originalne vrednosti su zaštićene transformacijom u kriptat. Dekripcija je jedini metod da se pročita podatak iz kriptovanog sadržaja.

Ključ je primenjiv za dekripciju u sesiji ako postoji u bazi ključeva i ako je aktivan za sesiju. Kako kontorlisati dostupnost ključevima od strane korisnika? Ključevi mogu biti deaktivirani i aktivirani. Samo aktiviran ključ može se upotrebiti za dekripciju. Pri ostvarivanju sesije aktiviraju se svi ključevi na koje korisnik ima pravo (dostupni).

Uvodimo pojam "ključ aktivan u sesiji". Za autorizaciju pristupa mora se konsultovati, na osnovu relacije poretka i obeležja korisnika u sesiji, koji su ključevi aktivni za sesiju.

Primetimo da zaštita zavisi od egzistencije ključa (neophodan uslov) i ne samo od egzistencije (jer bi ga mogao upotrebiti svako ko ima pravo pistupa bazi ključeva) nego i od dinamičkog statusa ključa "da li je aktivan za sesiju".

Isti ključ koji je kreiran u bazi ključeva može biti suspendovan ili ne, što važi za sve korisnike i sesije. Ako nije suspendovan, u nekim sesijama može biti aktivan, a u drugim da nije.

Ključevi ne smeju da se aktiviraju direktno od strane nekog korisnika niti administratora, ali se aktiviraju na početku sesije indirektno pozivanjem procedure za aktivaciju ključa.

Procedura za aktivaciju jednog ključa radi obaveznu proveru uslova da li korisnik iz sesije koja izvršava proceduru ima ovlaštenje nad tim ključem.

S obzirom da je više ključeva korišteno za zaštitu atributa $A_i$ kriptovanjem, da bi se postigla autorizacija nad celom kolonom moraju da se aktiviraju iterativno prethodnom procedurom svi dostupni ključevi.

Da bi sve radilo po pitanju autorizacije, neophodno je na početku sesije ili svake transakcije aktivirati sve ključeve nad čijim obeležjima korisnik ima ovlaštenje.

Primena enkripcije i dekripcije kod operacija i upita podrazumeva da u operacijama za insert i update koristi se poziv funkcije za kriptovanje nove vrednosti sa aktivnim ključem koji je pridružen obeležju korisnika u sesiji ili podrazumevanom obeležju $LA_i^d$. Pri izboru obeležja i pripadajućeg ključa za kriptovanje podataka zaštićene

kolone potrebno je pridržavati se utvrđenog kriterijuma upisa uz kriptovanje ključem tekućeg ovlaštenja korisnika ili podrazumevanim ključem koji je pridružen labeli $LA_i^d$ atributa $A_i$. Može se koristiti ključ obeležja koje definiše manju osetljivost od tekućeg ovlaštenja korisnika.

### 7.1. Zaštita i tajnost baze ključeva

Mehanizam pristupa ključevima mora obezbediti zaštićenu i autorizovanu kontrolu operacija i upita nad bazom ključeva. Ključevi ne smeju biti dostupni običnom korisniku. Koristi se baza ključeva odvojena od baze obeležja ključeva. Samo specijalni korisnik može pristupiti ključevima, koji je vlasnik scheme baze ključeva. Svi ostali korisnici pristupaju preko procedura i funkcija za rad sa ključevima samo sa pravima za izvršenje tih procedura i funkcija. Samo taj specijalni korisnik, vlasnik baze ključeva, ima privilegije nad bazom ključeva (tabelama modela).

Kod operacija nad datotekama baze podataka podaci atributa $A_i$ u tabeli *T* su zaštićeni kriptovanjem. Bazu ključeva možemo kriptovati nekim izabranim algoritmom da zaštitimo deo datoteke baze podataka u kojem je baza ključeva.

## 8. Napredne tehnike i istraživanja vezana za vertikalnu autorizaciju

Prethodnom diskusijom o vertikalnoj autorizaciji je ponuđeno rešenje koje, uz izložene prednosti, otvara nova pitanja. Upotreba ključeva za enkripciju u istoj bazi u kojoj su sakriveni podaci mora biti obezbeđena da bi se opasnosti kriptoanalize svele na minimum. Otvorena je lista pitanja za dalje istraživanje pojma vertikalne autorizacije, o pouzdanosti višestruke primene ključeva iz baze ključeva za enkripciju podataka prema raspodeli obeležja, generisanju ključeva i sakrivanju baze ključeva, performansama i standardizaciji njene primene u bazama podataka. Sledi lista tema za istraživanje:

**a)** Višestruka primena istog ključa za enkripciju nad kolonama otvara opasnosti otkrivanja ključa. Potrebna je raditi zaštitu od napada za detekciju ključa.

**b)** Rešavanje dekripcije u bazi uklanjanjem drugog argumenta u pozivu funkcije dekripcije i implicitna primena ključa dekripcije.

**c)** Za sprečavanja napada skeniranja ključa Key Scan Attack značajna je u rešenju primena sugestije da funkcija dekripcije $D(E_i)$ vraća *NULL* vrednost umesto $E_i$ ako korisnik nije imao ovlaštenje za tu ćeliju podataka. Važiće $D(E_i) = A_i$ ako korisnik ima ovlaštenje nad obeležjem ključa koji je u $E_i$, inače je *NULL*. Zaštita od otkrivanja ključa kriptovanja pogađanjem je postignuta time što funkcija za dekripciju $D(E_i)$ koristi implicitni ključ. Ako nema implicitni ključ vraća *NULL* vrednost. Nije moguće proslediti eksplicitno ključ za dekriptovanje.

**d)** Ako bi se za neautorizovan pristup sadržaju ćelija u izrazu $D(E_i)$ vraćala vrednost $E_i$ umesto *NULL* imali bismo neravnomernu distribuciju primene ključeva koje koristimo. Tada moramo primeniti tehnike odbrane od napada za otkrivanje ključa skeniranjem kod vertikalne autorizacije za smanjenje opasnosti od napada za detekciju ključa (ciphertext search attack). Primena *DENNINGOVE SCHEME* je dalje istraživanje za zaštitu od ove vrste napada i upoređivanje sa rezultatima vertikalne autorizacije za efikasnost zaštite od ove vrste napada.

**e)** Napredne strategije za zaštitu baze ključeva.

**f)** Istraživanje uticaja kriptografije na performanse upita i operacija u *BP* za različite klase opterećenja i kardinalnostima kolekcija podataka. Ispitivanje promenljivosti performansi zavisno od izbora algoritma ili dužine ključa.

Listi pitanja se mogu dodati pitanja specifičnosti primene u XML bazama podataka i noSQL bazama podataka.

## 9. Zaključak

Kriptografija se može koristiti za skrivanje podataka u tabelama na nivou ćelija transformacijom u kriptovane vrednosti. Sama enkripcija na uobičajen način korištena za kriptovanje vrednosti nad atributom ne daje mehanizam vertikalne autorizacije. Tek zajedno sa uvođenjem obeležja ključeva i relacija poretka nad njima koje će se

primenjivati na podatke i mehanizmom upravljanja obeležjima ključeva možemo primeniti kriptografiju za autorizaciju dostupnosti podacima u ćelijama. Potreban je mehanizam upravljanja bazom ključeva preko njihovih obeležja i relacije poretka nad njima. Za skrivanje ključeva od neovlaštene upotrebe koristi se zaštićeno skladište ključeva. Za dekripciju vrednosti koristimo mehanizam aktiviranja ključeva dostupnih korisniku na početku sesije.

Simetrična enkripcija sa *AES* algoritmom trenutno zadovoljava zahteve za kriptovanjem svih sadržaja u *BP*. Moguće je pri definiciji i realizaciji funkcija $E_i = E(A_i, KLA_i)$ i $D(E_i)$ koristiti bilo koji algoritam koji imamo na raspolaganju u *SUBP*.

## Bibliografija


[1] **Chien-Yuan Chena, Cheng-Yuan Kub, D.C. Yenc.** Cryptographic relational algebra for databases using the field Authenticator. *Computers and Mathematics with Applications 54*, 2007, 54, 38-44.

[2] **M. Vidić.** Autorizacija pristupa pojedinačnim podacima u bazi podataka - definicije i rešenja. *Konferencija o bezbednosti informacija BISEC 2014*, 2014.

[3] **O.S. Faragallah, E.M. El-Rabaie, F.E. Abd El-Samie, A.I. Sallam, H.S. El-Sayed.** Multilevel Security for Relational Databases. *CRC Press by Auerbach Publications ISBN 9781482205398 - CAT#K21447*, 2014.

[4] **E. Shmuelia, R. Vaisenbergb, E. Gudesc, Y. Elovicid.** Implementing a database encryption solution, design and implementation issues. *Computers & Security, Volume 44*, 2014, 44.

[5] **D. Cherry.** Securing SQL Server. *DOI:10.1016/B978-0-12-801275-8.00016-6*, 2015.

[6] **E. Schaefer.** An introduction to cryptography and cryptanalysis. *Santa Clara University*.

[7] **M. Živković.** Kriptografija. *Matematički fakultet Beograd - Preprint*, 2012.

[8] **D. Solomon.** Elements of Computer Security. *Springer*, 2011.

[9] **B. Schneier.** Applied Cryptography: Protocols, Algorithms and Source Code in C. *Wiley*, 2015.

[10] **H. Delfs, H. Knebl.** Introduction to Cryptography - Principles and Applications. *Springer*, 2015.

[11] **J. Katz, Y. Lindell.** Introduction to Modern Cryptography. *CRC Press*, 2015.

[12] **G. Pavlović-Lažetić.** Osnove relacionih baza podataka. *Matematički fakultet Beograd*, 1999.

[13] **D.E. Newton.** Encyclopedia of Cryptology.

[14] **A.J. Menezes, P.C. van Oorschot, S.A. Vanstone.** Handbook of Applied Cryptography. *CRC Press*, 2011.

[15] https://www.khanacademy.org/computing/computer-science/cryptography. *Journey into cryptography*, 2016.

Simpozijum MATEMATIKA I PRIMENE, Matematički fakultet, Univerzitet u Beogradu, 2015 ,Vol. VI(1)

# Kompleks presjeka ideala

**Nela Milošević**

*Fakultet za informacione sisteme i tehnologije*
*Univerzitet Donja Gorica*
*81000 Podgorica, Crna Gora*
*e-mail: nela.milosevic@udg.edu.me*

**Apstrakt.** U ovom radu izučava se topologija simplicijalnog kompleksa koji je pridružen komutativnom prstenu sa jedinicom na sledeći način: tjemena u kompleksu su svi pravi ne-nula ideali u prstenu, a skup ideala čini simpleks ako i samo ako njihov presjek nije trivijalan. Homotopski tip ovog kompleksa određuje se za generalni slučaj komutativnih prstena sa jedinicom, odnosno pokazuje se da je kompleks presjeka ideala kontraktibilan za svaki prsten osim kada je prsten izomorfan proizvodu konačno mnogo polja - u tom slučaju kompleks je homotopno ekvivalentan sferi odgovarajuće dimenzije.

**Ključne reči:** simplicijalni kompleksi; homotopski tip; graf presjeka ideala; komutativni prsteni;

## 1. Uvod

Analiziranje algebarskih svojstava pomoću kombinatornih objekata posebnu popularnost steklo je poslednjih godina kada su se pojavili mnogi rezultati u radovima koji povezuju komutativnu algebru i teoriju grafova. Inspirisani ovim pristupom, dolazimo do ideje pridruživanja simplicijalnih kompleksa komutativnim prstenima sa jedinicom. Naime, ako posmatramo grafove kao jednodimenzionalne simplicijalne komplekse, prirodno je uopštiti ovakvu analizu na generalne simplicijalne komplekse. Korist pridruživanja simplicijalnih kompleksa komutativnim prstenima je to što možemo računati homologiju i pokušati da odredimo homotopski tip.

Za neke od grafova koji su pridruženi prstenima, sam graf ili komplement tog grafa može se prirodno uopštiti na simplicijalni kompleks. U radu [ 1] izučava se graf presjeka ideala u prstenu, gdje su tjemena grafa svi pravi ne-nula ideali u prstenu, a dva tjemena su susjedna ako i samo ako je njihov presjek netrivijalan. Ovakvu analizu uopštavamo na simplicijalni kompleks pridružen komutativnim prstenima i određujemo njegov homotopski tip za generalni slučaj.

Za dalje radove koji se bave pridruživanjem simplicijalnih kompleksa komutativnim prstenima, čitalac može pogledati [ 3, 4, 5].

## 2. Definicije

U ovoj sekciji navodimo potrebne definicije. Za temeljnija objašnjenja, čitalac može pogledati [ 2, 6]

**Definicija 1.** *Apstraktan simplicijalni kompleks* $K$ je skup $A$ zajedno sa kolekcijom $K$ konačnih nepraznih podskupova u $A$ tako da ako je $X \in K$ i $Y \subseteq X$, onda je $Y \in K$.

Element $v \in A$ takav da je $\{v\} \in K$ nazivamo *tjeme* i skup svih tjemena obilježavamo sa $V(K)$. Elementi u $K$ nazivaju se *simpleksi* i obično se obilježavaju sa $\sigma$. Dimenzija simpleksa $\sigma$ je $|\sigma| - 1$, gdje je $|\sigma|$ kardinalnost skupa $\sigma$. Neprazan podskup simpleksa nazivamo *lice* tog simpleksa; lica su takođe simpleksi. Simpleksi koji nisu lica nijednog drugog simpleksa u $K$ zovu se *maksimalni simpleksi*. Za simpleks $\tau$ koji je sadržan u samo jednom maksimalnom simpleksu $\sigma$ u kompleksu, kažemo da je $\tau$ *slobodno lice* simpleksa $\sigma$.

**Definicija 2.** Neka su $K$ i $L$ dva apstraktna simplicijalna kompleksa. *Simplicijalno preslikavanje* od $K$ do $L$ je preslikavanje dato na tjemenima $f\colon V(K) \to V(L)$ takvo da ako je $\{x_0, \ldots, x_n\}$ simpleks u $K$, onda je $\{f(x_0), \ldots, f(x_n)\}$ simpleks u $L$. Za takvo preslikavanje pišemo $f\colon K \to L$.

**Definicija 3.** *Geometrijski* $n$*-simpleks* je konveksna ljuska $n+1$ afino nezavisnih tačaka, $\sigma = \text{conv}\{x_0, x_1, \ldots, x_n\}$. Njegova *dimenzija* je $\dim \sigma = n$. *Standardan* geometrijski $n$-simpleks $\Delta^n$ je određen na sledeći način:

$$\Delta^n := \{(x_0, x_1, \ldots, x_n) \in \mathbb{R}_+^{n+1} : x_0 + x_1 + \cdots + x_n = 1\},$$

gdje je $\mathbb{R}_+$ skup svih ne-negativnih realnih brojeva.

Postoji afina bijekcija izmedju bilo kojeg geometrijskog $n$-simpleksa i standardnog geometrijskog $n$-simpleksa.

Sa $\mathbb{R}^{\oplus J}$ obilježimo direktnu sumu $|J|$ (gdje $J$ može biti beskonačno; $|J|$ je kardinalnost skupa $J$) kopija $\mathbb{R}$ (pa je prema tome podskup u $\mathbb{R}^J$ kojeg čine sve tačke $x = (x_j)_{j\in J} \in \mathbb{R}^J$ takve da je $x_j = 0$, za sve osim konačno mnogo $j \in J$).

(Geometrijski) simplicijalni kompleks $K$ u $\mathbb{R}^{\oplus J}$ je kolekcija simpleksa u $\mathbb{R}^{\oplus J}$ koja zadovoljava sledeća dva uslova:

(1) Svako lice simpleksa u $K$ je simpleks u $K$.

(2) Neprazan presjek dva simpleksa u $K$ je lice svakog od ta dva simpleksa.

*Dimenzija* kompleksa $K$ je maksimalna dimenzija njegovih simpleksa. Sa $|K|$ obilježavamo uniju svih simpleksa u $K$. Ovom skupu data je topologija na sledeći način: skup $F \subset |K|$ je zatvoren ako i samo ako je $F \cap \sigma$ zatvoren u $\sigma$ za svaki simpleks $\sigma \in K$ (sam simpleks $\sigma$ naslijeđuje topologiju koju indukuje $n$-dimenzionalna ravan koju određuju njegova tjemena). Topološki prostor $|K|$ naziva se *geometrijska realizacija* i određen je do na homeomorfizam.

Simplicijalno preslikavanje $f\colon K \to L$ može se produžiti na neprekidno preslikavanje $|f|\colon |K| \to |L|$ topoloških prostora. Kada je iz konteksta jasno da se radi o preslikavanju topoloških prostora, to preslikavanje ćemo označiti sa $f$ umjesto $|f|$.

**Definicija 4.** Simplicijalni kompleks $K$ sa skupom tjemena $V(K)$ je *konus* sa vrhom $v \in V(K)$ ako za svaki simpleks $\sigma \in K$ takođe imamo da je $\sigma \cup \{v\} \in K$.

Ako je simplicijalni kompleks $K$ konus sa vrhom $v$ onda je njegova geometrijska realizacija kontraktibilna.

Biće nam potrebna i sledeća lema iz [ 6].

**Lema 1** (Lemma 2.5, [ 6])**.** *Ako je $K$ konačan simplicijalni kompleks, onda je topološki prostor $|K|$ kompaktan. Obrnuto, ako je podskup $A$ od $|K|$ kompaktan, onda je $A \subset |K_0|$ za neki konačan potkompleks $K_0$ u $K$.*

Treba obratiti pažnju da je ova lema u [ 6] formulisana za simplicijalne komplekse koji se nalaze u $\mathbb{R}^N$ za neko $N$, što ograničava kardinalnost kompleksa $K$ i dimenziju njegovih simpleksa. Međutim, u sledećem poglavlju u [ 6], autor uklanja ova ograničenja, tako da ovaj rezultat važi u generalnom slučaju što ćemo i koristiti.

## 3. Kompleks presjeka ideala

Neka je $R$ komutativan prsten sa jedinicom, i neka je $I^*(R)$ skup svih pravih ne-nula ideala u $R$. Kompleks presjeka ideala $\mathcal{K}(R)$ sa skupom tjemena $I^*(R)$ definiše se na sledeći način:

$$\{I_0, I_1, \ldots, I_n\} \in \mathcal{K}(R) \text{ ako i samo ako } I_0 \cap I_1 \cap \cdots \cap I_n \neq 0$$

Ovo prirodno čini simplicijalni kompleks jer ako presjek nekog broja ideala nije trivijalan onda ni presjek nekog manjeg broja tih ideala nije trivijalan. Primjer na slici 1 ilustruje takav kompleks za prsten $R = \mathbb{Z}_{30}$.

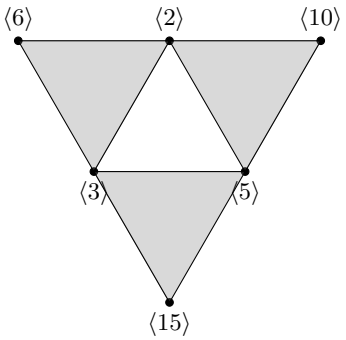
Homotopski tip ovog kompleksa određujemo tako što posmatramo slučajeve kada je prsten $R$ $(1)$ lokalan, $(2)$ semilokalan, i $(3)$ kada ima beskonačno mnogo maksimalnih ideala.

**Tvrđenje 1.** *Ako je prsten $R$ lokalan sa maksimalnim idealom $M$, onda je $|\mathcal{K}(R)|$ kontraktibilan.*

*Dokaz.* Svaki pravi ideal u prstenu sadržan je u maksimalnom idealu $M$, pa stoga $M$ netrivijalno siječe svaki pravi ideal u prstenu. Prema tome kompleks presjeka ideala $\mathcal{K}(R)$ je konus sa vrhom $M$ pa možemo zaključiti da je $|\mathcal{K}(R)|$ kontraktibilan. □

Sada pretpostavimo da je prsten $R$ semilokalan. Prvo ćemo pokazati povezanost kompleksa.

**Tvrđenje 2.** *Neka je $R$ semilokalan prsten sa $n > 1$ maksimalnih idela, tako da $R$ nije izomorfan proizvodu dva polja. Tada je kompleks $|\mathcal{K}(R)|$ povezan.*




**Slika 1.** Kompleks presjeka ideala za $R = \mathbb{Z}_{30}$

*Dokaz.* Presjek bilo koja dva maksimalna ideala u prstenu nije trivijalan. U suprotnom prsten bi bio izomorfan proizvodu dva polja što je kontradiktorno pretpostavci. Posmatramo graf koji je jednodimenzionalni skelet kompleksa. Za netrivijalne prave ideale $I$ i $J$ u prstenu, putanja na grafu koja pokazuje povezanost je $I, M_I, M_I \cap M_J, M_J, J$, gdje je $M_I$ bilo koji maksimalan ideal koji sadrži $I$, odnosno $M_J$ bilo koji maksimalan ideal koji sadrži $J$. □

Kada je prsten $R$ izomorfan proizvodu dva polja, $R \cong F_1 \times F_2$, geometrijska realizacija kompleksa sastoji se od dvije nepovezane tačke, $\{0 \times F_2\}$ i $\{F_1 \times 0\}$.

Kako bi odredili homotopski tip kompleksa za semilokalne prstene, prvo ćemo posmatrati prstene koji imaju netrivijalan Džekobsonov radikal.

**Tvrđenje 3.** *Neka je $R$ semilokalan prsten. Ako $J(R) \neq \{0\}$, onda je $|\mathcal{K}(R)|$ kontraktibilan.*

*Dokaz.* Pokazaćemo da postoji maksimalan ideal $M$ u prstenu $R$ takav da $M \cap I \neq \{0\}$ za svaki ne-nula ideal $I$ u prstenu. Time ćemo pokazati da za svaki simpleks $\{I_0, I_1, \ldots, I_n\} \in \mathcal{K}(R)$ važi da je $\{I_0, I_1, \ldots, I_n\} \cup \{M\} \in \mathcal{K}(R)$, odnosno time je $\mathcal{K}(R)$ konus sa vrhom $M$, pa je $|\mathcal{K}(R)|$ kontraktibilan.

Neka je $x$ bilo koji ne-nula element u $J(R)$. Tvrdimo da je ideal $\langle x \rangle + \mathrm{Ann}(x)$ pravi ne-nula ideal. Naime, ako je $\langle x \rangle + \mathrm{Ann}(x) = R$, onda je $1 = rx + a$, za neki element $r \in R$ i $a \in \mathrm{Ann}(x)$. S obzirom da je $x \in J(R)$, po dobro poznatom svojstvu Džekobsonovog radikala, $a = 1 - rx \in U(R)$. Kako je $ax = 0$ i $a$ invertibilan, onda slijedi da je $x = 0$, što dovodi do kontradikcije.

Svaki ne-nula pravi ideal je sadržan u nekom maksimalnom idealu u prstenu, pa neka je $M$ maksimalan ideal koji sadrži $\langle x \rangle + \mathrm{Ann}(x)$. Dalje, neka je $I$ bilo koji ne-nula ideal, i neka je $t$ bilo koji ne-nula element u idealu $I$. Razmatramo element $tx$.

1) Ako je $tx = 0$, onda je $t \in \mathrm{Ann}(x)$, pa je $t \in I \cap M$. Prema tome imamo i $I \cap M \neq \{0\}$.

2) Ako $tx \neq 0$, onda je $tx$ ne-nula element u $I \cap M$, pa imamo $I \cap M \neq \{0\}$. □

**Tvrđenje 4.** *Neka je $R$ semilokalan prsten. Ako je $|\mathrm{Max}(R)| = n > 1$ i $J(R) = \{0\}$, onda je $|\mathcal{K}(R)| \simeq \dot{\Delta}^{n-1}$.*

*Dokaz.* S obzirom da $J(R) = \{0\}$, Kineska teorema o ostacima nam govori da je $R$ izomorfan direktnom proizvodu konačno mnogo polja, $R \cong F_1 \times \cdots \times F_n$. Prema tome, svaki ideal u prstenu je presjek nekog broja maksimalnih ideala, odnosno $M_S = \bigcap_{i \in S} M_i$ gdje je $S \subset [m] = \{1, 2, \ldots, m\}$. Primjetimo da imamo $n$ maksimalnih simpleksa u kompleksu, gdje takav simpleks čine tjemena $M_T$, $T \subseteq S$ gdje je $S = [m] \setminus \{j\}$ za neko $j \in [m]$.

Neka je $K$ potkompleks u $\mathcal{K}(R)$ kojeg čine tjemena $M_1, \ldots, M_n$. Geometrijska realizacija ovog kompleksa je granica $n$-simpleksa, s obzirom da presjek bilo kojeg broja maksimalnih ideala (osim presjeka svih) nije trivijalan. Definišemo neprekidno preslikavanje $f\colon |\mathcal{K}(R)| \to |K|$ tako što svako tjeme $M_S$ slikamo u baricentar simpleksa čija su tjemena $M_i$ za svako $i \in S$. Na ovaj način projektujemo svaki maksimalan simpleks na njegovo odgovarajuće lice, i svaka tačka unutar tog maksimalnog simpleksa projektuje se duž linije unutar simpleksa. Prema tome preslikavanje $f$ je jak deformacioni retrakt pa je stoga $|\mathcal{K}(R)| \simeq |K|$, odnosno, $|\mathcal{K}(R)| \simeq \dot{\Delta}^{n-1}$.

□

Da bismo odredili homotopski tip kompleksa presjeka ideala za prstene sa beskonačno mnogo maksimalnih ideala, prvo ćemo dokazati sledeću lemu.

**Lema 2.** *Neka je prsten $R$ takav da je skup maksimalnih ideala $\mathrm{Max}(R)$ beskonačan. Ako je $K_0$ konačan potkompleks u $\mathcal{K}(R)$, onda postoji potkompleks $K_1$ takav da je $K_0$ potkompleks u $K_1$ i $|K_1|$ je kontraktibilan.*

*Dokaz.* Neka je $K_0$ bilo koji konačan potkompleks u $\mathcal{K}(R)$ , i neka je $\{I_1, \ldots, I_m\}$ skup svih tjemena u $K_0$. Pokazaćemo da postoji maksimalan ideal $M$ u prstenu $R$ takav da $I_k \cap M \neq \{0\}$ za svako $1 \leqslant k \leqslant m$. Time ćemo i ujedno pokazati da je $|\mathcal{K}(R)|$ povezan.

Za svaki indeks $1 \leqslant k \leqslant m$, neka je $M_k$ maksimalan ideal koji sadrži $\mathrm{Ann}(I_k)$ i neka je $M$ maksimalan ideal takav da $M \neq M_k$ za svako $1 \leqslant k \leqslant m$. Tvrdimo da $M \cap I_k \neq \{0\}$.

Po teoremi o izbjegavanju prostih ideala imamo da $M \not\subseteq M_1 \cup \cdots \cup M_m$ pa možemo odabrati element $x \in M \setminus (M_1 \cup \cdots \cup M_m)$. Neka je $k \in \{1, \ldots, m\}$. S obzirom da $x \notin \mathrm{Ann}(I_k)$, postoji element $t_k \in I_k$ takav da $xt_k \neq 0$. Prema tome, element $xt_k$ je ne-nula element u $M \cap I_k$. Dakle, maksimalan ideal $M$ netrivijalno siječe svaki ideal u $K_0$, odnosno $M \cap I_k \neq \{0\}$ za svako $1 \leqslant k \leqslant m$. Neka je $K_1$ potkompleks u $\mathcal{K}(R)$ kojeg čine tjemena $\{M, I_1, \ldots, I_m\}$ i odgovarajući simpleksi. Onda je $K_1$ konus sa vrhom $M$ pa je stoga kontraktibilan. □

Nakon dokaza ove leme možemo odrediti homotopski tip kompleksa presjeka ideala $|\mathcal{K}(R)|$ za prsten $R$ sa beskonačno mnogo maksimalnih ideala.

**Tvrđenje 5.** *Ako je skup maksimalnih ideala $\mathrm{Max}(R)$ beskonačan, onda je $|\mathcal{K}(R)|$ kontraktibilan.*

*Dokaz.* S obzirom da $|\mathcal{K}(R)|$ ima homotopski tip CW-kompleksa, možemo koristiti teoremu Dž.H.K. Vajtheda koja nam govori da je preslikavanje između CW-kompleksa koje indukuje izomorfizme na svim homotopskim grupama homotopska ekvivalencija, odnosno da ako su sve homotopske grupe trivijalne da je kompleks kontraktibilan. Pretpostavimo da je $n \geq 1$ i da je $g\colon S^n \to |\mathcal{K}(R)|$ neprekidno preslikavanje. Slika $g[S^n]$ je kompaktna, pa prema lemi 1, postoji konačan potkompleks $K_0$ takav da je $g[S^n] \subseteq |K_0|$. Lema 2 nam govori da postoji potkompleks $K_1$ takav da je $K_0 \subset K_1$ i $|K_1|$ kontraktibilan. Stoga, preslikavanje $g$ možemo faktorisati kroz kontraktibilan prostor $|K_1|$ pa je homotopski trivijalno. Zaključujemo da je $\pi_n(|\mathcal{K}(R)|, *)$ trivijalna. Ovo važi za svako $n$, pa prema teoremi Vajtheda zaključujemo da je $|\mathcal{K}(R)|$ kontraktibilan. □

## Bibliografija


[1] **I. Chakrabarty, S. Ghosh, T. K. Mukherjee, M. K. Sen.** Intersection graphs of ideals of rings. *Discrete Math.*, 309 (2009), no. 17, 5381 - 5392.

[2] **D. Kozlov.** Combinatorial Algebraic Topology. *Algorithms and Computation in Mathematics, 21, Springer, Berlin*, 2008.

[3] **N. Milošević.** Independence complexes of comaximal graphs of commutative rings with identity. *Publications de l'Institut Mathématique*, 98(112) (2015) 109 -117

[4] **N. Milošević, Z. Z. Petrović.** Ideal zero-divisor complex. *J. Commut. Algebra*, to appear.

[5] **N. Milošević, Z. Z. Petrović.** Order complex of ideals in a commutative ring with identity. *Czechoslovak Mathematical Journal*, 2015, Vol. 65, No. 4, pp 947 - 952.

[6] **J. R. Munkres.** Elements of Algebraic Topology. *Addison-Wesley Publishing Company, Menlo Park, California*, 1984.

Simpozijum MATEMATIKA I PRIMENE, Matematički fakultet, Univerzitet u Beogradu, 2015 ,Vol. VI(1)

# Metaheuristička metoda optimizacije kolonijom pčela: teorijske osnove i primene

**Tatjana Davidović**

*Matematički institut SANU, Knez-Mihailova 36/III, Beograd*
*e-mail: tanjad@mi.sanu.ac.rs*

**Sažetak** Algoritam optimizacije kolonijom pčela (Bee Colony Optimization, BCO) prvi put su predložili 2001. godine P. Lučić i D. Teodorović. BCO i njegove brojne varijante spadaju u klasu metaheurističkih metoda inspirisanih prirodom, preciznije ponašanjem pčela u potrazi za hranom. To je lako razumljiva metoda, jednostavna za implementaciju i do sada je uspešno primenjena na mnoge probleme optimizacije. U ovom preglednom radu opisan je najpre prirodni proces prikupljanja nektara i jezik kojim pčele komuniciraju, zatim je prikazan osnovni BCO algoritam i njegove modifikacije, uključujući strategije paralelizacije i hibridizacije. U drugom delu rada, navedene su primene metode na različite teške probleme kombinatorne optimizacije, uglavnom u oblasti transporta, lokacije i raspoređivanja, kao i neke skorije primene u kontinualnoj optimizaciji. Navedene su i doktorske disertacije koje se bave razvojem i primenama ove metode, a odbranjene su na univerzitetima u Srbiji. Osnovni cilj rada je da se BCO metoda promoviše među domaćim istraživačima, jer bez obzira na jednostavnost i efikasnost, do sada nije dovoljno promovisana. Samim tim, primat u primenama drže neke novije i složenije metode koje su imale uspešniji marketing. Želja autora je da tu situaciju preokrene u korist BCO metode i podstakne istraživače da je primenjuju na probleme kombinatorne i globalne optimizacije.

**Keywords:** Optimizacioni problemi, metaheurističke metode, algoritmi inspirisani prirodom, inteligencija roja.

## 1. Uvod

Prirodni procesi inspirišu razvoj optimizacionih algoritama već više od četrdeset godina. Počevši od simuliranog kaljenja (Simulated Annealing, SA) [ 1, 2, 3], genetskog i drugih evolutivnih algoritama (Genetic Algorithms, GA, Evolutionary Algorithms, EA) [ 4, 5, 6, 7, 8, 9], preko optimizacije mravljim kolonijama (Ant Colony Optimization, ACO) [ 10, 11, 12], rojem čestica (Particle Swarm Optimization, PSO) [ 13, 14] i drugih algoritama inteligencije roja (Swarm Intelligence, SI) [ 15, 16, 17], sve do metode zasnovane na veštačkim imunim sistemima (Aritificial Immune System, AIS) [ 18, 19] i mnogih drugih. Trenutno postoje na hiljade različitih metoda inspirisanih prirodnim procesima koje se klasifikuju kao tehnike računarske inteligencije (Computational Intelligence Techniques) [ 20, 21, 22, 23, 24].

Metoda optimizacije kolonijom pčela (BCO) je metaheuristika inspirisana ponašanjem pčela u potrazi za hranom. To je jedan od prvih algoritama koji koristi osnovne principe kolektivne inteligencije pčela u rešavanju problema kombinatorne optimizacije. BCO je prvobitno predložen za primenu na poznate teške probleme kombinatorne optimizacije kao što su problem trgovačkog putnika [ 25, 26, 27] i rutiranje vozila [ 28]. To je stohastička tehnika (tehnika slučajne pretrage) koja radi nad populacijom rešenja. Inspiracija za razvoj algoritma dobijena je uspostavljanjem analogije između ponašanja pčela u potrazi za hranom i ponašanja optimizacionog algoritma tokom pretraživanja prostora rešenja datog kombinatornog problema [ 29]. Osnovna ideja metode je formiranje sistema multi-agenta (kolonije veštačkih pčela) koji bi se efikasno primenjivao na teške probleme optimizacije. Veštačke pčele pretražuju prostor rešenja u potrazi za dopustivim rešenjima datog problema. U cilju popravljanja kvaliteta generisanih rešenja, autonomne veštačke pčele sarađuju i razmenjuju informacije. Deljenjem dostupnih informacija i korišćenjem kolektivnog znanja, veštačke pčele koncentrišu pretragu na oblasti koje potencijalno sadrže bolja rešenja, dok istovremeno napuštaju rešenja slabijeg kvaliteta. Ovakvim potupkom veštačke pčele kolektivno generišu sve bolja i bolja rešenja.

BCO algoritam radi u iteracijama sve do zadovoljenja nekog unapred definisanog kriterijuma zaustavljanja. Najčešće korišćeni kriterijumi zaustavljanja su maksimalno dozvoljeno vreme, maksimalni broj iteracija, maksimalni broj iteracija bez poboljšanja trenutno najboljeg rešenja, naksimalni broj izračunavanja vrednosti funkcije cilja i mnogi drugi, a ponekad se koriste i kombinacije više kriterijuma. Postoje dve osnovne varijante BCO metode, konstruktivna (koja gradi rešenja primenjujući neku vrstu stohastičkog pohlepnog izbora novih komponenti) i varijanta sa popravkom (koja polazi od nekih kompletnih rešenja i pokušava da ih poboljša primenom stohastičkih pravila zamene komponenti). Varijanta sa popravkom je u literaturi poznata kao improvement BCO (BCOi) [ 30, 31, 32].

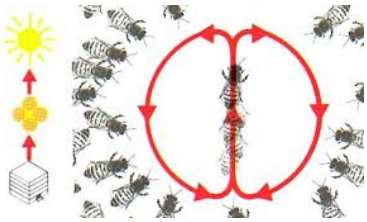
Ovaj rad je nastavak nedavno objavljenih preglednih radova [ 29, 33, 34]. Sastoji se iz opisa ponašanja pčela u prirodi, prikaza BCO algoritma, njegovih varijacija i modifikacija, kao i klasifikacije i analize njegovih skorijih primena. U novijoj literaturi BCO se uspešno koristi za modeliranje raznih složenih optimizacionih problema. Međutim, BCO metoda i dalje nije rasprostranjena među istraživačima, pa je osnovni cilj ovog rada njena popularizacija u krugovima istraživača koji se bave optimizacijom. Metoda je teorijski verifikovana i utvrđeni su uslovi koji moraju da se zadovolje prilikom implementacije da bi se obezbedila konvergencija generisanih rešenja ka željenom optimumu. Dakle, postoje i teorijski preduslovi da se za svaki konkretan razmatrani problem može razviti efikasna implementacija BCO metode što je od posebnog značaja istraživačma koji rade u praksi.

Rad je organizovan na sledeći način. Prirodni proces prikupljanja nektara i jezik kojim pčele komuniciraju opisani su u odeljku 2. Odeljak 3 sadrži kratak prikaz osnovne varijante BCO algoritma i sprovedenu teorijsku analizu metode. Različite modifikacije opisane su u odeljku 4. Odeljak 5 posvećen je primenama BCO metode, dok poslednji odeljak sadrži neke zaključke i smernice za dalja istraživanja u vezi sa BCO metaheuristikom.

## 2. Pčele u prorodi

Prilikom razvoja algoritama zasnovanih na SI principima, istraživači koriste modele ponašanja roja u prirodi. Veštački sistemi obično ne uključuju potpunu imitaciju prirodnih procesa, već preuzimaju neke karakteristične postupke i prilagođavaju ih konkretnoj implementaciji. Proces kojim pčele tragaju za nektarom [ 60] inspirisao je nekoliko grupa autora da razviju optimizacione algoritme koji imitiraju to ponašanje [ 25, 61, 62, 63]. Različiti autori interpretirali su na razne načine organizaciju populacije pčela i njihov ples u košnici koji predstavlja veoma razvijeni jezik za komunikaciju među pčelama. Više detalja o tome dato je u nastavku ovog odeljka.

Tokom potrage za hranom, pčele istražuju polja u blizini košnice. One skupljaju i skladište nektar koji će kasnije koristiti za ishranu. Uobičajeno je da u potragu za hranom ne idu sve pčele, već samo neke od njih, tzv. *pčele izviđači*. Po povratku u košnicu, oni izviđači koji su pronašli kvalitetne izvore hrane (cvetne livade bogate nektarom), obaveštavaju o tome ostale pčele izvodeći takozvani *ritualni ples* (engl. waggle dance) (Sl. 1). Tim plesom kodirani su podaci o lokaciji (pravcu i daljini) i količini pronađenog nektara u oblastima koje su ispitale pčele izviđači. Ostale pčele posmatraju taj ples i na osnovu dobijenih uputstava i same kreću u skupljanje nektara. Takve pčele nazivaju se *neopredeljene* ili *sledbenici*.

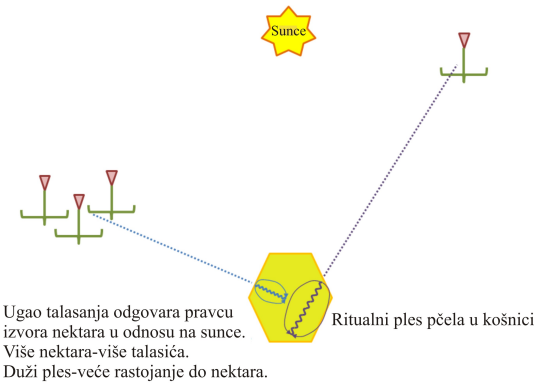


**Slika 1.** Ilustracija plesa kojim pčele kominiciraju (preuzeto sa www.pcelarski-inkubator.vup.hr)

Ritualni ples sastoji se od pokreta kojim pčele formiraju cifru osmice sa talasićima u sredini. Značenje ovog plesa otkrio je Karl von Frisch [ 64] (Sl. 2). Smer (ugao) središnjeg dela osmice u odnosu na položaj sunca usko je povezan sa pravcem izvora hrane reklamiranog od strane pčele izviđača. Udaljenost između košnice i izvora hrane kodiran je brojem talasića u ritualnom plesu. Brzina talasanja označava količinu otkrivenog nektara, a kvalitet nektara pčele proveravaju probanjem uzoraka koje donose izviđači. Ritualni ples je uvek usklađen sa položajem sunca, ugao talasanja se menja kako sunce putuje preko neba tokom dana. To znači da su sledbenici uvek ispravno usmereni ka izvoru hrane.

Kako je moguće da se desi da nekoliko pčela istovremeno plešu i pokušavaju da regrutuju sledbenike, nije jasno kako se neopredeljene pčele odlučuju kog će izviđača da prate. Jedino što su naučnici do sada uspeli da utvrde je da ta odluka direktno zavisi od količine i kvaliteta hrane na izvoru [ 60]. Opisani proces ponavlja se sve dok pčele u košnici akumuliraju nektar i/ili istražuju nove oblasti sa potencijalnim izvorima hrane.

Ako neopredeljena pčela odluči da napusti košnicu i sakuplja nektar, ona će slediti jednog od izviđača koji je prethodno otkrio izvor hrane. Po povratku u košnicu, pčela nosi sakupljeni nektar i odlaže ga u skladište hrane. U tom trenutku pčela bila između sledećih opcija: (1) odustaje od izvora hrane i vraća se ulozi neopredeljenog sledbenika; (2) nastavlja sa eksploatacijom izvora nektara bez regrutovanja ostatka kolonije; ili (3) pokušava




**Slika 2.** Prevod jezika pčela (adaptirano sa https://factismals.com/tag/etymology/)

regrutaciju ostalih neopredeljenih pčela izvodeći ritualni ples pre povratka na izabranu lokaciju hrane. Pčela se opredeljuje za jednu od navedenih opcija u zavisnosti od doba dana i količine preostalog nektara na toj lokaciji.

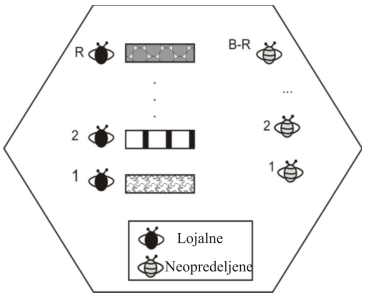
Ovde je opisan dsmo deo pčelinjeg jezika koji je relevantan za razvoj i primene BCO metode. Komunikacija među pčelama daleko je složenija, a samoorganizacija i odgovornost svake jedinke u košnici potpuno su fascinantni. Detaljnije o životu pčela može se naći u publikacijama iz melitologije, ali i na Internetu (https://animalwise.org/2011/08/25/the-honeybee-waggle-dance-%E2%80%93-is-it-a-language/)

## 3. Opis BCO algoritma

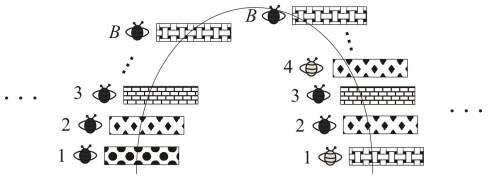
Populacija veštačkih pčela sastoji se od $B$ agenata koji zajedno tragaju za rešenjem razmatranog optimizacionog problema. Svaka veštačka pčela generiše po jedno rešenje za dati problem [ 40]. Rad BCO algoritma odvija se kroz iteracije, a svaka iteracija se sastoji od $NC$ koraka. Korak BCO algoritma deli se na dve faze: let unapred (engl. forward pass) i let unazad (engl. backward pass). Dakle, u okviru jedne iteracije ove dve faze se smenjuju $NC$ puta, tj. dok se ne izgenerišu kompletna rešenja, u konstruktivnoj verziji, ili dok se ne izvrši $NC$ modifikacija početnih rešenja, u verziji sa popravkom. Tokom leta unapred sve veštačke pčele su uključene u istraživanje prostora rešenja. Po analogiji sa prirodnim procesima, kvalitet (parcijalnog) rešenja može se poistovetiti sa količinom sakupljenog nektara i/ili sa udaljenošću košnice od izvora hrane, a proces sakupljanja nektara u prirodi odgovara fazi leta unapred u BCO algoritmu. Ova faza algoritma direktno zavisi od problema koji se rešava i mora se prilikom implementacije maksimalno prilagoditi tom problemu.

U fazi leta unazad veštačke pčele razmenjuju informacije o kvalitetu pronađenih rešenja. Ritualni ples je u algoritmu pretraživanja zamenjen procenjivanjem (evaluiranjem) kvaliteta svakog generisanog (parcijalnog) rešenja. Kvalitet rešenja iskazuje se realnim brojem koji pripada intervalu [0,1]. Najkvalitetnije rešenje među svim pčelama dobija oznaku kvaliteta 1, a najlošije 0 (ili neku vrednost blisku nuli). Kada se sva rešenja evaluiraju, svaka pčela treba da odredi da li će ostati *lojalna* svom (parcijalnom) rešenju. Ta odluka donosi se na osnovu verovatnoće, koja zavisi od odnosa kvaliteta trenutnog rešenja date pčele i kvaliteta rešenja „najbolje" pčele. Najbolja pčela je ona koja trenutno poseduje najkvalitetnije (parcijalno) rešenje. Ona je uvek lojalna tom svom rešenju, jer je njena verovatnoća lojalnosti uvek jednaka jedinici. Pčela koja napusti (odbaci) svoje (parcijalno) rešenje postaje *nelojalna* i mora da preuzme neko od rešenja lojalnih pčela. Pčele sa boljim rešenjem imaju više šansi da ga zadrže i reklamiraju ostalima. Za razliku od pčela u prirodi, koje ne moraju da izvode ritualni ples i reklamiraju svoje izvore hrane drugima, veštačke pčele koje ostanu lojalne svom (parcijalnom) rešenju u isto vreme postaju i *regruteri*, odnosno, među njima će nelojalne pčele birati zamenu za rešenja koja su odbacile. U tom momentu izdvajaju se dve vrste pčela (slika 3): $R$ regrutera i preostalih $B - R$ nelojalnih (neopredeljenih) pčela [ 40]. Odluku koje od reklamiranih rešenja će preuzeti, neopredeljena pčela donosi sa verovatnoćom koja je proporcionalna kvalitetu odgovarajućeg reklamiranog rešenja i ta odluka se realizuje pomoću ruleta.

Proces regrutacije ilustrovan je na slici 4. Pčela 1 odlučila je da napusti svoje (parcijalno) rešenje i da se pridruži pčeli $B$. To znači da će pčela 1 kopirati (parcijalno) rešenje pčele $B$. Nakon toga, obe pčele donose samostalnu odluku o tome na koji način će izvesti naredni deo konstrukcije/modifikacije. Na slici 4 ilustrovana je situacija gde pčela 4 kopira (parcijalno) rešenje pčele 2, dok su pčele 2, 3 i $B$ ostale lojalne svojim prethodnim (parcijalnim) rešenjima.




**Slika 3.** Ilustracija procesa utvrđivanja lojalnosti (adaptirano iz [ 30]).




**Slika 4.** Ilustracija procesa regrutacije (adaptirano iz [ 30]).

Kao što je već pomenuto, dve faze algoritma pretraživanja (faza leta unapred i leta unazad) smenjuju se $NC$ puta. Parameter $NC$ služi za definisanje učestalosti razmene informacija među pčelama. Po završetku $NC$ koraka, određuje se najbolje od svih $B$ ponuđenih rešenja. Ono se potom koristi za ažuriranje najboljeg globalnog rešenja, čime je jedna iteracija BCO algoritma završena. U ovom trenutku dobijena rešenja za svaku pčelu mogu se izbrisati kako bi u sledećoj iteraciji pčele gradile/modifikovale nova rešenja. Iteracije BCO algoritma izvršavaju se sve dok se ne zadovolji definisani kriterijum zaustavljanja. Po zadovoljenju kriterijuma zaustavljanja, ispisuje se najbolje pronađeno rešenje (tzv. trenutno globalno najbolje rešenje).

### 3.1. Pseudo kod BCO algoritma

Jedna od glavnih prednosti BCO metode je veoma mali broj parametara. BCO ima dva osnovna parametra i to su broj pčela ($B$), i broj letova unapred/unazad u jednoj iteraciji ($NC$). Algoritmom 1 prikazan je pseudo-kod BCO algoritma. Na početku svake iteracije, svim pčelama se dodeljuje neko početno rešenje. Prve dve petlje po $b$ su zavisne od konkretnog problema i potrebno ih je prilagoditi prilikom svake implementacije BCO algoritma. Za preostale korake postoje formule na osnovu kojih svaka pčela donosi odgovarajuće odluke i one su opisane u nastavku ovog odeljka.

Let unazad započinje evaluacijom svih (parcijalnih) rešenja koja su generisana u prethodnom letu unapred. Kvalitet rešenja izražava se realnim brojem iz intervala $[0, 1]$ koji se dobija normalizacijom ocene vrednosti funkcije cilja, tav. fitnesa, (parcijalnog) rešenja svake pčele. Verovatnoća da će $b$-ta pčela ostati lojalna svom prethodno pronađenom parcijalnom rešenju računa se na sledeći način:

$$p_b = e^{-\frac{1-O_b}{u}}, \quad b = 1, 2, ..., B, \tag{1}$$

pri čemu je $O_b$ kvalitet rešenja $b$-te pčele, a $u$ trenutna vrednost brojača letova unapred/unazad. Ova formula predložena je za konstruktivne varijante imajući na umu dva osnovna cilja. Prvi cilj je da se obezbedi veća verovatnoća da pčela koja je generisala bolje parcijalno rešenje ostane lojalna. Kako veća vrednost $O_b$ odgovara boljem parcijalnom rešenju, to formula (1) obezbeđuje veću verovatnoću da takva pčela ostane lojalna. Drugi cilj je povećanje uticaja već uloženog truda u generisanje parcijalnih rešenja. Preciznije, na početku pretrage pčele se lakše odlučuju da napuste trenutna rešenja, kaže se da su „hrabrije" pri eksploataciji prostora rešenja. Kako više truda ulažu u generisanje rešenja, tj. kako broj letova unapred raste, pčele su više fokusirane na postojeća rešenja i teže se odlučuju da ih napuste. Ovo je izraženo faktorom $u$ u imeniocu eksponenta u formuli (1).

Kod BCOi metode, prethodno angažovanje nad trenutnim rešenjima gubi na značaju, pa je u nekim novijim radovima [ 31, 43] $u$ zanemareno prilikom određivanja lojalnosti, tj. korišćena je funkcija $p_1(O_b)$. U radu [ 65] predložene su i analizirane još dve nove formule za određivanje verovatnoće lojalnosti ($p_2(O_b)$,$p_4(O_b, u)$), dok

**Algorithm 1** Pseudokod za BCO metodu

**procedure** BCO
  INICIJALIZACIJA(Ulaz za problem, $B, NC, STOP$)
  **while** kriterijum zaustavljanja nije zadovoljen **do**
    **for** $b \leftarrow 1, B$ **do**
      $Resenje(b) \leftarrow$ IZABERIRESENJE()
    **end for**
    **for** $u \leftarrow 1, NC$ **do**
      **for** $b \leftarrow 1, B$ **do**
        PROCENIPOTEZ($Resenje(b)$)
        IZABERIPOTEZ($Resenje(b)$)
      **end for**
      **for** $b \leftarrow 1, B$ **do**
        LOJALNOST($Resenje(b)$)
      **end for**
      **for** $b \leftarrow 1, B$ **do**
        **if** $b$ nije lojalna **then**
          REGRUTACIJA($Resenje(b)$)
        **end if**
      **end for**
    **end for**
    AZURIRAJ($x_{best}, f(x_{best})$)
  **end while**
  RETURN($x_{best}, f(x_{best})$)
**end procedure**

je u [ 37] uvedeno još šest funkcija koje imaju dobre osobine pri određivanju verovatnoće lojalnosti veštačkih pčela. Dakle, u literaturi je do sada razmatrano 10 potencijalnih funkcija lojalnosti i to:

$$
\begin{array}{ll}
p_0(O_b,u) = e^{-(1-O_b)/u}, & p_5(O_b,u) = e^{-(1-O_b)\sqrt{u}/\sqrt{u+1}}, \\
p_1(O_b) = e^{-(1-O_b)}, & p_6(O_b,u) = e^{-(1-O_b)/\log u}, \\
p_2(O_b) = O_b, & p_7(O_b,u) = e^{-(1-O_b)/u\log(u+1)}, \\
p_3(O_b,n_{it}) = e^{-(1-O_b)/n_{it}}, & p_8(O_b) = e^{-2(1-O_b)}, \\
p_4(O_b,u) = e^{-(1-O_b)/\sqrt{u}}, & p_9(O_b,u) = e^{-(1-O_b)\log(u+1)/\log(u+2)}.
\end{array}
$$

Na osnovu [ 37, 65, 31], najbolji rezultati su primenom $p_1(O_b)$, $p_2(O_b)$ i $p_8(O_b)$, ali se do tih zaključaka došlo eksperimentima na konkretnim primerima. Dalje analize i nove primene možda bi dovele do drugačijih zaključaka, pa ova tema svakako zaslužuje pažnju budućih istraživača. Trenutno se u literaturi najviše koriste $p_0(O_b,u)$ i $p_1(O_b)$.

U poslednjem delu leta unazad, izbor regrutera za svaku nelojalnu pčelu vrši se na osnovu verovatnoće koja se računa po formuli:

$$pr_b = \frac{O_b}{\sum_{k=1}^{R} O_k}, \quad b = 1, 2, ..., R, \tag{2}$$

gde $O_k$ predstavlja kvalitet $k$-tog reklamiranog (parcijalnog) rešenja, a $R$ broj regrutera, odnosno broj reklamiranih rešenja. Formula (2) ustvari predstavlja verovatnoću da će rešenje regrutera $b$ biti izabrano od strane bilo koje nelojalne pčele. Pomoću formule (2) i generatora slučajnih brojeva svaka nelojalna pčela će se pridružiti jednom od regrutera. U tom momentu (parcijalno) rešenje izabranog regrutera kopira se u odgovarajuće strukture podataka nelojalne pčele (ili više njih koje su izabrale istog regrutera), što znači da će regruter i njegovi sledbenici nastaviti dalju pretragu prostora rešenja iz iste početne tačke.

### 3.2. Teorijska analiza BCO metode

Efikasnost BCO metode ilustrovana je empirijski kroz brojne uspešne primene. Osim toga, postoje neki noviji radovi koji se bave empirijskom evaluacijom [ 31] i kalibracijom parametara [ 65] BCO metode. Međutim, kvalitet

konačnog rešenja (dobijenog izvršavanjem BCO metode do kriterijuma zaustavljanja) nemoguće je utvrditi na taj način, ukoliko je optimalno rešenje nepoznato. Da li je dobijeno rešenje optimalno, a ako nije – koliko je od njega udaljeno, nije moguće utvrditi samo eksperimentalnom analizom izvršavanja BCO metode. Moguće je jedino povećati broj iteracija (ili dozvoljeno vreme za rad) u nadi da će se eventualno, dobiti bolje konačno rešenje. Ovaj zaključak ne odnosi se samo na BCO metodu, nego na sve metaheuristike.

Prema tome, teorijska analiza metaheuristika postala je veoma popularna istraživačka tema. Metodološki okvir za ispitivanje konvergencije metaheurističkih metoda postavljen je u [ 66]. Teorijski (matematički) temelji konstruktivnog BCO algoritma dati su u [ 67, 68, 69]. U [ 67] identifikovani su neophodni uslovi koji obezbeđuju da se optimalno rešenje može generisati od strane bilo koje pčele kada je broj iteracija dovoljno veliki. Uz te uslove, dokazana je *konvergencija najboljeg rešenja*, tzv. best-so-far konvergencija BCO metode. Pokazano je da, ukoliko je verovatnoća da će neka od pčela generisati bilo koje dopustivo rešenje razmatranog problema (pa i optimalno) strogo veća od nule, trenutno najbolje rešenje konvergira sa verovatnoćom jedan ka nekom od optimalnih rešenja, kada se broj iteracija povećava. Ova vrsta konvergencije je sasvim uobičajena i važi čak i za neke jednostavne stohastičke metode optimizacije, kao što je na primer, metoda slučajnog hoda (random walk). Sofisticiranija, tzv. *konvergencija po modelu*, konstruktivne BCO metode razmotrena je u [ 68, 69]. Ova vrsta konvergencije pretpostavlja učenje iz prethodnog iskustva i stoga se može razmatrati samo za varijante gde postoji neki mehanizam razmene znanja između iteracija BCO metode. U tim slučajevima, verovatnoća izbora komponenti tokom leta unapred (korak (2)(i)) nije konstantna u svim iteracijama, već se menja po zakonitostima koje su utvrđene u [ 68]. Teorijska analiza BCOi varijante razmatrana je u [ 37].

## 4. Varijante BCO algoritma

Primena BCO metode na različite teške probleme optimizacije zahteva njeno prilagođavanje karakteristikama problema koji se razmatra. Stoga, svaka implementacija predstavlja zapravo razvoj i modifikaciju originalnog BCO algoritma. Prve verzije BCO metode [ 25, 26, 27, 28] bile su konstruktivne i imale su više sličnosti sa ponašanjem pčela u prirodi nego kasnije varijante algoritma. Osnovne karakteristike ovih verzija su: (1) košnica je imala važnu ulogu i njena pozicija je bila značajna za izvršavanje metode (imala je preciziranu lokaciju (na primer, početni čvor u procesu pretrage ili prva odabrana komponenta rešenja), iako je takođe mogla da promeni svoju poziciju u procesu pretrage; (2) nisu sve pčele bile angažovane na početku procesa pretrage (postojale su, tzv. pčele izviđači (scout bees) koje bi započinjale pretragu, a u svakoj fazi nove pčele su se pridruživale procesom regrutovanja); (3) verovatnoća izbora sledeće komponente računala se na osnovu Logit modela [ 70] (dok se u novijim verzijama izbor vrši na osnovu ruleta, turnira, rangiranja ili poremećaja).

U većini ranih primena, BCO metoda bila je konstruktivna [ 39, 41, 42, 44, 45, 46, 47, 48, 51, 52, 53, 56, 57, 58, 59, 71, 72, 73, 74, 75]. Za svaku pčelu rešenje je izgrađivano od početka. Polazilo se od praznog rešenja i, korak po korak, primenom nekih stohastičkih, heurističkih pravila koja su zavisila od razmatranog problema dodavane su komponente. Slučajnost izazvana ovim stohastičkim konstrukcijama obezbeđivala je raznolikost pretrage. Odluke o lojalnosti i regrutaciji donošene su na osnovu ocene parcijalnih rešenja i procene kvaliteta potencijalnih konačnih rešenja. Očigledno je da se preciznost procene ne može uvek kontrolisati dovoljno dobro. Pored toga, nakon regrutacije, grupa pčela je imala ista parcijalna rešenja i zbog toga se raznovrsnost konačnih rešenja smanjivala. U okviru svake iteracije generisano je $B$ rešenja i najbolje od njih je korišćeno za ažuriranje trenutnog globalno najboljeg rešenja. Ponekad, globalno znanje je korišćeno za usmeravanje procesa konstrukcije ka kvalitetnijim rešenjima (potencijalno boljim od trenutno globalno najboljeg rešenja do tada generisanog od strane pčela). Svaka iteracija započinjala je sa $B$ praznih rešenja, a završavala se sa $B$ novih konačnih rešenja, među kojima je traženo novo globalno najbolje.

Najznačajnija promena originalnog BCO algoritma svakako je uvođenje transformisanja (poboljšanja) kompletnih rešenja (nasuprot konstrukcijama). Ova varijanta, nazvana BCOi, korišćena je prvi put u rešavanju problema $p$-centara [ 30], a kasnije je postala dominantna varijanta koja se koristi u literaturi [ 31, 32, 43, 55] (za detaljniji pregled, videti [ 29, 33, 34]). BCOi algoritam se može opisati na sledeći način. Na početku svake iteracije, pčelama se dodele neka kompletna rešenja polaznog problema, koja se transformišu (modifikuju, popravljaju) kroz letove unapred. Početna rešenja iteracije mogu se generisati slučajno [ 30, 43], biti birana među već postojećim rešenjima (globalno najbolje rešenje [ 76], konačna rešenja iz prethodne iteracije [ 31, 32], neka od rešenja koja se čuvaju u skupu dobrih rešenja namenjenih za dalje popravljanje [ 77], itd.). Transformacija rešenja mora biti stohastička i ne treba da sadrži vremenski zahtevno lokalno pretraživanje. Slučajnost je potrebna kako bi se osiguralo da svaka pčela izvršava različite transformacije i generiše čitav niz novih i potencijalno boljih

rešenja. Konstruktivna varijanta je još uvek dominantna, primenjena je u oko 70% radova koji se bave primenom BCO metode, ali BCOi se ipak primenjuje deseak godina manje.

Nove implementacije BCO metode uvek pretpostavljaju da su sve pčele uključene u proces pretraživanja. Međutim, za razliku od ranijih varijanti u kojoj su sve pčele obavljale isti zadatak (izgradnju ili poboljšanje), novije implementacije pokazuju tendenciju dodele pčelama različitih uloga [ 43, 55, 76, 77, 78, 79]. Stoga, možemo razlikovati varijante BCO metode sa homogenim i heterogenim pčelama. Varijante BCO metode u savremenoj literaturi razlikuju se i po načinu određivanja verovatnoće lojalnosti, o čemu je bilo reči u prethodnom odeljku.

Algoritmi zasnovani na inteligenciji roja generalno su pogodni za paralelizaciju jer koriste populaciju rešenja. Preciznije, oni su zamišljeni kao sistem sa više agenata koji rade samostalno, ali sarađuju. Zbog toga SI algoritmi pružaju dobru osnovu za paralelizaciju i to na različitim nivoima. Paralelizacija visokog nivoa pretpostavlja krupnu granulaciju zadataka i može se primeniti na iteracije SI metoda. Manji delovi algoritama mogu takođe sadržati mnogo nezavisnih operacija, a pogodni su za paralelizaciju nižeg nivoa. Strategije paralelizacije BCO metode na višeprocesorske sisteme sa distribuiranom memorijom predložene su u [ 80, 81], a opisane su i u [ 33, 29, 82]. Za prenos podataka i komunikaciju između procesora korišćena je MPI (Message Passing Interface) biblioteka. Paralelizacija za sisteme sa deljenom memorijom pod OpenMP paradigmom, analizirana je u [ 37]. Utvrđeno je da je taj način paralelizacije prilično jednostavan i da se uvek može očekivati linearno ubrzanje. Jedini nedostatak ovog pristupa su hardverska ograničenja: broj procesora koji imaju pristup zajedničkoj memoriji je obično veoma mali. Na osnovu toga može se zaključiti da je interesantan pravac daljeg istraživanja hibridizacija MPI i OpenMP pristupa.

U cilju povećanja efikasnosti BCO metode ili mogućnosti primene na stohastičkim ili problemima višekriterijumske optimizacije, razvijani su hibridi BCO metode sa odgovarajućim tehnikama. Primeri takvih hibridnih metoda su: (1) kombinacija BCO metode sa kompromisnim programiranjem [ 45]; (2) kombinacija BCO i fazi logike [ 83] može se naći u [ 28, 46, 47, 84]; (3) kombinacija BCO metode sa lokalnim pretraživanjem (LS) [ 51, 73, 59]. LS nije prikladno za sistematsku upotrebu u metaheuristikama koje koriste populaciju rešenja jer zahteva puno vremena za izvršavanje. Međutim, ponekad se kombinuje sa ovim metodama, na primer kada treba dodatno poboljšati trenutno najbolje rešenje na kraju izvršavanja. U [ 54], kao lokalno pretraživanje korišćeno je tabu pretraživanje (Tabu Search, TS), a u [ 85] primenjeni su SA i varijanta metode spusta (Late Acceptance Hill Climbing, LAHC).

Asinhrona komunikacija, koja se pojavljuje kod pčela u prirodi, još nije razmatrana u okviru BCO metode. To bi podrazumevalo da svaka pčela odlučuje da li će učestvovati u letu unazad ili ne. Pčela koja ne učestvuje u letu unazad, sigurno ostaje lojalna svom rešenju, nastavlja da ga transformiše (možda i više puta) pre nego što dozvoli drugim pčelama da „vide" njeno rešenje, tj. uključi se u ispitivanje lojalnosti i proces regrutacije. U tom slučaju, broj rešenja koja se evaluiraju i porede varirao bi u svakom letu unazad. Ovaj vid komunikacije među pčelama usložnjava implemenaciju BCO metode, ali potencijalno pruža nove mogućnosti u procesu pretrage prostora rešenja. Zbog toga svakako zavređuje pažnju budućih istraživača koji će raditi na razvoju i primenama BCO metode.

## 5. Primena BCO metode

U ovom odeljku sumirane su dosadašnje primene BCO metode i njenih varijanti koje su bile dostupne autoru. U skorije vreme, odbranjeno je nekoliko doktorskih disertacija [ 35, 36, 37, 38] čija je glavna tema bila razvoj i/ili primena BCO metode. Pored uspešnih primena publikovanih od tvorca metode Teodorovića i njegovih koautora [ 30, 39, 40, 41, 25, 26, 27, 28, 42, 31, 43, 44, 45, 46, 47, 48, 49], sve češće i drugi autori koriste ovu metodu [ 50, 51, 52, 53, 54, 55, 56, 57, 58, 59]. Primene opisane u radovima [ 33, 34] mogu se klasifikovati na sledeći način:

- Rutiranje: problem trgovačkog putnika [ 25, 59, 73], rutiranje vozila [ 28], rutiranje vozila sa vremenskim ograničenjima [ 76], rutiranje vozila u situacijama sa neočekivano velikim zahtevima [ 79], dodela frekvencija u optičkim mrežama [ 42];
- Lokacijski problemi: problem $p$-medijane [ 48], lociranje saobraćajnih senzora na autoputevima [ 44], postavljanje inspekcijskih stanica u transportnim mrežama [ 45], lokacijski problem anti-pokrivanja [ 41], problem $p$-centara [ 30], lociranje distribuiranih računarskih resursa [ 56], lociranje fabričkih postrojenja uz ograničenja kapaciteta [ 51];
- Problemi raspoređivnja: problem statičke raspodele nezavisnih zadataka na identične procesore [ 40, 86], raspoređivanje poslova na mašine [ 87], problem zajedničkih vožnji[ 46, 47], raspoređivanje zavisnih

poslova na mašine [ 54], raspoređivanje programa u grid okruženju [ 52], planiranje operacija čuvanja podataka u računarskim sistemima (backup) [ 53], problem raspoređivanja vezova brodovima u luci [ 50];
- Mrežni problemi: projektovanje transportnih mreža [ 43, 78, 88];
- Problemi izbora: izbor objekata zadatih svojstava [ 89];
- Problemi kontinualne i mešovite optimizacije: minimizacija numeričkih funkcija [ 31], problem zadovoljivosti u verovatnosnoj logici [ 32], određivanje visina naknada u železničkom transportu [ 77];
- Optimizacioni problemi u hemiji i medicini: optimizacija hemijskih procesa [ 55], doziranje terapije u tretmanu pacijenata obolelih od raka [ 49].

Kako su svi ovi radovi detaljno opisani u [ 33, 34], u nastavku ovog odeljka navode se samo noviji rezultati, koji do sada nisu analizirani.

Da bi poboljšali performanse konstruktive BCO metode za problem trgovačkog putnika, Wong i njegovi koautori su u nekoliko radova predlagali njenu hibridizaciju sa lokalnim pretraživanjem i raznim strategijama odsecanja stabla pretrage. U [ 90] predložena je nova strategija potkresivanja stabla pretrage koja je zasnovana na dvosmernom prepoznavanju obrazaca. Samo one pčele koje su generisale rešenja sa velikim brojem prepoznatih obrazaca smatraju se perspektivnima i podvrgavaju se algoritmu lokalne pretrage radi dodatnog poboljšanja njihovih rešenja. Implementacija je testirana na 18 primera simetričnog problema trgovačkog putnika kod kojih broj gradova varira između 318 i 1291. Dobijeni eksperimentalini rezultati pokazali su da se vreme potrebno za nalaženje trenutno najboljih rešenja može, u proseku, skratiti za preko 20%.

Problem preraspodele raspoloživih autobusa u javnom saobraćaju u slučaju otkaza nekih od vozila razmatran je u [ 91]. Osnovni cilj je da se minimizira vreme čekanja putnika. Kako se otkazi dešavaju u realnom vremenu, dispečeri treba da reaguju u roku od nekoliko minuta. Korišćena je BCOi varijanta kod koje se kao početna rešenja svake iteracije uzimaju ona kod kojih nema preraspodele, dakle, čim se smanjio broj autobusa jednostavno se obrišu polasci nedostajućih vozila i ono što ostane predstavlja početno rešenje za svaku pčelu. Tokom leta unapred, pčele pokušavaju da modifikovanjem postojećih polazaka smanje kašnjenja u transportu putnika. Razmatrano je nekoliko načina da se to uradi, a najbolji od njih upoređen je sa SA metodom na 20 test primera u kojima broj linija varira od 32 do 61, a broj stanica od 200 do 350. Pokazano je da je BCOi u proseku bolja metoda po oba kriterijuma evaluacije: kvalitet rešenja i vreme izvršavanja.

Konstruktivna varijanta BCO metode primenjena je u [ 92] na problem klasifikacije dokumenata. U cilju poboljšanja performansi originalnog algoritma, autori su izvršili nekoliko modifikacija. Najpre su dozvolili da u regrutaciji učestvuju i parcijalna rešenja nelojalnih pčela i uveli su dodatnu pčelu (tzv. klona) koja, u svakom letu unapred, simulira ponašanje pčele koja je generisala do sada najbolje rešenje. Nakon toga, hibridizovali su BCO sa k-means heuristikom i to na više načina. Eksperimentalno je identfikovana najbolja hibridna varijanta i pokazano je da daje bolje rezultate od GA.

U [ 85] modifikovani BCOi primenjen je na problem određivanja rasporeda polaganja ispita. Prva modifikacija sastoji se u adaptaciji transformacija rešenja. Autori su predložili 4 vrste transformacija i kombinuju ih tako da se češće koriste one koje vode poboljšanju trenutnih rešenja. Naredna modifikacija odnosi se na hibridizaciju sa lokalnim pretraživanjem: u letu unapred, SA i LAHC se koriste za poboljšanje rešenja lojalnih pčela. Konačno, u procesu regrutacije testirano je 4 metode izbora (rulet, rangiranje, turnir, poremećaj) i pokazano je da poslednja daje najbolje rezultate. Rezultujući algoritam ravnopravan je sa trenutno najboljim metodama za dati problem, a dobijeno je i jedno novo najbolje rešenje među standardnim test primerima.

Problem predviđanja vremena putovanja u srednje opterećenim putnim mrežama razmatran je u [ 93]. Primenjena je konstruktivna verzija BCO metode prilagođena dinamičkoj prirodi problema. Sioux-Falls City putna mreža korišćena je za poređenje BCO pristupa sa popularnim komercijalnim softverom za isti problem. Pokazano je da BCO realnije prdviđa vreme putovanja i prosečnu brzinu kretanja vozila.

Hibridizacija BCO metode i fazi logike predložena je u [ 84] za projektovanje efikasnih fazi kontrolera. Zanimljivo je da su verovatnoće izbora preuzimane iz ACO metode, a ne iz originalnog BCO algoritma. Fazi pravila korišćena su za dinamičko podešavanje parametara u formulama za izračunavanje verovatnoća. Na nekoliko primera pokazano je da modifikovani BCO radi bolje od polaznog algoritma.

U radu [ 94] razmatratno je automatsko pogađanje značenja (smisla) rečenice na osnovu reči koje se u njoj pojavljuju. Korišćena je nova verzija konstruktivne BCO metode nazvana D-Bees. U ovoj verziji uloga košnice je velika, ona treba da predstavlja ključnu reč na osnovu koje će se ispitivati smisao cele rečenice. Košnica se postavlja u reč koja ima najmanje različitih značenja, ali nije nedvosmislena. Broj pčela uvek je jednak broju značenja košnice. Svaka pčela generiše skup značenja reči u rečenici i izračunava vrednost smisla u tom skupu tako što slučajno (sa verovatnoćom proporcionalnom učestalosti značenja) odabire smisao za ostale reči (koje nisu

košnica). Vrednost smisla definisana je kao suma sličnosti značenja analiziranih reči. Izmene postoje i u procesu regrutacije, samo podskup lojalnih pčela koje poseduju najbolja parcijalna rešenja mogu da postanu regruteri. Poređenja sa GA, SA i ACO pokazala su da D-Bees daje bolje rezultate na 7 standardnih primera iz korpusa engleskog jezika.

## 6. Zaključak

Algoritam optimizacije kolonijom pčela (Bee Colony Optimization, BCO), je metaheuristička metoda inspirisana ponašanjem pčela u potrazi za hranom i pripada klasi tehnika inteligencije roja. Ona predstavlja opšti algoritamski okvir koji može da se primenjuje na različite optimizacione probleme u kombinatornoj/kontinualnoj optimizaciji i inženjerstvu. BCO metoda se zasniva na konceptu saradnje, čime se povećava efikasnost veštačkih pčela i omogućava postizanje ciljeva koje samostalni agenti ne mogu ostvariti. Kroz proces razmene informacija i regrutovanje, BCO ima sposobnost da intenzivira pretragu u regionima prostora rešenja koji sadrže potencijalno kvalitetna rešenja. BCO postaje veoma popularana metoda zbog svoje jednostavnosti: lako je razumljiva i ima mali broj parametara (broj pčela i broja transformacija tokom jedne iteracije). Međutim, nije rasprostranjena jer je autori, pa ni ostali istraživači koji je koriste nisu dovoljno reklamirali.

Ovaj rad predstavlja BCO metodu kao efikasnu optimizacionu metodu i sadrži pregled novije literature u vezi sa razvojem i primenama BCO metode na kombinatorne i kontinualne optimizacione probleme. Glavni cilj rada je promovisanje ove metode kao prve metaheuristike koja koristi inteligenciju pčela u razvoju optimizacionog algoritma. Glavna konkurencija BCO metodi, optimizacija veštačkim pčelinjim kolonijama (Artificial Bee Colony, ABC) pojavila se znatno kasnije [ 61], ali je mnogo poznatija i rasprostranjenija. U [ 95] procenjeno je da se BCO koristi samo u 13% radova koji promovišu pčelinje algoritme.

Predstavljeni pregled radova sigurno nije iscrpan, jer su mogućnosti za nove primene beskrajne. Osim toga, pogodnost za paralelizaciju BCO metode otvara ne samo novi pravac istraživanja, već i neke nove potencijalne primene. Nove varijante utemeljene na principima BCO (autonomija, distribuirano funkcionisanje, samoorganizovanje, razmena informacija, kolaboracija), a razvijene na osnovu do sada postignutih rezultata i stečenog iskustva, verovatno će značajno doprineti rešavanju složenih optimizacionih problema u inženjerstvu, transportu, upravljanju i mnogim drugim oblastima svakodnevnog života.

**Zahvalnica.** *Autor se zahvaljuje Ministarstvu prosvete, nauke i tehnološkog razvoja preko projekta br. 174033.*

### Bibliografija

[1] E.H. Aarts, H.M. Korst, and P.J. van Laarhoven. Simulated annealing. In E. Aarts and J.K. Lenstra, editors, *Local Search in Combinatorial Optimization*, pages 121–136. Wiley, Chichester, 1997.

[2] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[3] A. G. Nikolaev and S. H. Jacobson. Simulated annealing. In M. Gendreau and J-Y. Potvin, editors, *Handbook of Metaheuristics*, pages 1–39. (second edition) Springer, 2010.

[4] T. Bäck, D. B. Fogel, and Z. Michalewicz. *Evolutionary computation 1: Basic algorithms and operators*. Taylor & Francis Group, LLC, CRC Press, 2000.

[5] T. Bäck, D. B. Fogel, and Z. Michalewicz. *Evolutionary computation 2: Advanced Algorithms and Operators*. Bristol, Philadelphia: Institute of Physics Publishing, 2000.

[6] K. A. Dowsland. Genetic algorithms – a tool for OR? *Journal of the Operational Research Society*, 47:550–561, 1996.

[7] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1989.

[8] H. Mühlenbein. Genetic algorithms. In E. Aarts and J. K. Lenstra, editors, *Local Search in Combinatorial optimization*, pages 137–171. John Wiley & Sons Ltd., 1997.

[9] C. R. Reeves. Genetic algorithms. In M. Gendreau and J-Y. Potvin, editors, *Handbook of Metaheuristics*, pages 109–139. (second edition) Springer, 2010.

[10] M. Dorigo and G. Di Caro. Ant colony optimization: a new meta-heuristic. In *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*, volume 2. IEEE, 1999.

[11] M. Dorigo and T. Stützle. *Ant Colony Optimization*. The MIT Press, 2004.

[12] M. Dorigo and T. Stützle. Ant colony optimization: Overview and recent advances. In M. Gendreau and J-Y. Potvin, editors, *Handbook of Metaheuristics*, pages 227–263. (second edition) Springer, 2010.

[13] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks IV*, pages 1942–1948, 1995.

[14] E Garcia-Gonzalo and JL Fernandez-Martinez. A brief historical review of particle swarm optimization (pso). *Journal of Bioinformatics and Intelligent Control*, 1(1):3–16, 2012.
[15] G. Beni. The concept of cellular robotic system. In *Proceedings of the 1988 IEEE International Symposium on Intelligent Control*, pages 57–62, IEEE Computer Society Press, Los Alamitos, CA, 1988.
[16] G. Beni and J. Wang. Swarm intelligence. In *Proceedings of the Seventh Annual Meeting of the Robotics Society of Japan*, pages 425–428, RSJ Press, Tokyo, 1989.
[17] E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence*. Oxford University Press, Oxford, 1997.
[18] J. Greensmith, A. Whitbrook, and U. Aickelin. Artificial immune systems. In M. Gendreau and J-Y. Potvin, editors, *Handbook of Metaheuristics*, pages 421–448. (second edition) Springer, New York Dordrecht Heidelberg London, 2010.
[19] J. Timmis, P. Andrews, and E. Hart, editors. *Swarm Intelligence: Special issue on artificial immune systems*, volume 4. Springer, 2010.
[20] R. Chiong, editor. *Nature-inspired algorithms for optimization*. Springer, 2009.
[21] G. Rozenberg, T. Bäck, and J. N. Kok, editors. *Handbook of Natural Computing*, volume I-IV. Springer, 2012.
[22] B. Xing and W.-J. Gao. *Innovative computational intelligence: a rough guide to 134 clever algorithms*. Springer, 2014.
[23] X.-S. Yang. *Nature-inspired metaheuristic algorithms*. Luniver press, 2010.
[24] X.-S. Yang. *Nature-inspired optimization algorithms*. Elsevier, 2014.
[25] P. Lučić and D. Teodorović. Bee system: modeling combinatorial optimization transportation engineering problems by swarm intelligence. In *Preprints of the TRISTAN IV Triennial Symposium on Transportation Analysis*, pages 441–445. Sao Miguel, Azores Islands, 2001.
[26] P. Lučić and D. Teodorović. Transportation modeling: an artificial life approach. In *Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence*, pages 216–223, Washington, DC, 2002.
[27] P. Lučić and D. Teodorović. Computing with bees: attacking complex transportation engineering problems. *International Journal on Artificial Intelligence Tools*, 12(3):375–394, 2003.
[28] P. Lučić and D. Teodorović. Vehicle routing problem with uncertain demand at nodes: the bee system and fuzzy logic approach. In J. L. Verdegay, editor, *Fuzzy Sets based Heuristics for Optimization*, pages 67–82. Physica Verlag, 2003.
[29] T. Davidović, D. Teodorović, and M. Šelmić. Bee colony optimization Part I: The algorithm overview. *Yugoslav Journal of Operational Research*, 25(1):33–56, 2015.
[30] T. Davidović, D. Ramljak, M. Šelmić, and D. Teodorović. Bee colony optimization for the p-center problem. *Computers and Operations Research*, 38(10):1367–1376, 2011.
[31] M. Nikolić and D. Teodorović. Empirical study of the bee colony optimization (BCO) algorithm. *Expert Systems with Applications*, 40(11):4609–4620, 2013.
[32] T. Stojanović, T. Davidović, and Z. Ognjanović. Bee-colony optimization for the satisfiability problem in probabilistic logic. *Applied Soft Computing*, 31:339–347, 2015.
[33] T. Davidović. Bee colony optimization: Recent developments and applications. In *Proc. XI Balkan Conference on Operational Research*, pages 225–235, BALCOR 2015, (plenary talk) Constanta, Romania, 2015.
[34] D. Teodorović, M. Šelmić, and T. Davidović. Bee colony optimization Part II: The application survey. *Yugoslav Journal of Operational Research*, 25(2):185–219, 2015.
[35] M. Šelmić. *Location problems on transport networks by computational intelligence methods*. PhD thesis, Faculty of Traffic and Transportation, University of Beograde, 2011.
[36] M. Nikolić. *Disruption management in transportation by the Bee Colony Optimization mataheuristic*. PhD thesis, Faculty of Traffic and Transportation, University of Beograde, 2015.
[37] T. Jakšić Krüger. *Development, implementation, and theoretical analysis of the Bee Colony Optimization (BCO) metaheuristic method*. PhD thesis, Faculty of Techincal Science, University of Novi Sad, 2016.
[38] T. Stojanović. *The development and analisys of metaheuristics for satisfiability in probabilistic logics*. PhD thesis, Faculty of Science, University of Kragujevac, 2016.
[39] T. Davidović, M. Šelmić, and D. Teodorović. Scheduling independent tasks: Bee colony optimization approach. In *Proc. 17th Mediterranean Conference on Control and Automation*, pages 1020–1025, Makedonia Palace, Thessaloniki, Greece, 2009.
[40] T. Davidović, M. Šelmić, D. Teodorović, and D. Ramljak. Bee colony optimization for scheduling independent tasks to identical processors. *J. Heur.*, 18(4):549–569, 2012.
[41] B. Dimitrijević, D. Teodorović, V. Simić, and M. Šelmić. Bee colony optimization approach to solving the anticovering location problem. *Journal of Computing in Civil Engineering*, 26(6):759–768, 2011.
[42] G. Marković, D. Teodorović, and V. Aćimović-Raspopović. Routing and wavelength assignment in all-optical networks based on the bee colony optimization. *AI Commun.*, 20(4):273–285, 2007.
[43] M. Nikolić and D. Teodorović. Transit network design by bee colony optimization. *Expert Systems with Applications*, 40(15):5945–5955, 2013.
[44] M. Šelmić, P. Edara, and D. Teodorović. Bee colony optimization approach to optimize locations of traffic sensors on highways. *Tehnika*, 6 (in Serbian):9–15, 2008.
[45] M. Šelmić, D. Teodorović, and K. Vukadinović. Locating inspection facilities in traffic networks: an artificial intelligence

approach. *Transport. Plan. Techn.*, 33(6):481–493, 2010.

[46] D. Teodorović and M. Dell'Orco. Bee colony optimization - a cooperative learning approach to complex transportation problems. In *Advanced OR and AI Methods in Transportation. Proceedings of the 10th Meeting of the EURO Working Group on Transportation*, pages 51–60, Poznan, Poland, 2005.

[47] D. Teodorović and M. Dell'Orco. Mitigating traffic congestion: solving the ride-matching problem by bee colony optimization. *Transport. Plan. Techn.*, 31(2):135–152, 2008.

[48] D. Teodorović and M. Šelmić. The BCO algorithm for the $p$-median problem. In *Proceedings of the XXXIV Serbian Operations Research Conferece*, pages 417–420, Zlatibor, Serbia (in Serbian), 2007.

[49] D. Teodorović, M. Šelmić, and Lj. Mijatović-Teodorović. Combining case–based reasoning with bee colony optimization for dose planning in well differentiated thyroid cancer treatment. *Expert Systems with Applications*, 40(6):2147–2155, 2013.

[50] N. Kovač. Bee colony optimization algorithm for the minimum cost berth allocation problem. In *XI Balkan Conference on Operational Research*, pages 245–254, (BALCOR, 2013), Beograd–Zlatibor, 2013.

[51] T. V. Levanova and E. A. Tkachuk. Development of a bee colony optimization algorithm for the capacitated plant location problem. In *II International conference, Optimization and applications (OPTIMA-2011)*, pages 153–156, Petrovac, Montenegro, 2011.

[52] Z. Mousavinasab, R. Entezari-Maleki, and A. Movaghar. A bee colony task scheduling algorithm in computational grids. In *Proc. Int. Conf. Digital Information Processing and Communications, ICDIPC 2011, Part I*, pages 200–210. Springer, Ostrava, Czech Republic, July 7-9, 2011.

[53] R. Nedeljković, S. Mitrović, and D. Drenovac. Bee colony optimization meta-heuristic for backup allocation problem. In *Proc. PosTel XXVII*, pages 115–122, (in Serbian), Beograd, Serbia, 2009.

[54] A. P. P. Pertiwi and P. Suyanto. Globally evolved dynamic bee colony optimization. In *Proc. 15th Int. Conf. Knowledge-Based and Intelligent Information and Engineering Systems, KES 2011, Part I*, pages 52–61. Springer, Kaiserslautern, Germany, Sept. 12–14, 2011.

[55] M. Sa'idi, N. Mostoufi, and R. Sotudeh-Gharebagh. Modelling and optimisation of continuous catalytic regeneration process using bee colony algorithm. *The Canadian Journal of Chemical Engineering*, 91(7):1256–1269, 2013.

[56] M. F. Sohi, M. Shirdel, and A. Javidaneh. Applying bco algorithm to solve the optimal dg placement and sizing problem. In *Power Engineering and Optimization Conference (PEOCO), 2011 5th International*, pages 71–76. IEEE, 2011.

[57] L.-P. Wong, M. Y. Hean Low, and C. S. Chong. A bee colony optimization algorithm for traveling salesman problem. In *2-nd Asia International Conference on Modelling & Simulation*, pages 818–823, 2008.

[58] L.-P. Wong, M. Y. Hean Low, and C. S. Chong. An efficient bee colony optimization algorithm for traveling salesman problem using frequency-based pruning. In *7-th IEEE International Conference on Industrial Informatics*, pages 775–782, 2009.

[59] L-P. Wong, M. Y. Hean Low, and C. S. Chong. Bee colony optimization with local search for traveling salesman problem. *International Journal on Artificial Intelligence Tools*, 19(03):305–334, 2010.

[60] S. Camazine and J. Sneyd. A model of collective nectar source by honey bees: self-organization through simple rules. *Journal of Theoretical Biology*, 149:547–571, 1991.

[61] D. Karaboga. An idea based on honey bee swarm for numerical optimization. Technical report, Erciyes University, Engineering Faculty Computer Engineering Department Kayseri/Turkiye, 2005.

[62] D. T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, and M. Zaidi. The bees algorithm - a novel tool for complex optimisation problems. In *Proceedings of the 2nd Virtual International Conference on Intelligent Production Machines and Systems (IPROMS 2006)*, pages 454–459, Elsevier, Cardiff, 2006.

[63] D. T Pham, A. J. Soroka, A. Ghanbarzadeh, and E. Koc. Optimising neural networks for identification of wood defects using the bees algorithm. In *Proceedings of the IEEE International Conference on Industrial Informatics*, pages 1346–1351, Singapore, 2006.

[64] K. Von Frisch. *The dance language and orientation of bees.* Harvard University Press, 1967.

[65] P. Maksimović and T. Davidović. Parameter calibration in the bee colony optimization algorithm. In *XI Balcan Conference on Operational Research*, pages 263–272, BALCOR 2013, Beograd-Zlatibor, Serbia, 2013.

[66] W. J. Gutjahr. Convergence analysis of metaheuristics. In V. Maniezzo, T. Stèutzle, and S. Voss, editors, *Matheuristics: hybridizing metaheuristics and mathematical programming*, volume 10, pages 159–187. Springer, 2009.

[67] T. Jakšić Krüger. On the convergence of the bee colony optimization meta-heuristic. In *The fourth Symposium "Mathematics and Applications"*, volume IV(1), Faculty of Mathematics, University of Belgrade, Serbia (in Serbian), 2013.

[68] T. Jakšić Krüger, T. Davidović, D. Teodorović, and M. Šelmić. The bee colony optimization algorithm and its convergence. *International Journal of Bio-Inspired Computation*, (accepted), 2014.

[69] Jakšić Krüger, T. and Davidović, T. Model convergence properties of the constructive bee colony optimization algorithm. In *XLI Symposium on Operations Research*, pages 340–345, SYM-OP-IS 2014, Divčibare, Serbia, 2014.

[70] D. McFadden. Conditional logit analysis of quanlitative choice behavior. In P. Zaremmbka, editor, *Frontier of Econometrics*, pages 105–142. Academic Press, New York, 1973.

[71] D. Teodorović, P. Lučić, G. Marković, and M. Dell'Orco. Bee colony optimization: principles and applications. In

B. Reljin and S. Stanković, editors, *Proceedings of the Eight Seminar on Neural Network Applications in Electrical Engineering - NEUREL 2006*, pages 151–156, University of Belgrade, Belgrade, 2006.

[72] P. Edara, M. Šelmić, and D. Teodorović. Heuristic solution algorithms for a traffic sensor optimization problem. In *INFORMS 2008*, Washington D.C., 2008.

[73] L.-P. Wong, M. Y. Hean Low, and C. S. Chong. Bee colony optimization with local search for traveling salesman problem. In *6-th IEEE International Conference on Industrial Informatics*, pages 1019–1025, 2008.

[74] D. Teodorović. Bee colony optimization (BCO). In C. P. Lim, L. C. Jain, and S. Dehuri, editors, *Innovations in Swarm Intelligence*, pages 39–60. Springer-Verlag, Berlin Heidelberg, 2009.

[75] L.-P. Wong, C. Y. Puan, M. Y. H. Low, Y. W. Wong, and C. S. Chong. Bee colony optimisation algorithm with big valley landscape exploitation for job shop scheduling problems. *International Journal of Bio-Inspired Computation*, 2(2):85–99, 2010.

[76] M. Nikolić, D. Teodorović, and M. Šelmić. Solving the vehicle routing problem with time windows by bee colony optimization metaheuristic. In *Proc. 1st Logistics International Conference*, pages 44–48, Belgrade, Serbia, 2013.

[77] M. Bugarinović, T. Davidović, and B. Bošković. Management of the access charges level for the use of railway infrastructure by bee colony optimization. In *18th Euro Working Group on Transportation, EWGT 2015*, Delft, The Netherlands, 2015.

[78] M. Nikolić and D. Teodorović. A simultaneous transit network design and frequency setting: Computing with bees. *Expert Systems with Applications*, 41(16):7200–7209, 2014.

[79] M. Nikolić and D. Teodorović. Vehicle rerouting in the case of unexpectedly high demand in distribution systems. *Transportation Research Part C: Emerging Technologies*, 55:535–545, 2015.

[80] T. Davidović, D. Ramljak, M. Šelmić, and D. Teodorović. Mpi parallelization of bee colony optimization. In *Proc. 1st International Symposium & 10th Balkan Conference on Operational Research*, volume 2, pages 193–200, Thessaloniki, Greece, 2011.

[81] T. Davidović, T. Jakšić, D. Ramljak, M. Šelmić, and D. Teodorović. Mpi parallelization strategies for bee colony optimization. *Optimization, Special Issue entitled "Advances in Discrete Optimization" dedicated to BALCOR 2011*, 62(8):1113–1142, 2013. DOI:10.1080/02331934.2012.749258.

[82] T. G. Crainic, T. Davidović, and D. Ramljak. Designing parallel meta-heuristic methods. In M. Despotović-Zrakić, V. Milutinović, and A. Belić, editors, *High Performance and Cloud Computing in Science and Education*, pages 260–280. IGI-Global, 2014.

[83] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[84] C. Caraveo, F. Valdez, and O. Castillo. Optimization of fuzzy controller design using a new bee colony algorithm with fuzzy dynamic parameter adaptation. *Applied Soft Computing*, 43:131–142, 2016.

[85] M. Alzaqebah and S. Abdullah. Hybrid bee colony optimization for examination timetabling problems. *Computers & Operations Research*, 54:142–154, 2015.

[86] V. Arabnejad, A. Moeini, and N. Moghadam. Using bee colony optimization to solve the task scheduling problem in homogenous systems. *IJCSI International Journal of Computer Science Issues*, 8(5):348–353, 2011.

[87] Y.-M. Huang and J.-C. Lin. A new bee colony optimization algorithm with idle-time-based filtering scheme for open shop-scheduling problems. *Expert Systems with Applications*, 38(5):5438–5447, 2011.

[88] H. Zhanga, P. Zhaoa, J. Gaoa, C. Zhugeb, and X. Yaob. An effective intelligent method for optimal urban transit network design? *Journal of Information & Computational Science*, 12(6):2177–2184, 2015.

[89] A. Moayedikia, R. Jensen, U. K. Wiil, and R. Forsati. Weighted bee colony algorithm for discrete optimization problems with application to feature selection. *Engineering Applications of Artificial Intelligence*, 44:153–167, 2015.

[90] L-P. Wong and S. S. Choong. A bee colony optimization algorithm with frequent-closed-pattern-based pruning strategy for traveling salesman problem. In *IEEE Conference on Technologies and Applications of Artificial Intelligence* (TAAI), pages 308–314. IEEE, 2015.

[91] M. Nikolić, D. Teodorović, and K. Vukadinović. Disruption management in public transit: the bee colony optimization approach. *Transportation Planning and Technology*, 38(2):162–180, 2015.

[92] R. Forsati, A. Keikha, and M. Shamsfard. An improved bee colony optimization algorithm with an application to document clustering. *Neurocomputing*, 159:9–26, 2015.

[93] M. Dell'Orco, M. Marinelli, and M. A. Silgu. Bee colony optimization for innovative travel time estimation, based on a mesoscopic traffic assignment model. *Transportation Research Part C: Emerging Technologies*, 66:48–60, 2016.

[94] S. Abualhaija and K-H. Zimmermann. D-bees: A novel method inspired by bee colony optimization for solving word sense disambiguation. *Swarm and Evolutionary Computation*, 27:188–195, 2016.

[95] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga. A comprehensive survey: artificial bee colony (abc) algorithm and applications. *Artificial Intelligence Review*, 42:21–57, 2014.

Симпозијум МАТЕМАТИКА И ПРИМЕНЕ, Математички факултет, Универзитет у Београду, 2015 ,Vol. VI(1)

# Расподела по модулу 1 збира степена Пизоових и Салемових бројева

**Драган Станков**

*Рударско-геолошки факултет Универзитета у Београду, Ђушина 7*
*e-mail: dstankov@rgf.bg.ac.rs*

**Апстракт.** Добро је познато да за било који ирационалан реалан број $\alpha$, низ $(n\alpha)_{n\geq 1}$ има униформну расподелу по модулу 1. За поједине алгебарске бројеве одређујемо подниз овог низа такав да разломљени делови чланова подниза конвергирају ка 0. Такође је добро познато да за било који Салемов број $\theta$, низ $(\theta^n)_{n\geq 1}$ је густ на сегменту $[0,1]$ али нема униформну расподелу по модулу 1. За Салемове бројеве показујемо како одредити подниз низа $(n\theta)_{n\geq 1}$ такав да разломљени делови чланова подниза немају униформну расподелу.

**Кључне речи:** Пизоов број; Салемов Број; расподела по модулу 1.

## 1. Увод

Већ дуже време се проучава расподела по модулу 1 (расподела разломљеног дела) степена неког фиксираног реалног броја $\theta$ већег од 1. У својој монографији [7], Салем је разматрао одређене специјалне алгебарске целе бројеве. Он је показао да низ $(\theta^n)_{n\geq 1}$ тежи ка нули у $\mathbb{R}/\mathbb{Z}$ када је $\theta$ Пизоов (Pisot) број. Ако је $\theta$ Салемов број тада $(\theta^n)_{n\geq 1}$ је густ у $\mathbb{R}/\mathbb{Z}$, т.ј. разломљени делови $\theta^n$ су густи у јединичном интервалу $[0,\ 1]$ али немају униформну расподелу. (Видети [2], с. 87-89.) Штавише Салемови бројеви су једини познати бројеви чији су степени густи у $\mathbb{R}/\mathbb{Z}$.

Користићемо следећу нотацију где $x$, $x'$ означавају реалне бројеве:

1. Цео део броја: $[x] = \max\{n \in Z : n \leq x\}$.
2. Разломљени део броја: $\{x\} = x - [x]$.
3. Конгруенција по модулу 1: $x \equiv x' \bmod 1 \Leftrightarrow x - x' \in \mathbb{Z}$.
4. Растојање броја $x$ од најближег целог броја: $\|x\| = \min\{|x - n| : n \in \mathbb{Z}\}$.

Наводимо дефиниције које ће нам бити потребне.

**Дефиниција 1.** Пизоов број је реални алгебарски цео број $\theta$ већи од 1, чији сви алгебарски конјугати, осим самог $\theta$, имају модуо строго мањи од 1.

**Дефиниција 2.** Салемов број је реални алгебарски цео број $\theta$ већи од 1, чији сви алгебарски конјугати, осим самог $\theta$, имају модуо мањи или једнак од 1, и бар један има модуо који је једнак 1.

Лако се показује да Салемов број има тачно један коњугат, а то је $\theta^{-1}$, унутар јединичног диска, док су сви остали на његовој граници, тј. на јединичном кругу. Степен Салемовог броја, означимо га са $2t$, мора бити паран и већи или једнак од 4.

**Дефиниција 3.** Нека је $(u_n)_{n\geq 1}$ низ реалних бројева и нека је $x \in [0,1]$. Тада се гранична вредност $f(x) = \lim_{N\to\infty} \frac{\text{card}\{n<N|\{u_n\}<x\}}{N}$, када она постоји назива функција расподеле низа $(u_n)_{n\geq 1}$ за аргумент $x$.

Овде разматрамо само оне аргументе $x$ за које постоје функција $f(x)$ и њен први извод скоро свуда.

Претпостављамо да је $\theta$ Салемов број. Означимо конјугате броја $\theta$ са $\theta^{-1}$, $\exp(\pm 2i\pi\omega_1)$, $\ldots$, $\exp(\pm 2i\pi\omega_{t-1})$. Како је сума $n$-их степена ма ког алгебарског броја и његових конјугата цео број, за све $n \in \mathbb{N}$, коришћењем Де Моаврових формула закључујема да мора бити $\theta^n + \theta^{-n} + 2\sum_{j=1}^{t-1} \cos 2\pi n\omega_j \equiv 0(\bmod 1)$ тако да расподела $\theta^n \bmod 1$ је заправо расподела $-2\sum_{j=1}^{t-1} \cos 2\pi n\omega_j$. Ако је $\theta$ Салемов број четвртог степена Џупа (Dupain) [6] је експлицитно одредио функцију расподеле за $(\theta^n)_{n\geq 1}$, по модулу 1. Наиме,

$$f(x) = \frac{5}{2} - \frac{1}{\pi}\left(\arccos\frac{x-2}{2} + \arccos\frac{x-1}{2} + \arccos\frac{x}{2} + \arccos\frac{x+1}{2}\right).$$

Из овога следи да је

$$f'(x)=\frac{1}{2\pi}\left(\frac{1}{\sqrt{1-\left(\frac{x-2}{2}\right)^2}}+\frac{1}{\sqrt{1-\left(\frac{x-1}{2}\right)^2}}+\frac{1}{\sqrt{1-\left(\frac{x}{2}\right)^2}}+\frac{1}{\sqrt{1-\left(\frac{x+1}{2}\right)^2}}\right).$$

Ако је $\theta$ Салемов број степена $2t$, $t\geq 2$, Дош(Doche), Мандес Франс (Mendès France) и Руш(Ruch) [5] су одредили функцију густине за $(\theta^n)_{n\geq 1}$, по модулу 1:

$$f'(x)=1+2\sum_{k=1}^{\infty}J_0(4k\pi)^{t-1}\cos 2\pi kx \tag{1}$$

на интервалу $(0,1)$. Овде је $J_0(\cdot)$ Беселова функција прве врсте са индексом 0.

## 2. Главни резултати

Најважније резултате ћемо изложити у наредне две теореме.

**Теорема 1.** *Нека је $\theta$ Пизоов број чији је минимални полином $P(x)=x^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0$. Нека су његови конјугати $\theta_2,\theta_3,\ldots,\theta_m$. Нека је $(F_n)_{n\geq 1}$ низ дефинисан линеарном рекурентном формулом: $F_n=\theta^n+\theta_2^n+\theta_3^n+\cdots+\theta_m^n$, за $n=1,2,\ldots,m$ и $F_n=-b_{m-1}F_{n-1}-b_{m-2}F_{n-2}-\cdots-b_1F_{n-m+1}-b_0F_{n-m}$ за $n>m$. Тада низ $(F_n\theta)_{n\geq 1}$ конвергира ка 0 по модулу 1.*

*Доказ.* Користимо теорију диференцних једначина. Рекурентном формулом је дата диференцна једначина са карактеристичним полиномом $P(x)$. Из датих почетних услова следи да је $F_n=\theta^n+\theta_2^n+\theta_3^n+\cdots+\theta_m^n$, за све $n=1,2,\ldots$, при чему је јасно да су $F_n$ цели бројеви. Покажимо да $F_n\theta$ може бити за довољно велико $n$ по вољи близак целом броју. Како су $|\theta_i|<1$ за $i=2,3,\ldots,m$ следи да $\theta_2^n+\theta_3^n+\cdots+\theta_m^n$ тежи нули када $n$ тежи бесконачно. Зато за произвољно $\varepsilon>0$ можемо одабрати $n_0$ тако да је за свако $n\geq n_0$ је $\theta_2^n+\theta_3^n+\cdots+\theta_m^n<\frac{\varepsilon}{2\theta}$. Сада имамо да је $F_n\theta=\theta^{n+1}+(\theta_2^n+\theta_3^n+\cdots+\theta_m^n)\theta=\theta^{n+1}+\sum_{i=2}^m\theta_i^{n+1}-\sum_{i=2}^m\theta_i^{n+1}+(\theta_2^n+\theta_3^n+\cdots+\theta_m^n)\theta=F_{n+1}-\sum_{i=2}^m\theta_i^{n+1}+(\theta_2^n+\theta_3^n+\cdots+\theta_m^n)\theta$. Следи да је $|F_n\theta-F_{n+1}|=|-\sum_{i=2}^m\theta_i^{n+1}+(\theta_2^n+\theta_3^n+\cdots+\theta_m^n)\theta|<\frac{\varepsilon}{2\theta}+\frac{\varepsilon}{2}<\varepsilon$ када је $n\geq n_0$. □

**Лема 1.** *За сваки квадратни корен из природног броја $N$ постоји бесконачно парова природних бројева $(p,q)$ таквих да је $p+q\sqrt{N}$ Пизоов број.*

*Доказ.* Искористићемо добро познат Дирихлеов резултат да за било који реалан број $\xi$ постоји бесконачно много рационалних бројева $p/q$ таквих да је

$$|\xi-\frac{p}{q}|<\frac{1}{q^2}.$$

Ако за $\xi=\sqrt{N}$ одаберемо разломак $p/q$ тако да важи претходна неједнакост онда ће важити и $|q\sqrt{N}-p|<1$. Следи да је $\theta=p+q\sqrt{N}$ Пизоов број чији је минимални полином $x^2-2px+p^2-Nq^2$. □

**Последица 1.** *За сваки квадратни корен из природног броја $N$ постоји низ целих бројева $(G_n)_{n\geq 1}$, дефинисан рекурентном формулом другог степена, тако да $G_n\sqrt{N}$ конвергира нули по модулу 1.*

*Доказ.* На основу претходне леме најпре налазимо бројеве $p$, $q$ такве да је $q\sqrt{N}+p$ Пизоов број. Сада можемо применити претходну теорему и наћи целобројни низ $(F_n)_{n\geq 1}$ дефинисан рекурентном формулом другог степена тако да је $(F_n(q\sqrt{N}+p))_{n\geq 1}$ конвергира ка 0 по модулу 1. Како су $F_nq$, $F_np$ целобројни ако узмемо да је $G_n=F_nq$ онда $G_n\sqrt{N}$ конвергира нули по модулу 1. □

**Пример 1.** *Ако је* $N = 3$ *тада је* $|2\sqrt{3} - 3| < 1$ *па је* $3 + 2\sqrt{3}$ *Пизоов број чији је минимални полином* $x^2 - 6x - 3$. *Пошто је* $F_1 = 6$, $F_2 = 42$, $F_n = 6F_{n-1} + 3F_{n-2}$, *налазимо да је* $F_{20} = 16224207714897426$, *и да је* $2 \cdot F_{20} \cdot \sqrt{3} = 56202304149506592,000001488957204$ *веома близак целом броју.*

**Последица 2.** *Нека су* $\theta$, $P(x)$ *и* $F_n$ *дефинисани као у претходној теореми и нека је* $x \in [0,1]$. *Тада за свако* $\varepsilon$ *постоје* $x' \in [0,1]$ *и цео број* $k$ *такви да низ* $(F_n + k)\theta$ *конвергира ка* $x'$ *по модулу 1 и да је* $|x - x'| < \varepsilon$.

*Доказ.* Пошто је низ $(\{n\theta\})_{n\geq 1}$ свуда густ на интервалу $[0,1]$ постоји $x' \in [0,1]$ и цео број $k$ тако да је $\{k\theta\} = x'$ и да је $|x - x'| < \varepsilon$. Како је $(F_n + k)\theta = F_n\theta + k\theta$ и како $F_n\theta$ конвергира ка 0 по модулу 1, следи да $(F_n + k)\theta$ конвергира ка $x'$ по модулу 1. □

**Теорема 2.** *Нека је* $\theta$ *Салемов број чији је минимални полином* $P(x) = x^m + b_{m-1}x^{m-1} + \cdots + b_1x + b_0$. *Нека је* $(F_n)_{n\geq 1}$ *низ дефинисан као збир* $n$*-тих степена Салемовог броја* $\theta$ *и његових конјугата, за* $n = 1, 2, \ldots, m$ *и* $F_n = -b_{m-1}F_{n-1} - b_{m-2}F_{n-2} - \cdots - b_1F_{n-m+1} - b_0F_{n-m}$ *за* $n > m$. *Тада је низ* $(F_n\theta)_{n\geq 1}$ *свуда густ на* $[0,1]$ *али није униформно расподељен. Ако је* $\theta$ *Салемов број четвртог степена и нека је* $\exp(\pm 2i\pi\omega)$ *пар његових конјугата на јединичном кругу. Тада се може експлицитно одредити функција расподеле:*

$$f(x) = \sum_{i=-M}^{M} (g(x+i) - g(i)),$$

*где је* $M = [2\theta] + 3$, $a = 2\pi\omega$ *и при том је*

$$g(x) = \frac{1}{\pi} \arccos \frac{\pm \sin a\sqrt{4\theta^2 - 8\theta\cos a - x^2 + 4} - x\cos a + \theta x}{2(\theta^2 - 2\cos a\theta + 1)}. \tag{2}$$

*Такође се може експлицитно одредити функција густине:*

$$f'(x) = \sum_{i=-M}^{M} g'(x+i),$$

*где је*

$$g'(x) = \frac{\cos a - \theta + \frac{x\sin a}{\sqrt{4\theta^2 - 8\theta\cos a - x^2 + 4}}}{\sqrt{4(\theta^2 - 2\theta\cos a + 1)^2 - (\sin a\sqrt{4\theta^2 - 8\theta\cos a - x^2 + 4} - x\cos a + \theta x)^2}}. \tag{3}$$

*Доказ.* Означимо конјугате броја $\theta$ са $\theta^{-1}, \exp(\pm 2i\pi\omega_1), \ldots, \exp(\pm 2i\pi\omega_{t-1})$ где је $m = 2t$. Како је сума $n$-тих степена ма ког алгебарског броја и његових конјугата цео број, за свако $n \in \mathbb{N}$, мора бити $\theta^n + \theta^{-n} + 2\sum_{j=1}^{t-1} \cos 2\pi n\omega_j \equiv 0 (\bmod 1)$. Користимо теорију диференцних једначина. Рекурентном формулом је дата диференцна једначина са карактеристичним полиномом $P(x)$. Из датих почетних услова следи да је $F_n = \theta^n + \theta^{-n} + 2\sum_{j=1}^{t-1} \cos 2\pi n\omega_j$, за свако $n = 1, 2, \ldots$. Сада имамо да је $F_n\theta = \theta^{n+1} + \theta^{-n+1} + 2\theta\sum_{j=1}^{t-1} \cos 2\pi n\omega_j = \theta^{n+1} + \theta^{-n-1} + 2\sum_{j=1}^{t-1} \cos 2\pi(n+1)\omega_j - 2\sum_{j=1}^{t-1} \cos 2\pi(n+1)\omega_j - \theta^{-n-1} + \theta^{-n+1} + 2\theta\sum_{j=1}^{t-1} \cos 2\pi n\omega_j = F_{n+1} - 2\sum_{j=1}^{t-1} \cos 2\pi(n+1)\omega_j - \theta^{-n-1} + \theta^{-n+1} + 2\theta\sum_{j=1}^{t-1} \cos 2\pi n\omega_j$. Следи да је $\{F_n\theta - F_{n+1}\} = \{-2\sum_{j=1}^{t-1} \cos 2\pi(n+1)\omega_j - \theta^{-n-1} + \theta^{-n+1} + 2\theta\sum_{j=1}^{t-1} \cos 2\pi n\omega_j\}$, тако да је расподела $F_n\theta$ заправо расподела $-2\sum_{j=1}^{t-1} \cos 2\pi(n+1)\omega_j + 2\theta\sum_{j=1}^{t-1} \cos 2\pi n\omega_j$ по модулу један. На основу [1] можемо закључити да та расподела није униформна али да за велико $t$ тежи униформној расподели.

Ако је $\theta$ Салемов број степена четири нека је $\exp(\pm 2i\pi\omega)$ пар његових конјугата на јединичном кругу. Тада је расподела $F_n\theta$ заправо расподела $-2\cos 2\pi(n+1)\omega + 2\theta\cos 2\pi n\omega$ по модулу један. Ово можемо једноставније записати ако уведемо ознаке $w = 2\pi n\omega$ и $a = 2\pi\omega$ као $Q(w) = -2\cos(w + a) + 2\theta\cos w$. Очигледно је да је $Q(w)$ периодична са периодом $2\pi$ и

да је $Q(w) = -Q(w+\pi)$ тако да је функција $Q(w)$ веома налик синусоиди помереној по $w$-оси. Екстремне вредности ова функција има када је $Q'(w) = 0$. Одредићемо нуле функције $Q'(w) = 2\sin(w+a) - 2\theta\sin w$. Добијамо једначину $\sin w\cos a + \cos w\sin a - \theta\sin w = 0$, чије једно решење је $w_1 = \arctan(\frac{\sin a}{q-\cos a})$. Искористићемо алгоритам за одређивање функције густине који је овај аутор изложио у [9] у оквиру Теореме 2.1 на основу којег можемо закључити да је $x = \{Q(w_1)\}$ вертикална асимптота графика функције густине. Може се показати да је $Q(w_1 + w) = Q(w_1 - w)$: искористимо формуле косинуса збира и разлике, пребацимо све чланове на леву стану знака = и након поништавања одређених чланова добијемо $\sin wQ'(w_1)$ што је очигледно једнако нули. Из овога закључујемо да, према поменутом алгоритму, је довољно узети само једну грану функције $Q$ где је она монотоно опадајућа, а то је или на $[w_1, w_1+\pi]$ или на $[w_1-\pi, w_1]$ и формирати инверзну функцију.

За налажење инверзне функције полазимо од $Q(w) = -2\cos(w+a) + 2\theta\cos w = -2\cos w\cos a + 2\sin w\sin a + 2\theta\cos w = x$. Одавде се добија $2\sin w\sin a = x + 2\cos w(\cos a - \theta)$. Пошто је $\sin w = \pm\sqrt{1-\cos^2 w}$ треба квадрирати леву и десну страну тако да добијамо квадратну једначину по $\cos w$: $4(1-\cos^2 w)\sin^2 a = x^2 + 4x\cos w(\cos a - \theta) + 4\cos^2 w(\cos a - \theta)^2$. Решавањем ове једначине добијамо:

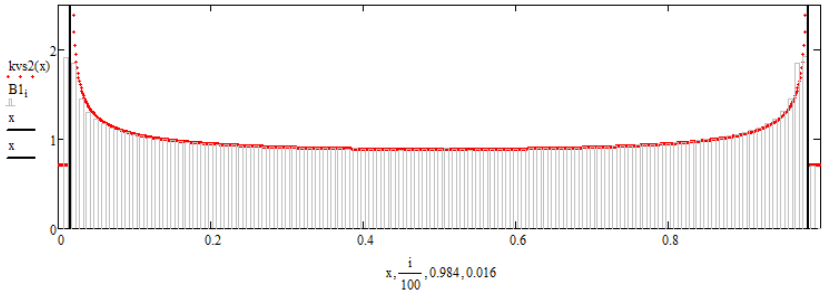
$$\cos w = \frac{\pm\sin a\sqrt{4\theta^2 - 8\theta\cos a - x^2 + 4} - x\cos a + \theta x}{2(\theta^2 - 2\cos a\theta + 1)}.$$

Функцију $g(x)$ (2) добијамо налажењем аркус косинуса десне стране а функцију $g'(x)$ (3) као први извод функције $g(x)$. Најзад из [9] закључујемо да је $f'(x) = \sum_{i=-M}^{M} g'(x+i)$. □

*Напомена* 1. График функције густине има вертикалну асимптоту $x = \{Q(w_1)\}$ и он је осно симетричан у односу на симетралу јединичног интервала, из чега следи да има још једну вертикалну асимптоту $x = 1 - \{Q(w_1)\}$.

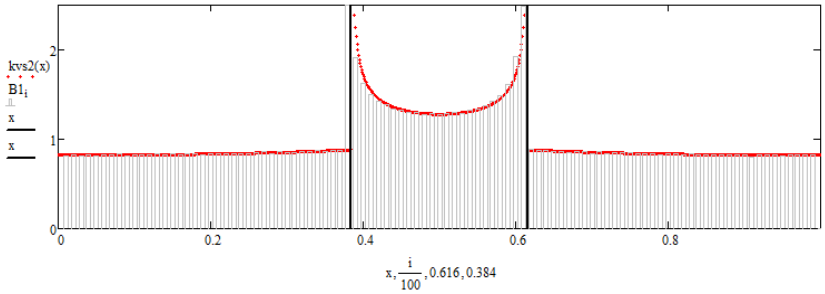
Као илустрацију Теореме 2 дајемо примере Салемових бројева четвртог степена. Ако се подсетимо дефиниције функције расподеле $f(x)$ и познате чињенице да је први извод функције на малом сегменту $[a,b]$ близак са коначном разликом $f(b) - f(a)$ подељеном са $b-a$, можемо апроксимирати $f'(x)$. Јединични сегмент поделимо на $p$ делова. Рачунамо разломљени део $F_n\theta$ за $1 \leqslant n \leqslant N$, и за сваки подинтервал одређујемо колико има бројева $n$ таквих да разломљени део $F_n\theta$ упада у тај подинтервал. Вертикална оса приказује број таквих $n$ подељених са $N/p$. Као резултат ове нормализације се добија релативни хистограм који се у највећој мери поклапа са $f'(x)$, функцијом густине одређеном коришћењем Теореме 2 и која је приказана на сликама црвеном бојом.

**Пример 2.** *Ако је* $\theta = 1,7220838...$ *Салемов број чији је минимални полином* $x^4 - x^3 - x^2 - x + 1$, *онда је* $a = 2,280208...$, $w_1 = 0,30941$, $Q(w_1) = 4,983598$ *па су* $x = 0,984$ *и* $x = 0,016$ *вертикалне асимптоте графика функције густине приказаног на Слици 1.*




**Слика 1.** Расподела низа $(F_n\theta)_{n\geq 1}$ по модулу 1 где је $\theta$ Салемов број чији је минимални полином $x^4 - x^3 - x^2 - x + 1$

**Пример 3.** *Ако је* $\theta = 1,8832...$ *Салемов број чији је минимални полином* $x^4 - 2x^3 + x^2 - 2x + 1$, *онда је* $a = 4,503772...$, $w_1 = -0,437742$, $Q(w_1) = 4,615844$ *па су* $x = 0,616$ *и* $x = 0,384$ *вертикалне асимптоте графика функције густине приказаног на Слици 2.*




**Слика 2.** Расподела низа $(F_n\theta)_{n\geq 1}$ по модулу 1 где је $\theta$ Салемов број чији је минимални полином $x^4 - 2x^3 + x^2 - 2x + 1$

## 3. Закључак

На основу изложеног намеће нам се више праваца за будуће истраживање. За које бројеве $\alpha$ постоји низ целих бројева $(G_n)_{n\geq 1}$ дефинисан рекурентном формулом тако да $G_n\alpha$ конвергира нули по модулу 1? Из [1] очекујемо да што је Салемов број $\theta$ већег степена расподела $F_n\theta$ по модулу 1 је све ближа униформној. Можемо ли то доказати и како?

**Захвалница.** Ово истраживање је подржано од стране Министарства просвете, науке и технолошког развоја Републике Србије (пројекат 174032: Анализа и алгебра са применама).

## Библиографија

[1] **S. Akiyama, Y. Tanigawa.** Salem numbers and uniform distribution modulo 1. *Publicationes Mathematicae Debrecen*, 2004, 64(3–4), 329–341.
[2] **M.-J. Bertin, A. Decomps-Guilloux, M. Grandet-Hugot, M. Pathiaux-Delefosse, J.-P. Schreiber.** Pisot and Salem numbers. *Birkhäuser, Basel*, 1992.
[3] **D. W. Boyd.** Salem numbers of degree four have periodic expansions. *Théorie des nombres, Number Theory, Walter de Gruyter, Berlin, New York*, 1989, 57–64.
[4] **Y. Bugeaud.** Distribution modulo one and diophantine approximation. *Cambridge Tracts in Mathematics 193, Cambridge University Press, Cambridge*, 2012.
[5] **C. Doche, M. Mendès France, J.-J. Ruch.** Equidistribution modulo 1 and Salem numbers. *Functiones et Approximatio, Commentarii Mathematici*, 2008, 39, part 2, 261–271.
[6] **Y. Dupain.** Répartition et discrépance. *PhD thesis, Université Bordeaux I*, 1978.
[7] **R. Salem.** Power series with integral coefficients. *Duke Mathematical Journal*, 1945, 12, 153–172.
[8] **C. Smyth.** Seventy years of Salem numbers. *Bulletin of the London Mathematical Society*, 2015, 47 (3), 379–395.
[9] **D. Stankov.** On the distribution modulo 1 of the sum of powers of a Salem number. *Comptes rendus - Mathématique*, 2016, 354, 569–576.
[10] **D. Stankov.** On linear combinations of Chebyshev polynomials. *Publications de l'Institut Mathématique*, 2015, 97, 57–67.
[11] **T. Zaïmi.** Comments on the distribution modulo one of powers of Pisot and Salem numbers. *Publicationes Mathematicae Debrecen*, 2012, 80, 417–426.

Simpozijum MATEMATIKA I PRIMENE, Matematički fakultet, Univerzitet u Beogradu, 2015 ,Vol. VI(1)

# Ocenjivanje parametara Bajesovih mreža za sisteme preporuke

**Dobrica Ćosić**

*Elektrotehnički fakultet Univerziteta u Beogradu, Bulevar kralja Aleksandra 73*
*e-mail: cosic_dobrica@yahoo.com*

**Apstrakt.** Rastući trend dostupnosti brzog interneta sve više usmerava njegovu upotrebu sa akademskih ka komercijalnim sadržajima. Sa druge strane, zahvaljujući tom trendu, istraživači i kompanije su u poziciji da, skladištenjem podataka o transakcijama i pronalaženjem parametara od značaja kojima se mogu opisati njihove međusobne veze, pokušaju da efikasno predvide korisničke želje i potrebe. Algoritam kojim od skupa diskretnih promenljivih koje karakterišu jednu odluku dolazimo do predikcije sledeće odluke u nizu se naziva sistemom preporuke (en. *recommender system*). Zadatak ovakvog sistema je da, nakon prikupljanja podataka od aktuelnog korisnika o proizvodima koji ga interesuju, i upoređivanja sa aktivnostima njemu sličnih korisnika iz baze, kreira model njegovog ponašanja nad određenim domenom i odredi meru predikcije za naredne korake. Ovaj rad pruža pregled klasifikacije sistema preporuke, kao i motiva i izazova za njihovo korišćenje. Radi mogućnosti analize kvaliteta predloženih rešenja, čitalac će dobiti uvid u niz statističkih osnova ovih sistema, kao i tehnike parametrizacije i redukcije velikih skupova podataka u cilju ubrzanja odziva i štednje resursa. Kao pogodan način, kako za kvalitativnu tako i za kvantitativnu reprezentaciju jednog takvog sistema i njegovih parametara, prikazano je postupno modeliranje hibridnog sistema preporuke pomoću Bajesove mreže. Određivanjem raspodela verovatnoće direktno se opisuje veza između korisnika i objašnjava njihov zajednički uticaj na formiranje konačne preporuke.

**Ključne reči:** sistemi preporuke, Bajesove mreže, estimacija

## 1. Uvod

Velike korporacije u ovom trenutku ulažu mnogo novca i napora kako bi znatna količina podataka koja im je na raspolaganju dobila određeni smisao, dajući im prednost u trci za osvajanje tržišta. Internet napušta tzv. „doba pretrage" i ulazi u „doba otkrića", kada je značajno da informacije koje su nedovoljno poznate ili potpuno nepoznate same pronađu put do potencijalnog korisnika [1]. Većina današnjih algoritama za formiranje sistema preporuke je bila u upotrebi i pre desetak godina [2], [3], s tim da trenutna ekspanzija prediktivne analitike kao nauke, razvoj računara nove generacije i brzih diskova, kao i programskih paketa optimizovanih za paralelno procesiranje, omogućava znatna proširenja u pogledu „čišćenja" podataka i brzine odziva sistema. Da bi ovakav sistem radio precizno i efikasno, neophodno je i dobro poznavati osobine skupa podataka čijom obradom se dolazi do preporuke, kako bi se lakše odabrali odgovarajući uzorci i metode klasifikacije [4].

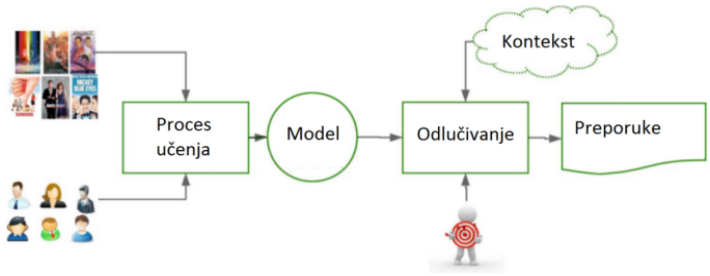
Preporuke se danas u velikoj meri dobijaju deljenjem informacija na društvenim mrežama, pri čemu se intuitivna mera kvaliteta preporuke predstavlja skupom uslovnih verovatnoća formiranih pregledom međusobne istorije dvoje ili više „umreženih" korisnika [5]. Ovakav vid razmene informacija je utemeljen na mnogo većem poverenju od onoga koje korisnici pojedinačno gaje prema standardnim medijskim servisima [6].

Motivi za komercijalnu upotrebu preporuka su raznoliki, počev od povećanja prodaje, dugoročne analize profita, diverziteta ponude, povećanog zadovoljstva korisnika... U pozadini svega je mehanizam koji, prikupljajući podatke od korisnika, pomaže pri donošenju odluka o daljim koracima na tržištu. Iskustvo nas je naučilo tome da ljudi vole da čuju tuđe mišljenje, kao i da njihovo mišljenje postane svrsishodno, pogotovo ako ga iznose iz udobnosti svog doma ili kancelarije.

Rangiranje i preporučivanje narednih predmeta korisniku se najčešće vrši na osnovu informacija o prethodnim ocenama koje je korisnik dodelio, kao i na osnovu obrazaca ponašanja sličnih korisnika [7], pri čemu trenutni trendovi favorizuju korišćenje tzv. hibridnih algoritama [8] koji kombinuju prethodno navedene tehnike. Detaljniji pregled ostalih klasifikacija vodećih algoritama i primena sistema preporuke, od kojih će nekoliko biti pomenuto u ovom radu, uključujući i status razvoja istih, dat je u publikacijama [9] i [10].

### 1.1. Izazovi i problemi

Proces formiranja i odabira kvalitetne preporuke donosi sa sobom i određene poteškoće. Uspešnost sistema preporuke najviše zavisi od aktivnosti korisnika, bilo da je u pitanju direktna kupovina, pisanje komentara i recenzija, pa čak i dodeljivanje jedne ocene ili klik na *„like"* dugme. Zbog toga su baze podataka ovih sistema,




**Slika 1.** Proces stvaranja preporuke

iako velikih dimenzija, veoma oskudne podacima o međusobnim vezama korisnika i proizvoda u trenutku preporučivanja. Pored toga, korišćenje statističkih mera i metoda za evaluaciju sistema [11] često ne uzima u obzir naglašenu važnost jedne određene osobine proizvoda, što je sa druge strane u koliziji sa ciljem maksimizacije profita [12] koji čini osnovni koncept svih komercijalnih rešenja. Ponekad je neophodna pomoć eksperata iz oblasti proizvoda koji se preporučuje, kako bi informacije koje ulaze u sistem bile razvrstane prema karakteristikama od značaja. Naravno, nikako se ne sme zanemariti nekonzistentnost u korisničkim preferencijama koja može da zbuni i navede na neželjen izbor.

Najveći problem je svakako ukomponovati odnos brzine i kvaliteta, a s tim u vezi rešiti sve poteškoće izazvane veličinom baze podataka, oskudnošću iste, množenjem velikih matrica i kvalitetnom selekcijom parametara ukoliko nam resursi nisu na zadovoljavajućem nivou.

### 1.2. Zadatak

Sistem koji će biti prikazan u ovom radu je modifikacija sistema iz publikacije [8]. U osnovi baze dat je skup proizvoda $I = \{i_1, i_2, \ldots, i_m\}$ koji mogu biti opisani skupom karakteristika $F = \{f_1, f_2, \ldots, f_l\}$. Na primeru arhive naučnih radova, karakteristike mogu biti vrsta istraživačkog rada, naučna disciplina, godina... U ovom slučaju, sardžaj je predstavljen raštrkanom *m* x *l* matricom **D**, gde $d_{i,j}$ označava da proizvod *i* može biti opisan karakteristikom *j*.

**Tabela 1.** Baza karakteristika proizvoda, **D**

| I | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $i_1$ | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $i_2$ | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| $i_3$ | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| $i_4$ | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| $i_5$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| . | . | . | . | . | . | . | . | . | . | . |

Sa druge strane, dat je skup korisnika $U = \{u_1, u_2, \ldots, u_n\}$, koji su, na eksplicitan način ili ne, dodelili ocenu određenom proizvodu, *r*. Radi jednostavnosti interpretacije, usvojena je pretpostavka da $r \in \mathbb{N}$, te da je *max(r) = 2*. Ocene se zatim smeštaju u sličnu matricu **R**, dimenzija *n* x *m* , gde $r_{a,j}$ predstavlja ocenu koju je korisnik $u_a$ dodelio objektu $i_j$ .

Navedeni primer se može modelovati Bajesovom mrežom, moćnim alatom koji, pored olakšica pri vizuelizaciji problema i uzročno-posledičnih veza koji ga sačinjavaju, dozvoljava lako donošenje kvalitativnih zaključaka o trenutnom stanju promenljivih od značaja i mogućim adaptacijama, u zavisnosti od količine dostupnih informacija. Kao konačan rezultat, dobija se procena parametara tzv. hibridnog sistema preporuke, uz postavljen okvir za predviđanje konačne ocene akcije korisnika nad traženim proizvodom.

Analiza ovakvog sistema je svakako pogodna za nadogradnju, pogotovo usled neizbežno smanjene efikasnosti statičkog modela nakon aktivacije priliva novih podataka i obučavanja parametara direktno na mreži [13]. Opšti prikaz rezultata primene ostalih Bajesovskih metoda za pronalaženje optimalne predikcije se može naći u radu [14], pri čemu bi autor kao posebno zanimljivu istakao analizu posvećenu društvenim mrežama [15].

**Tabela 2.** Baza ocena korisnika, **R**

| U | $i_1$ | $i_2$ | $i_3$ | $i_4$ | $i_5$ |
|---|---|---|---|---|---|
| $u_1$ | 2 | 2 | 0 | 1 | 0 |
| $u_2$ | 0 | 0 | 1 | 2 | 0 |
| $u_3$ | 2 | 2 | 0 | 0 | 0 |
| $u_4$ | 2 | 1 | 0 | 0 | 0 |
| $u_5$ | 0 | 0 | 0 | 0 | 2 |
| . | . | . | . | . | . |

## 2. Osnovni pojmovi

### 2.1. Notacija

Dva osnovna entiteta sistema preporuke su *korisnik* i *proizvod*. Problem preporuke se može objasniti na sledeći način: *Za svakog korisnika $u \in U$ želimo da odaberemo proizvod $i \in I$ koji maksimizira funkciju koristi preporuke korisniku, r, koja je najšešće predstavljena ocenom ili rangom (mada može biti bilo koja funkcija)*, to jest:

$$\forall u \in U, i'_u = argmax(r(u,i)), \quad i \in I. \tag{1}$$

Ulazna promenljiva sistema preporuke zavisi od algoritama filtriranja, i najčešće spada u jednu od tri kategorije: ocene, demografski podaci i tekstualni sadržaj. Radi lakše manipulacije i reprezentacije, u ovom radu se pretpostavlja da sve ulazne promenljive nakon filtriranja mogu biti predstavljene celobrojnom pozitivnom vrednošću. Neka je *n* broj korisnika, *U = {$u_1$, $u_2$, …, $u_n$}*, a *m* broj proizvoda koji se ocenjuju, *I = {$i_1$, $i_2$, …, $i_m$}*. Svaki korisnik $u_i$ je direktno povezan sa listom proizvoda $I_{u_i}$ o kojima je izneo svoje mišljenje, pri čemu $I_{u_i} \subseteq I$. Sve ocene se sakupljaju u matricu dimenzija *m* x *n*, označenu sa *R*. Izlaz sistema se, sa druge strane, prikazuje u vidu predikcije ili preporuke. Predikcija predstavlja interpretaciju konkretne numeričke vrednosti $r_{a,j}$ kojom se označava očekivano mišljenje aktivnog korisnika $u_a$ o proizvodu $i_j$, dok se preporuka najčešće interpretira kao uređena lista *M* proizvoda, $M \leqslant m$ koji će korisniku najverovatnije da se svide.

### 2.2. Preciznost predikcija i mere sličnosti

Svakako najvažnija osobina sistema preporuke je preciznost (tačnost) njegovih predikcija. Sistem, u zavisnosti od primene, uglavnom predviđa mišljenje korisnika (ocena), ili verovatnoću da će korisnik načiniti akciju nad proizvodom (akvizicija). Najčešća mera greške predikcije je srednja kvadratna greška (*RMSE*), i za predviđeno $\hat{r}_{u,i}$ iznosi:

$$RMSE = \sqrt{\frac{1}{\mid \tau \mid} \sum_{(u,i)\in\tau} (\hat{r}_{u,i} - r_{u,i})^2}. \tag{2}$$

Uobičajene alternative, ukoliko je cilj odbaciti elemente sa većom greškom nad određenim parametrima, su korišćenje srednje apsolutne greške:

$$MAE = \sqrt{\frac{1}{\mid \tau \mid} \sum_{(u,i)\in\tau} \mid \hat{r}_{u,i} - r_{u,i} \mid}, \tag{3}$$

kao i prosečne kvadratne greške:

$$ARMSE = \sqrt{\sum_{(u,i)\in\tau} w_i(\hat{r}_{u,i} - r_{u,i})^2}, \tag{4}$$

gde $w_i > 0$ predstavlja težinu značaja proizvoda *i*, pri čemu je $\Sigma w_i = 1$.

Da bi sistem imao brži odziv, ponekad se kao validna uzimaju samo poređenja korisnika i proizvoda čija je mera sličnosti iznad definisane granice. Jedna od mera koja se, pored dobro poznatog Euklidskog rastojanja, često pojavljuje u literaturi je *kosinusna sličnost*:

$$cos(r_u, r_v) = \frac{r_u^T r_v}{|| \, r_u \, ||| \, r_v \, ||}, \tag{5}$$

dok će u analizi u petom poglavlju ovog rada biti korišcen *Pirsonov koeficijent korelacije*:

$$PC(u,v) = \frac{\sum_{i \in I_{u,v}} (r_{ui} - \bar{r_u})(r_{vi} - \bar{r_v})}{\sqrt{\sum_{i \in I_{u,v}} (r_{ui} - \bar{r_u})^2 \sum_{i \in I_{u,v}} (r_{vi} - \bar{r_v})^2}} \tag{6}$$

kao količnik kovarijanse dve varijable i proizvoda njihovih standardnih devijacija, pri čemu sa $I_{u,v}$ označavamo skup proivoda koji su ocenjeni od strane oba korisnika čiju korelaciju merimo.

### 2.3. Priprema podataka i redukcija dimenzija

Podaci koji služe za stvaranje preporuka često greškom ili spletom okolnosti ispadaju iz očekivanih okvira, nisu usklađeni sa drugima, ili nisu dovoljno verodostojni. Zbog toga je pre njihove primene korisno uočiti ove anomalije, a uz to obratiti pažnju na dimenzije koje su jako oskudne validnim vrednostima. Osnovna intuitivna tehnika preprocesiranja podataka koji su rezultat pretrage („rudarenja", en. *Data Mining*) velikih baza je kombinovana metoda srednjih vrednosti. Ideja je da se kao polazna predikcija preporuke postavi srednja ocena korisnika $u_i$ nad proivodom $i_j$, a da se ona zatim koriguje prema specifičnostima drugih korisnika. Predikcija ocene na poziciji *i,j* je, dakle, data sa:

$$r_{i,j} = \begin{cases} \bar{r_i} + \dfrac{\sum_{p=1}^{n} \delta_p}{n}, & ako\, korisnik\, i\, nije\, ocenio\, proizvod\, j \\ r, & ako\, je\, korisnik\, i\, dodelio\, ocenu\, proizvodu\, j \end{cases}, \tag{7}$$

gde je sa $\overline{r_i}$ predstavljena srednja ocena korisnika *i*, a sa $\delta_n = r_{n,j} - \overline{r_n}$ njena korekcija.

Tokom razvoja algoritma, očekivano je suočavanje sa nepopunjenim matricama velikih dimenzija, i u takvoj postavci mere udaljenosti i sličnosti prestaju da budu validne. Zato se preporučuje primena tehnika za redukciju dimenzionalnosti, kao što je Analiza osnovnih komponenti (en. *Principal Component Analysis*), ortogonalna linearna transformacija u novi koordinantni sistem sa dimenzijama poređanim prema opadajućoj vrednosti varijanse. Iako se nakon toga dimenzionalnost sistema lako redukuje zanemarivanjem komponenti sa malim uticajem na ukupnu varijansu, pokazano je da ovaj pristup ne daje smislene rezultate ukoliko podaci ne prate Gausovu raspodelu. Kao alternativa, preporučuje se korišćenje principa Dekompozicije singularnih vrednosti (en. *Singular Value Decomposition*). Proizvoljnu matricu *A* je, naime, skoro uvek moguće predstaviti u obliku $A = U\lambda V^T$, pri čemu je $\lambda$ dijagonalna matrica pozitivnih singularnih vrednosti poređanih u opadajućem redosledu, $U$ matrica-kolona sopstvenih vektora $AA^T$, a $V$ matrica-kolona sopstvenih vektora $A^T A$.

Ukoliko prvi pogled na podatke nagoveštava moguću separabilnost prema parametru od značaja, korisno je primeniti neku metodu klasterizacije radi grupisanja tačaka koje su dovoljno udaljene jedne od drugih prema utvrđenoj meri sličnosti. Jedna od najpopularnijih metoda particionisanja podataka je algoritam *k-srednjih vrednosti*, dodeljivanje elementa jednom od disjunktnih podskupova definisanih centroidom $\lambda$. Funkcija koju je potrebno minimizovati u ovom procesu optimizacije predstavlja sumu rastojanja svih članova grupe (*klastera*) do odgovarajućeg centroida:

$$E = \sum_{1}^{k} \sum_{n \in S_j} d(x_n, \lambda_j), \tag{8}$$

gde *d* označava odabranu meru rastojanja.

## 3. Pristupi formiranju sistema preporuke

Kao što je navedeno u uvodnom poglavlju, osnovna podela sistema preporuke je na *sadržinske* i *sisteme saradnje*. Sadržinski sistemi (en. *content-based*) skladište informacije o objektima koji odlaze na preporuku, a zatim ih koriste za procenu njihovih zajedničkih osobina i međusobne sličnosti, kao i sličnosti sa preferencijama korisnika (koje se takođe izražavaju podskupom osobina iz domena proizvoda). Prednost ovih sistema je u tome što daju rezultate nezavisno od ponašanja ostalih korisnika, i omogućavaju preporučivanje objekata koji prethodno nisu bili ocenjeni. Analiza sadržaja je kod njih ipak jako ograničena, pa su zbog toga u stanju da preporuče

jedino objekte koji su slični postojećim, ali ne i ostale sa dobrom ocenom.

Sa druge strane, sistemi saradnje (en. *collaborative filtering*) povezuju profil aktuelnog korisnika sa njemu sličnim prethodnicima tako što mu predlažu njihove preporučene sadržaje koje on nije pogledao. Pretpostavka je da će korisnici sa sličnim ukusom u prošlosti pokazivati tu sličnost i ubuduće, i zbog toga ovi sistemi obuhvataju sve vrste proizoda iz korisničkog profila, bez obzira na njihovu međusobnu sličnost. Ipak, kod novih objekata i korisnika imamo problem „preskakanja", zbog nedovoljne količine informacija.

Osim ove podele, značajno je znati da li je odnos sistema prema korisniku aktivan ili pasivan, to jest da li su ulazni korisnički podaci preuzeti implicitno na osnovu nekih drugih akcija, ili popunjavanjem unapred određenih i ograničenih formi. Većina metoda koje se baziraju na kombinovanju navedenih pristupa se naziva hibridnim. Kombinacije mogu nastati uzimanjem srednje vrednosti izlaza različitih algoritama ili linearne kombinacije istih sa težinskim faktorima, zatim smenjivanjem preporučenih izlaza prema unapred utvrđenom redosledu dok se ne dobije rezultat zadovoljavajuće preciznosti. Formiranje sistema sa podacima datim u poglavlju 1.2. će upravo biti bazirano na principima kako sadržinskih tako i sistema saradnje, što rezultuje preciznijim preporukama za potencijalno potpuno nove i neocenjene proizvode.

## 4. O Bajesovim mrežama

Potreba za korišćenjem verovatnoća proizilazi iz nedostatka potpune slike o ishodu događaja, gde nam sa druge strane iskustvo dozvoljava da pretpostavimo neko očekivanje sa određenom dozom sigurnosti. Rešavanje problema sa kojima se čovek svakodnevno susreće zahteva izvođenje niza uzročno posledičnih veza nad skupom mogućih ishoda. Skup takvih veza se može vizuelizovati mrežom (grafom), pri čemu je svaka pojedinačna veza jednoznačno opisana težinom koja označava meru njenog uticaja na sistem.

*Bajesova mreža* predstavlja grafički model verovatnoća, prikazan usmerenim acikličnim grafom, koji određuje transformaciju združene raspodele verovatnoće $V = \{X_1, X_2, \ldots, X_n\}$ u skup lokalnih raspodela verovatnoće, pri čemu se svakoj promenljivoj dodeljuje tačno jedna raspodela. Njena struktura se definiše *čvorovima* i skupom usmerenih linija, tzv. *granama*. Čvor se u mreži obeležava imenom odgovarajuće promenljive upisanim u krug, dok grane prikazuju zavisnost između promenljivih, sa strelicom koja nam ukazuje na to koji čvor direktno utiče na koji. Struktura acikličnog grafa obezbeđuje da nijedan čvor ne može biti svoj prethodnik (ne postoji direktna putanja unazad do njega) ni sledbenik (direktnim kretanjem unapred nije moguće vratiti se u isti čvor).

Svaka promenljiva u Bajesovoj mreži, zajedno sa odgovarajućom raspodelom verovatnoće, zavisi isključivo od svojih nesledbenika u grafu, za definisano stanje svojih prethodnika. Ova *Markovljeva* osobina se koristi za veoma značajno redukovanje broja parametara kojima je opisana raspodela verovatnoće promenljive. Tako, za datu strukturu *S*, raspodela ima faktorizovan oblik:

$$p(x) = \prod_{i=1}^{n} p(X_i \mid pa_i), \tag{9}$$

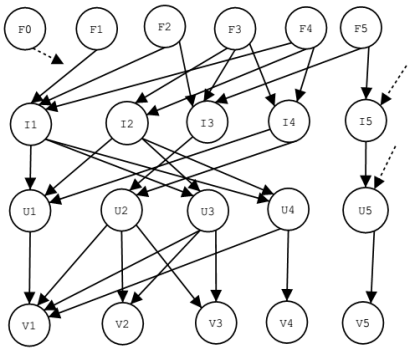
gde sa $pa_i$ obeležavamo skup prethodnika čvora $X_i$. Lokalna raspodela čvora koji nema prethodnike naziva se početnom (apriornom). Uz datu strukturu koja predstavlja kvalitativni opis sistema, potrebno je kvantitativno definisati i njegove parametre, upravo procenama navedenih lokalnih raspodela verovatnoće $p(X_i \mid pa_i)$. Iako je u ovom koraku, u slučaju potpuno povezanog grafa, potrebno definisati $2^n - 1$ vrednosti, struktura usmerenog acikličnog grafa dozvoljava potpun opis modela sa svega $n + 2m$ verovatnoća, po jednom početnom za svaki čvor i po dvema za svaku uzročno-posledičnu vezu, u slučajevima pozitivnog i negativnog ishoda.

Ukoliko je raspodela Bajesove mreže izvedena kao u (9), tj. ukoliko je jasno vidljiv pojedinačni doprinos čvorova, moguće je uraditi evaluaciju svih potencijalnih linija zaključivanja metodom marginalizacije. Potvrdu konkretnih rezultata možemo dobiti na dva načina: propagacijom unapred (*prediktivna podrška*), koja je zasnovana na analizi stanja u čvorovima prethodnicima, i propagacijom unazad (*dijagnostička podrška*). U velikom broju praktičnih postavki, Bajesova mreža je dobrim delom nepoznata i potrebno je obučiti je na osnovu poznatih podataka, te na osnovu iskustva i poznatih veza uraditi procenu topologije grafa. Obučavanje mreže je značajno teži zadatak od estimacije samih parametara, a najveću prepreku predstavlja parcijalna opservabilnost u slučaju skrivenih čvorova ili nepotpunih skupova podataka.

### 4.1. Kreiranje preporuke preko Bajesove mreže

Radi lakše interpretacije potpuno definisanog sistema predstavljenog u poglavlju 1.2., potrebno je za početak usvojiti skup linearnih relacija među opisanim izvorima informacija kao $F \to I \to U \to V$ , gde je prva relacija

opis objekta njegovim osobinama (u datoj mreži se manifestuje preko grana koje povezuju konkretnu osobinu iz predodređenog skupa i odgovarajući proizvod), druga baza ocena, a treća definiše vezu konačne predikcije aktivnom korisniku sa glasovima njemu sličnih korisnika. Veze između korisnika bi, logično, trebalo modelirati usmerenim granama u samoj mreži koje dobijamo kao rezultat algoritma obučavanja. Pošto su te veze u ovom slučaju uzajamno simetrične, a ciklične strukture nisu dozvoljenje u topologiji Bajesovih mreža, skup *V* se formira na osnovu udruženih glasova. Čvorovi *V = {*$V_1$*,* $V_2$*, . . . ,* $V_n$*}* će biti iskorišćeni za estimaciju raspodela verovatnoće glasanja, pa će vrednosti koje uzimaju biti iz istog domena kao i *U*, $\{0, 1, \ldots, \#r\}$.




**Slika 2.** Topologija Bajesove mreže sistema preporuke

Skup prethodnika čvora $V_a$ , $P_a(V_a)$ se može odrediti analizom baze glasova R. Skup će sadržati korisničku promenljivu $U_b \in U$, ukoliko je zaključeno da zadovoljava graničnu meru sličnosti sa $U_a$ i njenim dodavanjem ne prelazimo maksimalnu dozvoljenu veličinu skupa. Preporučeno je korišćenje modifikacije Pirsonovog korelacionog koeficijenta iz (6):

$$sim(U_a, U_b) = \frac{\sum\limits_j (r_{a,j} - \bar{r_a})(r_{b,j} - \bar{r_b})}{\sqrt{\sum\limits_j (r_{a,j} - \bar{r_a})^2 \sum\limits_j (r_{b,j} - \bar{r_b})^2}}, \tag{10}$$

gde $\bar{r_a}$ predstavlja srednju ocenu korisnika $U_a$

$$\bar{r_a} = \frac{1}{\mid Pa(U_a) \mid} \sum_{I_k \in Pa(U_a)} r_{a,k}, \tag{11}$$

uz napomenu da se sumiranje po *j* vrši samo nad objektima za koje su oba korisnika *a* i *b* glasali, tj. za elemente skupa $P_a(U_a) \cap P_a(U_b)$ date mreže.

## 5. Estimacija parametara Bajesove mreže

Nakon formiranja strukture Bajesove mreže, potrebno je pristupiti podacima iz datih tabela radi procene parametara, uslovnih verovatnoća od značaja za dati model. I dok je za prvi nivo to lako uraditi, i to prostim linearnim sumiranjem nad granama odgovarajućeg čvora, dalje se račun komplikuje. Radi efikasnosti, preporučuje se upotreba kanoničkog modela verovatnoća. Za dati čvor $X_i$ u stanju *j* i sa konfiguracijom prethodnika $pa(X_i)$, verovatnoća se predstavlja sumom:

$$Pr(x_{i,j} \mid pa(X_i)) = \sum_{Y_k \in Pa(X_i)} w(y_{k,l}, x_{i,j}), \tag{12}$$

gde je $y_{k,l}$ vrednost koju promenljiva $Y_k$ uzima iz konfiguracije $pa(X_i)$, a $w(y_{k,l}, x_{i,j})$ težinski faktor koji određuje kako *l*-ta vrednost $Y_k$ opisuje dato stanje čvora. Pri tome, moraju biti zadovoljeni uslovi normalizacije:

$$w(y_{k,.}, x_{i,j}) \geqslant 0, \quad \sum_{Y_k \in Pa(X_i)} w(y_{k,.}, x_{i,j}) \leqslant 1. \tag{13}$$

Sve težine se mogu proceniti na osnovu vrednosti iz postojećih tabela sledećim algoritmom:
**1)** Za svaku osobinu $F_k$, kao čvora koji nema svoje prethodnike, $Pr(f_{k,1}) = n_k/m$, što predstavlja količnik broja pojava osobine u skupu i ukupnog broja proizvoda, pri čemu je $Pr(f_{k,0}) = 1 - Pr(f_{k,1})$. Iz tabele 1 na ovaj način možemo da izračunamo i $Pr(f_{2,1}) = 0.2$, $Pr(f_{5,1}) = 0.4, \ldots$
**2)** Za čvor proizvoda $I_j$, obzirom da se radi o reprezentaciji promenljive sa dve vrednosti, potrebno je definisati samo težinske faktore za $Pr(i_{j,1} \mid pa(I_j))$ zbog toga što važi jednakost $Pr(i_{j,0} \mid pa(I_j)) = 1 - Pr(i_{j,1} \mid pa(I_j))$, i to kao:

$$\begin{aligned} w(f_{k,1}, i_{j,1}) &= \frac{log((m/n_k)+1)}{log(m+1)M(I_j)} \quad , \\ w(f_{k,0}, i_{j,1}) &= 0 \end{aligned} \tag{14}$$

pri čemu $M(I_j) = \sum_{F_k \in P_a(I_j)} log((m/n_k)+1)/log(m+1)$ predstavlja normalizacioni faktor za dati čvor. Pomenuti izraz $log((m/n_k)+1)/log(m+1)$ služi za određivanje mere važnosti osobine proizvoda za ceo skup. Što vise ovakvih osobina postoji u skupu, veća je verovatnoća da se i sam proizvod $I_j$ često preporučuje. Gledajući tabelu 2, neke od vrednosti su $M(I_j) = 2.447$, $w = (f_{1,1}, i_{1,1}) = 0.445$, $w = (f_{2,0}, i_{1,1}) = 0$, $w = (f_{4,1}, i_{1,1}) = 0.243$, $Pr(i_{1,1} \mid f_{1,1}, f_{2,0}, f_{4,1}) = 0.706$.
**3)** Kod čvorova *U* meri se uticaj proizvoda $I_k$ na šemu glasova određenog korisnika $U_a$. Ako je poznato da je korisnik dao ocenu $r_{a,k} = s$, i u zavisnosti od toga da li konfiguracija prethodnika proizvoda $pa(U_a)$ ukazuje na interesovanje korisnika ka objektu ili ne, težinski faktori su određeni sa:

$$\begin{aligned} &w(i_{k,1}, u_{a,s}) = 1/ \mid Pa(U_a) \mid, \\ &w(i_{k,1}, u_{a,t}) = 0, t \neq s, 0 \leqslant t \leqslant \#r, \\ &w(i_{k,0}, u_{a,0}) = 1/ \mid Pa(U_a) \mid, \\ &w(i_{k,0}, u_{a,t}) = 0, 1 \leqslant t \leqslant \#r. \end{aligned} \tag{15}$$

gde je $\mid pa(U_a) \mid$ ukupan broj glasova koje je korisnik $U_a$ dodelio. Za $pa(U_1) = \{i_{1,1}, i_{2,0}, i_{4,0}\}$, gledajući ponovo podatke iz tabele 2, vrednosti raspodela su $Pr(u_{1,k} \mid pa(U_1)) = 0 + 1/3 + 0$, $Pr(u_{1,1} \mid pa(U_1)) = 0 + 0 + 1/3$, $Pr(u_{1,2} \mid pa(U_1)) = 1/3 + 0 + 0$.
**4)** Na kraju je moguće doći do konačne procene ocene $V_a$ koja se dodeljuje korisniku, odvojenom analizom težinskih faktora kojima korisnik $U_a$ doprinosi oceni od onih koji su uzeti od sličnih korisnika $U_b \in pa(V_a)$, pa važi da je:

$$\begin{aligned} &w(u_{a,s}, v_{a,s}) = \alpha \\ &w(u_{a,s}, v_{a,s}) = 0, t \neq s, 0 \leqslant t \leqslant \#r. \end{aligned} \tag{16}$$

Za $0 \leqslant t, s \leqslant \#r$:

$$w(u_{b,t}, v_{a,s}) = \frac{1-\alpha}{\mid Pa(V_a) \mid -1} \frac{N(u_{b,t}, v_{a,s}) + \beta q_s}{N(u_{b,t}) + \beta}, \tag{17}$$

gde $0 \leqslant \alpha \leqslant 1$, na način takav da je za veće $\alpha$ veća i težina koja se dodeljuje prethodnim glasovima samog korisnika. Vrednost $N(u_{b,t}, v_{a,s})$ označava broj proizvoda iz skupa $Pa(U_a) \bigcup Pa(U_b)$ koji su od korisnika *A* i *B* dobili ocene *s* i *t*, respektivno. $\beta$ i $q_s$ su parametri početne raspodele ocena, a težine $w(u_{b,t}, v_{a,s})$ su proporcionalne maksimalnom posteriornom (MAP) estimatoru $Pr(v_{a,s}, u_{b,t})$. Na primer, za parametre $\alpha = 0.5, \beta = 1, q = 1/3$ i podatke iz tabele 2, uz konfiguraciju $pa(V_1) = \{u_{1,2}, u_{2,0}, u_{3,1}, u_{4,1}\}$, raspodele verovatnoća za $V_1$ iznose:

$$\begin{aligned} &Pr(v_{1,0} \mid Pa(V_1)) = Pr(v_{1,1} \mid pa(V_1)) = 0 + 0.02 + 0.06 + 0.03 = 0.11, \\ &Pr(v_{1,2} \mid Pa(V_1)) = 0.5 + 0.12 + 0.05 + 0.11 = 0.78. \end{aligned} \tag{18}$$

U cilju dobijanja konačne predikcije korisnosti nepregledanog sadržaja korisniku $U_a$, potrebno je za $0 \leqslant s \leqslant \#r$ odrediti verovatnoću:

$$Pr(v_{a,s} \mid ev) = \sum_{F,I,U} Pr(v_{a,s}, F_l, I_k, U_j \mid ev). \tag{19}$$

Pošto se u datom primeru svaki čvor formira nezavisno od prethodnog sloja, pri čemu su prednosti kanoničkih modela za raspodele uslovnih verovatnoća mreže uspešno primenjeni, konačne verovatnoće se najefikasnije mogu dobiti algoritmom propagacije nadole. Bez detaljnijeg izvođenja, ova verovatnoća je određena sledećom sumom:

$$Pr(x_{a,s} \mid ev) = \sum_{j=1}^{m_{X_s}} \sum_{k=1}^{l_{Y_j}} w(y_{j,k}, x_{a,s}) Pr(y_{j,k} \mid ev). \tag{20}$$

## 6. Zaključak

Kreiranje efikasnog sistema preporuke za određenu oblast zaista nije lak zadatak. Čak i kada je u potpunosti baziran na povratnim informacija od korisnika, ne postoji garancija da će se konačni rezultat obrade korisniku zaista svideti. Ovaj rad čitaocu daje postupan prikaz jednog vizuleno-matematičkog pristupa rešavanju opšte grupe problema stvaranja preporuke Bajesovom mrežom, uz primenu odgovarajućih mera validacije i pripreme podataka koji figuriraju kao ulazne promenljive sistema. Izvedene formule za procene težina afiniteta pojedinih aktivnih korisnika vode ka boljem upoznavanju neocenjenih proizvoda i postavljaju okvir za adaptaciju sistema potpuno novim i neispitanim korisnicima i proizvodima. Kao moguće nadogradnje, rezultate iz ovog rada je potrebno testirati na odabranim specijalizovanim oblastima u kojima preporuke imaju vidljiv komercijalni potencijal, a zatim i postaviti okvir samoobučavajućeg adaptivnog mehanizma koji brzo reaguje na priliv velike količine novih informacija preko interneta.

**Zahvalnica.** *Motivacija za pisanje rada je potekla iz istraživačkog projekta sa prve godine doktorskih studija modula Primenjene matematike na Elektrotehničkom fakultetu u Beogradu, u saradnji sa profesorom Željkom Đurovićem, kome se zahvaljujem na stručnoj pomoći i korisnim smernicama. Veliku zahvalnost na podršci i saradnji u prethodne dve uspešne godine doktorskih studija dugujem i svom mentoru, profesoru Nenadu Cakiću, kao i profesoru Zoranu Popoviću koji je pročitao rad pre publikovanja i ukazao mi na njegove nedostatke. Ovim putem bih se zahvalio i roditeljima Mariji i Nenadu, bratu Aleksandru, bliskim prijateljima i kolegama na ohrabrenju za nastavak studija. Na kraju, izražavam neizmernu zahvalnost supruzi i koleginici Milici na brojnim pregledima ovog rada, stručnim sugestijama, kao i na razumevanju zbog vremena provedenog na izradi istog.*

## Bibliografija

[1] **X. Amatriain.** Collaborative Filtering and other approaches. *Machine Learning Summer School*, 2014.
[2] **E. Vozalis, K. G. Margaritis.** Analysis of recommender systems algorithms. *The 6th Hellenic European Conference on Computer Mathematics and its Applications.*, 2003.
[3] **G. Adomavicius, A. Tuzhilin.** Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering.*, 2005, pp. 734-739.
[4] **X. Amatriain, A.Jaimes, N.Oliver, J.M.Puyol.** Data mining methods for recommender systems. *Recommender Systems Handbook, Springer US*, 2011, pp. 39-71.
[5] **X. Yang, Y. Guo, Y. Liu.** Bayesian-Inference-Based Recommendation in Online Social Networks. *IEEE transactions on Parallel and Distributed Systems vol. 24 no. 4*, April 2013, pp. 642-651.
[6] **J. Leskovec.** How users evaluate each other in social media. *Stanford University*, 2014.
[7] **B. Martin.** Collaborative filtering: A machine learning perspective. *Diss. University of Toronto*, 2004.
[8] **L.M. de Campos, J. M. FernÃąndez-Luna, J. F. Huete.** A Bayesian network approach to hybrid recommending systems. *Eleventh International Conference of Information Processing*, 2006.
[9] **J. Xu, K. Johnson-Wahrmann, S. Li.** The development, status and trends of recommender systems a comprehensive and critical literature review. *Mathematics and Computers in Science and Industry* , 2014., pp. 117-122.
[10] **D.H. Park, H.K. Kim, I.Y.Choi, J.K.Kim.** A Review and Classification of Recommender Systems Research. *2011 International Conference on Social Science and Humanity IPEDR vol.5*, 2011., pp. 290-294.
[11] **L. Kidzinski.** Statistical foundations of recommender systems. *Diss. Master Thesis submitted in Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw*, 2011.
[12] **A. Das, C. Mathieu, and D. Ricketts.** Maximizing profit using recommender systems. *arXiv preprint arXiv 0908.3633* , 2009.
[13] **S-Z. Zhang, L. Liu, Y-Z.Dong.** An Online Personalized Recommendation Model Based on Bayesian Networks. *Research and Practical Issues of Enterprise Information Systems II. Springer US*, 2008., pp. 1575-1584.
[14] **S.Guo.** Bayesian Recommender Systems, Models and Algorithms. *Australian Narional University*, 2011.
[15] **M. Gartrell, U. Paquet, R. Herbrich.** A bayesian treatment of social links in recommender systems. *CU Technical Report CU-CS-1092-12*, 2012.

Симпозијум МАТЕМАТИКА И ПРИМЕНЕ, Математички факултет, Универзитет у Београду, 2015 ,Vol. VI(1)

# Најчешће грешке при статистичкој анализи у истраживањима

**Марија Минић**

*Пољопривредни факултет, Универзитет у Београду*
*e-mail: minic.m.marija@gmail.com*

**Зоран Видовић**

*Учитељски факултет, Универзитет у Београду*
*e-mail: zoran.vidovic@uf.bg.ac.rs*

**Апстракт.** Статистика представља значајан део модерних научних истраживања из различитих области. Међутим, често долази до њене погрешне употребе, било због незнања или из намере. На овај начин добијају се погрешни резултати, стварају се лажна истраживања и троше драгоцени ресурси. Уколико дође до објаве рада који садржи грешку у статистичкој анализи, негативне последице могу осетити и истраживач, чији се кредибилитет доводи у питање, и читаоци који те информације на даље користе и употребљавају. Поред грешака које се јављају као последица незнања или непажње истраживача, постоје и ,,намерне грешке". Истраживач може да злоупотреби статистику тако да добије резултате које жели. У раду су приказани примери грешака статистичке анализе у публикованим радовима и примери злоупотребе статистике.

**Кључне речи:** статистика; грешка; злоупотреба.

## 1. Увод

Статистика[1] је млада наука, али која је још у време Старе Грчке имала своју неприметну и наизглед подразумевану примену, која тек 1933. добија своје признање као посебна научна дисциплина математике. Првобитно је статистика имала улогу у прикупљању података који су били од велике важности за државу (попис становништва, војне опреме, попис умрлих од куге, итд.), док данас њена улога и примена се манифестују у релативно свим наукама где је неопходно извршити закључивање о читавој популацији на основу узорка из популације.

Данас ретко која научна област не примењује статистику. Упоредо са достигнућима и развојем информационих технологија, развијала се и статистика, првенствено због повећане брзине прикупљања података и детаљније и поузданије анализе. Постоје многобројни софтвери који као примарну улогу имају статистичку обраду података, као што су R, SPSS и SAS, али и софтвери чија примарна улога није статистичка анализа, али који имају уграђене пакете за статистичку обраду података као што су Microsoft Excel, MatLab и многи други. Услед појаве ,,кориснички настројених" статистичких софтвера, где корисници без познавања елементарних особина, чињеница и методологије статистике, могу да изведу одређене закључке базиране на узорку из популације, учесталија је појава грешака при статистичкој анализи.

Грешке при статистичкој анализи нису ретка појава. У [3] могу се пронаћи двадесет најчешћих уочених грешака при статистичкој анализи у биомедицинским истраживањима. Према [1], наглашава се забринутост да велики број публикованих радова садрже бар једну грешку статистичке природе. Сматра се да вероватноћа да резултат статистичког закључивања буде коректан расте уколико постоји неколико индиција о коректности самог истраживања. Пожељно је да постоје истраживања на исту или сличну тему, специјална контрола примене статистичких тестова над опсервацијама, коректна и прецизна презентација дескриптивних статистика, са главним циљем реализације валидних статистичких закључака.

Уколико се, пак, објави рад који садржи неку грешку, последице могу да буду енормне и ненадокнадиве. Истраживач губи свој кредибилитет, док целокупно истраживање, сви

[1] Добила је назив од италијанске речи *stato*, што значи држава.

њени резултати и закључци, се проглашавају невалидним. Овај рад се састоји из неколико секција. У другој секцији су приказане најчешће грешке статистичке природе које су представљене у радовима [1, 10]. У секцији Материјали и резултати приказане су грешке уочене од стране самих аутора из скупа часописа над којим је вршено истраживање.

## 2. Најчешће грешке

Не постоји подручје истраживања где није могуће начинити грешку при статитичкој анализи. Грешке се могу јавити у различитим фазама истраживања: при дизајну истраживачке студије, анализи података, при презентацији као и при интерпретацији резултата. Предлаже се присуство и консултација са статистичарем у свим фазама истраживања као предуслов коректне и поуздане статистичке анализе, према [1]. Међутим, истраживачи, да би добили резултате који одговарају њиховим очекивањима, лако могу да ,,наместе" податке. Тако настају ,,намерне" грешке. Да би се спречило такво недолично поступање предлаже се да, при слању рада у часопис, уједно пошаљу и оригиналне податке, што најчешће није случај , из нпр. [7, 4]. Поред представљања статистичких метода у раду, неопходно је омогућити читаоцу оригиналне податке над којима је извршено статистичко закључивање све у циљу поновног потврђивања резултата.

Најчешће грешке које се јављају при самом дизајну студије јесте да циљеви истраживања нису јасно дефинисани, а самим тим и иницијалне статистичке хипотезе нису јасно дефинисане. Научници често погрешно одаберу контролну групу (групу испитаника у експерименталном истараживању која је лишена утицаја експерименталне величине и која има улогу контроле утицаја величине која се надгледа) неувиђајући да нарушавају целокупну коректност истраживања.

Прост случајан узорак представља основу за коректну и поуздану примену многобројних статистичких метода, стога ако се у истраживању наруши ова претпоставка статистичко закључивање постаје неосновано, а самим тим и неупотребљиво. Непристрасност је особина која представља једнаку вероватноћу одабира сваког појединачног члана популације у узорку. Често научници не разликују пристрасан и непристрасан узорак. Специјална област статистике која се бави одабиром узорка у истраживању зове се теорија узорка [8]. Узорак, пре свега, мора да репрезентује читаву популацију, те се такав узорак зове репрезентативан узорак. Уколико то није случај, кажемо да је узорак нерепрезентативан. Истраживачи често нису свесни да погрешно бирају узорак, наглашавајући да су изабрали узорак погодан за њихово истраживање, што је, у најмању руку, двосмислено.

Приликом анализе података у истраживањима неретко се примети неразумевање према одабиру статистичког теста, из [1, 4]. Научници без тестирања претпостављене нормалности, графички или применом директних формалних тестова нормалности, претпостављају да њихови подаци потичу из популације са нормалном расподелом. Самим тим, интервали поверења за вредности параметара обележја популације нису тачни, стога директна последица је невалидност свих генерисаних статистичких резултата. Сваки тест има одређене претпоставке које морају да буду задовољене да би методологија теста била важећа, стога те претпоставке морају бити истакнуте и потврђене. Уколико претпоставке методологије коју тест користи нису испуњене онда нема смисла интерпретирати резултате теста. Статистички тест се бира у зависности од врсте варијабли, нормалности обележја популације, независности и обима узорка и других претпоставки.

Поређењем узорака из различитих популација услед непотпуних и неиспуњених претпоставки наилазимо на проблем нагомилавања грешке прве врсте, што има неоспоран утицај на коректност статистичког закључивања, на основу [1].

Често се деси да аутори потежу за тестовима који нису често примењивани у пракси, а да не истакну њихове особине и претпоставке под којима врше статистичку анализу. Такође, непотпуне и неодређене вредности се лако игноришу или замењују вредностима које сматрају прикладним, без претходног објашњења, према [3].

Подаци могу да поседују квантитативна или квалитативна својства. Некада се квантитативне величине могу посматрати и као квалитативне, и обрнуто, док постоји извесна једнозначна кореспонденција између њих. Међутим, неопходно је јасно назначити који су

критеријуми трансформације података, иначе веома лако може да дође до грешки које имају утицај на даљи напредак истраживања [3].

Недостајуће вредности су велики проблем у истраживачком раду. Из најразличитијих разлога број опсервација није исти на почетку и на крају истраживачке студије. Такође се неретко дешава да одређени подаци буду проглашени аутлајерима и буду изостављени из студије без коментара. Све такве случајеве треба уредно документовати због предострожности и нивоа квалитета статистичке анализе.

Уколико постоји параметарска и непараметарска верзија неког статистичког теста, неопходно је нагласити која верзија је коришћена и које су иницијалне претпоставке теста.

Презентација представља веома битан део истраживања. Циљ презентације је да се учесници упознају са самим истраживањем, његовим резултатима и циљевима.

Најпознатије дескриптивне статистике су аритметичка средина и стандардно одступање. Међутим, ове мере дају адеквантне и довољне информације о расподели само ако је у питању нормална расподела. Код нормалне расподеле, на основу $3\sigma$ правила, приближно 68% популације се налази на удаљености од једне стандардне девијације од аритметичке средине, око 95% на две стандардне девијације и око 99% на три стандардне девијације. Правило $3\sigma$ не важи код популација које немају нормалну расподелу. Зато, уколико се користе аритметичка средина и стандардна девијација као једине дескриптивне сатистике функције расподеле узорка, треба проверити да ли расподела обележја задовољава услове нормалности. Како, у општем случају, подаци често потичу из популације са расподелом обележја помереном улево (удесно), показатељи као што су медијана, интервал варијације и интерквартилна разлика често представљају бољи избор за добијање информација о расподели обележја. У таквим ситуацијама, у реду је приказати и аритметичку средину, али без стандардне девијације. Аритметичка средина није добар показатељ и у случају јако малог узорка. Још један проблем са стандардном девијацијом може да настане уколико се замени са стандардном грешком средње вредности. Како је стандардна грешка средње вредности увек мања од стандардне девијације, неки истраживачи је наводе у циљу приказивања мање дисперзије. Међутим, стандардна грешка није дескриптивна статистика.

Што се тиче $p$ вредности, боље је написати која је била тачна $p$ вредност, него у раду представити ознаке $p > 0,05$ или $p < 0,05$. Уколико је, на пример, $p$ вредност била $0,049$, та информација је много садржајнија него $p < 0,05$. Изузетак је уколико је $p$ вредност јако мала, тада само треба написати да је мања од неке мале вредности ($p < 0,001$). Није довољно приказати само $p$ вредност. Пожељно је написати и број параметара слободе, 95% интервал поверења као и вредност тест статистике теста који се примењује ([6]).

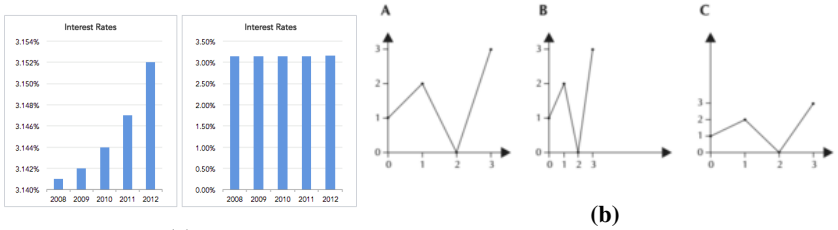
Нема потребе приказивати податке са свим децималама које су познате. Према [3], већина људи интуитивно најбоље разуме бројеве са једном до две децимале. Стога, набрајање даљих децимала може само да збуни читаоца. Ово нарочито важи уколико су бројеви релативно велики у односу на број децимала. Такође, при рачунарским симулацијама, број децимала не сме бити већи од броја итерација.

Многи људи лакше разумеју уколико су подаци графички приказани него уколико су дати табеларно или ако су текстуално објашњени. Колико графички приказ може бити користан, толико може имати и потпуно супротан ефекат ако се неправилно користи.

Један од најлакших начина да се лажно представе подаци је мењање оса координатног система. Углавном, $y$-оса је обележена од нуле до највеће вредности међу подацима. Некада се $y$-оса скраћује, тако да не почиње од нуле, у циљу истицања разлика међу подацима. У екстремним случајевима ова метода може да максимизира или минимизира разлике међу подацима, што доводи до неадекватног разумевања разлика између података (Пример 1б). Графици $A$, $B$ и $C$ приказују исте податке. Уколико, на пример, $x$-оса представља време, изгледа као да се промена брже десила на графику $B$, него на графицима $A$ и $C$. Сабијање $y$-осе чини да су се промене нагло десиле. Пожељно је да обе осе имају интервале једнаке дужине.

Такође, графици могу да презентују исте податке, али са различитим вредностима означеним на $y$-оси (Пример 1а). На $y$-оси се налазе вредности $3,140$% до $3,154$%, док на десном $y$-оса почиње нулом. На овај начин добија се утисак, на левој слици, да су разлике у ка-

матама огромне. Чини се да камата у 2012. години вишеструко већа него камата у 2008. години, иако то није случај. Уколико истраживач жели да скрати осу, неопходно је да то на графику и нагласи.




(a) (b)

**Пример 1.** Пример неисправне презентације дескриптивних статистичких алата[2].

Корелација представља међусобну повезаност између две величине које су представљене вредностима двеју промењивих. То заправо значи да је вредност једне промењиве могуће са одређеном вероватноћом предвидети на основу вредности друге промењиве. При том, та веза може бити позитивна или негативна. Примери корелисаности су количина падавина и принос усева, унос слане хране и висина крвног притиска и друго. Међутим, корелација веома лако може да се интерпретира погрешно, увиђајући везе између неповезаних величина, и самим тим лако долази до статистичких грешки. Детаљније о овом питању може да се нађе у [9].

## 3. Материјали и резултати

Према категоризацији домаћих научних часописа за друштвено-хуманистичке науке према листи Министарства просвете која је објављена 2013. године, последња доступна издања на интернету часописа из групе Психологија, педагогија, андрагогија и специјално васпитање представљају популацију над којом је вршено испитивање фреквентности грешака насталих приликом статистичке анализе. Узети су у обзир само часописи који имају доступна потпуна последња издања на интернету, издата закључно до октобра 2015. године. Радове који су садржали било какав вид примењене статистике су прегледани од стране самих аутора.

Од укупно 22 часописа, само 12 часописа имају доступна последња издања на интернету. Прегледано је 125 радова, од којих 39 радова садржи статистичке резултате (32.2%).

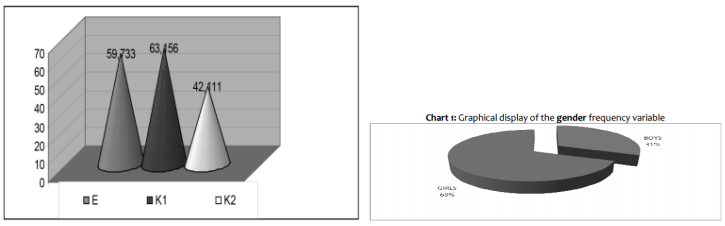
Грешке које су уочене: 1) неисправана презентација $p$ вредости (10); 2) неисправна презентација дескриптивних статистика (4); 3) неисправна дескриптивна анализа (1).

Истоветно као у [6], ниједан рад не садржи анализу моћи, тј. анализу неопходности величине узорка у циљу адекватне поузданости стастистичких резултата.

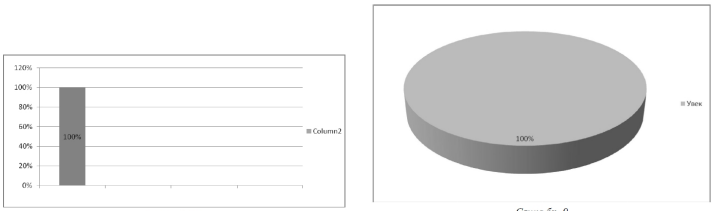
Приказаћемо неке примере статистичких грешака које су уочене.

Често аутори користе тродимензионалне дескриптивне статистике где је довољно њихово дводимензионално представљање. Тим поступком аутори подстичу неразумевање читаоца који нису упућени са датим радом и његовим резултатима истраживања. Као примере наводимо тродимензионалне хистограме и ,,пите" где трећа димензија не ,,носи" никакву информацију (Пример 2). Такође, у Примеру 3 се види да коришћење ових графика нема смисла, с обзиром да не пружају никакву информацију.

У току истраживања је неопходно праћење обима узорка због адекватне статистичке анализе. Уколико постоји разлика у првобитном обиму узорка на почетку истраживања и обиму узорка сагледаног на крају статитичке анализе неопходно је нагласити да је реч о евентулним избацивањима чланова узорка из анализе, због њиховог проглашавања као аутлајерима или слично (Пример 4).




**Пример 2.** Пример погрешног графичког приказа.




**Пример 3.** Примери непотребне употребе дескриптивних статистика.

*Abstract.* The purpose of the present study was to examine the influence of the "Competence-based didactic units" in vocational education on students' motivation and self-regulated learning. The sample consisted of 115 males and 133 females (n=250) who were attending secondary vocational or technical schools in Slovenia.

**Uzorak i postupak**

Istraživanje je sprovedeno na uzorku od 599 pripadnika opšte populacije (348 ženskog pola), uzrasta od 16 do 72 godine. Prosečna starost ispitanika je 33 godine (*SD* = 12.77). Usled neujednačenosti uzorka u odnosu na stepen obrazovanja, ispitanici su kategorisani u dve grupe. Prva grupa sačinjena je od 208 ispitanika sa 12 ili manje godina školovanja, dok u drugu grupu spada 390 ispitanika sa više od 12 godina školovanja. Među ispitanicima iz prve grupe najobrazovaniji imaju završenu srednju školu, a drugu grupu čine studenti, kao i ispitanici koji imaju završeno više ili visoko obrazovanje, master i doktorske diplome.

**Пример 4.** Грешке у калкулацији обиму узорка.

Најчешћа грешка приликом статистичке анализе у објављеним радовима из посматране групе часописа представља погрешно приказивање $p$ вредности одговарајућег статистичког теста (Пример 5). У другој секцији овог рада је објашњено како на исправан начин презентовати $p$ вредност датог теста.

*Узорак и поступак*

Узорак је чинило 643 ученика средњих школа из урбане средине (61,7% мушког пола) при чему је 33% похађало други разред, 31,1% трећи разред и 35,9% је похађало четврти разред (нису забележене полне разлике у односу на разред, $\chi^2(2,N=643)=0,04$, $p>,05$). Узорак је обухватио три гимназије и пет средњих стручних школа (економска, машинска, елект-

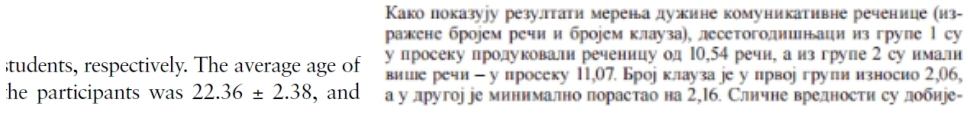
**Пример 5.** Пример погрешног приказивања $p$ вредности теста.

Приликом сваког истраживања мора да се води рачуна о екстремним вредностима које могу да се појаве. Уколико се одлучи да се оне не узимају у разматрање неопходно је нагласити разлог томе и истакнути његово адекватно обрзложење. Наредни пример показује неадекватно образложење избацивања аутлајера из даље анализе (Пример 6).

vrednosti dobijene na zadacima za vežbu. U narednom koraku analiza distribucije ukazala je na jedan mali broj ekstremnih vrednosti tj. veoma dugih vremena reakcije, većih od 2.5 standardne devijacije. Uobičajeno za ovu vrstu istraživanja, one su takođe uklonjene iz dalje analize. Situacije gde su ispitanici dali pogrešne

**Пример 6.** Пример одбацивања аутлајера из узорка без адекватног објашњења. Фраза ,,уобичајено".

Непотребна прецизност је честа појава у публикованим радовима. Сматра се да рад добија на значајности уколико се подаци представљају са великом прецизношћу. Међутим, то може само да оптерети читаоца (Пример 7).




**Пример 7.** Пример оптерећивања читаоца непотребним информацијама.

Недовољно подржан статистистички закључак може да настане уколико се сматра да одређене нумеричке вредности дескриптивних статистика једнозначно одређују функцију расподеле, што није случај (Пример 8).

На основу вредности *скјуниса* и *куртозиса*, које су у распону од 2 до -2, може да се тврди да је расподела нормална, односно није ни спљоштена ни померана ка једном крају.

**Пример 8.** Пример извођења закључака на основу непотпуних података.

Није ретко да се уоче недовољно јасни статистички закључци, где читалац нема адекватну и коректну информацију о предмету истраживања и статистичког закључивања (Пример 9).

REZULTATI ISTRAŽIVANJA SA DISKUSIJOM

Analiza rezultata procene crtanja prema modelu i reprezentacione dimenzije crteža prikazana je u Tabeli 2.

*Tabela 2 - Rezultati procene crtanja kod dece mlađeg školskog uzrasta*

| SUBTEST | Min | Max | AS | SD | Var. | r | p |
|---|---|---|---|---|---|---|---|
| Crtanje oblika | 0 | 20 | 13.18 | 4.173 | 17.418 | 0.320 | 0.000 |
| Crtanje | 3 | 20 | 14.63 | 2.843 | 8.085 | | |

Postignuća u oblasti crtanja, posmatrano iz ugla raspona i srednjih vrednosti rezultata, kod naših ispitanika su nešto bolja (ali ne i statistički značajno – p>0.05) od postignuća na zadacima precrtavanja oblika, gde se zapaža i izraženija disperzija rezultata (detaljnije u Tabeli 2). Utvrđena je značajna korelacija između posmatranih parametara (r=0.320, p<0.000).

**Пример 9.** Пример представљања недовољно јасних закључака.

## 4. Закључак

По броју уочених грешака од стране аутора можемо потврдити да приликом процеса публикације радова изразита пажња се поклања испитивању коректности статистичких резултата из истраживања. Сматрамо да је то показатељ повећања степена залагања за тачност, прецизност и коректност садржаја публикованих радова.

Овај рад не представља систематизацију грешака статистичке природе које се могу јавити у публикованим радовима из дате класе часописа. Такође, претходне наведене грешке су уочене од стране аутора, што не говори у прилог томе да су сви могући случајеви грешака насталих приликом статистичке анализе обухваћени.

Главни циљ овог рада је истицање неких основних статистичких грешака које се појављују у садржају ових радова као додатна упуства при припреми радова за публикацију. Свака сугестија која допринси побољшању етичности, коректности и професионализму се сматра пожељном, те оправдава циљ и идеју овог рада.

## Библиографија


[1] **J.P.A. Ioannidis.** Why most published research findings are false. *Essay-Freely avaible online*, 2005, 2(8), e124.

[2] **H.W. Walker.** Degrees of Freedom. *Journal of Educational Psychology*,1940, 31(4), 253–269.

[3] **T. Lang.** Twenty statistical errors even you can find in Biomedical research articles. *Croatian Medical Journal*, 2004, 45(4), 361–370.

[4] **S.J. White.** Statistical errors in papers in the British Journal of Psychiatry. *The British Journal of Psychiatry*, 1979, 135(4), 336–342.

[5] **E.C. Carter.** A standard error: Distinguishing standard error. *Diabetes*, 2013, 62, e15.

[6] **A. Simundic, U. Nikolac** Statistical errors in manuscripts submitted to Biochemia Medica journal. *Biochemia Medica*, 2009.

[7] **U. Simonsohn.** Just post it:The lessons from two cases of fabricated data detected by statistics alone. *The Wharton School, University of Pennsylvania*, 2013.

[8] **Lj. Petrović.** Teorija uzoraka i planiranje eksperimenata. *Ekonomski fakultet, Beograd*,2000.

[9] **Y.H. Kuo.** Extrapolation of correlation between 2 variables in 4 general medical journals. *JAMA*, 2002; 287, 2815-7.

[10] **T.H. Holmes.** Ten categories of statistical errors: a guide for research in endocrinology and metabolism. *American Journal of Physiology - Endocrinology and Metabolism* ,2004, 286(4): E495-E501.

Symposium MATHEMATICS AND APPLICATIONS, Faculty of Mathematics, University of Belgrade, 2015 ,Vol. VI(1)

# Delay and stochastic differential equations as models of seismogenic fault motion

**Srđan Kostić**

*Institute for Development of Water Resources "Jaroslav Černi", Jaroslava Černog 80, 11226 Belgrade*
*e-mail: srdjan.kostic@jcerni.co.rs*

**Nebojša Vasović**

*University of Belgrade Faculty of Mining and Geology, Đušina 7, 11000 Belgrade*
*e-mail: nebojsa.vasovic@rgf.bg.ac.rs*

**Kristina Todorović**

*University of Belgrade Faculty of Pharmacy, Vojvode Stepe 450, 11226 Belgrade*
*e-mail: kisi@pharmacy.bg.ac.rs*

**Abstract.** In present paper we analyze two different mechanical models of fault motion, whose dynamics is governed by sets of delay and stochastic differential equations. In the first case, we analyze dynamics of a single block, by numerically solving a set of three first-order delay differential equations. Results of the performed analysis indicate that for certain values of fault friction parameters, and by introducing time delay, solutions of the analyzed set are represented by irregular aperiodic oscillations, which denote the onset of deterministic chaos. In the second case, we examine dynamics of 100 globally-coupled blocks, by solving a set of stochastic delay differential equations and by applying the mean-field approximation. In this case, obtained results indicate that solutions of examined set are irregular aperiodic oscillations for parameter values near the bifurcation curve, under the sole effect of seismic noise or due to effect of noise and global bifurcation.

**Keywords:** fault motion; mechanical model; time delay; seismic noise; differential equation

## 1. Introduction

From the mechanical viewpoint, tectonic movement along seismogenic faults in Earth's crust can be modeled by displacement of a block moving along the rough surface [1]. Such dynamical system is commonly described by a set of differential equations, whose specific solutions for some parameter values could be ascribed to different regimes of fault motion. In particular, equilibrium point corresponds to inactive fault, while periodic and quasiperiodic oscillations could be ascribed to aseismic fault creeping. Irregular aperiodic oscillations denote the onset of co-seismic fault motion. Results of previous studies on this topic indicated the occurrence of different dynamical regimes under the perturbation of control parameters or as a consequence of the effect of the newly introduced influential factors. For example, Galvanetto [2] studied the two-block model and showed that several periodic, quasi-periodic and chaotic attractors can coexist in this simple system. Erickson et al. [3] analyzed dynamics of a single-block model with Dieterich-Ruina friction law and found that system undergoes a Hopf bifurcation to a periodic orbit, with further occurrence of a strange attractor through period doubling cascade. In the first phase of the present research, we examined the set of delay differential equations describing the motion of a single-block model, by assuming friction as a function of velocity with a time delay, which was previously shown to be in the range 3-9 ms in laboratory conditions [4], depending on the viscosity and contact load. In this way we are modelling the well-known frictional memory effect [5]. Such approach is similar to our previous work [6].

In the second phase of the research, we examine a set of stochastic differential equations describing the motion of 100 interconnected blocks. Stochastic nature of the analyzed system arises from the assumption that incoherent seismic noise also affect the fault motion, even leading to transition between different dynamical regimes. Such random uncorrelated seismic noise originates from small-scale faulting, different irregularities and inhomogeneities or it is of undefined origin [7]. In present paper, such stochastic system is examined by studying its simplified version, i.e. by deriving and solving the set of deterministic mean-field equations, which exhibits the same dynamics as the starting stochastic system, for the same parameter values. This way of solving the stochastic differential equations is very powerful, since it enables reduction of 200 stochastic differential equations to only 5 mean-field deterministic equations, which represents a system easier to analyze. The applied

approach is similar to our previous work on this topic [8].One should note that in this case time delay describes the delayed interaction among the blocs of the model.

The paper is organized as follows. In Section 2 we briefly describe the applied methods. In Section 3 and 4 we present results of the analysis of the models with included time delay and random noise. In the final section we provide a brief review on the obtained results, with suggestions for future research.

## 2. Applied methods

In the first phase of the research, in order to analyze the system of delay differential equations, we utilize the backward differentiation formula method, which represents a linear multistep method that for the given function $y(t)$ and the moment $t_n$ approximates the function's derivative in terms of the $y(t)$ values at $t_n$ and the earlier times [3]. For the considered system, we have implemented the second order algorithm, which links the function values at the given moment with the ones at two prior iteration steps. Analysis of local bifurcations, as qualitative changes of dynamics, is conducted by analytically solving the system of delay differential equations, while the obtained results are numerically corroborated in DDE-BIFTOOL [9].

In the second phase of the research, we apply the method of mean-field approximation that replaces a many component system by a simpler system described by a small number of average macroscopic properties [10]. Local bifurcation analysis of the approximated model is conducted numerically also using DDE-BIFTOOL.

## 3. Model incorporating the effect of time delay - delay differential equations

We start from equations of motion coupled with Dieterich- Ruina rate-and state-dependent friction law originally given by [3]:

$$\begin{aligned} \dot{\theta} &= -\left(\frac{v}{L}\right)\left(\theta + B\log\frac{v}{v_0}\right), \\ \dot{u} &= v - v_0, \\ \dot{v} &= -\left(\frac{1}{M}\right)\left(ku + \theta + A\log\frac{v}{v_0}\right). \end{aligned} \tag{1}$$

System (1) describes the motion of a single block attached to a driver plate and moving in one direction along the rough surface of the lower plate. Parameter $M$ is the block mass and the spring stiffness $k$ could be ascribed to the linear elastic properties of the rock mass surrounding the fault [11].Parameters $A$ and $B$ are empirical constants, while $L$ denotes the critical slip distance, i.e. block displacement in a single slip phase. Variables $\theta$, $u$ and $v$ stand for state of the contact surface, displacement and velocity of the block, respectively. For convenience, system (1) could be non-dimensionalized by defining the new variables $\theta'$, $v'$, $u'$ and $t'$ in the following way: $\theta = A\theta'$, $v = v_0 v'$, $u = Lu'$, $t = (L/v_0)t'$, after which we return to the use of $\theta$, $v$, $u$ and $t$, putting the system (1) into the following form:

$$\begin{aligned} \dot{\theta} &= -v(\theta + (1+\epsilon)\log v), \\ \dot{u} &= v - 1, \\ \dot{v} &= -\gamma^2\left(u + \left(\frac{1}{\xi}\right)(\theta + \log v)\right), \end{aligned} \tag{2}$$

where $\epsilon = (B-A)/A$ measures the sensitivity of the velocity relaxation, $\xi = (kL)/A$ is the nondimensional spring constant, and $\gamma = (k/M)(1/2)(L/v_0)$ is the nondimensional frequency [3]. In present analysis, we add the time delay $\tau$ in the first equation of the system (2), such that the evolution of the state variable is modified:

$$\begin{aligned} \dot{\theta} &= -v(\theta + (1+\epsilon)\log(v(t-\tau)), \\ \dot{u} &= v - 1, \\ \dot{v} &= -\gamma^2\left(u + \left(\frac{1}{\xi}\right)(\theta + \log v)\right). \end{aligned} \tag{3}$$

We proceed by analyzing the local stability of the stationary state, which depends on the roots of the characteristic equation for the system (3), and which is obtained after appropriate linearization:

$$-\lambda^3 - \lambda^2\left(\frac{\gamma^2}{\xi} + 1\right) - \lambda\gamma^2\left(\frac{1}{\xi} + 1\right) - \gamma^2 + \lambda(1+\epsilon)\frac{\gamma^2}{\xi}e^{-\lambda t} = 0. \tag{4}$$

The equation (4) secures the existence of a nontrivial solution for the system of algebraic equations obtained after the system (1) was linearized. Bifurcations of the stationary state take place for the parameter values where the roots of (4) cross the imaginary axis. Since for $\gamma \neq 0$, the solution is ruled out, we look for the purely imaginary roots given by $\lambda = i\omega$, for real and positive $\omega$. Concerning this, we further proceed with the analysis of the following equation:

$$\frac{i\omega^3 + \omega^2\left(\frac{\gamma^2}{\xi}+1\right) - i\omega\gamma^2\left(\frac{1+\xi}{\xi}\right) - \gamma^2}{i\omega(1+\epsilon)\frac{\gamma^2}{\xi}} = -(cos\omega t - isin\omega t). \tag{5}$$

obtained when $\lambda = i\omega$ is replaced in (4). Results of the performed analysis provide the parametric representations among $\tau$ and the control parameters $\epsilon$, $\gamma$ and $\xi$ at the bifurcation values $\lambda = i\omega$:

$$\epsilon = -1 + \sqrt{\frac{\left(\omega^3 - \omega\gamma^2\left(\frac{1+\xi}{\xi}\right)\right)^2\left(\omega^2\left(\frac{\gamma^2}{\xi}+1\right) - \gamma^2\right)^2}{\left(\omega\frac{\gamma^2}{\xi}\right)}}, \tag{6}$$

and:

$$\tau = \tau_c = \frac{1}{\omega}\left(arctg\left(\frac{-\omega^2\left(\frac{\gamma^2}{\xi}+1\right)+\gamma^2}{-\omega^3+\omega\gamma^2\left(\frac{1+\xi}{\xi}\right)}\right) + k\pi\right), \tag{7}$$

where $k$ is any nonnegative integer such that $\tau_{c,k}$ holds. These relations define the corresponding bifurcation curves in the appropriate parameter planes. In particular, parametric equations for $\epsilon$, $\xi$ and $\tau$ coincide with the Hopf bifurcation curves illustrated in Figure 1 where the bifurcation curves $\tau(\epsilon)$ are shown for the fixed parameter values $\xi = 0.5$ and $\gamma = 0.8$. The system (4) admits both the direct and inverse Hopf bifurcations [12], resulting in creation and annihilation of an unstable plane in the system's state space when the corresponding curve is crossed, respectively.

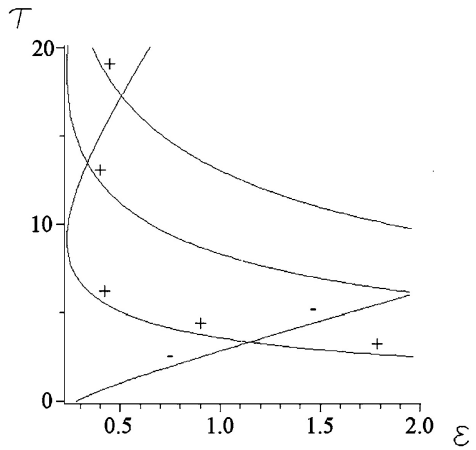



**Figure 1.** Hopf bifurcation curves $\tau(\epsilon)$, for the fixed values of parameters $\xi = 0.5$, and $\gamma$=0.8. The signs +/- denote the destabilizing or the stabilizing Hopf bifurcations, respectively, with increase of $\tau$.

From Figure 1 one can infer about the effect of the introduced time delay on dynamics of the analyzed spring-block model.Apparently, only by increasing the time-lag $\tau$, e.g. by setting $\tau = 0$, $\tau = 10$, $\tau = 13$ to $\tau = 20$, and by slightly changing the other parameter values, solutions of starting system (3) change from the equilibrium (fixed point), over the limit cycle oscillation (first Hopf bifurcation) and torus (second Hopf bifurcation) to deterministic chaos. The single peak in power spectrum indicates the oscillatory behavior, while

the existence of the second peak implies that the system evolution takes place on a torus. The broadband noise suggests the emergence of the strange attractor (Figure 2). One should note that the negative derivatives along the bifurcation curves indicate inverse Hopf bifurcations meaning that solutions of system (3) change from oscillatory behavior to equilibrium state (the so-called oscillation death) or they exhibit a change of oscillation amplitudes (the so-called amplitude death) [13].

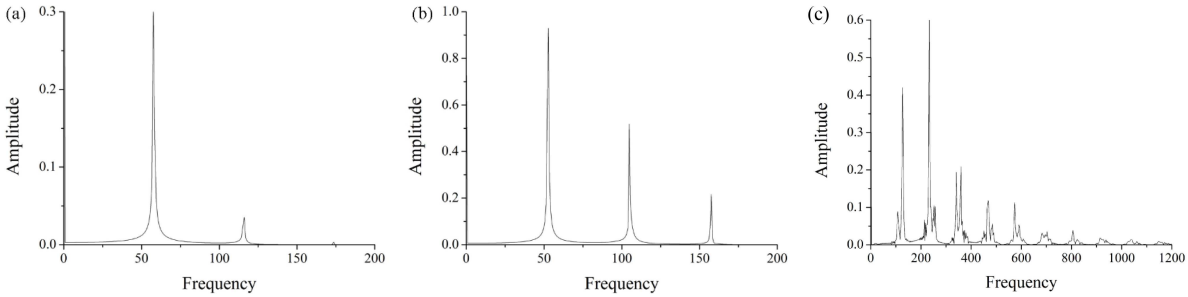



**Figure 2.** Fourier power spectrum for solutions of system (3) with increasing time delay:(a) single peak indicates oscillations ($\tau = 10$, $\epsilon = 0.3$, $\xi = 0.5$ and $\gamma = 0.8$); (b) Two peaks imply the appearance of second Hopf bifurcation ($\tau = 13$, $\epsilon = 0.5$, $\xi = 0.5$ and $\gamma = 0.8$); (c) The broadband noise confirms the onset of deterministic chaos ($\tau = 20$, $\epsilon = 0.5$, $\xi = 0.5$ and $\gamma = 0.8$).

## 4. Model incorporating the effect of noise and delayed interaction - stochastic delay differential equations

Model which includes the effect of seismic noise and delayed interaction between different fault segments is based on the mono-block model, originally suggested in [14], and modified here using the convenient coordinate transformation:

$$\begin{aligned} \dot{U}_{1i} &= U_{2i}(t), \\ dU_{2i}(t) &= \left( -U_{1i}(t) + \phi(U_{2i}) + \nu - \phi(\nu) + \frac{k}{N}\sum_{j=1}^{N} U_{1j}(t-\tau) - U_{1i}(t) \right) dt + \sqrt{2D}dW_i, \end{aligned} \tag{8}$$

where $U_{1i}$ and $U_{2i}$ represent displacement and velocity of the $i$-th block, respectively, $k$ is constant of spring connecting the blocks, $\phi$ stands for the friction force, $\tau$ is time delay (delayed interaction) and $\nu$ is a nondimensional pulling background velocity. Terms $\sqrt{2D}dW_i$ represent stochastic increments of independent Wiener process, i.e. $dW_i$ satisfy: $E(dW_i) = 0$, $E(dW_i dW_j) = \delta_{ij}dt$, where $E()$ denotes the expectation over many realizations of the stochastic process and $D$ is intensity of additive local noise. Each of $i = 1, 2, ..., N$ units in (8) is coupled with each other unit. In present case, we examine system of 100 units ($N$=100). By deriving the Taylor expansion of $\phi(U_2(t) + \nu)$ in the vicinity of the mean values:

$$(\langle U_1 \rangle, \langle U_2 \rangle) = \left( \lim_{\to\infty} \frac{1}{N}\sum_{i=1}^{N} U_{1i}(t), \lim_{\to\infty} \frac{1}{N}\sum_{i=1}^{N} U_{2i}(t) \right) = (m_{U1}, m_{U2}), \tag{9}$$

system (8) for $N \to \infty$ becomes:

$$
\begin{aligned}
dU_{1i}(t) &= U_{2i}(t)dt, \\
dU_{2i}(t) &= \Bigg(-U_{1i}(t) + \phi(m_{U_2}+\nu) - \phi(\nu) + \frac{1}{1!}\left(\phi'(m_{U_2}+\nu)\right)\left(U_2(t)-m_{U_2}\right) \\
&\quad + \frac{1}{2!}\left(\phi''(m_{U_2}+\nu)\right)\left(U_2(t)-m_{U_2}\right)^2 \\
&\quad + \frac{1}{3!}\left(\phi'''(m_{U_2}+\nu)\right)\left(U_2(t)-m_{U_2}\right)^3 \\
&\quad + \frac{1}{4!}\left(\phi^{(4)}(m_{U_2}+\nu)\right)\left(U_2(t)-m_{U_2}\right)^4 \\
&\quad + k(m_{U_1}(t-\tau) - U_1(t))dt + \sqrt{2D}dW_i.
\end{aligned}
\tag{10}
$$

In order to derive mean-field approximate dynamical equations for starting system (8), we suppose that distributions of $U_{1i}$ and $U_{2i}$ are Gaussian and that, for large $N$, the average over local random variables is given by the expectation with respect to the corresponding distribution. Following the procedure from Burić et al. [10], starting system (8) of 200 stochastic delay differential equations is reduced to the system of only 5 deterministic delay differential equations for the global variables and global centered moments:

$$
m_{U_1}(t) = \langle U_1(t)\rangle, m_{U_2}(t) = \langle U_2(t)\rangle, s_{U_1}(t) = \langle n^2_{U_1}(t)\rangle, s_{U_2}(t) = \langle n^2_{U_2}(t)\rangle, s_{U_1U_2}(t) = \langle n_{U_2}n_{U_1}\rangle \tag{11}
$$

where $n_{U_j}(t) = m_{U_j}(t) - U_{ji}(t), j = 1, 2$. The final mean-field approximated model with the general form of friction term $\phi$ is given in the following way:

$$
\begin{aligned}
\dot{m}_{U_1}(t) &= m_{U_2}(t), \\
\dot{m}_{U_2}(t) &= -m_{U_1}(t) + \phi(m_{U_2}+\nu) - \phi(\nu) + \frac{1}{2}(\phi''(m_{U_2}+\nu))s_{U_2} + \frac{1}{24}(\phi^(4)(m_{U2}+\nu)3s^2_{U_2} \\
&\quad + k(m_{U_1}(t-\tau) - m_{U1}(t)), \\
\frac{1}{2}\dot{s}_{U_1}(t) &= s_{U_1U_2}, \\
\frac{1}{2}\dot{s}_{U_2}(t) &= s_{U_2}\left(\phi'(m_{U_2}+\nu) + \frac{1}{2}\phi'''(m_{U_2}+\nu)s_{U_2}\right) - (k+1)s_{U_1U_2} + D, \\
\dot{s}_{U_1U_2} &= s_{U_1U_2}\left(\phi'(m_{U_2}+\nu) + \frac{1}{2}\phi'''(m_{U_2}+\nu)s_{U_2}\right) - (k+1)s_{U_1} + s_{U_2}(t).
\end{aligned}
\tag{12}
$$

The obtained results indicate a transition from equilibrium state to periodic oscillations for certain parameter values, as it is shown in Figure 3. Equilibrium state for the starting model(8) is represented by small fluctuations around the constant zero value of a mean-field approximated displacement for a mean-field model (12). On the other hand, when bifurcation curve is crossed, oscillation frequencies are the same for both the starting model and approximated system, while amplitude could be slightly different due to effect of the introduced random seismic noise in the stochastic model (8) (Figure 4).

One should note that when initial conditions are set away from the equilibrium point, mean-field model (12) could be in equilibrium state (Figure 5) or could exhibit oscillatory behavior due to effect of global bifurcation depending on the parameter values prior to local bifurcation (Figure 6). It should be emphasized that bifurcation curve in Figure (reffig:6) is captured only for the mean-field approximated model (12). These bifurcations in the real starting model (8) are captured only as significantly higher fluctuations of stochastic system when bifurcation curve is crossed.

It should ne emphasized that solutions of system (8) appear as irregular oscillations (stick-slip like motion) at the verge of equilibrium state regime under the impact of both incoherent noise and global attractor (Figure 7).

## 5. Conclusion

In present paper we analyze two different models of fault dynamics, whose motion is governed by delay and stochastic delay differential equations. Model with included time delay assumes rate-and state- dependent friction

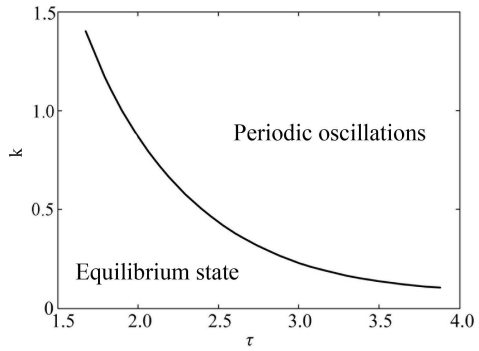



**Figure 3.** Parameter domain $(k, \tau)$ admitting equilibrium state or periodic oscillations of the mean-field approximated model (12). For a given parameter domain, other parameters are held constant at the following values: $\nu$=1.2, $D$=0.001, $a$=0.1.

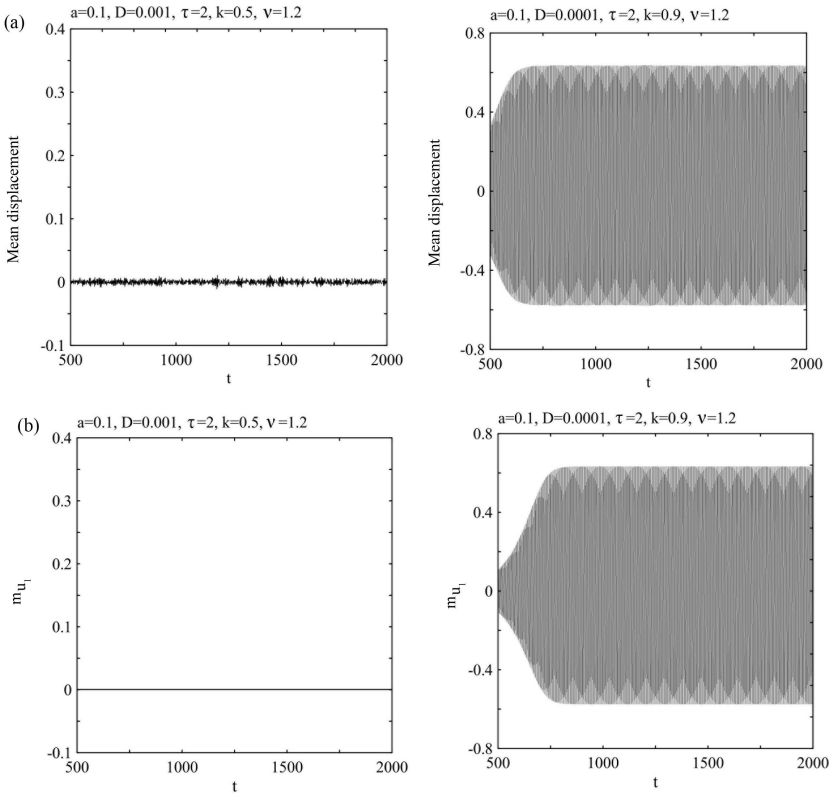



**Figure 4.** Time series of mean displacements of 100 blocks (a) and approximated mean-field displacements (b). Values of parameter $k$ are conveniently chosen for equilibrium state ($k$=0.5) and periodic oscillations ($k$=0.9). Other parameter values are held constant: $a$=0.1, $D$=0.001, $\tau$=2, $\nu$ =1.2. Oscillation frequency for both the starting stochastic model and its mean-field approximation is the same ($f$=0.33).

law along the contact of the moving block and rough surface of the lower plate. In contrast to previous studies, additional parameter is included in friction term, which involves delayed friction response on the change of block velocity. Results of the performed research indicate a transition from equilibrium state and regular periodic oscillations to quasiperiodic oscillations and deterministic chaos, for certain parameter values. One should note that deterministic chaos is observed for rather large values of time delay, which is rarely observed in laboratory conditions.

Model with included seismic noise and delayed interaction among the blocks assumes only rate-dependent

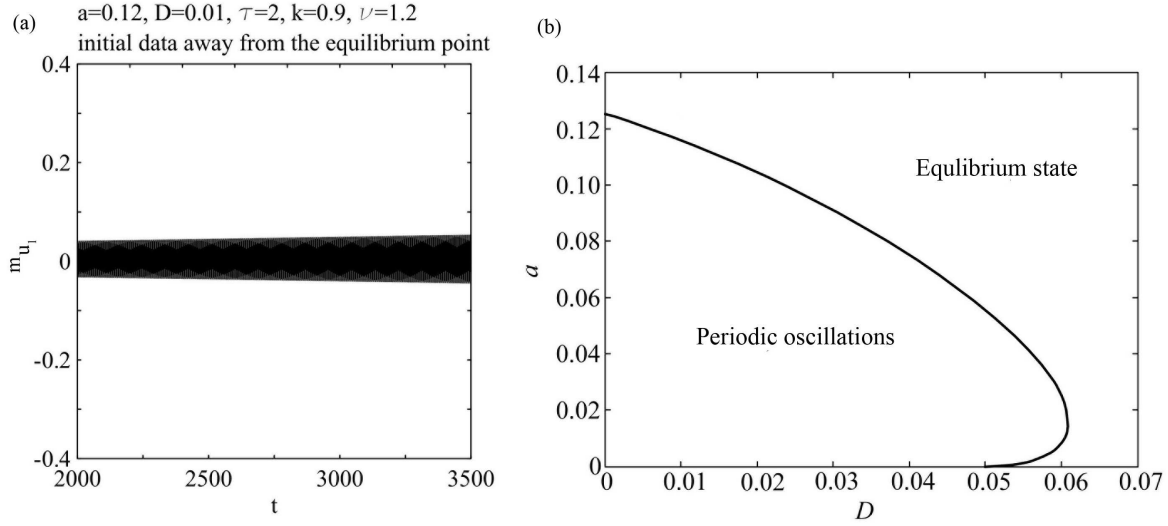



**Figure 5.** (a)Time series of mean-field displacement in an approximated model (12) for initial conditions away from the equilibrium point and for different values of material property $a$ and noise level $D$ prior the bifurcation; (b) Parameter domain $(a, D)$ admitting equilibrium state or periodic oscillations of the mean-field approximated model (12). For a given parameter domain, other parameters are held constant at the following values: $K$=0.9, $\nu = 1.2$, $\tau$=2.

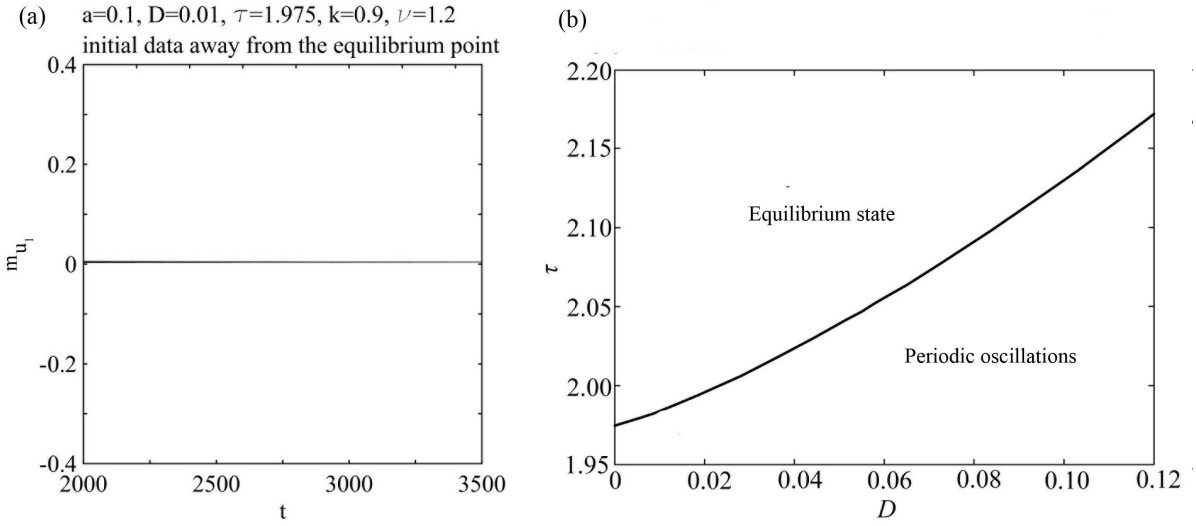



**Figure 6.** (a)Time series of mean-field displacement in an approximated model (12) for initial conditions away from the equilibrium point and for different values of noise level $D$ and time delay $\tau$ prior the bifurcation; (b) Parameter domain $(\tau, D)$ admitting equilibrium state or periodic oscillations of the mean-field approximated model (12). For a given parameter domain, other parameters are held constant at the following values: $a$=0.1, $K$=0.9, $\nu = 1.2$.

friction law. In present case, analysis is conducted for the model composed of 100 interconnected blocks, meaning that one needs to examine a system of 200 stochastic delay differential equations. For convenience, in present paper we applied mean-field approximation method, by which we obtained a simplified system of 5 deterministic delay differential equations, whose solutions for different parameter values are qualitatively similar with the solutions of the starting stochastic system. In this case, it is observed that complex irregular oscillations arise due to effect of seismic noise near the bifurcation point, or as a consequence of the bistability, with co-existence of

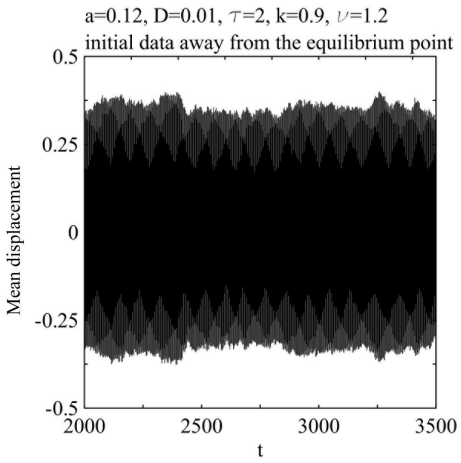



**Figure 7.** Time series of mean displacement in a starting system (8) for initial conditions away from the equilibrium point. It is clear that oscillation amplitudes are higher than introduced noise level ($D$=0.01).

two stable attractors, which aperiodically attract the solutions to different basins.

It should be emphasized that delayed friction response to velocity change, seismic noise and delayed interaction among interacting blocks of the model are for the first time introduced in the analyzed models. Moreover, method of mean-field approximation is for the first time applied in the analysis of dynamics of a spring-block model, indicating that irregular displacement of the starting stochastic (”real”) system (with oscillation amplitude higher than introduced noise level) near the transition from equilibrium state to periodic oscillations could occur either due to sole effect of seismic noise or under the impact of seismic noise in the presence of local and global attractor.

Regarding the further research on this topic, subsequent analysis should focus on possible excitable dynamics of the spring-block models, or their analogs. If found, such models with excitable behavior could provide the most accurate phenomenological description of seismogenesis, since earthquakes represent sudden phenomena, whose occurrence is interspersed with relatively long periods of stable stationary fault motion.

**Acknowledgements.** *This work is partially supported by the Ministry of Education, Science and Technological Development of Republic of Serbia (Contracts No. 176016 and 171017).*

## References

[1] **R. Burridge, L. Knopoff.** Model and theoretical seismicity. *Bulletin of Seismological Society of America*, 1967, 57, 341 - 371.

[2] **U. Galvanetto.** Some remarks on the two-block symmetric Burridge-Knopoff model. *Physics Letters A*, 2002, 293, 251 - 259.

[3] **B. Erickson, B. Birnir, D. Lavallee.** A model for aperiodicity in earthquakes. *Nonlinear Processes in Geophysics*, 2008, 15, 1 - 12.

[4] **D.P. Hess, A. Soom.** Friction at a lubricated line contact operating at oscillating sliding velocities.*Journal of tribology*, 1990, 112, 147 - 152.

[5] **E. Berger.** Friction modeling for dynamic system simulation. *Applied Mechanics Review*, 2002, 55, 535 -577.

[6] **S. Kostić, I. Franović, K. Todorović, N. Vasović.** Friction memory effect in complex dynamics of earthquake model.*Nonlinear Dynamics*, 2013, 73, 1933-1943.

[7] **V.B. Ryabov, A.M. Correig, M. Urquizu, A.A. Zaikin.** Microseism oscillations: from deterministic to noise-driven models. *Chaos Solitons and Fractals*, 2003, 16, 195-210.

[8] **N. Vasović, S. Kostić, I. Franović, K. Todorović.** Earthquake nucleation in a stochastic fault model of globally coupled units with interaction delays.*Communications in Nonlinear Science and Numerical Simulation*, 2016, 38, 117-129.

[9] **K. Engelborghs, T. Luzyanina, G. Samaey.** DDE-BIFTOOL v. 2.03: a MATLAB package for bifurcation analysis of delay differential equations. *Technical Report TW-330, Department of Computer Science, K.U. Leuven. Leuven, Belgium*, 2000.

[10] **N. Burić, D. Ranković, K. Todorović, N. Vasović.** Mean field approximation for noisy delay coupled excitable neurons.*Physica A*,2010, 389, 3956-64.

[11] **C.H. Scholz.** The mechanics of earthquakes and faulting. *Cambridge University Press,Cambridge*, 2002.
[12] **S. Wiggins.** Introduction to Applied Nonlinear Dynamical Systems and Chaos.*Springer,New York*,2000.
[13] **D.V. Ramana Reddy, A. Sen, G.L. Johnson.** Time Delay Induced Death in Coupled Limit Cycle Oscillators. *Physical Review Letters*,1998, 80, 5109-5112.
[14] **M. De Sousa Vieira.** Chaos and synchronized chaos in an earthquake model. *Physical Review Letters*, 1999, 82, 201 - 204.

Symposium MATHEMATICS AND APPLICATION, Faculty of Mathematics, University of Belgrade, 2015 ,Vol. VI(1)

# Moment matching discretization of a stochastic integral

**Tatjana Bajić**

*Higher Education School of Professional Studies for Preschool Teachers, Dobropoljska 5, Šabac, Serbia*
*e-mail: ttanja.bajic@gmail.com*

**Abstract.** Bearing in mind that the moment matching idea can be fruitful in many practical problems, we consider the moment matching discretization of a stochastic integral of deterministic function, as a random variable having a certain absolutely continuous probability distribution on an interval $I \subseteq \mathbb{R}$. On the basis of one generalization of a mean value theorem for systems of deterministic integrals, we prove that there exists a discrete probability distribution on $I \subseteq \mathbb{R}$ such that the corresponding discrete random variable matches moments of the given stochastic integral. Moreover, the moment matching discretization of a stochastic integral can be considered from a viewpoint of $n$-point Gaussian quadrature rule. Due to the connection between the normalized Hermite polynomial and the classical one, we show that the Gauss-Hermite quadrature rule can be used to carry out the moment matching discretization of an arbitrary Gaussian distribution by the normalized Hermite polynomials. Applying this result, we obtain a system of discrete random variables having the same moments to a certain order as the corresponding stochastic integrals defined with respect to a Gaussian orthogonal stochastic measure. The case when stochastic integrals form a Gaussian process is discussed additionally.

**Keywords:** Stochastic integrals, Moment matching,Gauss-Hermite quadrature.

## 1. Introduction

Under certain conditions, an absolutely continuous probability distribution can be discretized in such a way that the discrete distribution matches moments of the given absolutely continuous distribution, e.g., expected value and variance. As the moment matching approach is a flexible and intuitively appealing way of discretization of random variables, this idea is quite common, for example, in stochastic optimization and dynamic programming (see e.g. [3], Chapter 10, Chapter 11 and [8]).

Recall that for a positive measure $\mu$ on an interval $I \subseteq \mathbb{R}$ the $k$-th moment is defined as

$$\int_I x^k d\mu(x)$$

– provided the integral exists. If we suppose that $\{m_k\}_{k\geqslant 0}$ is a given sequence of real numbers, then the classical moment problem consists of solving the following: Does there exist a positive measure on an interval $I \subseteq \mathbb{R}$ with moments $\{m_k\}_{k\geqslant 0}$? For a more detailed discussion one is referred, for example, to [2].

However, without loss of generality we can always assume that $m_0 = 1$. Then the positive measure $\mu$ is a probability measure concentrated on $I \subseteq \mathbb{R}$. Accordingly, for the given finite moments $\{m_k\}_{k\geqslant 0}$ of an absolutely continuous probability distribution concentrated on $I \subseteq \mathbb{R}$, the moment matching discretization to some order $N \geqslant 0$ comes down to finding a discrete probability distribution on $I$ with moments $\{m_k\}_{0\leqslant k\leqslant N}$ .

On the other side, the moments of a stochastic integral of deterministic function, as the moments of corresponding random variable, depend on the choice of integrand. Keeping that in mind, in this paper we consider the moment matching discretization of stochastic integral which has an absolutely continuous probability distribution concetrated on some interval $I \subseteq \mathbb{R}$.

## 2. Moment matching discretization of a stochastic integral

### 2.1. A stochastic integral. Setting up the problem

For a left-continuous increasing function $F(t)$ on $\mathbb{R}$, let $\eta = \eta(t)$, $t \geqslant t_0$, be a real-valued random process $\eta$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$,

$$\eta(t_0) = 0, \text{ a.s.} \quad E\eta(t) = 0, \quad E\left[\eta(t)\right]^2 = F(t) < +\infty, \quad t > t_0,$$

corresponding to a left-continuous function in the Hilbert space $\mathcal{L}^2(\Omega) = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ with orthogonal increments

$$\Delta\eta = \eta(t) - \eta(s)$$

on disjoint intervals $\Delta = [s, t)$ , $s < t$, such that

$$E\Delta\eta = 0 \quad \text{and} \quad E\left[\Delta\eta\right]^2 = E\left[\eta(t) - \eta(s)\right]^2 = F(t) - F(s) = \Delta F.$$

For an interval $T \subseteq [t_0, +\infty)$ , denote by $\mathcal{B}(T)$ the Borel sigma-field of $T$. Keeping in mind the properties of increments of the random process $\eta$, we have that

$$\langle \eta(\Delta_1), \eta(\Delta_2) \rangle_{\mathcal{L}_2(\Omega)} = 0 \quad \text{when} \quad \Delta_1 \cap \Delta_2 = \varnothing, \quad \Delta_1, \Delta_2 \in \mathcal{B}(T) \tag{i}$$

$$\eta(\Delta) = \sum_{k=1}^{n} \eta(\Delta_k), \quad \text{a.s. when} \quad \Delta = \bigsqcup_{k=1}^{n} \Delta_k \in \mathcal{B}(T), \quad \Delta_i \cap \Delta_j = \emptyset \quad \text{for } i \neq j, \tag{ii}$$

and

$$\|\eta(\Delta)\|^2_{\mathcal{L}_2(\Omega)} = \Delta F < +\infty \quad \text{for all} \quad \Delta \in \mathcal{B}(T). \tag{iii}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{L}^2(\Omega)}$ and $\|\cdot\|_{\mathcal{L}^2(\Omega)}$ are the inner product and the norm in the space $\mathcal{L}^2(\Omega)$, respectively. Then with

$$\eta(\Delta) = \Delta\eta, \quad \Delta \in \mathcal{B}(T) \tag{1}$$

is defined an *orthogonal stochastic measure* on $\mathcal{B}(T)$ which takes values in $\mathcal{L}^2(\Omega)$. In view of the equality (iii), the measure $\eta(\Delta)$ is characterized by the finite positive measure $dF$ on $\mathcal{B}(T)$, known as structure measure of measure $\eta(\Delta)$, such that

$$\|\eta(\Delta)\|^2_{\mathcal{L}_2(\Omega)} = \Delta F = \int_{\Delta} dF < +\infty \quad \text{for all} \quad \Delta \in \mathcal{B}(T).$$

Then, for any square integrable function $f$ with respect to the measure $dF$ on $\mathcal{B}(T)$, one can define a stochastic integral

$$\xi^T(f) = \int_T f(u) d\eta(u) \tag{2}$$

with respect to the orthogonal stochastic measure (1) on $\mathcal{B}(T)$ as a linear functional on the Hilbert space $\mathcal{L}^2(T, dF)$ ([9], 212-214).

On the other side, the stochastic integrals in the form (2) generate a system $\mathcal{H}^T = \left\{\xi^T(f), f \in \mathcal{L}^2(T, dF)\right\}$ of real-valued random variables with zero mean, variance

$$E\left[\xi^T(f)\right]^2 = \|f\|^2_{\mathcal{L}^2(T,dF)} = \int_T f^2(u) dF(u) = \mathcal{J}^T(f^2)$$

and covariance

$$E\left[\xi^T(f)\xi^T(g)\right] = \langle f, g \rangle_{\mathcal{L}^2(T,dF)} = \int_T f(u)g(u) dF(u) = \mathcal{J}^T(fg)$$

$f, g \in \mathcal{L}^2(T, dF)$, where $\|\cdot\|_{\mathcal{L}^2(T,dF)}$ and $\langle \cdot, \cdot \rangle_{\mathcal{L}_2(T,dF)}$ denote the norm and the inner product in $\mathcal{L}^2(T, dF)$, respectively. The deterministic integrals $\mathcal{J}^T(f^2)$ and $\mathcal{J}^T(fg)$, on the right-hand side of above equalities, are the Lebesgue-Stieltjes integrals with respect to the corresponding measure $dF$. In particular, if $f = 0$ a.e. on $\mathcal{B}(T)$, then $\mathcal{J}^T(f^2) = 0$ and $\xi^T(f)$ is a degenerate random variable. Otherwise, $\mathcal{J}^T(f^2) > 0$.

If we denote by $\overline{\mathcal{L}}_{hull}(\eta^T)$ the closed linear hull spanned by

$$\eta^T = \{\eta(\Delta), \Delta \in \mathcal{B}(T)\},$$

then the system $\mathcal{H}^T$ is equal to $\overline{\mathcal{L}}_{hull}(\eta^T)$. Hence, for any $f \in \mathcal{L}^2(T, dF)$, random variable $\xi^T(f)$ is defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as the orthogonal stochastic measure (1), and $\xi^T(f)$ is measurable with

respect to the sigma algebra $\mathcal{F}_T \subseteq \mathcal{F}$ generated by the system $\eta^T$. Further, since the stochastic integral (2) depends on the choice of function $f \in \mathcal{L}^2(T, dF)$, the probability measure induced by random variable $\xi^T(f)$ on the Borel sigma-field of $\mathbb{R}$ will be denoted by $\mu_f$.

In order to carry out a moment matching discretization of the stochastic integral (2), we will suppose that the random variable $\xi^T(f)$, $f \in \mathcal{L}^2(T, dF)$, satisfies the following assumptions:

1. The random variable $\xi^T(f)$, $f \in \mathcal{L}^2(T, dF)$ has an absolutely continuous probability distribution $\mu_f$ concentrated on some interval $I \subseteq \mathbb{R}$.

2. The random variable $\xi^T(f)$, $f \in \mathcal{L}^2(T, dF)$ possesses finite moments of all orders.

Under these conditions we consider the next problem: For the given moments of $\xi^T(f)$, we should to find a discrete probability distribution $\mu_f^{(n)}$ on $I$, $n \geqslant 1$, i.e. the values $r_1, ..., r_n \in I$ and the corresponding probabilities $\lambda_1, ..., \lambda_n$ of a discrete random variable $\zeta_n^T(f)$ such that

$$E\left[\xi^T(f)\right]^k = \int_I x^k d\mu_f(x) = (r_1)^k\lambda_1 + ... + (r_n)^k\lambda_n = E\left[\zeta_n^T(f)\right]^k, \tag{3}$$

to an order $N$, $0 \leqslant k \leqslant N$.

### 2.2. A mean value theorem for systems of integrals and the moment matching discretization of a stochastic integral

The next theorem, representing one *generalization* of Kowalewski's theorem entitled "A mean value theorem for a system of $n$ integrals" (see [5]), gives us the existence of solution of the problem.

**Theorem 1.** *For an interval $I \subseteq \mathbb{R}$, let $\mu$ be a finite positive measure on the Borel sigma-field of $I$. Let $g_k$, $k = 1, ..., n$, $n \geqslant 1$, be continuous functions on $I$, integrable on $I$ with respect to the measure $\mu$. Then there exist points $r_1, ..., r_n$ in $I$, and non-negative numbers $\lambda_1, ..., \lambda_n$, with $\lambda_1 + ... + \lambda_n = \mu(I)$, such that*

$$\int_I g_k(x)d\mu(x) = g_k(r_1)\lambda_1 + ... + g_k(r_n)\lambda_n, \quad k = 1, ..., n. \tag{4}$$

Namely, if $\mu$ be a probability measure on the Borel sigma-field of $I$ then, on the basis of Theorem 2.1, we can prove the next statement regarding the moment matching discretization of the stochastic integral (2).

**Proposition 1.** *For an interval $T \subseteq [t_0, +\infty)$, let $\xi^T(f)$, $f \in \mathcal{L}^2(T, dF)$, be the stochastic integral (2) satisfying the assumptions 1 and 2. Then, there exists a discrete random variable $\zeta_n^T(f)$ taking at most $n \geqslant 1$ values, $r_1(f), ..., r_n(f) \in I$, with corresponding probabilities $\lambda_1(f), ..., \lambda_n(f)$, such that*

$$E\left[\xi^T(f)\right]^k = E\left[\zeta_n^T(f)\right]^k, \quad k = 0, 1, 2, ..., n. \tag{5}$$

*Proof.* As $\xi^T(f)$ has an absolutely continuous probability distribution $\mu_f$ concentrated on some interval $I \subseteq \mathbb{R}$ ($\mu_f(I) = 1$) with finite moments of all orders, it follows that

$$E\left[\xi^T(f)\right]^k = \int_I x^k d\mu_f(x) < +\infty, \quad k \geqslant 1.$$

The functions $g_k(x) = x^k$, $k \geqslant 0$, are continuous on any interval of $\mathbb{R}$. Therefore, according to Theorem 2.1, there exist $n \geqslant 1$ points $r_1(f), ..., r_n(f)$ in $I$, and non-negative numbers $\lambda_1(f), ..., \lambda_n(f)$, with $\lambda_1(f)+...+\lambda_n(f) = 1$, such that

$$\int_I x^k d\mu_f(x) = (r_1(f))^k \cdot \lambda_1(f) + ... + (r_n(f))^k \cdot \lambda_n(f), \quad k = 1, ..., n. \tag{6}$$

Denote by $\delta_{r_i(f)}$ the unit mass located at $r_i(f) \in I$, $i = 1, ..., n$, and by $\mathcal{B}(I)$ the sigma-field on $I \subseteq \mathbb{R}$. As $\lambda_i(f) \geqslant 0$, $i = 1, ..., n$, satisfies $\lambda_1(f)+...+\lambda_n(f) = 1$, it follows that $\mu_f^{(n)} := \lambda_1(f)\cdot\delta_{r_1(f)}+...+\lambda_n(f)\cdot\delta_{r_n(f)}$ defines a probability measure on $\mathcal{B}(I)$. In other words, there exists a discrete random variable $\zeta_n^T(f)$ taking at most $n$ values $r_1(f), ..., r_n(f) \in I$, with corresponding probabilities $\lambda_1(f), ..., \lambda_n(f)$. Thus, from (6) it follows (5).

Further, for functions $f$, $g \in \mathcal{L}^2(T, dF)$ such that $g \neq f$ a.e. on $\mathcal{B}(T)$, we have that

$$E\left[\xi^T(f)\right]^2 = \mathcal{J}^T(f^2) \neq \mathcal{J}^T(g^2) = E\left[\xi^T(g)\right]^2.$$

Therefore, the discrete probability measure $\mu_f^{(n)}$, i.e. the values $r_1(f), ..., r_n(f) \in I$ and the corresponding probabilities $\lambda_1(f), ..., \lambda_n(f)$, depend on the choice of function $f \in \mathcal{L}^2(T, dF)$■

Unfortunately, it is important to note that the discretization of integrals, given by (4) in Theorem 2.1, is not unique ([5], 336) and because of that, for a fixed function $f \in \mathcal{L}^2(T, dF)$, the values $r_1(f), ..., r_n(f) \in I$ and the corresponding probabilities $\lambda_1(f), ..., \lambda_n(f)$ from Proposition 2.1 are also not unique.

## 3. Gauss-Hermite quadrature and the moment matching discretization of a stochastic integral

Theorem 2.1 actually claims that, for a given any set of continuous functions on $I$, and a finite measure $\mu$ on $I$, there exists an $n$-point quadrature rule which is exact for those functions ([5], 336). However, as we have already noted, this quadrature rule is not unique.

On the other side, an $n$-point Gaussian quadrature rule with respect to the system of functions

$$g_k(x) = x^k, \quad x \in I \subseteq \mathbb{R}, \quad k = 0, ..., 2n-1. \tag{7}$$

integrates exactly all of these functions in a unique way ([6], 973). Therefore, the moment matching discretization of stochastic integral (2), given by the equality (5), can be considered from a viewpoint of $n$-point Gaussian quadrature rule.

In many practical problems, one of the most common absolutely continuous probability distribution with finite moments of all orders is the Gaussian distribution. On the other side, a Gaussian quadrature is based on the fact that polynomials of a certain class are orthogonal with respect to an appropriate weight function on an interval $I \subseteq \mathbb{R}$. As the classical Hermite polynomials ([7], 61) are orthogonal with respect to the weight function $e^{-x^2}$ on $\mathbb{R}$, we will consider the moment matching discretization of a Gaussian random variable from a viewpoint of the well known Gauss-Hermite quadrature rule.

Denote by $\mathcal{B}(\mathbb{R})$ the Borel sigma-field of $\mathbb{R}$. The moment matching discretization of a Gaussian distribution includes integration with respect to the corresponding Gaussian measure on $\mathbb{R}$. As the normalized Hermite polynomials

$$\mathfrak{h}_n(x) = \frac{(-1)^n}{\sqrt{n!}} e^{\frac{x^2}{2}} \frac{d^n e^{-\frac{x^2}{2}}}{dx^n}, \quad x \in \mathbb{R}, \quad n = 0, 1, 2, ... \tag{8}$$

form a complete orthonormal system in the space $\mathcal{L}^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$, where $P = \mathcal{N}(0, 1)$ is the standard Gaussian measure on $\mathcal{B}(\mathbb{R})$, we will deal with the polynomials (8) instead the classical ones. Namely, due to the connection between the normalized Hermite polynomial and the classical one, we can use the Gauss-Hermite quadrature rule to prove the next statement regarding the moment matching discretization of a Gaussian distribution.

**Proposition 2.** *Let $r_1, ..., r_n \in \mathbb{R}$ be the roots of normalized Hermite polynomial $\mathfrak{h}_n(x) \in \mathcal{L}^2(\mathbb{R}, P)$ of degree $n \geqslant 1$. If $\widetilde{P} = \mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$, is an arbitrary Gaussian measure on $\mathcal{B}(\mathbb{R})$, then it holds*

$$\int_R \widetilde{x}^k d\widetilde{P}(\widetilde{x}) = (\sigma r_1 + m)^k \cdot \lambda_1 + ... + (\sigma r_n + m)^k \cdot \lambda_n, \quad k = 0, 1, ..., 2n-1, \tag{9}$$

*where $\lambda_i = 1/(n\left[\mathfrak{h}_{n-1}(r_i)\right]^2)$, $i = 1, ..., n$, and $\lambda_1 + ... + \lambda_n = 1$.*

*Proof.* It is well known that $n$-point Gauss-Hermite quadrature rule integrates exactly all of functions from system (7) on $\mathbb{R}$,

$$\int_{-\infty}^{+\infty} x^k e^{-x^2} dx = \sum_{i=1}^{n} (r_i^{\#})^k \lambda_i^{\#}, \quad k = 0, 1, ..., 2n-1, \tag{10}$$

where the nodes $r_1^{\#}, ..., r_n^{\#} \in \mathbb{R}$ are roots of the classical Hermite polynomial $h_n^{\#}(x)$ of the $n$-th order and the weights $\lambda_i^{\#} > 0$ are equal to $\lambda_i^{\#} = \left\langle h_{n-1}^{\#}, h_{n-1}^{\#} \right\rangle / (n\left[h_{n-1}^{\#}(u_i)\right]^2)$, $i = 1, ..., n$ ([1]: 890, [3], [8]). Thus,

keeping in mind that standard Gaussian measure $P$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and using an appropriate change in variable $x = \sqrt{2}r$, we get

$$\int_R x^k dP(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^k e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \left(\sqrt{2}r\right)^k e^{-r^2} dr = \sum_{i=1}^{n} \left(\sqrt{2}r_i^{\#}\right)^k \frac{\lambda_i^{\#}}{\sqrt{\pi}}, \tag{11}$$

for $k = 0, 1, ..., 2n-1$. Considering the connection between Hermite polynomial $h_n^{\#}$ and normalized Hermite polynomial $\mathfrak{h}_n$, expressed with

$$\mathfrak{h}_n(\sqrt{2}r) = \frac{h_n^{\#}(r)}{\sqrt{2^n n!}},$$

it follows that $r_i = \sqrt{2}r_i^{\#}$, $i = 1, ..., n$, are roots of normalized Hermite polynomial $\mathfrak{h}_n$ of the $n$-th order and the corresponding weights are

$$\begin{aligned}\lambda_i &= \frac{\lambda_i^{\#}}{\sqrt{\pi}} = \frac{\left\langle h_{n-1}^{\#}, h_{n-1}^{\#} \right\rangle}{\sqrt{\pi} n \left[h_{n-1}^{\#}(r_i^*)\right]^2} = \frac{2^{n-1}(n-1)!\sqrt{\pi}}{\sqrt{\pi} n \left[h_{n-1}^{\#}(r_i^*)\right]^2} = \frac{1}{n\left[\frac{h_{n-1}^{\#}(r_i^{\#})}{\sqrt{2^{n-1}(n-1)!}}\right]^2} \\ &= \frac{1}{n\left[\mathfrak{h}_{n-1}(\sqrt{2}r_i^{\#})\right]^2} = \frac{1}{n\left[\mathfrak{h}_{n-1}(r_i)\right]^2} > 0, \quad i = 1, ..., n.\end{aligned}$$

Replacing $\sqrt{2}r_i^*$ and $\lambda_i^*/\sqrt{\pi}$ with $r_i$ and $\lambda_i$ in (11), respectively, we obtain

$$\int_R x^k dP(x) = (r_1)^k \lambda_1 + ... + (r_n)^k \lambda_n, \quad k = 0, 1, ..., 2n-1. \tag{12}$$

Denote by $\delta_{r_i}$ the unit mass located at $r_i$, $i = 1, ..., n$. As the transformed Gauss-Hermite quadrature rule (11) holds for the constant functions, we get that the weights $\lambda_i > 0$, $i = 1, ..., n$, satisfy

$$\lambda_1 + ... + \lambda_n = \frac{\lambda_1^{\#}}{\sqrt{\pi}} + ... + \frac{\lambda_n^{\#}}{\sqrt{\pi}} = 1. \tag{13}$$

Therefore, with

$$P_n := \lambda_1 \delta_{r_1} + ... + \lambda_n \delta_{r_n}, \tag{14}$$

is defined a probability measure on $\mathcal{B}(\mathbb{R})$. Due to the fact that *an n-point quadrature rule is exact for an arbitrary polynomial of order $N$ if and only if it is exact for all functions $f_k(x) = x^k$, $k = 0, 1, ..., N$*, the equality (12) is exact for all polynomials $p_k(x)$, $x \in \mathbb{R}$, of order $2n-1$ or less. Thus, the Weierstrass approximation theorem indicates that the measures $P_n$ converge weakly to $P$ as $n \nearrow \infty$.

If $\widetilde{P} = \mathcal{N}(m, \sigma^2)$ an arbitrary Gaussian measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then it is related to the measure $P = \mathcal{N}(0, 1)$ through a linear transformation. Thus, introducing linear mapping $L : \mathbb{R} \to \mathbb{R}$, with $L(x) = \sigma x + m$, $\sigma > 0$, $x \in \mathbb{R}$, we obtain that for any $k = 0, 1, ..., 2n-1$, it holds

$$\int_R \widetilde{x}^k d\widetilde{P}(\widetilde{x}) = \int_R (\sigma x + m)^k dP(x).$$

On the other hand, the probability measure $\widetilde{P} = PL^{-1}$ is induced on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ in a unique way by the standard Gaussian measure $P$ and by the linear transformation $L$. Thus, the probabilities measures $P_n$ induce on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the corresponding unique probabilities measures $\widetilde{P}_n = P_n L^{-1}$. If we denote by $\delta_{\sigma r_i + m}$ the unit mass located at $\sigma r_i + m$, then from (14) it follows that

$$\widetilde{P}_n = P_n L^{-1} := \lambda_1 \delta_{r_1} L^{-1} + ... + \lambda_n \delta_{r_n} L^{-1} = \lambda_1 \delta_{\sigma r_1 + m} + ... + \lambda_n \delta_{\sigma r_n + m}, \tag{15}$$

is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. As $L$ is continuous mapping and $P_n$ converge weakly to $P$ as $n \nearrow \infty$, we have that $\widetilde{P}_n = P_n L^{-1}$ converge weakly to $\widetilde{P} = PL^{-1}$ as $n \nearrow \infty$ ([4], Theorem 5.1). Therefore, for any fixed $n \geqslant 1$ the equality (9) holds■

In other words, Proposition 3.1 claims that if $\widetilde{\xi} \sim \mathcal{N}(m, \sigma^2)$ is an arbitrary Gaussian random variable, then there exists a discrete random variable $\widetilde{\zeta}_n$ which values $r_i^* = \sigma r_i + m \in \mathbb{R}$ and the corresponding probabilities $\lambda_i = 1/(n\left[\mathfrak{h}_{n-1}(r_i)\right]^2)$, $i = 1, ..., n$, are computed by the roots $r_1, ..., r_n \in \mathbb{R}$ of normalized Hermite polynomial $\mathfrak{h}_n(x) \in \mathcal{L}^2(\mathbb{R}, P)$ of degree $n \geqslant 1$, such that $E\widetilde{\xi}^k = E\widetilde{\zeta}_n^k$, $k = 0, 1, ..., 2n-1$. Therefore, on the basis of Proposition 3.1, one can carry out a moment matching discretization of a system of Gaussian random variables. Notice that for different Gaussian distributions, the values of probabilities $\lambda_i = 1/(n\left[\mathfrak{h}_{n-1}(r_i)\right]^2)$, $i = 1, ..., n$, of the corresponding discrete distributions are the same. Moreover, as the equality (9) holds for all polynomials $p_k(x)$, $x \in \mathbb{R}$, of order $2n-1$ or less instead the functions $x^k$, $k = 0, 1, ..., 2n-1$, the Proposition 3.1 can be used to directly approximate the expectation of some continuous function $g$ on $\mathbb{R}$ under an arbitrary Gaussian distribution $\widetilde{P} = \mathcal{N}(m, \sigma^2)$ such that

$$E(g(\widetilde{\xi})) = \int_R g(\widetilde{x}) d\widetilde{P}(\widetilde{x}) \approx g(\sigma r_1 + m)\lambda_1 + ... + g(\sigma r_n + m)\lambda_n.$$

In order to discretize stochastic integral (2) by Proposition 3.1, the orthogonal stochastic measure (1) has to be Gaussian. Then the stochastic integral (2) is a Gaussian random variable $\xi^T(f)$ with Gaussian probability distribution $\mu_f = P_{\sigma_f^2} = \mathcal{N}(0, \sigma_f^2)$ on $\mathcal{B}(\mathbb{R})$, where

$$\sigma_f^2 = E\left[\xi^T(f)\right]^2 = \mathcal{J}^T(f^2) \tag{16}$$

depends on the choice of $f \in \mathcal{L}^2(T, dF)$. Consequently, the system $\mathcal{H}^T$ is a Gaussian with the Gaussian distributions

$$\left\{P_{\sigma_f^2} = \mathcal{N}(0, \sigma_f^2),\ f \in \mathcal{L}^2(T, dF)\right\} \tag{17}$$

and in accordance with the Proposition 3.1, we obtain the following statement.

**Corollary 1.** *For an interval $T \subseteq [t_0, +\infty)$, let $\mathcal{H}^T = \left\{\xi^T(f),\ f \in \mathcal{L}^2(T, dF)\right\}$ be a system of stochastic integrals (2) with the Gaussian distributions (17). Then, if $r_1, ..., r_n \in \mathbb{R}$ are the roots of normalized Hermite polynomial $\mathfrak{h}_n(x) \in \mathcal{L}^2(\mathbb{R}, P)$ of degree $n \geqslant 1$, there exists a system of discrete random variables $\left\{\zeta_n^T(f),\ f \in \mathcal{L}^2(T, dF)\right\}$, where for any $f \in \mathcal{L}^2(T, dF)$, random variable $\zeta_n^T(f)$ takes $n$ values, $r_i(f) = \sigma_f \cdot r_i \in \mathbb{R}$, $i = 1, ..., n$, with the probabilities $\lambda_i = 1/(n\left[\mathfrak{h}_{n-1}(r_i)\right]^2)$, $i = 1, ..., n$, such that*

$$E\left[\xi^T(f)\right]^k = E\left[\zeta_n^T(f)\right]^k, \quad k = 0, 1, 2, ..., 2n-1, \tag{18}$$

*and $\zeta_n^T(f)$ converges in distribution to the corresponding Gaussian random variable $\xi^T(f)$ as $n \nearrow \infty$.*

*Proof.* As for any $f \in \mathcal{L}^2(T, dF)$, the stochastic integral (2) is a Gaussian random variable $\xi^T(f)$ with Gaussian probability distribution $P_{\sigma_f^2} = \mathcal{N}(0, \sigma_f^2)$ on $\mathcal{B}(\mathbb{R})$, from the Proposition 3.1 it follows that

$$\int_R x^k dP_{\sigma_f^2}(x) = (\sigma_f \cdot r_1)^k \lambda_1 + ... + (\sigma_f \cdot r_n)^k \lambda_n, \quad k = 0, 1, ..., 2n-1, \tag{19}$$

where $r_1, ..., r_n \in \mathbb{R}$ are the roots of normalized Hermite polynomial $\mathfrak{h}_n(x) \in \mathcal{L}^2(\mathbb{R}, P)$ of degree $n \geqslant 1$, and $\lambda_i = 1/(n\left[\mathfrak{h}_{n-1}(r_i)\right]^2)$, $i = 1, ..., n$.

Taking into account the proof of Proposition 3.1, from (19) it follows that for any $f \in \mathcal{L}^2(T, dF)$ the equality (18) holds and the discrete random variable $\zeta_n^T(f)$ converges in distribution to the corresponding Gaussian random variable $\xi^T(f)$ as $n \nearrow \infty$■

Moreover, if $T = [t_0, +\infty)$, then for any $[t_0, t) \subseteq T$ and $f \in \mathcal{L}^2(T, dF)$, on the basis of stochastic integral (2), we can define

$$\xi^t(f) = \int_{t_0}^t f(u) d\eta(u) = \int_T f(u) \chi_{[t_0,t)} d\eta(u), \quad \chi_{[t_0,t)}(u) = \begin{cases} 1, & u \in [t_0, t) \\ 0, & u \notin [t_0, t) \end{cases}. \tag{20}$$

The random variable $\xi^t(f)$, $t > t_0$, is defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as orthogonal stochastic measure (1) and is measurable with respect to the sigma-field $\mathcal{F}_t \subseteq \mathcal{F}$ generated by the system $\eta^t = \{\eta(\Delta), \Delta \in \mathcal{B}([t_0, t))\}$.

Assume that the orthogonal stochastic measure (1) is a Gaussian. Then, for a fixed function $f \in \mathcal{L}^2(T, dF)$ such that $f \neq 0$ a.e. on $\mathcal{B}(T)$, the stochastic integral (20), as random variable $\xi^t(f)$, has a Gaussian probability distribution $P_{\sigma_t^2} = \mathcal{N}(0, \sigma_t^2)$ on the Borel sigma-field of $\mathbb{R}$, where

$$\sigma_t^2 = E\left[\xi^t(f)\right]^2 = \int_{t_0}^{t} f^2(u) dF(u) = \mathcal{J}^t(f^2) \tag{21}$$

depends on $t > t_0$. Having in mind that $\eta(t_0) = 0$ a.s., it follows that $\xi^{t_0}(f) = 0$ a.s.. Moreover, the obtained system $\{\xi^t(f),\ t \in T\}$ of the stochastic integrals (20), is actually a centered Gaussian random process defined on $(\Omega, \{\mathcal{F}_t, t \in T\}, \mathbb{P})$. In this case, on the basis of Proposition 3.1 again, we get the next statement.

**Corollary 2.** *For a fixed function $f \in L^2(T, dF)$, where $T = [t_0, +\infty)$ and $f \neq 0$ a.e. on $\mathcal{B}(T)$, let $\{\xi^t(f),\ t \in T\}$ be a centered Gaussian random process, such that $\xi^{t_0}(f) = 0$ a.s. and for any fixed $t > t_0$, $\xi^t(f)$ is the stochastic integral defined with (20). If $r_1, ..., r_n \in \mathbb{R}$ are the roots of normalized Hermite polynomial $\mathfrak{h}_n(x) \in L^2(\mathbb{R}, P)$ of degree $n \geqslant 1$, then there exists a system of discrete random variables $\{\zeta_n^t(f),\ t \in T\}$ where $\zeta_n^{t_0}(f) = 0$ a.s. and for any fixed $t > t_0$, random variable $\zeta_n^t(f)$ takes $n$ values $r_i(t) = \sigma_t \cdot r_i \in \mathbb{R}$, $i = 1, ..., n$, with the probabilities $\lambda_i = 1/(n\left[\mathfrak{h}_{n-1}(r_i)\right]^2)$, $i = 1, ..., n$, such that*

$$E\left[\xi^t(f)\right]^k = E\left[\zeta_n^t(f)\right]^k, \quad k = 0, 1, 2, ..., 2n-1.$$

*For any $t > t_0$, the values $\sigma_t > 0$ are determined with (21) and $\zeta_n^t(f)$ converges in distribution to the Gaussian random variable $\xi^t(f)$ as $n \nearrow \infty$.*

As opposed to Corollary 3.1, where for a fixed interval $T$, the obtained discrete random variables are indexed with the functions $f \in L^2(T, dF)$, in Corollary 3.2, for a fixed function $f \in L^2(T, dF)$, the obtained discrete random variables are indexed with the parameter $t \in T$ such that $[t_0, t) \subseteq T$. However, in both cases, for the given roots of normalized Hermite polynomial $h_n(x) \in L^2(\mathbb{R}, P)$ of some degree $n \geqslant 1$, and for the known values (16), i.e. (21), one can carry out a moment matching discretization of the system of stochastic integrals (2), i.e. (20), respectively.

## References


[1] **M. Abramowitz, I. A Stegun** (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. Dover, New York, 1972.

[2] **N. I. Akhiezer**. *The classical moment problem and some related questions in analysis*. Hafner Publishing Co., New York, 1965.

[3] **P. Brandimarte**. *Numerical Methods in Finance and Economics*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[4] **P. Billingsley**. *Convergence of Probability Measures*. John Wiley & Sons, Inc., New York, 1968

[5] **S. Janković, M. Merkle**. A mean value theorem for systems of integrals. *Journal of Mathematical Analysis and Applications,* 342, 334-339, 2008.

[6] **J. Ma, V. Rokhlin, S. Wandzura**. Generalized Gaussian quadrature rules for systems of arbitrary functions. *SIAM J. Numer. Anal.* 33/3, 971-996, 1996.

[7] **S. Mitrinović, R. Janić**. *Uvod u specijalne funkcije.* Građevinska knjiga, Beograd, 1975.

[8] **T. Pennanen, M. Koivu**. Epi-convergent discretizations of stochastic programs via integration quadratures. *Numer. Math.* 100, 141-163, 2005.

[9] **Yu. A. Rozanov**. *Probability Theory, Random Processes and Mathematical Statistics.* Springer Science + Business Media, Dordrecht, 1995.

Simpozijum MATEMATIKA I PRIMENE, Matematički fakultet, Univerzitet u Beogradu, 2015 ,Vol. VI(1)

# O razvoju geometrijskog mišljenja u nastavi matematike prema van Hiele-ovoj teoriji

**Nives Baranović**

*Filozofski fakultet u Splitu*
*e-mail: nives@ffst.hr*

**Apstrakt.** Iako su geometrijska znanja privilegirano sredstvo za razvoj matematičkog mišljenja, a uz to su i društveno korisna, u nastavnoj praksi geometrija odmiče prema marginama: mnogi učenici je ne vole učiti, a ima i nastavnika koji je ne vole poučavati.

Baveći se tom problematikom, brojna istraživanja o učenju i poučavanju geometrije zadnjih desetljeća otkrivaju zašto mnogi učenici imaju teškoće pri učenju geometrijskih sadržaja te daju smjernice koje bi mogle pomoći u savladavanju tih teškoća. Jedna od istaknutih teorija, koja se time bavi, van Hiele-ova teorija, postavljena je krajem pedesetih godina prošlog stoljeća i stalno se razvija kroz razne vrste novih istraživanja.

U radu će biti prikazane glavne ideje van Hiele-ove teorije s primjenom na učenje geometrije: model razvoja procesa mišljenja kroz pet razina, osnovne karakteristike tog modela te preporučena strategija poučavanja za uspješno napredovanje prema opisanom modelu.

Osim opisa same teorije, biti će prikazani rezultati jednog eksperimentalnog istraživanja utemeljenog na van Hiele-ovoj teoriji koji potvrđuju opisano.

**Ključne reči:** geometrija; razvoj geometrijskog mišljenja; strategija poučavanja; van Hiele-ova teorija

## 1. Uvod

Geometrijsko mišljenje je apsolutno nužno potrebno u svakoj grani matematike. Gledajući s povijesnog aspekta, upravo je geometrija zaslužna za izgradnju aksiomatskih sustava, a zahvaljujući geometrijskom pogledu na stvari osiguran je točan i ispravan uvid u mnoga istraživanja, npr. razvoj kompleksne analize (vidjeti [5], str. 389).

Osim toga, geometrijsko mišljenje je od vitalne važnosti za svakoga od nas. U svakodnevnom životu razmještamo namještaj po kući kako bi prostor bio optimalno iskorišten, sami sastavljamo namještaj kupljen u dijelovima, snalazimo se u nepoznatim gradovima čitajući lokacije s raznih vrsta mapa, koristimo se GPS-om na putovanjima kroz nepoznate dijelove zemlje, itd. Sve to ostvarujemo, s većom ili manjom uspješnošću, zahvaljujući geometrijskim znanjima i služeći se geometrijskim mišljenjem koje smo na temelju tih znanja razvili.

Smatra se da djeca, kada krenu u školu već imaju potreban potencijal da razmišljaju geometrijski i da svijet oko sebe doživljavaju i vide matematički. No, kako bi kroz obrazovanje svoj potencijal učenici uspješno razvijali, treba im osigurati odgovarajuću potporu (vidjeti [9])

Ako zaista želimo osnažiti naše učenike za život poslije škole, trebamo ih pripremiti da koriste, razumiju, kontroliraju i mijenjaju nešto što možda još i ne postoji. To je moguće ako im osiguramo da istinski razviju matematički način mišljenja, čiji je važan sastavni dio i geometrijsko mišljenje. Geometrijska znanja koja se uče kroz nastavu matematike su važna, ali još važniji su procesi mišljenja, način gledanja na stvari, navike uma (vidjeti [5], str. 401).

No, još prije 2000 godina Euklid je svojom rečenicom „Nema kraljevskih putova u geometriji" dao naslutiti da učenje i poučavanje geometrije nije nimalo jednostavan posao pa time ni razvoj geometrijskog mišljenja. Iako do danas još nitko nije pronašao magični štapić koji bi promijenio istinitost Euklidove tvrdnje, ipak su brojna istraživanja u matematičkom obrazovanju, koja su intenzivirana zadnjih 60-tak godina, otkrila moguće uzroke teškoća i dala smjernice kojima bi se premostile neke od njih.

Osim teškoća uzrokovanih samom prirodom stvari, iz raznih razloga (preopširan plan i program rada, mala satnica, brojnost učenika po razredima itd.), nastavnici se često stavljaju u poziciju da se u nastavi geometrije više bave sadržajima nego procesima, a učenici u takvim okolnostima geometrijske sadržaje (definicije pojmova i njihova svojstva, formule ...) najčešće uče na pamet, dok njihovu primjenu i algoritme rješavanja savladavaju mehanički bez razumijevanja.

## 2. O nastanku i razvoju teorije van Hiele

Suočavajući se s raznim teškoćama pri učenju i poučavanju matematike učenika srednjoškolske dobi, nizozemski nastavnici Dina van Hiele-Geldof i Pierre Marie van Hiele, njezin suprug, došli su do značajnih otkrića zašto učenici imaju teškoće i kako bi ih trebalo poučavati da te teškoće savladaju.

Svoja zapažanja opisali su u doktorskom radu koji je objavljen 1957. godine na nizozemskom jeziku. Najvažniji doprinosi tog rada su u postavljanju teorije o razvoju procesa mišljenja kroz hijerarhijsku strukturu te faze učenja koje osiguravaju napredovanje po toj strukturi. Kako je Dina uskoro nakon objave doktorskog rada umrla, Pierre (umro je 2010. godine, u svojoj 100. godini života) je teoriju sam dodatno pojasnio i unaprijedio. Njima u čast, teorija se naziva van Hiele-ova teorija (vidjeti [3]).

Ubrzo nakon izlaska njihova rada, 1960-tih godina, Sovjetski savez je nakon opširnog istraživanja utemeljenog na van Hiele-ovoj teoriji, dao preraditi aktualni kurikulum geometrije. Širenje utjecaja na druge dijelove svijeta odvijalo se sporo vrlo vjerojatno zbog toga što je rad pisan na nizozemskom jeziku te je bio dostupan samo manjem broju svjetske populacije. Nakon Sovjetskog saveza, o van Hiele-ovoj teoriji se najprije počelo pisati u Sjevernoj Americi početkom 1970-tih godina, a veći interes se pojavio tek 1980-tih godina, kada se rad preveo na engleski jezik, pod nazivom *Structure and insight*, iako je i tada broj tiskanih primjeraka bio nedovoljan pa time i nemogućnost dolaska do izvornog dijela (vidjeti [6], [11]).

Nažalost, bilo je i onih koji su privilegij dostupnosti djelu (na nizozemskom ili engleskom jeziku) pomalo i zloupotrebljavali te su svjesno, ili zbog nerazumijevanja, iskrivljavali originalnu misao van Hiele-a, a neki su, negirajući dovršenost djela, stvarali svoj materijal kao novi doprinos (vidjeti [2], [3]).

Van Hiele procese mišljenja razmatra općenito, a geometriju uzima kao primjer na kojem testira svoju teoriju, vjerojatno zato što je geometrija vrlo prikladna za isticanje ključnih karakteristika pojedinih razina mišljenja kao i za odabir primjera kroz faze poučavanja. Kako mnogima nije bilo dostupno originalno djelo, s obzirom na to da je pisano na nizozemskom jeziku, a prijevodi nisu u potpunosti vjerni originalu, s vremenom se njegova teorija, koja je općeg karaktera i primijenjiva na različita područja, počela tumačiti kao teorija o razvoju geometrijskog mišljenja (vidjeti [2], [3]). Iz tih razloga većina istraživanja koja se temelje na van Hiele-ovoj teoriji primjere pronalaze upravo u geometriji, ali ima i oni koji teoriju van Hiele-a primjenjuju na druga područja matematike (vidjeti [6]).

Unatoč svim teškoćama i okolnostima, od tada do danas teorija je prolazila svoj razvojni put. Brojna obrazovna istraživanja s različitim uzrastom učenika potvrdila su njezinu valjanost, a kritički osvrti na samu teoriju doveli su do dodatnih pojašnjenja i suptilnih shvaćanja, kako same strukture tako i razvoja procesa mišljenja kroz opisanu strukturu (vidjeti [1], [3], [4], [11], [12]).

Mnogi predmetni kurikulumi i prateći udžbenici diljem svijeta nastoje poštivati preporuke opisane van Hiele-ovom teorijom, posebno dio vezan za geometriju. Neki istraživači u obrazovanju razvijaju različite vrste instrumenata u dijagnostičke svrhe (vidjeti [6], [12]), dok drugi pak razvijaju različite vrste primjera u svrhu planiranja odgovarajuće strategije poučavanja (vidjeti [1], [4], [6], [11]).

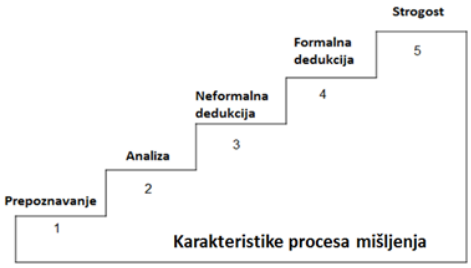
S obzirom da je van Hiele bio srednjoškolski nastavnik kojem se, ni nakon što je doktorirao, nije pružila mogućnost zapošljavanja na visokoškolskoj ustanovi, njegov interes je prvenstveno bio usmjeren na prve četiri razine mišljenja te je i većina istraživača manje pozornosti usmjeravala petoj razini (neki su proučavali samo prve tri), posebno što se većina istraživanja provodila u osnovnoškolskom i srednjoškolskom obrazovanju (vidjeti [4]).

## 3. Van Hiele model

Van Hiele-ova teorija hijerarhijski strukturira karakteristike procesa mišljenja kroz pet razina, a svaku razinu naziva prema ključnoj karakteristici: prepoznavanje, analiza, neformalna dedukcija, formalna dedukcija i strogost (slika 1).Tako na primjer, oni koji uče euklidsku geometriju da bi postigli odgovarajuću zrelost geometrijskog mišljenja, proces učenja bi trebali započeti prepoznavanjem određenih objekata, zatim uočavati svojstava promatranih objekata i stvarati veze među njima pa tek onda napredovati do izvođenja formalnih dokaza, a kao vrhunac učenja trebali bi postići razumijevanje i drugih, neeuklidskih geometrija (vidjeti [4], [10]).

Svakoj razini odgovara točno određeni proces mišljenja i jezik komuniciranja. Na svakoj razini učenici istražuju odgovarajuće objekte i pri tome razvijaju jezik primjeren toj razini. Kao rezultat istraživanja i odgovarajućih procesa mišljenja nastaje proizvod koji postaje predmetom proučavanja na sljedećoj razini.

Van Hiele je same razine u početku označavao brojevima od 0 do 4, ali se pokazalo praktičnije korištenje brojeva od 1 do 5 jer označavanjem od 0 do 4 može doći do nesporazuma kad se priča npr. o razini broj 2, koja je u tom slučaju zapravo treća razina mišljenja. Iz tog razloga se u ovom radu koristi numeracija od 1 do 5.




**Slika 1.** Van Hiele model mišljenja

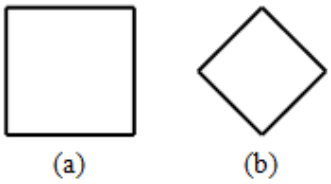
S obzirom da je tema ovog rada razvoj geometrijskog mišljenja, primjeri koji se navode uzimaju se iz geometrije, posebno što se na kraju želi dati kratki osvrt na jedno eksperimentalno istraživanje koje potvrđuje izneseno.

### 3.1. Razina prepoznavanja (eng. Recognition)

Na prvoj razini, učenici vizualno prepoznaju geometrijske objekte prema globalnom izgledu, ne identificirajući eksplicitno njihova svojstva, a prostor doživljavaju kao nešto čime su okruženi. Na isti način usvajaju i rječnik geometrijskih pojmova: imena likova uče na temelju njihovog oblika, a ne na temelju njihovih svojstava.

Neki istraživači ovu razinu nazivaju i razinom vizualizacije upravo zbog vizualnog prepoznavanja geometrijskih oblika, ali to nije u potpunosti korektno jer se time pojam vizualizacije sužava samo na prepoznavanje, dok je vizualizacija puno širi pojam.

Osoba koja funkcionira na ovoj razini može prepoznati kvadrat prema obliku u bilo kojem položaju (slika 2), a onaj tko nije savladao ovu razinu, kvadrat će prepoznati samo u standardnom položaju (slika 2, dio (a)), ali ne i u nekom drugom položaju (slika 2, dio (b)). U ovoj fazi, učenici mogu crtati odgovarajuće likove (na papiru, u mreži, na geoploči . . . ) prema nekom uzorku.




**Slika 2.** Kvadrati

Nakon procesa prepoznavanja, učenici postupno objekte počinju razvrstavati u grupe, ali opet samo prema njihovom izgledu. Tako na primjer, kvadrat i pravokutnik ili pravokutnik i romb neće staviti u istu grupu jer *ne izgledaju* jednako.

Dakle, na prvoj razini, objekti mišljenja su geometrijski oblici (likovi, tijela) u cjelini i njihov izgled, a proizvod mišljenja su grupe (klase) oblika koje izgledaju *slično*.

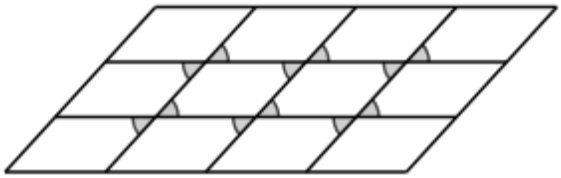
### 3.2. Razina analize (eng. Analysis)

Nakon prepoznavanja i grupiranja objekata, učenici postupno počinju uočavati i analizirati njihova svojstva (pojedinačno ili grupe), ali ih međusobno ne povezuju niti unutar istog objekta niti među objektima. Postupno savladavaju i terminologiju za opisivanje uočenih svojstava.

Tako uočavaju da jednakokračni trokut ima dvije stranice jednakih duljina i dva kuta jednakih veličina, ali još uvijek ne stvaraju vezu da se nasuprot jednakih stranica nalaze jednaki kutovi. Isto tako će uočiti da kvadrat ima sve četiri stranice jednake duljine i sva četiri kuta prava te da pravokutnik ima nasuprotne stranice jednakih duljine i sva četiri kuta prava, ali još uvijek ne povezuju da je kvadrat ujedno i pravokutnik jer ispunjava sva svojstva koja vrijede za pravokutnik.

Osoba koja funkcionira na ovoj razini, objekte razvrstava u grupe na temelju njihovih svojstava (najčešće samo jednog) te postupno stvara svoje definicije promatranih objekata opisivanjem uočenih svojstava, ne razlikujući još uvijek značenje nužnih i dovoljnih svojstava (vidjeti [6]).

Nakon dovoljno eksperimentalnog iskustva učenici mogu izvoditi i generalizacije, ali samo unutar određene grupe objekata. Na primjer, ako bi učenicima dali mrežu paralelograma, oni bi nakon isticanja vršnih kutova (koji su jednake veličine) mogli uspješno utvrditi da su i nasuprotni kutovi paralelograma jednake veličine. No to se još uvijek odnosi na tu grupu koju promatraju, a ne općenito za sve paralelograme (Slika 3). Više o ovoj vrsti aktivnosti može se pročitati u [6].



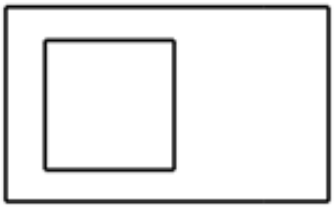
**Slika 3.** Mreža paralelograma

Na drugoj razini, objekti mišljenja su grupe (klase) oblika koje izgledaju slično, a proizvod mišljenja su svojstva oblika, određene grupe ili pojedinačnih.

### 3.3. Razina neformalne dedukcije (eng. Order)

Na trećoj razini, učenici uspostavljaju logičke odnose između svojstava jednog određenog objekta te između svojstva različitih objekata. Na temelju toga su u mogućnosti stvarati hijerarhijsku klasifikaciju objekata.

Na primjer, osobe koje funkcioniraju na ovoj razini u stanju su zaključiti da su zbog paralelnosti nasuprotnih stranica paralelograma i nasuprotni kutovi jednake veličine (povezivanje svojstava jednog objekta). Mogu razumjeti i da je svaki kvadrat pravokutnik jer ispunjava sva svojstva pravokutnika (povezivanje svojstava među objektima), ali i da svaki pravokutnik nije kvadrat (slika 4).



**Slika 4.** Kvadrat i pravokutnik

Na ovoj razini učenici počinju razmišljati što je nužno, a što dovoljno da se neki objekt opiše, prihvaćaju različite vrste definicija istog pojma te su u mogućnosti prepoznati nepotpune definicije i oblikovati ih u korektne. Prema van Hiele-u, izravno poučavanje formalnih definicija prije treće razine, a da nisu izvedene kao rezultat nekih prethodnih aktivnosti, unaprijed je osuđeno na neuspjeh. Učenike bi trebalo uključiti u proces definiranja i dopustiti im da stvaraju vlastite definicije na svakoj razini (vidjeti [6]).

Učenici su u mogućnosti pratiti proces dokazivanja, ali još uvijek nisu dovoljno samostalni da bi sami gradili svoj dokaz. Služe se logičkim oblikom izražavanja „ako je . . . onda je . . . ” pri izricanju određenih pravila, tvrdnji i sl., ali još uvijek nisu na dovoljno operativnoj razini. Uočavaju postojanje sustava definicija, aksioma, teorema i dokaza, ali još uvijek ne mogu u potpunosti razumjeti njihovu ulogu i funkcionalnost.

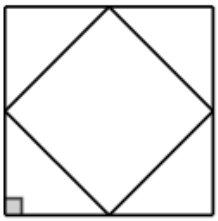
Na trećoj razini, objekti mišljenja su svojstva promatranih objekata, pojedinačnih ili određene grupe (klase), a proizvod mišljenja su odnosi između svojstava jednog objekta ili među objektima.

### 3.4. Razina formalne dedukcije (eng. Formal deduction)

Kada je proces mišljenja učenika dosegao ovu razinu zrelosti, onda su oni u mogućnosti shvatiti značenje i ulogu definicije, aksioma (postulata), teorema i dokaza unutar deduktivnog aksiomatskog sustava kao cjeline.

Shvaćaju značenje nužnog i dovoljnog uvjeta te su u mogućnosti izvesti obrat postavljene tvrdnje. Na ovoj razini mišljenja, učenici su u mogućnosti samostalno izvoditi dokaze određenih tvrdnji i to na više različitih načina. Još uvijek nisu u mogućnosti operativno provoditi indirektan dokaz i dokaz po kontrapoziciji.

Na primjer, da bi dokazali da je neki četverokut kvadrat znaju da treba dokazati da su sve četiri stranice jednake duljine, ali i da su sva četiri kuta prava. Za četverokut upisan u kvadrat (slika 5) bili bi u mogućnosti dokazati da je kvadrat služeći se različitim teoremima (o sukladnosti trokuta, o srednjici trokuta, koristeći Pitagorin poučak itd.).



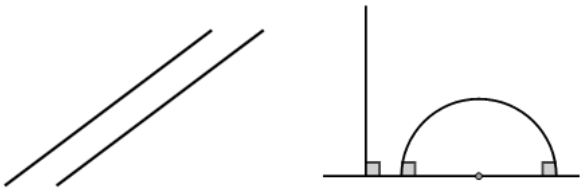
**Slika 5.** Četverokut upisan u kvadrat

Na četvrtoj razini, objekti mišljenja su odnosi između svojstava jednog objekta ili među objektima, a proizvod mišljenja je deduktivni aksiomatski sustav.

### 3.5. Razina strogosti (eng. Rigor)

Na ovoj razini mišljenja učenici mogu proučavati različite aksiomatske sustave i međusobno ih uspoređivati.

Učenici koji funkcioniraju na ovoj razini u stanju su shvatiti da postoje i drugi geometrijski sustavi osim euklidskog, mogu vidjeti sličnosti i razlike tih sustava i međusobno ih uspoređivati. Na primjer, iako vizualni prikaz paralelnih pravaca u euklidskoj i neeuklidskoj ravnini (slika 6) može izgledati kao da se radi o različitim stvarima, na ovoj razini učenici taj pojam mogu tumačiti s razumijevanjem u različitim geometrijskim sustavima te izvoditi odgovarajuće jednadžbe i odnose.



**Slika 6.** Paralelni pravci u euklidskoj i neeuklidskoj ravnini

Tek na ovoj razini učenici mogu razumjeti značenje konzistentnosti, nezavisnosti i potpunosti određenog aksiomatskog sustava.

Na petoj razini, objekti mišljenja su deduktivni aksiomatski sustavi, a proizvod mišljenja su odnosi (izomorfizam) između različitih sustava.

Osim prikazivanja samog modela procesa mišljenja kroz pet razina, van Hiele je istaknuo i neke njegove opće karakteristike. Te karakteristike su posebno važne nastavnicima jer im mogu poslužiti kao neka vrsta vodiča u fazi pripremanja materijala za nastavu.

## 4. Osnovne karakteristike van Hiele modela

Prije svega, opisani model razvoja mišljenja je hijerarhijski niz razina redoslijednog karaktera. To znači da svaki pojedinac razine mišljenja mora savladavati navedenim redom kako bi napredovao do određene zrelosti mišljenja. Da bi netko uspješno funkcionirao na određenoj razini, osnovni preduvjet je steći znanja, vještine i jezik prethodnih razina.

Na primjer, učenici nisu u mogućnosti samostalno izvoditi dokaze teorema, ako prethodno nisu savladali prve tri razine. Prema van Hiele-u, potrebno je gotovo dvije godine kontinuiranog poučavanja kako bi učenici iskusili vrijednost dedukcije, a zatim još dodatnog poučavanja kako bi shvatili značenje njezinog koncepta (vidjeti [11], str 213).

Drugo, napredovanje iz jedne razine u drugu ili izostanak napredovanja ne ovisi o starosnoj dobi niti sazrijevanju već više o načinu učenja, obrazovnom iskustvu, razumijevanju (vidjeti [10], [12]). Naime, neke nastavne metode omogućavaju učenicima da preskoče razinu, neke ubrzavaju napredak dok neke usporavaju ili čak sprečavaju kretanje između razina (vidjeti [4], str. 4).

Van Hiele je posebno naglašavao trijadu: nastavnik - učenik - tema, tj. važnost prilagođavanja sadržaja i načina poučavanja učenicima jer u suprotnom učenici ne samo da neće napredovati, već mogu čak i nazadovati (vidjeti [3], [13]).

Treće, proizvod mišljenja tekuće razine postaje predmet proučavanja sljedeće razine (Slika 7).


**Slika 7.** Prikaz razina prema objektima mišljenja

Dakle, učenici najprije prepoznaju određene oblike te ih na temelju globalnog izgleda razvrstavaju u grupe. Zatim formirane grupe analiziraju i uočavaju njihova svojstva, a tek onda stvaraju veze među tim svojstvima. Opisana svojstva strukturiraju u dobro uređen sustav, a zatim taj sustav uspoređuju sa sličnim sustavima.

Četvrto, svaka razina ima svoj način izražavanja (jezično i simbolički). Tako na primjer, na razini 2 dopušteno je razlikovati kvadrate i pravokutnike, ali na razini 3 za kvadrat se može koristiti i naziv pravokutnik, romb ili paralelogram.

Peto, ako se učenici nalaze na jednoj razini, a nastava se provodi na drugoj, višoj razini, željeno učenje i napredak se neće pojaviti. Naime, ako se nastavnik služi jezikom (govornim ili simboličkim), primjerima i sadržajem koji nadilazi razinu mišljenja njegovih učenika, učenici neće moći pratiti takav proces poučavanja.

Ni učenici koji se nalaze na različitim razinama razmišljanja ne mogu se razumjeti jer se koriste različitim jezikom i različitim načinima rješavanja problema te različito tumače iste riječi ili situacije (vidjeti [12], str. 4). Isto tako, kada nastavnik poučava sadržaje i služi se jezikom koji nije prilagođen razini mišljenja učenika, prema van Hiele-u, njegovi učenici ga ne mogu razumjeti niti uspješno pratiti takvu nastavu što dovodi do frustracija i obeshrabrenja (vidjeti [4]). Stoga je jako važno da nastavnici najprije ispitaju razinu mišljenja svojih učenika te na koji način interpretiraju određene sadržaje kako bi učinkovito s njima mogli komunicirati. U suprotnom, ako učenici ne mogu pratiti sadržaje koji se poučavaju, onda pribjegavaju učenju napamet, a takve sadržaje brzo zaboravljaju i nisu ih u stanju primijeniti (vidjeti [10]).

Na primjer, svaki nastavnik je vjerojatno u svojoj praksi doživio da učenik, nakon izračunavanja površine pravokutnika po formuli $P = ab$ (gdje su $a$ i $b$ duljine stranica pravokutnika), za računanje površine trokuta koristi formulu $P = abc$ (gdje su $a$, $b$ i $c$ duljine stranica trokuta). To je pokazatelj da je formula naučena napamet, bez razumijevanja. Ako učenici nisu iskustveno doživjeli kako su određene formule izvedene, već su ih dobili kao gotovi alat koji trebaju primijeniti, oni će te formule naučiti napamet, često će u njihovim primjenama griješiti, a vjerojatno će ih i zaboraviti.

Od trenutka prijevoda originalnog rada na engleski jezik, sve do danas provode se brojna istraživanja temeljena na van Hiele-ovom modelu i to sa različitim uzrastima. Neka od njih potvrđuju navedene karakteristike, a neka ih dodatno pojašnjavaju i nadopunjuju.

Tako neka istraživanja potvrđuju da se razine moraju savladavati od prve pa na dalje bez preskakanja (osim iznimno talentiranih učenika) te da svaka razina ima svoj specifičan jezik i posebnu interpretaciju istih sadržaja,

ali razine nisu diskretne, kako ih je razmatrao van Hiele na samom početku, već je moguće biti na prijelazu između razina (vidjeti [8], [12]). Osim toga, ako se nekim sadržajima netko bavi više i detaljnije nego drugima, onda razine mišljenja iste osobe za te sadržaje mogu biti različite (vidjeti [1], [8]).

Osim procesom mišljenja učenika, koji je prikazao kroz petodijelni hijerarhijski model, van Hiele se bavio i strategijom učenja koja bi osigurala učenicima napredak prema opisanom modelu, posebno naglašavajući važnost prilagođavanja sadržaja i strategije poučavanja. Zato bi nastavne metode, strategija nastave, sadržaji i nastavni materijali trebali stalno biti u fokusu razmatranja svakog nastavnika (vidjeti [4], str. 5).

## 5. Faze učenja

Za uspješno napredovanje kroz navedene razine razmišljanja, van Hiele-ova teorija preporučuje proces učenja u pet faza: faza informiranja, usmjerenog vođenja, objašnjavanja, aktivnosti otvorenog tipa, faza povezivanja (vidjeti [13]). Osim opisivanja pojedine faze procesa učenja, u ovom radu se daje i odgovarajući primjer aktivnosti radi ilustracije opisanog.

**Faza pitanja i informiranja (eng. Inquiry/Information)**

Postavljanjem ciljano odabranih pitanja, nastavnik učenike uvodi u raspravu i aktivnosti o određenoj temi. Time se postiže dvostruki cilj: kroz uvodnu raspravu, nastavnik otkriva što učenici znaju o određenoj temi i istodobno ih usmjerava na temu koja će se obrađivati.

Primjer aktivnosti: Učenike uvodimo u raspravu postavljanjem pitanja: Što je to kvadrat? Pravokutnik? Paralelogram? Romb? Po čemu su slični? Po čemu se razlikuju? Što mislite, bi li kvadrat mogao biti romb? Može li romb biti kvadrat? Itd. Nastavnik može imati pripremljene i modele ili slike odgovarajućih likova.

**Faza usmjerenog vođenja (eng. Directed orientation)**

U fazi usmjerenog vođenja učenici samostalno istražuju svojstva pojmova prema pažljivo strukturiranim zadacima: sami crtaju, mjere, izračunavaju, povezuju itd. kako bi otkrili ili razjasnili određene odnose koji im do tada nisu bili poznati ili su im bili nejasni. Preporuka je da zadaci budu kratki s ciljem dobivanja točno određenog odgovora. Pri izradi programiranog materijala potrebno je voditi računa o razinama mišljenja koje su učenici do tada savladali.

Primjer aktivnosti: Nastavnik može zadati učenicima da u mreži ili na geoploči nacrtaju/prikažu pravokutnik sa susjednim stranicama jednake duljine, romb s jednakim dijagonalama, paralelogram s dva prava kuta itd. Drugim riječima, zadatke bi trebalo osmisliti tako da učenike vode prema potrebnim zaključcima, ali do kojih trebaju samostalno doći.

**Faza objašnjavanja (eng. Explicitation)**

Nakon što su proveli zadane aktivnosti i izvršili postavljene zadatke, učenici svojim riječima opisuju ono što su radili te međusobno razmjenjuju i objašnjavaju zaključke do kojih su došli. Važno je da učenici najprije sami iznesu svoje zaključke bez utjecaja nastavnika. Nastavnik bi trebao pozorno pratiti izražavanje učenika te ih usmjeravati na korištenje točnog i primjerenog načina izražavanja.

Primjer aktivnosti: Ako su uspješno riješili postavljene zadatke, do izražaja bi trebali doći odnosi među likovima: pravokutnik kojemu su susjedne stranice jednake duljine je zapravo kvadrat, romb s jednakim dijagonalama je također kvadrat, paralelogram s dva prava kuta je pravokutnik itd. Neki od učenika će možda već u ovoj fazi doći do zaključka da je svaki kvadrat ujedno i romb, ali da svaki romb ne mora biti kvadrat.

**Faza aktivnosti otvorenog tipa (eng. Free orientation)**

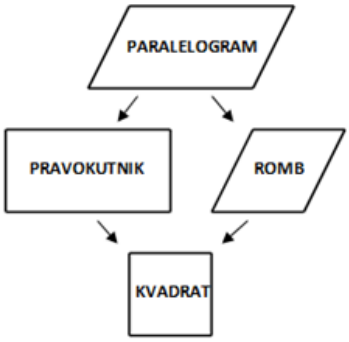
Nakon što su izveli odgovarajuće zaključke, učenici ih primjenjuju na rješavanje složenijih zadataka. U ovoj fazi su dobrodošli zadaci s više ključnih koraka, zadaci koji se mogu riješiti na više različitih načina, posebno zadaci otvorenog tipa. Cilj je da sami rješavaju zadatke na način koji je njima najprimjereniji, a za vrijeme dok su oni okupirani istraživanjem/rješavanjem, nastavnik ih može dodatno upoznati promatrajući njihove procese rješavanja postavljenih zadataka.

Primjer aktivnosti: Što možete reći o sjecištu dijagonala pravokutnika, kvadrata, romba, paralelograma? Usporedite kutove nastale presijecanjem dijagonala. Objasnite zašto se površina romba i kvadrata može odrediti kao polovica umnoška duljina dijagonala. Itd.

**Faza povezivanja (eng. Integration)**

Nakon provedenih aktivnosti izuzetno je važno da učenici razmotre i opišu čime su se bavili i do kojih zaključaka su došli. Potrebno je objediniti ono što su istražili i naučili te strukturirati jednu funkcionalnu mrežu objekata i relacija (vidjeti [4], [10], [12] ). Nastavnik ih u ovoj aktivnosti vodi i usmjerava radi sveobuhvatnosti, preciznosti i točnosti onoga što objedinjuju i strukturiraju.

Primjer aktivnosti: Nastavnik usmjerava učenike prema klasifikaciji likova koji su razmatrani, u ovom slučaju prema klasifikaciji paralelograma. Strukturiranje se može vršiti na različite načine, a posebno je koristan vizualni prikaz, na kojem će se istaknuti odgovarajuće veze (slika 8).




**Slika 8.** Klasifikacija paralelograma

Kada rade na ovaj način, učenici nemaju osjećaj da su učili nešto novo, a ipak su razvijali nove procese mišljenja koji će zamijeniti one stare, čime će zapravo doseći i novu razinu mišljenja prema van Hiele-u (vidjeti [4], [13]).

## 6. Rezultati jednog eksperimentalnog istraživanja

Eksperimentalno istraživanje je obuhvatilo sve studente učiteljskog studija u Splitu i Zadru koji su u ljetnom semestru 2014/2015. godine pohađali nastavu geometrije. Ukupno je sudjelovalo 90 studenata (52 studenata iz Splita činili su eksperimentalnu grupu, a 38 studenata iz Zadra kontrolnu grupu). Starosna dob studenata u Splitu je između 20 i 23 godine (srednja vrijednost je 21.8), a u Zadru između 19 i 23 godine (srednja vrijednost je 19.4). Dobna razlika između eksperimentalne i kontrolne grupe je bila neizbježna jer se u Splitu nastava geometrije održava u 6. semestru, a u Zadru u 2. semestru preddiplomskog studija. Istraživanjem je testirana određena strategija poučavanja pri učenju euklidske geometrije kroz 13 tjedana. Jedan od instrumenata, koji se proveo na početku i na kraju semestra, bio je van Hiele-ov test učenja geometrije (dalje u tekstu VH test). VH test je preuzet, uz dopuštenje, iz rada Usiskin Zalman, *Van Hiele Levels and Achievemnt in Secondary School Geometry*, CDASSG Project, 1982. by The University of Chicago. VH test se sastoji od 25 pitanja objektivnog tipa s jednim točnim odgovorom, (prvih 5 pitanja za prvu razinu mišljenja, sljedećih 5 pitanja za drugu razinu itd. do zadnjih 5 koji se odnose na petu razinu). Test je baždaren na 35 minuta. U ovom radu prikazuje se samo mali dio istraživanja koji se odnosi na temu rada.

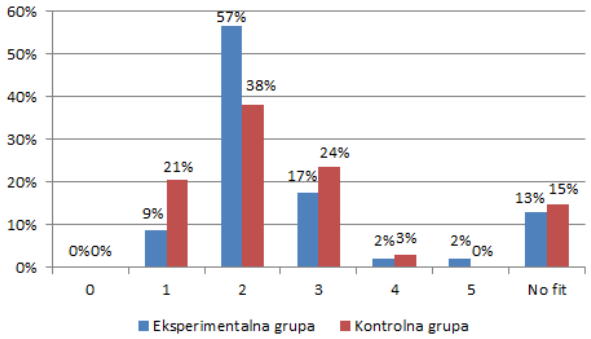
Prije početka učenja geometrije studenti su testirani VH testom radi ispitivanja postignutih razina mišljenja do tada, kako bi se u skladu s dobivenim rezultatima prilagodilo poučavanje planiranih geometrijskih sadržaja u eksperimentalnoj grupi. Usporednom analizom pokazano je da nema statistički značajnih razlika među grupama na početku semestra, što znači da se mogu razmatrati kao homogena cjeline.

Na slici 9. prikazana je raspodjela studenata po opisanim razinama mišljenja, pri čemu stupac No fit predstavlja one studente koji nisu raspoređeni ni na jednu razinu, što znači da se nalaze na prijelazu između određenih razina (vidjeti [12]). Stupac 0 predstavlja one studente koji nisu dosegli niti prvu razinu.

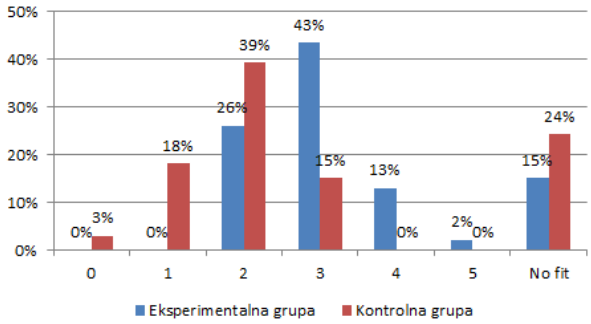
Sa slike 9 se može vidjeti da se dvije trećine studenata nalazi na prve dvije razine, petina studenata na trećoj razini, dok ih je na četvrtoj i petoj zanemarivo malo (po 1 student).

S obzirom da se euklidska geometrija na fakultetskoj razini poučava strogim aksiomatskim pristupom, za ove studente to bi značilo poučavanje bez uspjeha. Stoga su nastavne metode, strategija poučavanja i nastavni sadržaji u eksperimentalnoj grupi prilagođeni razini mogućnosti njihovog praćenja. Poučavanje u kontrolnoj grupi se nije prilagođavalo.

Nakon 13 tjedana poučavanja (planimetrija i stereometrija, bez trigonometrije), studenti su ponovno testirani istim testom s ciljem određivanja je li i u kojoj mjeri je provedeno poučavanje i učenje dovelo do napretka u svakoj od skupina. Rezultati su dani na slici 10.

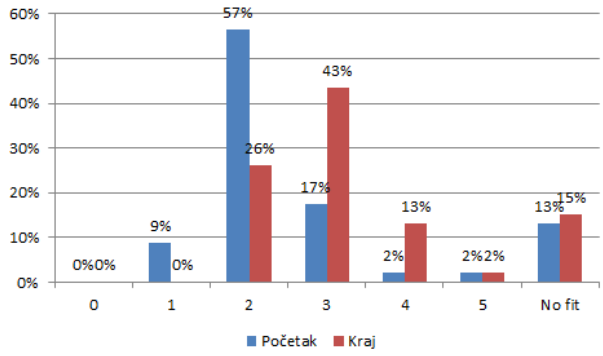



**Slika 9.** Raspodjela studenata prema van Hiele modelu, po grupama na početku




**Slika 10.** Raspodjela studenata prema van Hiele modelu, po grupama na kraju

Usporednom analizom utvrđena je statistički značajna razlika rezultata eksperimentalne grupe u odnosu na kontrolnu grupu čime je potvrđeno da ciljano odabrana strategija poučavanja utječe na razvoj geometrijskog mišljenja prema van Hiele-ovom modelu.




**Slika 11.** Raspodjela studenata eksperimentalne grupe, na početku i kraju

Uspoređujući samo rezultate eksperimentalne grupe (slika 11) vidljivo je da je većina studenata nakon semestra učenja geometrije napredovala na višu razinu geometrijskog mišljenja te da se skoro polovica njih sada nalazi na trećoj razini. Uvidom u detaljniji prikaz podataka može se reći da nitko od studenata nije nazadovao, što se ipak dogodilo u kontrolnoj grupi.

## 7. Zaključak

U radu je ukratko prikazan model van Hiele-ove teorije o procesu mišljenja s primjenom na razvoj geometrijskog mišljenja, osnovne karakteristike modela te preporučena strategija poučavanja u svrhu napredovanja prema tom modelu. Ipak, najbolji način razumijevanja opisane teorije postiže se upravo iskustveno, provodeći opisano u vlastitoj nastavnoj praksi, a ne samo čitanjem dostupne literature.

Upravo provedeno eksperimentalno istraživanje potvrđuje bitne karakteristike opisanog modela, posebno to da razvoj geometrijskog mišljenja ne ovisi o starosnoj dobi i sazrijevanju već prije svega o strategiji poučavanja, prilagođenoj učenicima, a time i poručuje da za učenje geometrije i razvoj geometrijskog mišljenja nikad nije kasno.

Osigurati učenicima da kroz opširan i zahtjevan nastavni plan i program razvijaju i geometrijsko mišljenje nije nimalo jednostavan posao, ali jest zadaća svakog nastavnika matematike. Iako ta zadaća pred nastavnike postavlja zahtjevnu obvezu i odgovornost, ona se ipak može uspješno provesti, što potvrđuju razna istraživanja utemeljena na van Hiele-ovoj teoriji, koja zasigurno u tom smislu ohrabruju i olakšavaju. Prema riječima van Hiele-a: „Kada pažljivo njegujete geometrijsko mišljenje učenika, oni će biti uspješniji u savladavanju i Euklidove matematike" (vidjeti [13], str. 316).

## Bibliografija


[1] **W.F.Burger, J.M.Shaughnessy.** Characterizing the van Hiele Levels of Development in Geometry. *Journal for Research in Mathematics Education*, 1986, Vol. 17, No. 1, 31-48. National Council of Teachers of Mathematics. Dostupno na: http://www.jstor.org/stable/749317 (5.10.2010.)

[2] **Th. Colignatus.** Freudenthal's "realistic mathematics education"appears to be a fraud, July 6, 2014, weblog, http://boycottholland.wordpress.com/2014/07/06/hansfreudenthal-s-fraud/ (15. 9. 2015.)

[3] **Th. Colignatus.** Pierre Van Hiele and David Tall: Getting the facts right. Paper presented at ARXIV, USA, 2014, Cornell University Library. Dostupno na: http://thomascool.eu/Papers/Math/2014-07-27-VanHieleTallGettingTheFactsRight.pdf (15.9.2015.)

[4] **M. L. Crowley.** The van Hiele model of development of geometric thought. In Learning and teaching geometry, K-12, 1-16, Yearbook of the National Council of Teachers of Mathematics. Edited by Mary Montgomery Lindquist, 1-16, 1987, Reston, VA: National Council of Teachers of Mathematics.

[5] **A. Cuoco, E.P. Goldenberg, J. Mark.** Habit of mind: an organizing principle for mathematics curricula. *Journal of mathematical behaviour*, 1996, Vol. 15, 375-402.

[6] **M. De Villers.** Some reflections on the Van Hiele theory. *Invited plenary presented at the 4th Congress of teachers of mathematics of the Croatian Mathematical Society*, Zagreb, 20 June - 2 July 2010.

[7] **D. Fuys, M. Geddes, R. Tischler.** English translation of selected writings of Dina van Hiele-Geldof and Pierre M. van Hiele. Brooklyn: Brooklyn Collage, School of Education, 1984.

[8] **D. Fuys, M. Geddes, R. Tischler.** The van Hiele model of thinking in geometry among adolescents. Monograph No. 3 of the Journal for Research in Mathematics Education. Reston, VA: National Council of Teachers of Mathematics, 1988.

[9] **S. Johnston-Wilder, J. Mason.** Developing thinking in geometry. A SAGE Publications Company. London, 2005.

[10] **M. Mason.** The van Hiele levels of Geometric understanding. *Professional Handbook for teachers, geometry: Explorations and applications*, 2002, Boston: McDougal Littell Inc.

[11] **A. Teppo.** Van Hiele levels of geometric thought revisited. *The Mathematics Teacher*, Vol. 84, No. 3 (March 1991), 210-221. National Council of Teachers of. Dostupno na: http://www.jstor.org/stable/27967094. (3.9.2013.)

[12] **Z. Usiskin.** Van Hiele Levels and Achievement in Secondary School Geometry. (Final Report of the Cognitive Development and Achievement in Secondary School Geometry Project.) Chicago, Illinois: University of Chicago, 1982.

[13] **P.M. Van Hiele.** Developing geometric thinking through activities that begin with play. *Teaching children mathematic*, 1999, 5(6), 310-316.

Симпозијум МАТЕМАТИКА И ПРИМЕНЕ, Математички факултет, Универзитет у Београду, 2015 ,Vol. VI(1)

# Видео материјали у настави математике

**Оливер Петковић**

*Покрајински секретаријат за спорт и омладину, Нови Сад*
*e-mail: oliver.petkovic@gmail.com*

**Мирослав Марић**

*Математички факултет, Универзитет у Београду*
*e-mail: maricm@matf.bg.ac.rs*

**Апстракт.** Улога интернета у функцији стицања, поделе и чувања знања у смислу добијања правих информација од пресудног је значаја за многе аспекте савременог образовања. Развој технологије и појава дигиталних видео-камера, софтвера за уређивање видео материјала и интернета пружили су могућност једноставне и јефтине продукције и дистрибуције видео садржаја. Због тога, видео материјали се намећу као изузетан алат за учење. Напредак технологије повећава доступност алата за њихову израду чиме се наставницима омогућава да их самостално креирају, прилагоде потребама наставе и да га путем интернета учине доступним свима. Употреба видео материјала у настави математике може обогатити наставу, учинити је занимљивом и динамичном, а ученике додатно мотивисати да самостално уче. Циљ рада је да представи поступак израде образовних видео материјала за наставу математике. Поред тога, у раду ће кроз примере бити приказано на који начин се, њиховом применом у наставном процесу, настава може унапредити и осавременити. Осим самог процеса израде, у раду ће бити представљени алати који су коришћени за њихову израду. Такође, биће представљени типови образовних видео материјала, приказане предности њиховог коришћења у настави, као и могућности њихове даље употребе.

**Кључне речи:** видео материјали у настави математике; иновација наставе математике.

## 1. Увод

Право на једнако и доступно образовање и васпитање без дискриминације један од основних принципа система образовања и васпитања у Србији, а дефинисано је Законом о основама система образовања. Због важности образовања за сваког појединца основно образовање је у нашој земљи обавезно. С обзиром на степен развоја данашњег друштва, науке и технологије, а имајући у виду да функционална писменост подразумева много више од познавања писма и рачуна, оно није довољно за његово нормално функционисање у савременом друштву. Међутим, много је примера који доказују да овај принцип доступности образовања није задовољен у потпуности, а према истраживањима Унеска у Србији у 2013. години 10.450 деце основношколског узраста није похађало школу. Узроци су различити, а најчешћи су недостатак новца, посебно код ромске популације, елементарне непогоде, недоступност образовања приликом хоспитализације и друго [1, 2].

## 2. Улога интернета у савременом образовању

Захваљујући интернету, процес образовања је значајно промењен у последње две деценије. Његова улога у функцији стицања, поделе и чувања знања у смислу добијања правих информација од пресудног је значаја за многе аспекте савременог образовања. Интернет је постао највећи ресурс информација, али што је још важније, једно од најбржих средства комуникације. Предности које интернет пружа у области образовања су вишеструке. Ученици, студенти, као и деца и одрасли ван образовног система, могу да прошире своје знање коришћењем електронске литературе, енциклопедија, речника, база података, којима могу бесплатно да приступе путем интернета, затим да похађају онлајн курсеве, учествују у заједничким пројектима са ученицима или студентима из других школа, универзитета, држава, дискутују о различитим проблемима са њима итд. Занимљиво је истаћи да је интернет осмишљен пре свега ради унапређења науке и образовања, а да је касније пронашао примену у свим сегментима савременог друштва.

У домену основног образовања у делу који се односи на квалитет процеса наставе и учења Стратегија развоја образовања у Србији до 2020. године предвиђа коришћење предности информационо-комуникационих технологија и различитих облика онлајн учења и налаже испитивање могућности и услова за коришћење неких видова наставе на даљину, пре свега за специфичне околности [3].

Због могућности које пружају информационо-комуникационе технологије, као и аудио-визуелне методе учења, коришћење образовних видео материјала у настави је један од начина унапређење наставе.

## 3. Израда образовних видео материјала

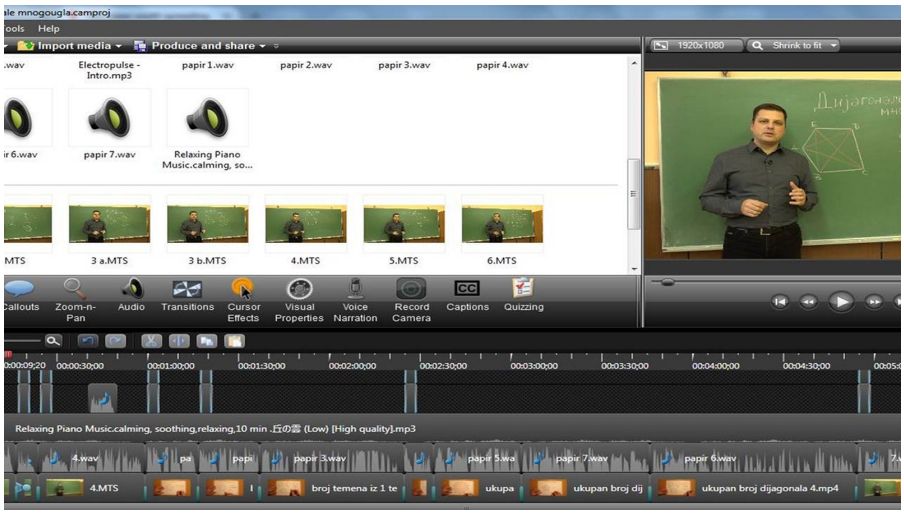
Процес израде образовних видео материјала обично се састоји од три кључне фазе: припрема за снимање, снимање видео материјала и обрада снимљеног материјала. Квалитетна припрема за снимање је предуслов за израду квалитетног образовног видео материјала и пре снимања је потребно изабрати намену и циљну групу, саставити сценарио (писану припрему за видео), одабрати софтверске алате за обраду итд.

Пре самог снимања потребно је припремити опрему и проверити квалитет слике и тона. У ту сврху може се направити неколико пробних снимака и проверити да ли су задовољавајућег квалитета. Снимање се врши у највишој резолуцији слике коју камера допушта, а накнадно се може смањити како би се образовни материјал прилагодио за публиковање путем интернета.

За снимање видео материјала потребно је одабрати простор и уредити га да буде погодан за ту намену. Неопходно је обратити пажњу на осветљеност простора, како би квалитет видео материјала био бољи, али и на рефлексије светлости које могу нарушити квалитет слике.

Образовне видео материјале је могуће снимити помоћу једноставнијих фотоапарата и у кућном амбијенту што наставницима олакшава процес израде.

Након снимања видео материјала, следи његова обрада. Снимљени видео материјали се обрађују помоћу различитих софтверских алата и сједињују се у једну композицију у складу са писаном припремом.




**Слика 1.** Обрада снимљеног видео материјала у програму *Camtasia Studio*

Након израде видео материјали се могу поставити на различите видео канале (*YouTube*, *Vimeo*, и сл.). На тај начин биће доступни ученицима, али и другим наставницима који би могли да их користе у раду.

## 4. Алати за израду видео материјала

Приликом избора алата за израду видео материјала потребно је водити рачуна о њиховој расположивости. Велики број програма који се могу користити за израду видео лекција су комерцијалног типа, међутим и међу бесплатним софтверским алатима постоји широк спектар корисних и квалитетних програма.

Цена софтвера за обраду видео материјала не утиче пресудно ни на његов квалитет нити на остварење његовог циља. Постоји велики број бесплатних програма помоћу којих је могуће скратити снимак до одређене величине, додати звук или наслове, применити прелазе или специјалне ефекте, те је потребно направити оптималан избор алата за израду доброг видео материјала.

Наставницима је на располагању и велики број квалитетних, како комерцијалних, тако и бесплатних програма за израду образовних видео материјала. *Camtasia Studio* је један комерцијалан програм намењен за израду образовних видео материјала, али се може користити и за не много захтевне обраде видео снимака. Овај програм омогућава једноставно исецање и зумирање видеа, комбиновање са сликом, а поседује и опцију слике у слици. Поред тога, наставници могу користити и бесплатне програме као што су на пример: *Screencast-O-Matic*, *Windows Movie Maker*, *VirtualDub*, *Wax*, *Avidemux*, *ZS4 Video Editor*, *Free Video Editor* и многи други [5].

Осим основног софтвера за обраду видео материјала за израду образовних видео материјала јављаће се потреба за коришћењем и других програма као што су на пример за обраду слике или звука, конвертовање аудио и видео фајлова у различите формате, а много је таквих програма бесплатно доступно путем интернета (*IrfanView*, *AudaCity*, *Format Factory*, итд.).

## 5. Типови образовних видео материјала

Постоји три типа образовних видео материјала: предавања снимана камером, видео материјали израђени коришћењем алата за снимање екрана (*Screencasting*) и комбиновани видео материјали.

### Предавања снимана камером

Предавања снимана камером су образовни видео материјали који се увек могу користити, а квалитет у највећој мери зависи од креативности и умешности самог предавача.

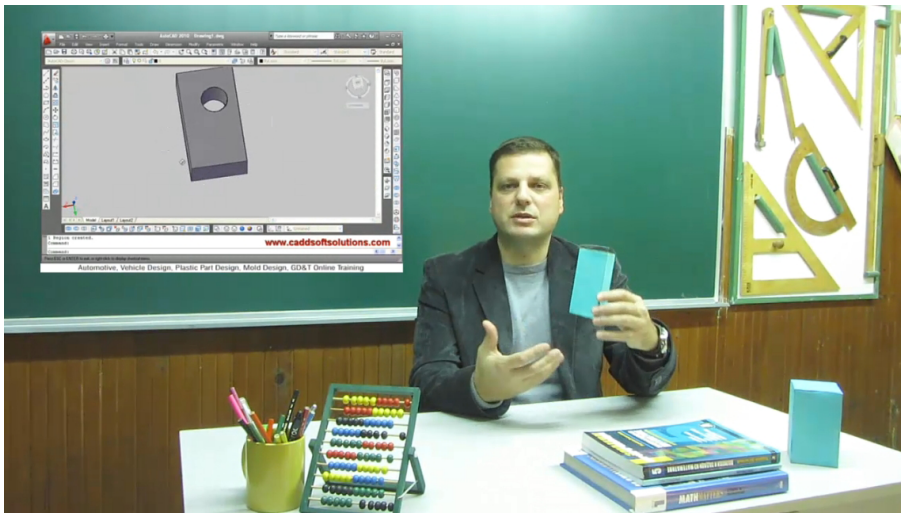
Предности се огледају у томе да ученици могу видети израз лица предавача, његову гестикулацију и покрете руком. Предавач може у току предавања да пише на табли или изводи експеримент. Такође, камера може зумирати на пример руке предавача док пише или црта на табли или на папиру. Све ово даје видео материјалу динамику што доприносу квалитету самог материјала.

Приликом израде оваквих видео материјала потребно је обратити пажњу на неколико сегмента, који у великој мери утичу на квалитет материјала. С обзиром на то да је наставни процес васпитно-образовни предавач би требало да осим образовних циљева подстиче и функционалне и васпитне циљеве. То се подстиче правилним изражавањем, угледним изгледом и сл. У техничком смислу, потребно је водити рачуна о његовој удаљености од камере и микрофона, осветљењу и уредности простора у коме се снима видео итд. Због удаљености предавача може се десити да је немогуће прочитати текст исписан на табли. У неким случајевима се тело или руке предавача могу наћи на путу онога што покушава да покаже или објасни.

### Видео материјали израђени коришћењем алата за снимање екрана

Видео материјали израђени коришћењем алата за снимање екрана (*Screencasting*) су видео материјали чији је садржај снимак свега онога што се приказује на екрану рачунара укључујући аудио или видео.

Једна од највећих предности израде оваквих образовних видео материјала огледа се у томе што је могуће користити било коју апликацију за креирање образовних материјала (презентације, апликације за обраду текста, табеле, калкулатор, алати за графичку




**Слика 2.** Образовни материјал за наставу математике

обраду, алати за цртање, итд.). Једноставно, може се снимити све што се приказује на екрану рачунара, а уз то програми су врло једноставни за подешавање и коришћење.

Међутим, ови видео материјали су ограничени на оно што се може показати на рачунару. Померање или окретање неких тродимензионалних модела на рачунару је ограничено покретима миша. Видео материјали израђени на овај начин личе на слајд презентације и недостаје им динамика.

**Комбиновани видео материјали**

Комбинацијом претходна два типа користе се њихове предности и надомештају се недостаци, тако да овакви видео материјали представљају најбоље решење за примену у наставном процесу.

## 6. Доступни образовни видео материјали

Образовне материјале, могу припремити сами наставници, али могу се користити већ припремљени образовни материјали који су доступни путем интернета. С обзиром на велики број доступних материјала, ради лакше претраге и проналажења квалитетних видео материјала следи преглед неколико онлајн ресурса.

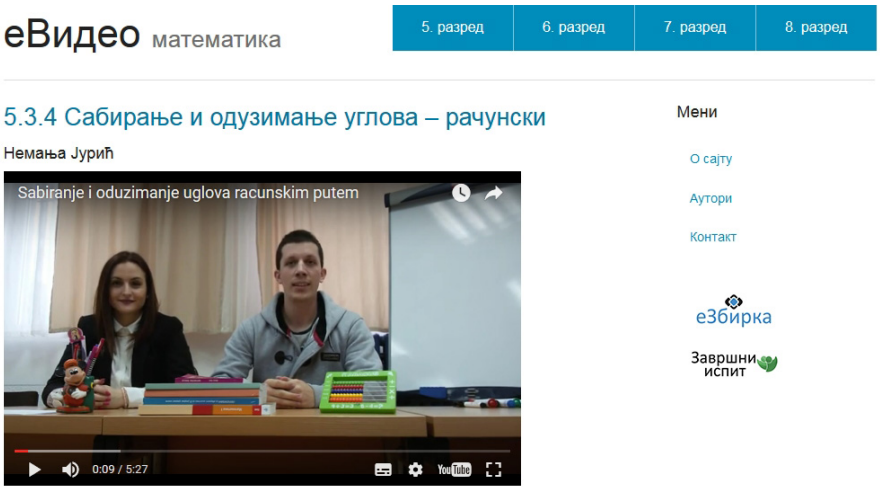
**еВидео**

Веб сајт ,,еВидео" (http://edusoft.math.rs/video) садржи значајан број видео материјала из математике за ученике од петог до осмог разреда основне. Стандардизован приступ израде видео материјала, усклађеност са планом и програмом предмета Математика прописаног од стране Министарства просвете је оно што овај сајт издваја од осталих.

**Виртуална школа Рајак**

Циљ ,,Виртуелне школе Рајак" (http://www.rajak.rs) је да помогне ученицима у да савладају градиво математике за основну школу, средњу школу и факултет. Осим математике на овом сајту има образовних материјала за друге наставне предмете као што су на пример физика, основи електротехнике, механике, статистике, програмирање итд.

Садржај сајта је бесплатан за коришћење и на самом почетку рада био је намењен ученицима који су желели да упишу средње школе и факултете. Садржаји су проширени многим видео материјалима из математике намењеним ученицима основне и средње школе.




**Слика 3.** Сајт ,,eВидео”

### Математирање

Сајт ,,Математирање” (http://matematiranje.in.rs) је почео са радом 2008. године. На сајту су најпре објављени наставни материјали који помажу средњошколцима да се припреме за полагање пријемног испита из математике на жељеним факултетима. С временом је, у складу са захтевима корисника, постављено много материјала са ширим градивом из математике, како за средње школе, тако и за факултете на којима се слуша математика.

Након увођења обавезног полагања завршног испита, односно мале матуре, у основној школи, сајт је проширен новим одељком ,,Мала матура” у коме су детаљно приказана решења задатака са претходних испита, а објављени су и видео материјали са тим решењима. На сајту се може наћи преко 320 електронских материјала са детаљно решеним задацима, употребљивим коментарима и кратким теоријским напоменама.

### Тони Милун

Портал ,,Тони Милун” (http://tonimilun.com) је настао на иницијативу студента Николе Мујдића којем се допао приступачан начин на који његов професор Тони Милун предаје математику и понудио му да снима видео лекције које ће објављивати на Интернету. Врло брзо су се пројекту придружили и други професори и волонтери.

Сајт је састављен од скупа видео лекција из математике објављених на Youtube каналу. Најпре су аутори објавили видео лекције на Youtube каналу, а затим су те снимке распоредили, селектовали и направили импозантан сајт. Инспирисани су многим друштвено корисним пројектима, а жеља им је да сниме видео материјале из математике и осталих предмета и понудити их свима, све уз мото: ,,Знање мора бити бесплатно и доступно свима!”

Објављен је велики број лекција из математике за основну, средњу школу, као и за више школе и факултете. Поред ових лекција, уз помоћ сарадника, Тони Милун је објавио видео лекције из физике, информатике и других области. Иако у својим видео лекцијама користи само флип-чарт таблу са папирима и фломастер, његови видео материјали имају велику посећеност, што указује на њихов квалитет и сврсисходност.

### Мала матура

,,Мала матура” (http://www.mala-matura.com) је пројекат бесплатне онлајн припремне наставе за завршни испит, настао под покровитељством Средње школе за информационе технологије ITHS.

Сајт садржи низ интерактивних, занимљивих лекција и видео материјала који могу помоћи да учење постане игра. Ученици могу почети са учењем од нуле или од делова који

им нису јасни. После сваке пређене области, знање се може проверити кроз мини тестове, а ту је и симулација завршног теста, са задацима који су налик онима који се очекују на завршном испиту.

## 7. Закључак

Добар наставник настоји да буде иноватор, али се иновирање наставе често поистовећује са употребом савремених наставних средстава. Наставник би требало да непрекидно унапређује наставу применом информационо-комуникационих технологија, али и применом нових наставних метода. Потребно је да проналази путеве за подстицање креативног приступа решавању проблемских задатака, те повезивању раније стечених знања из различитих наставних области и њиховој адекватној употреби.

Иако употреба аудио-визуелних средстава у настави није нови концепт, напредак савремених технологија и развој интернета доприноси њиховој широј и једноставнијој употреби. Предност наставних материјала који се у дигиталном облику креирају у данашњим условима је могућност интеграције са другим медијима и једноставне дистрибуције путем интернета. Напредак технологије повећава доступност алата за њихову израду чиме се наставницима омогућава да их самостално креирају, прилагоде потребама наставе и да га путем интернета учине доступним свима. Због тога, образовни видео материјали се намећу као моћан алат у наставном процесу.

## Библиографија


[1] **UNESCO Institute for Statistics**, Number of out-of-school children of primary school age, http://data.uis.unesco.org/Index.aspx?queryid=121

[2] **Закон о основама система образовања**, Службени гласник РС, број 72/2009, 52/2011, 55/2013, 35/2015 - аутентично тумачење и 68/2015

[3] **Стратегија развоја образовања у Србији до 2020. године**, Службени гласник РС, број 107/2012

[4] **В. Кулето, В. Дедић**. еУчење, LINK group DOO, http://www.valentinkuleto.com/wp-content/uploads/2014/02/eLearning2014.pdf

[5] **TechSmith Corporation**, Camtasia Studio 8.5, *Help File Document*, February 2015. https://www.techsmith.com/tutorial-camtasia.html

# Дигитализација српских службених новина 1813–2013

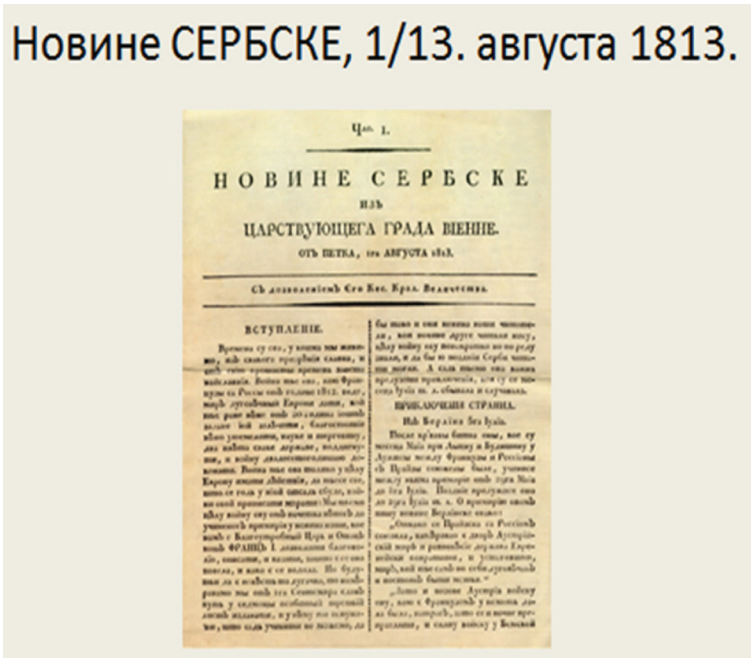
**Светлана Албијанић**

*ЈП Службени гласник, Јована Ристића 1, Београд*
e-mail: salbijanic@slglasnik.com

**Апстракт.** Од идеје да се текст "на новинској хартији" пребаци у електронски доступан текст кориснику до реализације те идеје и пројекта "Дигитализација српских службених новина 1813–2013". било је потребно много енергије, знања из различитих наука, времена, љубави и новца. Примењена математика и програмирање у последњих педесет година су саставни део свих научних истраживања. Како се пројекат базира на претраживању текста уз задате параметре нама у ЈП Службени гласник био је потребан одговарајући оперативни систем, апликативни сервер, база података, систем за претрагу и презентациони део за сваки од два периода – период славеносрпског (славјаносербског) језика и период савременог српског језика.

Дигитализована база српских службених новина 1813–1944. и дигитализоване периодике издања којима се баве библиотеке су нови извор информација у Републици Србији (прве базе су тек од пре неколико година доступне корисницима) и тек по овом скупу где су промовисане и сличним, сазнању учесника да постоје и приказу како могу да се користе, може да се очекује њихова употребљивост у научним истраживањима и настави.

**Кључне речи**: Службене новине, претрага, српске новине.




## 1. Увод

Дигитализована база података бесплатно је доступна корисницима.

### 1.1. Приступ бази

Приступ бази могућ је на адресама

1. www.sluzbenenovine.rs
2. www.службененовине.срб
3. Сајт www.sluzbeniglasnik.com па *Архив српских и југословенских службених издања*

### 1.2. Циљ ЈП Службени гласник и циљеви потенцијалних корисника

ЈП *Службени гласник* је у подухват дигитализације службених издања Србије и Југославије ушао желећи да омогући претраживање података из српске правне и културне историје у претходних 200 година излажења новина и тиме их учини доступним савременим корисницима, отргне од заборава и сачува.

Стручни скуп има за циљ унапређивање наставе и подизање опште културе и знања коришћењем дигитализованих српских службених новина и штампе:

- подизање опште културе и знања учесника
- ширење информација о постојању и доступности дигитализоване штампе и службених новина
- стручно усавршавање учесника у области одговарајуће научне дисциплине, методике наставе и образовне технологије (примене математике и програмирања)
- предочавање учесницима нових могућности за организацију часа (како подстицати ученика да савлада ново градиво или прошири постојеће коришћењем дигитализоване штампе као и српских службених новина) кроз примере и могућности коришћења базе за цео период 1813–1944. за наставне предмете српски језик и књижевност, историја, биологија, географија, астрономија, математика, социологија, грађанско васпитање, медицина, животна средина, право, економија, веронаука... О свему што се зидало (здања, школе, задужбине...), правио пут, развијала железница, забрањивало (клизање на залеђеним рекама, паљење дрвета на Калемегданском парку...), набављало за војску, када су се рађали Краљевићи, мењале Владе, могло је да се чита у новинама. Ту, у новинама, је текао живот: "Новине су очи, уши и душа сваког народа" су речи уредника Srpskih novina Милоша Поповића.

### 1.3. Из угла новинара

Недељник ВРЕМЕ од 10.09.2015. број 1288 штампао је текст о пројекту "Дигитализација српских службених новина 1813–2013" под називом *Огледало времена* у коме каже: "Дигитални архив српских и југословенских новина објављиваних током последња два века, који је 'Службени гласник' начинио и који је однедавно доступан свима, један је од најбољих извора српске државно-правне традиције, историје и културног наслеђа... Дигитализација српских службених новина 1813–2013' је државни капитални пројекат, пут и начин како се чува и шири идентитет. Нема разлога да ова реченица било ком зазвучи као празна фраза."

## 2. Из историјата пројекта

У трагању за погодним датумом за дан ЈП Службени гласник, дошло се до 1/13. августа када су Димитрије Давидовић и Димитрије Фрушић 1813. издали први број *Новина сербских* у Бечу. По усвајању овог датума, Службеном гласнику се наметнула велика одговорност: да као баштиник службених новина Кнежевине и Краљевине Србије, Краљевине Срба, Хрвата и Словенаца, Краљевине Југославије и социјалистичке Југославије кроз један пројекат учини доступним богату државно-правну традицију и културно наслеђе. Један од најбољих извора за то биле су управо службене новине. Одлучено је да се покрене пројекат дигитализације службених новина у Србији и Југославији од *Новина сербских* до 1945. и да овако добијена база буде архивска у односу на базе *Службеног гласника* које се односе на раздобље после 1945. године.

Приступило се фази писања пројекта и брзо дошло до сазнања да пројекат "Дигитализација српских службених новина 1813–2013" обухвата два периода у издавању српских службених новина. Први, који захвата период од 1813. до 1868. године, ограничен је на Новине српске, које су до 1868. излазиле на славеносрпском језику и штампане су предвуковском ортографијом. Други период (1869–2013) делимично обухвата *Новине српске* штампане Вуковом ортографијом, али и потоње службене новине које су имале различите називе.

Због природе ортографије и разлика у језику, два периода су захтевала различите приступе када је у питању њихова дигитализација. Пројекат "Дигитализација српских службених новина 1813–2013" подељен је на два потпројекта (1813–1868. и 1869–2013).

И тада, на почетку реализације пројекта, отпочињу планирани послови и настају некада видљиви а некада невидљиви, непредвиђени проблеми које је требало "у ходу" решити.

Општи послови за цео пројекат су:

- Проналажење новина
- Скенирање
- Оцеровање новина – програмско препознавање графема тј. графичких симбола (испитивање на узорку)
- Шифрирање скенова

За славеносрпски период:
Проблем: Како омогућити савременом кориснику да претражује текст:

- писан ортографијом коју не познаје и
- језиком којим не влада у потпуности?

Решење: Омогућити кориснику претрагу тако што ће:

- поставити упит (кључну реч) на савременом српском језику,
- добити све текстове у којима се појављује кључна реч (одредница)
  Кључна реч мора бити одредница, што значи да корпус текстова мора да буде лематизован (или делимично лематизован) како корисник не би добио примере који нису обухваћени упитом.
  Лематизација именица је процес у коме се издвајају: именица на савременом српском језику у номинативу, њој одговарајући запис речи на славеносрпском језику и њима се приписују сви остали облици речи (језик је флективан, има падеже).

Процес обраде СЛС текстова се одвијао у следећим фазама:

- Прекуцавање текстова (новина) уз помоћ апликације за тастатуру славеносрпских слова и симбола
- Тз. "сечење страница" (одвајање текста по новинским странама у посебан ворд фајл) и шифрирање сваког фајла
- Упаривање прекуцаног текста и текста на скену (процес индексације – програмска апликација за ручно маркирање текста у ворд-у и дела скена са одговарајућим текстом)
- Програмерска припрема за лематизацију
- Лематизација текста (овде само именица)
- Стручна редактура лема
- Прављење корисничке маске

За период Вукове ортографије фазе пројекта/послови које је требало урадити су:

- Направити листу доносилаца, одредити категорије, врсте докумената, области којима припадају текстови
- Извршити анотирање текстова користећи e7ел табеле – сваки наслов је имао: свој идентификациони код, назив гласила, број, датум издања, број стране на којој је текст–документ, редни број документа на страници, назив текста (ако постоји, а ако не постоји прављен је на основу садржаја текста коришћењем кључних речи), доносиоца, категорију (закон, остало/оглас), врсту документа, област.
- Стручна редактура (која се радила ручно)
- Исправити ручно све технички настале грешке, уз помоћ програма који их "хвата" – подаци о тексту мора да буду потпуно усаглашени да би га апликација учитала у

базу (нпр. број новина иде растућим редоследом, као и сви бројеви помоћу којих се текст позиционира у новинском издању или на страни, и слично)
- Учитати урађену табелу и шифриране скенове новина за одговарајућу годину помоћу одговарајуће апликације у базу. То поновити за сваку годину излажења новина и то посебно за редовна издања, ванредна издања и додатке.

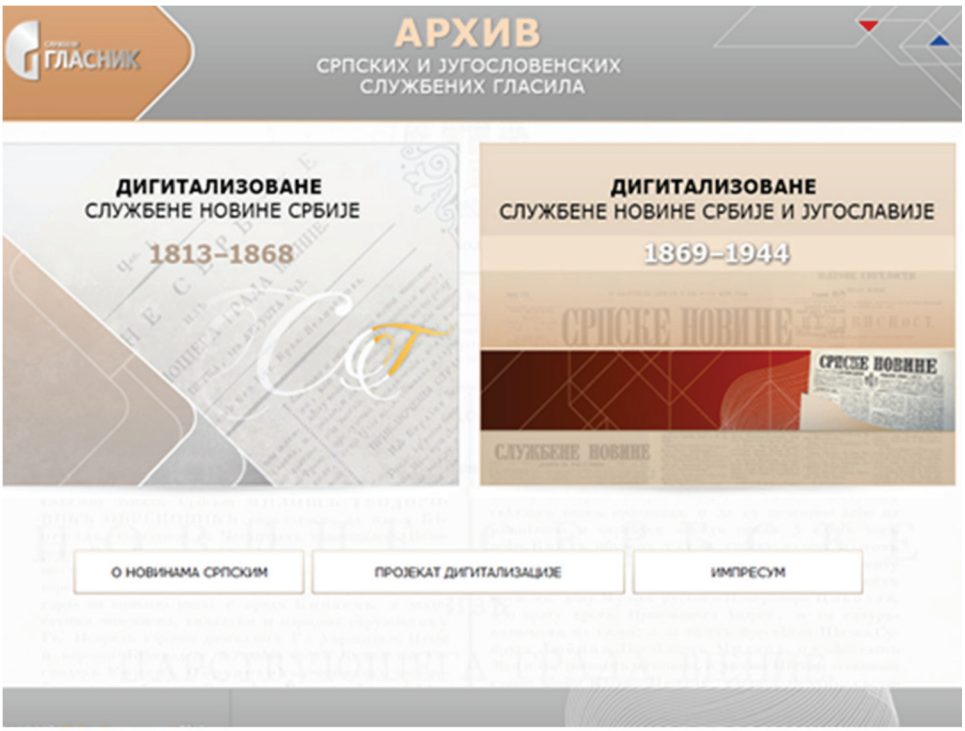
Сваку од фаза је требало контролисати да се грешке не би нагомилавале.

## 3. Корисничка апликација

Основни услов – да просечном кориснику буде

- једноставна и
- разумљива.

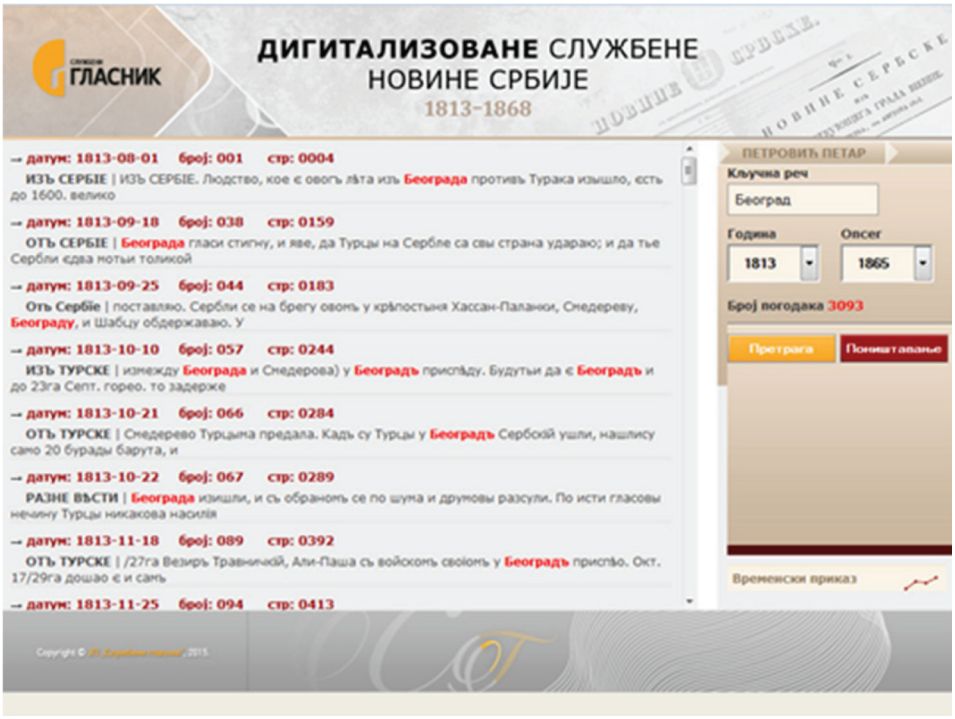
Уводи нас у два одвојена дела пројекта према запису језика, начину претраге и начину добијања резултата.



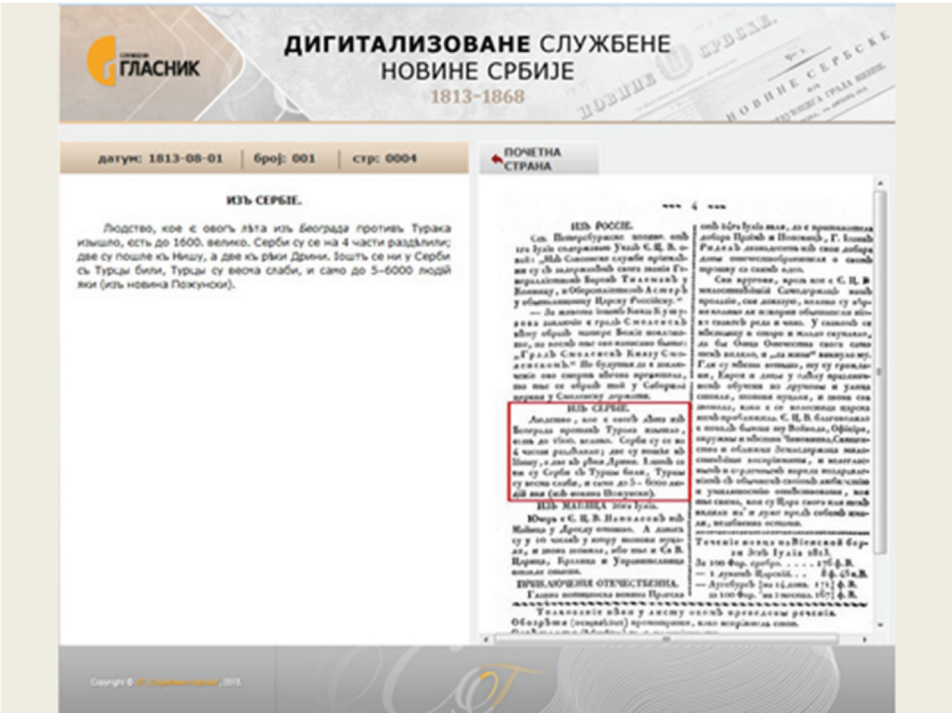

**Период 1813–1868.**

- Претрага се обавља по кључној речи која се уноси (одредница именице на савременом језику) и избору годишта или опсегу годишта.
- Програм кориснику ”даје” број пронађених текстова са свим речима истога значења а различитог записа, и списак свих издања у којима је текст са траженом речи.
- Списак садржи: датум издања–број издања–страницу на којој је текст и део текста где се задата реч ”види” (истакнута је и црвеном бојом).

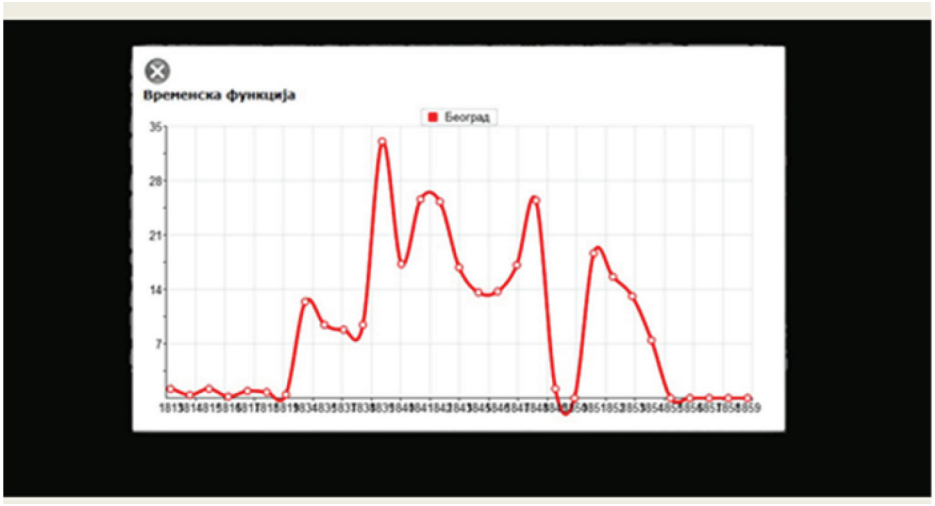
- А онда корисниче бирај текст који желиш да видиш и читаш!




- Избором текста добија се:
  - прекуцана верзија са обележеном траженом речју и
  - скен оригиналне стране Новина сербских са обележеним текстом.



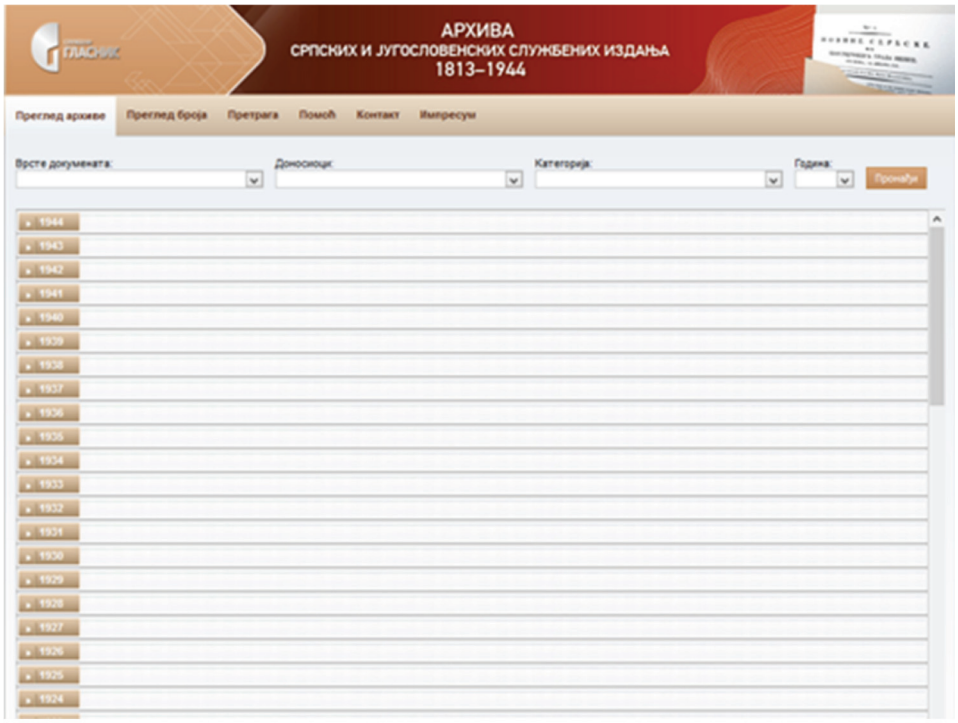
- Корисник може да добије и временску функцију појављивања текстова са траженом речју.



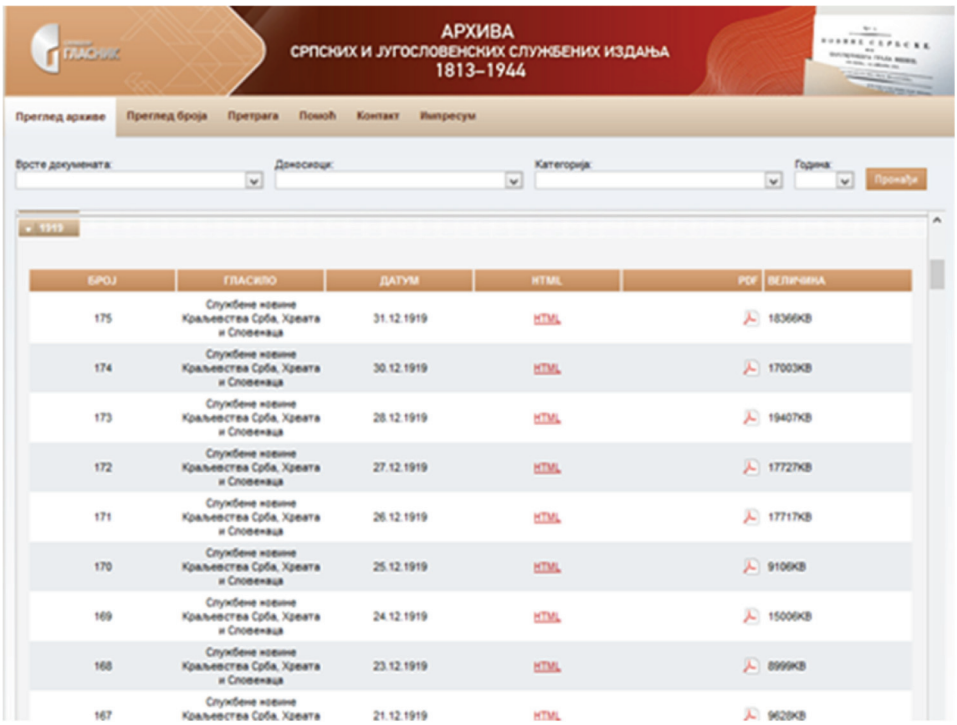

**Период 1869–1944.**

Период 1869–1944. год. је период службених новина Србије и Југославије који претходи већ постојећој бази ЈП Службени гласник – службена издања 1945–2016. и претрага је урађена на исти начин по годишту, по броју, по законском акту, огласу или остало, доносиоцу, врсти документа или области.

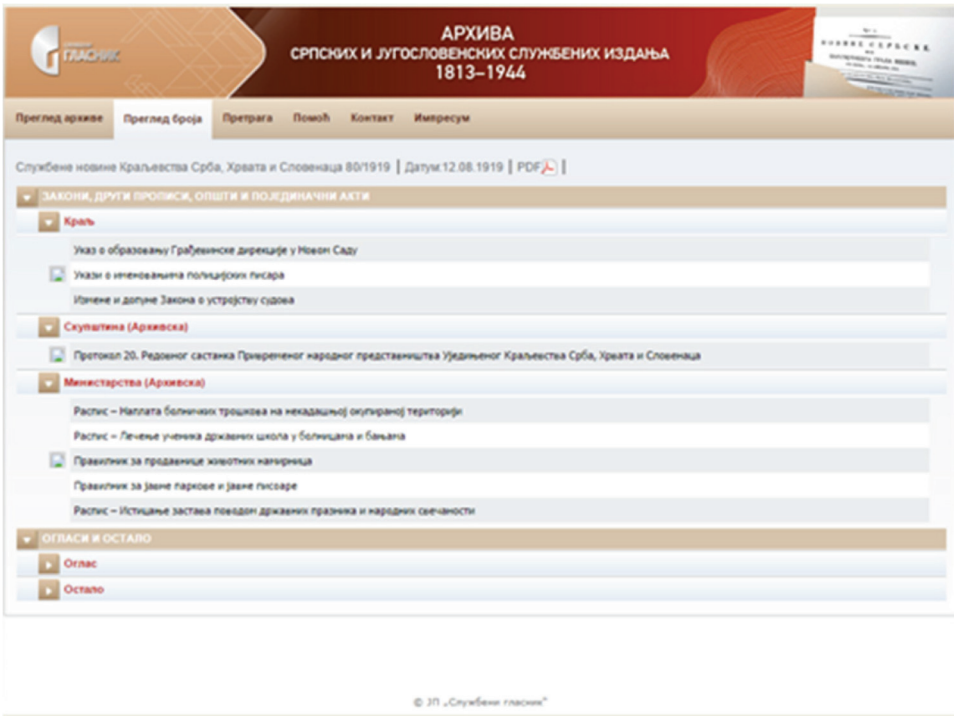
Преглед годишта



И избор, нпр. 1919. година

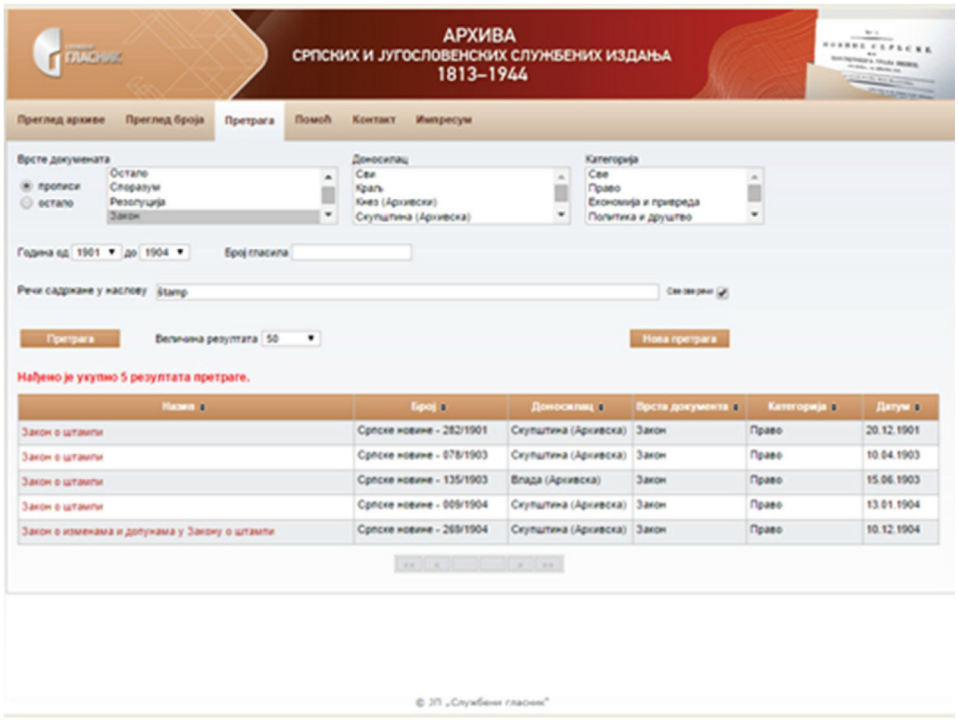



И преглед броја 80 из 1919. ако се траже *Закони* (разврстани према доносиоцу).



Преглед броја 80 из 1919. ако се тражи *Оглас* или *Остало*


И највише захтеван начин претраге је по *Речи садржане у наслову*. У оквиру могућности апликације Претрага задају се различити критеријуми (пропис/остало, врста прописа, доносилац, област, година или опсег година) и на тај начин сужава претрага тражених текстова.



## 4. Статистика

4.1. Период славеносрпског језика 1792–1794. и 1813–1868.

| ПЕРИОД | | ИЗДАЊА | | | | |
|---|---|---|---|---|---|---|
| | | поседујемо | | не поседујемо | | |
| 1792-1795 | | | | 208 | | |
| 1813 – 1868. | Бечки период | 1209 | 4335 | | | |
| | период излажења новина у Србији | 3126 | ≈74% | 1244 | 1452 | 1527 |
| | лоши скенови | 75 | | | | 26,05% |
| | | збирно: | | 4335+1527=5862 | 100% | |

Прегледом табеле долазимо до сазнања да је базу у овом периоду потребно допунити новинским издањима са 26% годишта (1820, 1821,1844, 1861. и 1866–1868. и годишта када су новине излазиле као Вједомости, такође на славеносрпском језику). Напоменимо да новине нису излазиле од 1822. до 1834. године.

4.2. Период Вукове ортографије 1869–1944.

| У бази се за 76 годишта на дан 9.7.2015. | налази | не налази |
|---|---|---|
| редовних издања | 19 451 | 50 (1869–1938)<br>526 (1939-1941) |
| додатака | 1 381 | 40 (1869–1939)<br>≈ 120 (1936 i 1940) |
| ванредних издања | 75 | немамо у<br>1930,1933 i 1935 |
| Забавник (Крф) | 18 | / |
| укупно издања | 20 925 | ≈ 736 |
| У процентима | 96,58% | 3,42% |

У бази се за ових 76 годишта налази 20 925 бројева службених издања тј. 96,58% свих издања (редовних бројева, додатака и ванредних издања). Обрађено је 221 000 наслова. Напоменимо да су у бази сва издања новина са Крфа као и свих 18 бројева Забавника.

### 4.1. Проблеми

Унапређивање базе:

- Комплетирање базе са издањима која недостају
- Отклонити проблеме великих издања
- Израда посебних менија

1) Посебни мени за претрагу Закона

2) Претрага устава
3) Претрага свих влада
4) Претрага књижевних дела објављених у службеним новинама
5) Претрага свих научних радова
6) Претрага некролога и читуља
7) Израда хронолошких таблица с главним догађајима из политичке и културне историје који су покривени у службеним новинама за период 1813-1822, 1834-1918.

## 5. Закључак


Овај корпус дигитализованих новина, као целина, представља огроман трезор података из српске и европске културе и правних норми. Зато је овај пројекат сведочанство о развоју српске културе и српске државности, као и о рецепцији и развоју правних норми у Србији. Службене новине су и дословно сведоци епохе, а податак да у Србији постоји континуитет од преко 180 година излажења оваквих новина (прве Службене новине појављују се у Кнежевини Србији 1834. године. у Крагујевцу) сам по себи говори о значају оваквог извора.

## Библиографија

[1] **Ђорђе Костић.** Квантитативни опис структуре српског језика СРПСКИ ЈЕЗИК ОД ШИИ ДО ШВИИИ ВЕКА (Фреквенцијски речник). *ЈП Службени гласник*, 2010.

**УНИВЕРЗИТЕТ У БЕОГРАДУ**
**МАТЕМАТИЧКИ ФАКУЛТЕТ**
**и**
**СРПСКА АКАДЕМИЈА НАУКА И УМЕТНОСТИ**
**ШЕСТИ СИМПОЗИЈУМ „МАТЕМАТИКА И ПРИМЕНЕ”**
**НАЦИОНАЛНИ СКУП СА МЕЂУНАРОДНИМ УЧЕШЋЕМ**

**ПРОГРАМ**

# 1. ДАН, ПЕТАК 16. ОКТОБАР 2015.

| |
|---|
| 10:00 – 11:45 |
| **Отварање скупа:**<br>Академик **Драгош Цветковић**, секретар одељења САНУ за математику, физику и геонауке |
| **Миодраг Матељевић,** Универзитет у Београду, Математички факултет, дописни члан САНУ - Председник програмског одбора Симпозијума |
| ***„Изопериметријска неједнакост - старо и ново“*** |
| **Зоран Петровић,** Универзитет у Београду, Математички факултет |
| ***„Гребнерове базе - од топологије до алгебарске комбинаторике“*** |

## I СЕКЦИЈА: МАТЕМАТИКА И ПРИМЕНЕ ДАНАС

| |
|---|
| 12:10 – 12:40<br>**Владимир Грујић,** Универзитет у Београду, Математички факултет |
| ***„Квазисиметричне функције уопштених пермутоедара“*** |
| 12:40 – 13:10<br>**Горан Ђанковић,** Универзитет у Београду, Математички факултет |
| ***„Теорија случајних матрица и фамилија L-функција придружених симетричним квадратним подизањима модуларних форми“*** |
| 13:10 – 13:40<br>**Петар Марковић,** Универзитет у Новом Саду, Департман за математику и информатику, ПМФ Нови Сад |
| ***„Computational complexity of the Constraint Satisfaction Problem, its significance and a brief survey of recent results“*** |

| |
|---|
| 14:10 – 14:40<br>**Милош Арсеновић,** Универзитет у Београду, Математички факултет |
| ***„Множитељи хармонијских функција у више димензија“*** |
| 14:40 – 15:10<br>**Бранислав Првуловић,** Универзитет у Београду, Математички факултет |
| ***„Modified Postnikov towers“*** |
| 15:10 – 15:40<br>**Миљан Кнежевић,** Универзитет у Београду, Математички факултет |
| ***„Hyperbolic derivatives and its applications“*** |
| 15:40 – 16:10<br>**Иван Аранђеловић,** Универзитет у Београду, Машински факултет<br>**Зоран Митровић,** Универзитет у Бања Луци, Електротехнички факултет |
| ***„Condensing KKM maps and its applications“*** |

| |
|---|
| 12:10 – 12:40<br>**Филип Марић,** Универзитет у Београду, Математички факултет |
| ***„Интерактивно доказивање теорема“*** |
| 12:40 – 13:10<br>**Татјана Давидовић,** Математичји институт САНУ, Београд |
| ***„Оптимизација колонијом пчела у првих петнаест година“*** |
| 13:10 – 13:40<br>**Бојана Милошевић,** Универзитет у Београду, Математички факултет<br>**Марко Обрадовић,** Универзитет у Београду, Математички факултет |
| ***„Тестови експоненцијалности засновани на емпиријским Лапласовим трансформацијама“*** |

| |
|---|
| 14:10 – 14:40<br>**Александра Делић,** Универзитет у Београду, Математички факултет |
| ***„Оцена брзине конвергенције диференцијских схема за једначине аномалне дифузије са концетрисаним капацитетом“*** |
| 14:40 – 15:10<br>**Душан Онић,** Универзитет у Београду, Математички факултет |
| ***„On the continuum radio spectra of Galactic supernova remnants: New insights from Planck“*** |
| 15:10 – 15:40<br>**Бранко Малешевић,** Универзитет у Београду, Електротехнички факултет<br>**Татјана Лутовац,** Универзитет у Београду, Електротехнички факултет |
| ***„Један допринос аутоматском доказивању неких класа аналитичких неједнакости“*** |
| 15:40 – 16:10<br>**Hans Hartmann,** Executive Vice President, OBJENTIS Software Integration GmbH, Vienna |
| ***„Applied Mathematics and the Internet of Things”*** |

## 2. ДАН, СУБОТА 17. ОКТОБАР 2015.

### I СЕКЦИЈА: МАТЕМАТИКА И ПРИМЕНЕ ДАНАС

| |
|---|
| 10:00 – 10:30<br>**Душко Јојић,** Универзитет у Бања Луци, Природно-математички факултет |
| ***„Комбинаторна и тополошка својства неких занимљивих комплекса“*** |
| 10:35 – 11:00<br>**Славко Моцоња**, Универзитет у Београду, Математички факултет |
| ***„Експанзије линеарних уређења“*** |
| 11:00 – 11:30<br>**Тијана Шукиловић,** Универзитет у Београду, Математички факултет |
| ***„Геодезијска еквивалентност Лоренцових метрика“*** |
| 11:30 – 12:00<br>**Миленко Пикула,** Универзитет у источном Сарајеву<br>**Владимир Владичић,** Универзитет у источном Сарајеву<br>**Исмет Калчо,** Универзитет у источном Сарајеву |
| ***„О инверзним проблемима диференцијалних оператора са отклоњеним аргументом“*** |

| |
|---|
| 12:30 – 13:00<br>**Нела Милошевић,** Факултет за информационе системе и технологије, Универзитет Доња Горица, Подгорица |
| ***„Комплекс пресјека идеала“*** |
| 13:00 – 13:30<br>**Миодраг Матељевић,** Универзитет у Београду, Математички факултет<br>**Милољуб Албијанић,** Универзитет Сингидунум и ФЕФА |
| ***„Конвексност и примене”*** |
| 13:30 – 14:00<br>**Драган Станков,** Универзитет у Београду, Рударско геолошки факултет |
| ***„Distribution modulo one of the sum of powers of a Salem number“*** |

| |
|---|
| 10:30 – 11:00<br>**Братислав Иричанин,** Универзитет у Београду, Електротехнички факултет |
| ***„Неки примери експлицитног решавања рационалних диференцних једначина и система рационалних диференцних једначина вишег реда“*** |
| 11:00 – 11:30<br>**Срђан Костић,** Институт за водопривреду „Јарослав Черни“, Београд<br>**Небојша Васовић,** Универзитет у Београду, Рударско геолошки факултет<br>**Кристина Тодоровић,** Универзитет у Београду, Фармацеутски факултет |
| ***„Delay and stochastic differential equations as models of seismogenic fault motion“*** |
| 11:30 – 12:00<br>**Младен Видић,** Саобраћајни факултет, Добој |
| ***„Примена симетричне енкрипције за вертикалну ауторизацију у базама података“*** |

| |
|---|
| 12:30 – 13:00<br>**Олга Јакшић, Ивана Јокић, Милош Франтловић, Дана Васиљевић-Радовић, Зоран Јакшић,** ITHM Centre of Microelectronic Technologies, University of Belgrade, Serbia |
| ***„Mathematical Modelling, Algorithms and Software for Investigations of Affinity Based Bio-Chemical Sensors“*** |
| 13:00 – 13:20<br>**Татјана Бајић**, Висока школа струковних студија за васпитаче, Шабац |
| ***„Moment matching discretization of a stochastic integral“*** |
| 13:20 – 13:40<br>**Душица Гавриловић,** Институт за онкологију и радиологију Србије, Пастерова 14, Београд |
| ***„Вишеструкости у биомедицини: Проблем инфлације грешке I врсте и нека решења“*** |
| 13:40 – 14:00<br>**Никола Перић,** Khaoticen, Београд, Србија |
| ***„Мрежно планирање у функцији евалуације пројеката“*** |

## II СЕКЦИЈА: МАТЕМАТИКА И ИНФОРМАТИКА У ОБРАЗОВАЊУ

| |
|---|
| 10:00 – 11:00 |
| **Миодраг Матељевић,** Математички факултет, Универзитет у Београду |
| ***„Површина, запремина и интеграли – елементарни приступ”*** |
| 11:00 – 11:20 |
| **Nives Baranović,** Filozofski fakultet u Splitu<br>**Maja Cindrić,** Odjel za izobrazbu učitelja i odgajatelja Sveučilišta u Zadru |
| ***„O razvoju geometrijskog mišljenja u nastavi matematike prema van Hieleovoj teoriji”*** |
| 11:20– 11:40 |
| **Ивана Ковачевић,** ОШ „Др Драган Херцог”, Београд<br>**Душан Џамић,** Факултет организационих наука, Универзитет у Београду<br>**Славиша Радовић,** ГеоГебра Центар Београд<br>**Мирослав Марић,** Математички факултет, Универзитет у Београду |
| ***„Информационо-комуникационе технологије и образовање ученика са развојним сметњама”*** |
| 11:40 – 12:00 |
| **Владимир Балтић,** Математичка гимназија, Београд |
| ***„Коришћење пакета Мејпл у средњошколској математици”*** |

| |
|---|
| 12:30 – 12:45 |
| **Оливер Петковић,** Покрајински секретаријат за спорт и омладину, Нови Сад<br>**Мирослав Марић,** Математички факултет, Универзитет у Београду |
| ***„Видео материјали у настави математике”*** |
| 12:45 – 13:00 |
| **Милорад Шуковић**, ОШ „Свети Сава”, Аранђеловац<br>**Зоран Ловрен**, ОШ „Свети Сава”, Аранђеловац |
| ***„Математичко моделирање. Линеарна функција”*** |
| 13:00 – 13:15 |
| **Ђурђица Такачи**, Природно математички факултет, Департман за математику и информатику, Нови Сад<br>**Валентина Костић**, Гимназија Пирот<br>**Тања Секулић**, Висока техничка школа струковних студија, Зрењанин |
| ***„Математичко моделовање проблема кретања у динамичком визуелном окружењу”*** |
| 13:15 – 13:30 |
| **Снежана Коњикушић,** Факултет за економију, финансије и администрацију – ФЕФА, Универзитет Сингидунум<br>**Данијела Миленковић,** Факултет за економију, финансије и администрацију – ФЕФА, Универзитет Сингидунум<br>**Милољуб Албијанић,** Факултет за економију, финансије и администрацију – ФЕФА, Универзитет Сингидунум |
| ***„Елементарна математика у методама обрачуна камата”*** |
| 13:30 – 13:45 |
| **Александра Росић,** Завод за вредновање квалитета образовања и васпитања |
| ***„Уџбеници за математику некад и сад”*** |
| 13:45 – 14:00 |
| **Светлана Албијанић**, ЈП Службени гласник, Београд |
| ***„Дигитализација српских службених новина”*** |

| |
|---|
| 14:30 – 14:45 |
| **Наталија Будински**, Основна и средња школа ,,Петро Кузмјак", Руски Крстур |
| ***„Упознавање ученика са применом математичког моделирања у технологијама добијања материјала"*** |
| 14:45 – 15:00 |
| **Марија Минић,** Пољопривредни факултет, Универзитет у Београду<br>**Зоран Видовић,** Учитељски факултет, Универзитет у Београду |
| ***„Најчешће статистичке грешке у истраживањима"*** |
| 15:00 – 15:15 |
| **Татјана Станковић,** Електротехничка школа ,,Никола Тесла", Панчево<br>**Јасна Бошковић,** Електротехничка школа ,,Никола Тесла", Панчево |
| ***„Греши, реши, па разреши!"*** |
| 15:15 – 15:30 |
| **Милан Живановић,** Висока школа струковних студија за образовање васпитача, Крушевац |
| ***„Једна особина Херонових троуглова без целобројних висина"*** |
| 15:30 – 15:45 |
| **Зорица Маринковић,** Земунска гимназија |
| ***„Пази, броји се"*** |
| 15:45 – 16:00 |
| **Верица Милутиновић,** Факултет педагошких наука Универзитета у Крагујевцу |
| ***„Моделовање намере употребе рачунара у настави математике код будућих учитеља и наставника математике"*** |

| |
|---|
| 16:00 – 16:15 |
| **Јелена Хаџи-Пурић,** Математички факултет, Универзитет у Београду |
| ***„Употреба програмског језика JAVA на такмичењима из програмирања-аргументи ЗА и ПРОТИВ"*** |
| 16:15 – 16:30 |
| **Радоје Кошанин,** ОШ ,,Вожд Карађорђе", Ниш |
| ***„Пројектна настава математике у основној школи (пример из праксе)"*** |
| 16:30 – 16:45 |
| **Вера Ивковић,** Осма београдска гимназија |
| ***„Математика и књижевност"*** |

## III СЕКЦИЈА: НАУЧНОИСТРАЖИВАЧКИ И СТРУЧНИ РАД СТУДЕНАТА

| |
|---|
| 10:00 – 10:20 |
| **Филип Живановић,** Математички факултет, Универзитет у Београду |
| ***„Учешће на летњој школи математике у Америци”*** |
| 10:20 – 10:40 |
| **Владимир Ћошовић,** Faculty of mathematics, Department of Astronomy<br>**Бојан Новаковић,** Faculty of mathematics, Department of Astronomy |
| ***„Dynamical evolution of the Seinajoki asteroid family”*** |
| 10:40 – 11:00 |
| **Вера Милер Јерковић,** Електротехнички факултет, Универзитет у Београду<br>**Бранко Малешевић** |
| ***„Примена блок репрезентације Moore-Penrose-овог инверза матрица код линеарне дискриминантне анализе”*** |
| 11:00 – 11:20 |
| **Марија Ненезић**, Електротехнички факултет, Универзитет у Београду<br>**Бојан Бањац,** Електротехнички факултет, Универзитет у Београду |
| ***„Рационалне апроксимације неке аналитичких функција које се користе у дигиталној обради сигнала”*** |
| 11:20 – 11:40 |
| **Милица Цветковић**, Електротхенички факултет, Универзитет у Београду |
| ***„Нумеричка реализација спрегнутих Грос-Питаевски једначина за нискотемпературне честице”*** |
| 11:40 – 12:00 |
| **Илија Суботић**, Математички факултет, Универзитет у Београду |
| ***„Неке карактеристичне особине Grötzsch-овог Томасеновог и Хершеловог графа”*** |

| |
|---|
| 12:30 – 12:50 |
| **Добрица Ћосић,** Електротехнички факултет, Универзитет у Београду |
| ***„Естимација параметара Бајесових мрежа за системе препоруке”*** |
| 12:50 – 13:10 |
| **Марина Нешовић,** Трећа београдска гимназија<br>**Душан Тошић,** Математички факултет, Универзитет у Београду<br>**Ђорђе Стакић,** Математички факултет, Универзитет у Београду |
| ***„Анализа информација из инфокутија српских књижевника на Википедији на српском језику”*** |
| 13:10 – 13:30 |
| **Димитрије Цицмиловић,** Математички факутет, Универзитет у Београду |
| ***„Студирати (математику)”*** |
| 13:30 – 13:50 |
| **Божидар Радивојевић,** Математички факултет, Универзитет у Београду |
| ***„Такмичење "Math Hackathon" виђено очима организатора”*** |
| 13:50 – 14:10 |
| **Миљан Колчић,** Математички факултет, Универзитет у Београду<br>**Стево Шеган,** Математички факултет, Универзитет у Београду |
| ***„Физичке ефемериде планета и сателита”*** |

**Одабране фотографије са VI Симпозијума „Математике и примене”**


ЈА
Математички
Факултет
www.matf.bg.ac.rs
1873